Cloning and Recombinant Expression of a 822 bp region of a Pf403 *Plasmodium falciparum* gene

By

Timothy Ian Smallie
B.Sc. Hons (Gene)

Submitted in fulfilment of the
academic requirements for the degree of
Master of Science
in
Biochemistry
in the
School of Molecular and Cellular Biosciences
University of Natal
Pietermaritzburg

October 2003

PREFACE

The experimental work described in this thesis was carried out in the School of Molecular and Cellular Biosciences, Department of Biochemistry, University of Natal, Pietermaritzburg from January 2002 to October 2003 under the supervision of Professor J.P. Dean Goldring. These studies represent original work by the author and have not been submitted in any other form to another university. Use of work of others has been duly acknowledged in the text.

Timothy Ian Smallie
October 2003

ABSTRACT

Malaria is a devastating parasitic disease in humans caused by species in the genus *Plasmodium*. With over 100 million cases and at least 1.5 million fatalities each year, the disease accounts for 4-5% of all fatalities in the world. A recent increase in the number of malaria cases in South Africa has imposed severe costs on the economy and public health.

Immunity to malaria is a multi-component system involving both B and T cell lymphocytes. Pc96 is a 96 kDa antigen identified in the mouse malaria model *Plasmodium chabaudi adami*. It is known to be associated with the outer membrane of mouse erythrocytes infected with the parasite and has shown protective roles in mice challenged with *P. chabaudi adami*. A specific T cell clone has been identified that adoptively provides protection to athymic mice infected with *P. chabaudi adami*. Antibodies raised against Pc96 identified proteins that induced the proliferation of the protective T cell clones. At least four other antigens of different species of malaria share at least one cross-reactive epitope.

In an attempt to identify a *Plasmodium falciparum* homologue of Pc96, the amino-acid sequence was used in a BLAST search of the *P. falciparum* genome database, identifying a 403 kDa protein with a high degree of homology to Pc96. Sequence alignments indicated a region spanning 90 amino acids in Pf403 that overlaps the Pc96 amino acid sequence. A 178 kDa protein in *P. yoelii yoelii* (Pyy178) was shown to be highly similar to Pc96. T-cell epitope prediction programs identified putative T cell epitopes in Pc96 which appear to be conserved in Pf403 and Pyy178. A casein kinase II phosphorylation site was also identified in this region and is conserved in both sequences. PCR primers were designed to amplify regions of the MAL3P6.11 gene coding for Pf403 from *P. falciparum* genomic DNA. An 817 bp region in the MAL3P6.11 gene was amplified. This codes for the region of Pf403 that shows high homology to Pc96 and contains the conserved T cell epitopes and casein kinase phophorylation site. A *Bam*HI site was incorporated into the forward primer to facilitate in-frame ligation with cloning vectors. The PCR product obtained was verified by restriction analysis using *Hind*III and *Eco*RI sites within the fragment.

The 817 bp PCR product was cloned into the pMOS*Blue* vector using a blunt-ended PCR cloning kit, and transformed into MOS*Blue* competent cells. Recombinants were identified using the α-

complementation system, and verified by PCR, plasmid DNA isolation, and restriction digestion analysis. The insert DNA in pMOS*Blue* was cut out with *Bam*HI and sub-cloned into the *Bam*HI site in the pMAL-C2x expression vector. Sequencing of the construct confirmed the identity of the cloned insert and showed the sequence to be in frame with the *malE* gene coding for maltose binding protein (MBP). The fusion protein, MBP-Pf32.5, was induced and expressed as a 75 kDa protein comprising of the 32.5 kDa region of Pf403, and MBP (42.5 kDa) and was detected by anti-MBP antibodies, by western blotting. This recombinant protein has many applications for further studies involving the characterisation of the Pf403 protein, and the determination of possible roles that the protein may have in stimulating an immune response during human malaria infections.

ACKNOWLEDGEMENTS

I wish to express my heartfelt appreciation to the following people for their substantial contributions to this study:

To Prof. J.P.D Goldring, for his excellent supervision, patience, humour and assistance with all aspects of this project.

To Dr Mervyn Beukes, for his invaluable assistance and technical advice, various resources, good humour and friendship.

To all the academic staff at the School of Molecular and Cellular Biosciences: Dr Edith Elliott, Professor Clive Dennison, Dr Trevor Anderson, Professor Theresa Coetzer, Dr Andrew Mitchell, Professor Annabel Fossey, Dr Mike Meusel and Dr Caroline Hancock for their involvement and help.

Thanks to Mrs Patricia Wallis for her kind and patient assistance with many aspects of my lab work.

To my lab colleagues Laura Huson, Tamara Hiltunen, Habtem Habte, Ike Achilonu, and Pamela Mkhize, thanks for your help and encouragement over the last few years.

To Janine Jeary, and all the administration staff, thanks for your help in the efficient ordering of reagents and equipment.

To my parents Janet and Ian Smallie for financial support, interest and encouragement.

To Annamarie Myers, for helping keep the balance between work and play.

To the National Research Foundation (NRF) for the Grant-Holders bursary, and the Scarce Skills Scholarship for financial assistance.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3D-PSSM	3 Dimensional - Position-Specific Scoring Matrix
A_{260}	absorbance at 260 nm
A_{280}	absorbance at 280 nm
ACC	acetyl-coA carboxylase
AIDS	acquired immunodeficiency syndrome
BCIP	5-bromo-4-chloro-3-indolyl phosphate
Bis-acrylamide	N,N'-methylenebisacrylamide
BLAST	Basic Local Alignment Search Tool
bp	base pairs
BSA	Bovine serum albumin
C-terminus	carboxy terminus
cDNA	complementary deoxyribonucleic acid
cfu	colony forming units
CIAA	Chloroform / iso amyl alcohol
CIAP	calf intestinal alkaline phosphatase
CRE	cAMP response element
CS	circumsporozoite
CSP	circumsporozoite protein
dATP	deoxy adenosine tri-phosphate
dCTP	deoxy cytosine tri-phosphate
dGTP	deoxy guanidine tri-phosphate
DDT	dichlorodiphenyltrichloroethane
DIG	Digoxigenin
DIG-ddUTP	digoxigenin-labelled dideoxyuridine-triphosphate
dist.H_2O	distilled water
DMF	dimethylformamide
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotide triphosphates
dNTPs	deoxy nucleotide tri-phosphate
DTT	dithiothreitol
dTTP	deoxy thymidine tri-phosphate
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factor
ELI	expression library immunization
ELISA	enzyme-linked immunosorbent assay
EST	expressed sequence tags

g	Relative centrifugal force
h	hour
HIV	human immunodeficiency virus
HRPO	horseraddish peroxidase.
ICAM-1	intercellular adhesion molecule 1
IFN-γ	gamma interferon
IgG	immunoglobulin G
IgM	immunoglobulin M
IgY	immunoglobulin Y
IL-2	interleukin 2
IPTG	isopropylthio-β-D-galactoside
kDa	kilo-Dalton
LSA-1	liver stage antigen-1
MBP	maltose-binding protein
MHC	major histocompatibility complex
min	minute
M_r	molecular mass
MSP1	merozoite surface protein-1
MSP2	merozoite surface protein-2
N-terminus	amino-terminus
NBT	nitroblue tetrazolium salt
NCBI	National Center for Biotechnology Information
OD	optical density
ORF	open reading frames
PCR	polymerase chain reaction
PSI-BLAST	Position Specific Iterated - Basic Local Alignment Search Tool
RESA	ring-infected erythrocyte surface antigen
rpm	rotations per minute
RT-PCR	reverse transcriptase polymerase chain reaction
SDS	sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
STARP	sporozoite threonine and asparagine rich protein
T_a	annealing temperature
TBS	Tris-buffered saline
TE	Tris, EDTA
TEMED	*N,N,N',N'*-tetramethylethylenediamine
TIGR	The Institute of Genomic Research
T_m	melting temperature

TNF	tumour necrosis factor
TRAP	thrombospondin related anonymous protein
Tricine	*N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine
Tris	2-amino-2-(hydroxymethyl)-1,3-propanediol
TSP	thrombospondin
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

CHAPTER 1
INTRODUCTION

1.1 Biology of malaria

Malaria is a protozoan disease of the group Sporozoa (Brock and Madigan, 1994). The human disease is caused by a parasite that has a highly specific requirement for man as a vertebrate host and the *Anopheles* mosquito as a vector. Malaria is seen by many to be the most important and influential disease in human history. It has played a significant role in the spread and development of culture and evolution of man and is thought to have originated in Africa up to 30 million years ago. The adaptive ness and success of the parasite as a pathogen indicates a long and close-knit relationship with humans. The severity of the malaria problem is appreciated when one considers that over 100 million malaria cases are estimated worldwide, with at least 1.5 million fatalities each year. This accounts for 4-5% of all fatalities in the world (Brock and Madigan, 1994).

There are four species of malaria that cause fever in humans: *Plasmodium vivax, P. ovale, P. malariae,* and *P. falciparum.* The most lethal, *Plasmodium falciparum,* is the most devastating of these species, commonly present throughout the many regions of the world, causing serious infection and cerebral malaria. Most of the deaths occur in young children and pregnant women in the sub-Saharan African regions. Newborn infants in these affected regions suffer complications from low birth weight caused by malaria infection during pregnancy. In most cases the patient will survive the illness yet will be sick for 10 to 20 days, placing hindrance on work productivity and education (Phillips, 2001).

The early 1960's saw drastic anti-malarial campaigns having a large effect on the eradication of malaria in many regions of the world, however modern times has seen an increase in the disease, re-affecting many of the regions thought to have been cleared. Several factors have caused this including recent wars and conflicts, a dramatic increase in migration from malarias areas, poor health facilities in developing countries and general poverty. The global ban on dichlorodiphenyltrichloroethane (DDT) prevented the total eradication of the disease from these regions. In general the rapid increase in global temperatures together with drug resistance makes effective control of the epidemic an emergent and immediate worldwide problem (Greenwood *et al.*, 1991). The distribution of drug-resistant malaria is shown in Figure 1.1.

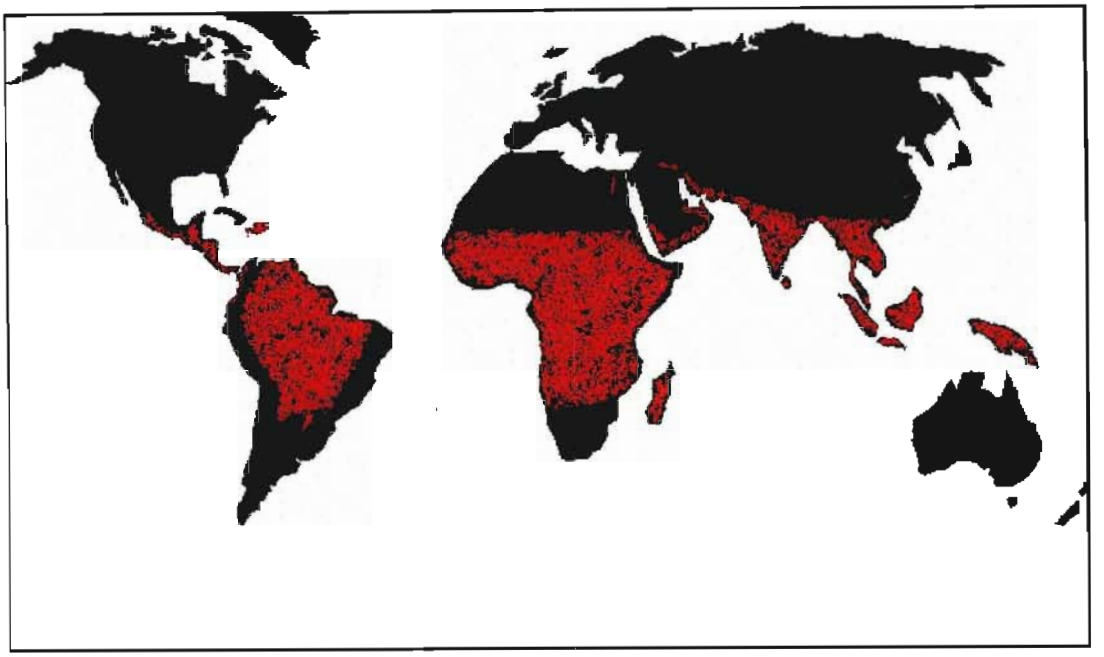


Figure 1.1 The worldwide distribution of chloroquine resistant *Plasmodium falciparum.*

One of the first goals of researchers is to rapidly develop new tools needed for the effective control of the disease. During the early 1990's many efforts at solving the problem included different approaches and strategies, such as the use of insecticide-impregnated bed nets. In more recent times, efforts largely have been focused on molecular and genetic strategies targeting both the host mosquito and parasite biology. The genetic biocontrol of the mosquito has been one of the main angles for research into control. The increase in knowledge of the disease due to the availability of the *Plasmodium falciparum* genome, has led to the discovery of vital features of the immune system with regard to the development of a protective vaccine against the disease (Collins and Paskewitz, 1995).

Although progress has been made slowly in this field, the development of a human malaria vaccine has proved to be a lot more complicated than previously believed. The immune mechanisms involved in malaria infections have been extensively studied although complicated by the extensive life cycle of the parasite, confining specific protective antigens to specific stages of development and pathogenesis (Zavala *et al.*, 1985). Some of the malaria proteins with vaccine potential, such as MSP-1 (malaria surface protein) have a high degree of antigenic variation

between clones, protecting only against specific strains in specific regions. Others may possess multiple copies of genes with different receptor specificities, altering the human host erythrocyte through the control and regulation of these genes (Miller *et al.*, 1994).

These challenges have indicated the need for a more multi-faceted approach towards the design of a vaccine against malaria. Recombinant proteins, attenuated vaccinia viruses, and DNA vaccines have been studied recently. Recombinant proteins derived from different antigens representing different stages of parasite growth, offer possible solutions to immunization against the parasites highly complex life cycles (Stanley, 1998). New techniques in the identification of potential vaccine antigens, along with molecular cloning and protein expression systems have allowed many promising research opportunities. The recent sequencing of the parasites genome has revealed a lot more information both from the protein and gene perspectives (Gardner et al., 2002).

The four main species of malaria parasites causing disease in humans (*P. falciparum, P. vivax, P. ovale* and *P. malariae*) contain several unique features suited to the disease they cause. The relapsing malarias caused by *P. vivax* and *P. ovale* can remain dormant in the liver for very long periods (hypnozoite stage). This capability allows the parasite to survive in regions that contain winters or periods where reproduction and transmittance is impossible. *P. malariae* is unique in its ability to survive for many years at extremely low parasite densities in the infected host (Collins and Paskewitz, 1995).

The plasmodial existence in the host involves a complex life cycle with three distinct phases in the mosquito and two in the human (Figure 1.2). The parasite is first transmitted to humans through the saliva of infected *Anopheles* mosquitoes, during a blood meal. Only a few hundred sporozoites pass directly from the mosquito salivary gland system and into the blood stream of humans where invasion of the liver takes place soon after infection (Phillips, 2001). The molecular basis for liver cell invasion is not entirely known, however the recognition and attachment to the cells has been shown to involve specific ligands on the outer sporozoite membrane such as circumsporozoite protein (CSP). After several weeks in the liver, the parasite develops into what is known as the schizont form (Oaks *et al.*, 1991). The schizont contains several thousand 'daughter' parasites known as merozoites, which are released from the liver into the blood where the infection and invasion of erythrocytes occurs (Phillips, 2001). Some of the sporozoite structures can remain dormant inside the liver cells for a number of years until they are reactivated and reinfection of the red blood cells occurs (depending on the particular species). The merozoite matures inside the infected erythrocyte forming a new schizont structure known to contain 8 to 32 new merozoites. The erythrocyte eventually ruptures after roughly 48 hours releasing the merozoites into the blood

stream where they are free and available to invade and infect other erythrocytes. It is at this point in the life cycle that fever associated with the clinical effects of malaria occurs (Oaks et al., 1991). The morbidity and mortality of the disease only appears at this stage (Phillips, 2001). Figure 1.2 shows the malaria life cycle, in the Anopheles mosquito and human host.

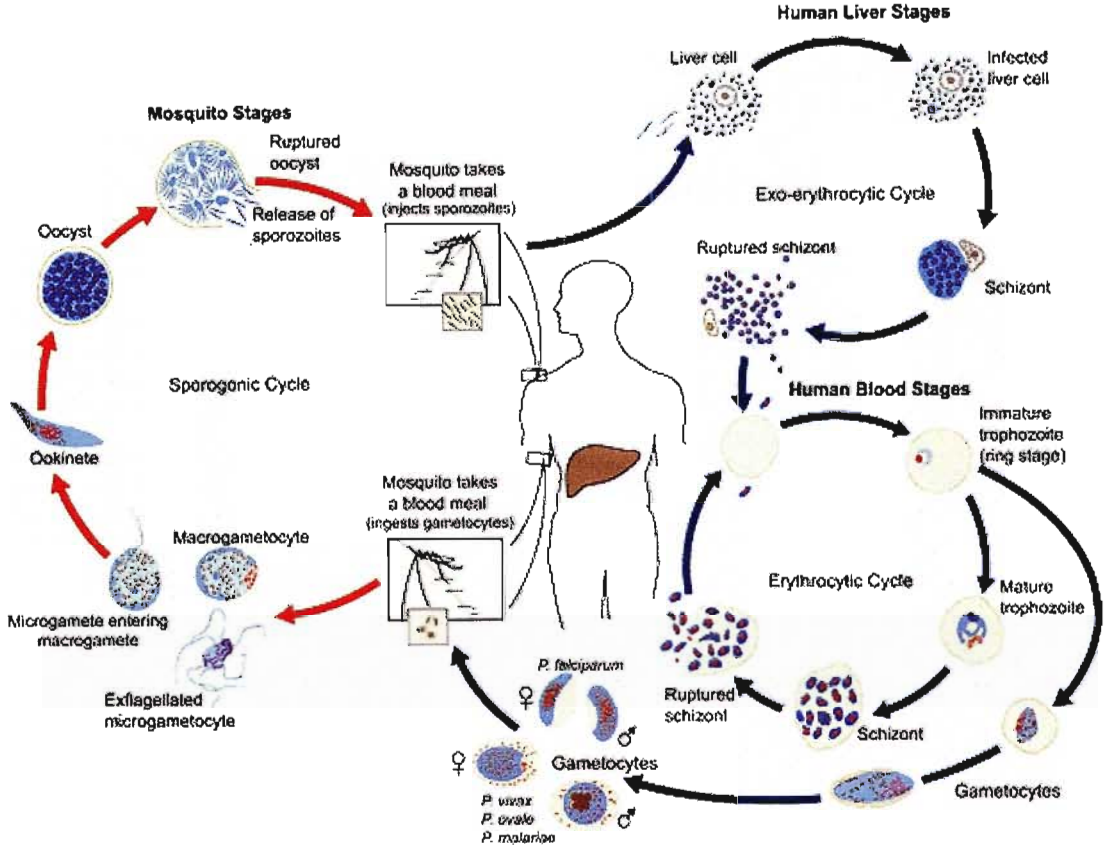


Figure 1.2 The life-cycle of the malaria parasite in man and the Anopheline mosquito host http://www.cdc.gov/malaria/biology/life_cycle.html

Several of these new merozoites will develop into sexual structures known as gametocytes, and if ingested by a mosquito then male and female gametes will fuse, forming a zygote. The zygote elongates over the next 12 to 48 hours forming an ookinete, which penetrates the insect's stomach wall and becomes an oocyst. This structure will enlarge over the next few days forming thousands of sporozoites. When the oocyst ruptures, the sporozoites move to the mosquito salivary glands, ready for the next cycle (Oaks *et al.*, 1991).

The cycle takes approximately 7 to 18 days after gametocyte ingestion, depending on factors related to the host-parasite combination and environmental conditions. A large proportion of these developmental stages are haploid, apart from the zygote, which is diploid (Billingsley and Sinden, 1997). It is at this stage where meiotic cell division occurs allowing genetic recombination.

1.2 Malaria control and diagnosis

The complexity of the malaria life cycle and the different stages of *Plasmodium* development provide an opportunity for a range of control measures. The biology of the vector does not lend itself to a simple or universal control method and the most effective form of control has been insecticide like DDT. Parasite development involves long term survival within the mosquito host, making the host vulnerable to insecticides before transmission of the parasite occurs. The modification of breeding habitats, such as removal of swamps and breeding grounds for *Anopheles*, has shown success in decreasing the disease in specific regions. Mosquito nets have been impregnated with chemicals, and biological control agents have shown some success in eliminating malaria from several regions (Collins and Paskewitz, 1995).

The prevalence of malaria in most of Europe and the USA declined in the early 1800's due to the massive uprising of agriculture and the sudden elimination of many drainage systems and areas such as wetlands and swamps. Eventually eradication of malaria in these regions was accomplished by the availability of quinine, along with an increase in education and economical status in the country, and better housing facilities. In the late 1940's the US Public Health Service implemented large scale spraying with DDT and outbreaks were limited to infected migrants and war veterans returning from infectious areas, such a Vietnam and the Middle East (Collins and Paskewitz, 1995). This caused many to think that the total eradication of malaria as a worldwide problem was a possible goal, with the use of DDT to control the vector.

The Global Malaria Eradication Program was established in 1955 in order to define specific worldwide goals for the elimination of the disease and one goal was the reduction in infected vector populations to such an extent as to interrupt further transmission and survival of the parasite. In Sri Lanka malaria incidence was reduced from 2.8 million cases in 1946 to 17 in 1963, yet when spraying was stopped an epidemic occurred in 1967 and a reintroduction of spraying was ineffective due to resistance of the vectors to DDT. The disease caused more than half a million cases by the end of 1967. By the early 1970's most resources were devoted to malaria control instead of eradication and DDT was removed from the US market because of environmental concerns. Agricultural spraying of insecticides for control of cotton and rice pests in El Salvador caused resistance in mosquitoes to several new compounds (Collins and Paskewitz, 1995).

There are a number of species of host vector capable of transmitting the parasite, *Anopheles gambiae* being the most common and abundant mosquito species. The insect primarily breeds in temporary rain puddles and so control at the larval level is difficult. The removal of swamp and

en-catchment areas does reduce populations of mosquitoes but the removal of small puddles after the rains is virtually impossible. In urbanized areas this control has been successful as the mosquito prefers to breed in more accessible structures where control is easier (Collins and Paskewitz, 1995).

Due to difficulty in the use of insecticides and the main environmental concerns that the chemicals pose, a lot of research is now focussed on biological control agents for controlling the spread and effect of malaria on specific populations. The most successful of these projects has been the use of larvivorous fish in the control of *Anopheles* larval stages present in water sources. The use of the North American fish, *Gambusia affinis*, in 3800 wells in India appeared to dramatically reduce mosquito densities and biting rates by 75% due to a reduction in larvae. Environmental issues have arisen due to the competition between *G. affinis* and indigenous fish in similar habitats. Several other biocontrol agents such as pathogens and parasites have been studied. This is however difficult due to the management and storage of these reagents, and successful use is proving impractical in the field (Collins and Paskewitz, 1995). In several poor countries where malaria poses a serious threat, the reduction and elimination of vector contact with the host has been the most simple and straightforward approach through the use of repellents or physical barriers such as bed nets. Bed nets are effective when impregnated with permethrin insecticides and produced a 70% reduction of clinical malaria in children though regions of Africa (Collins and Paskewitz, 1995).

In 1820, quinine was discovered in the bark of the South American *Cinchona* tree as an active ingredient. A high demand for the new drug quinine resulted in a massive depletion of the wild tree and soon after several *Cinchona* plantations were established. During World War II, several new compounds were discovered enabling the treatment and control of malaria infections, among them chloroquine and atabrine. Chloroquine is thought to interfere with the polymerisation of haeme molecules in the erythrocytes by the trophozoites due to the degradation of haemoglobin. Haeme molecules are usually polymerised into granules called hemozoin, a non-toxic product. The drug provided a cheap and effective solution for therapy yet resistance to chloroquine arose in the early 1960's in South America and South East Asia, and spread through Africa in the 1980's (Collins and Paskewitz, 1995). In 1989, strains of chloroquine resistant parasite infections were discovered in Papua New Guinea and soon spread to the Indonesian side of the island, indicating a spreading pattern and progression towards the rest of South East Asia (Baird, *et al.*, 1991). Other prophylactic drugs such as amodiaquine, pyrimethamine and proguanil have been used in chloroquine resistant areas and antibiotics such as doxycyline have shown effectiveness (Collins and Paskewitz, 1995).

It is now believed that a number of different approaches should be taken in vector control and research towards eradication of the *Anopheles* mosquito has moved more towards transgenic and genetic studies due to the development of drug resistance (Butler, 1997). The focus is now transgenic mosquitoes that are incapable of spreading malaria, and the spread of these genes involved in natural resistance throughout wild populations in the field. These genes are currently being mapped using linkage studies with resistant strains of mosquitoes. A collaborated effort to map the mosquito genome began in the late 1980's, which was of great benefit to these studies. Other approaches include the engineering of mosquitoes with foreign genes conferring protection against infection. The bigger challenge would be the introduction of transgenic mosquitoes into the environment. Engineered transposons can integrate and replicate into the genome of mosquitoes in wild-type populations, which should result in the effective transfer of genes in the environment, yet this is a controversial issue, as the safety of this procedure cannot be assured. The manipulation and release of genetically altered organisms into the field may have several unknown detrimental effects to the ecosystem and environment (Butler, 1997).

Malaria has become an increasing problem in South Africa over the last number of years. The high-risk areas include the northern regions on Mpumalanga, and Northern KwaZulu-Natal, including several of the game reserves alongside or near Mozambique (Kruger National Park). Malaria cases have been reported as far south as Durban, and rapid control measures are needed and an effective system in place for the prevention of further spread of the disease. In 1988, due to drug resistance, various regions of the country such Mpumalanga and Northern KwaZulu Natal, replaced the chloroquine drug with sulfadoxine- pyrimethamine (SP) for the successful treatment and therapy of *P. falciparum* infections (Sharp and Le Sueur, 1996). In 2001, artemether lumefantrine drugs replaced the pyrimethamine-based drugs, due to a rapid spread of SP resistant malaria (Bredenkamp *et al.*, 2001). Since then drug combinations including artemesinin derivatives have shown success in therapy and a decrease in transmittance in regions of South Africa due to the emergence of artemether lumefantrine resistance. The disease is seen as a huge problem in South Africa, from a clinical point of view, and in light of HIV infection rate. Similarly to other malarious regions in Africa and the rest of the world, a human malaria vaccine would be the most effective method of control of the disease.

1.3 Immunology of malaria

Immune responses in humans leading to the efficient defense against disease-causing pathogens, involves two separate mechanisms; humoral, and cell mediated immunity. All humoral

mechanisms involve the action of antibodies and antibody-producing B cells and their respective responses to infection. Cell mediated immunity is far more complex and involves the activation of T cells leading to the intracellular and extracellular killing of malaria parasites (Clark, 1987). Cell mediated immunity involves several attributes of the immune system including the interactions of the B-lymphocytes and their antibodies within these cellular mechanisms. Cell mediated and antibody responses in malaria infection are now seen as highly integrated functions. T cells appear to function in promotion of protective antibody synthesis and as inducer cells in cell mediated immune responses (Jayawardena and Murphy, 1982). The T-lymphocyte system plays an essential role in re-infection immunity to malaria, in human populations with resistance in Africa. The involvement of B cells in these infections is vital for the survival and resistance to chronic malaria infections (Grun and Weidanz, 1981).

1.3.1 Humoral responses to malaria

Most research-based efforts into the design of human malaria vaccines has been based on the initial observation that protection against the parasite is largely due to the involvement of the humoral immune system and antibodies (Brake, et al., 1988). Originally it was thought that antibodies were responsible for the phagocytic processes of parasitised erythrocytes by macrophages during infection (Allison and Eugui, 1983). The role of antibodies has since been studied more extensively, revealing more direct evidence. It has been shown that there are elevated levels of antibodies during malarial infection in the sera from adult Africans. These antibodies when transferred rapidly accelerated the recovery of African children from *Plasmodium falciparum* infections (Cohen *et al.,* 1961). In more recent studies, monoclonal antibodies have been used to identify and define stage-specific antigens and mechanisms for this immunity have been analysed (Schofield *et al.*, 1987).

There is growing evidence to suggest the role of immunoglobulin G (IgG) in protection against malaria in humans. Human antibodies have been shown to efficiently inhibit *in vitro P. falciparum* merozoite proliferation and the passive transfer of IgG have provided protection against falciparum malaria in South American monkeys (Aucan *et al.*, 2000). Some blood stage antigens may be the target of protective immunity and have been included in several vaccine trials in humans. In particular ring-infected erythrocyte surface antigen (RESA), merozoite surface protein-1 (MSP1) and merozoite surface protein-2 (MSP2) are of considerable interest as they have been the targets of protective immunity in animal models and are recognised by naturally acquired antibodies. These antigens consist of polymorphic and conserved B cell epitopes responsible for

the elicitation of immune response (Aucan *et al.*, 2000). More recent understandings of the immune mechanisms during malaria infection have revealed an integration of the cell-mediated responses, with regard to the involvement of T cells and their functions.

1.3.2 Cell Mediated Immunity

Mouse malaria models such as *Plasmodium chabaudi adami* and *P. yoelii yoelii* (both genomes have been sequenced recently), have provided vast amounts of information about the immunology of malaria infection. DNA sequencing has proven highly successful in the production of several significant discoveries into the involvement of cell mediated mechanisms, and immunology in general. Manipulation of the mouse immune system such as the removal of B and T cell components has provided the basis for much new knowledge gained. B cells in mice can be disabled by the continuous treatment of the mice with goat anti-mouse immunoglobulin M (IgM) serum (Roberts and Weidanz, 1979). The mice were totally incapable of producing any antibody, as indicated by their failure to elicit responses against standard antigens such as sheep erythrocytes. The involvement of B cells in infections was assessed using these models. Mice also can be rendered T cell deficient (athymic) and have been used extensively in this field to study the dependence of the mice on an intact thymus and cell mediated immune system.

Grun and Weidanz (1981) have shown that B-cell deficient mice challenged with *P. chabaudi adami*, resolved their infections in a similar manner to that seen in immunologically intact mice. The ability of B cell deficient mice to show immunity against parasite infection indicates that cell mediated immunity plays a significant role. The complexity of immune reactions involved in these experiments was demonstrated by challenge with exogenous parasites *P. berghei* and *P. yoelii,* producing fatal infections in B cell deficient mice. This suggests that different mechanisms of immunity exist in different species and perhaps more than one immune mechanism is required for resistance of infection with certain plasmodial species (Grun and Weidanz, 1981). Cell mediated mechanisms of immunity were further studied using T cell deficient mice subjected to infection with *P. chabaudi adami*. The failure of athymic (T cell deficient) mice to resolve *P. chabaudi adami* infections shows that the development of immunity to *P. chabaudi adami* infection is dependent on an intact thymus-dependent system (Grun and Weidanz, 1981).

It is possible that certain T cell epitopes on plasmodial antigen are the primary targets for protective cellular immune responses. These molecules can be seen as potential antigens and should be investigated for their potential as immunizing agents (Brake *et al.*, 1988). Multiple

antigen peptides containing potential and known B cell epitopes were coupled to T cell epitopes from tetanus toxin. The tetanus toxin is known to produce a strong cell mediated response. This, coupled to non-ionic block copolymer adjuvants, has shown an improved protective response in malaria (Wang *et al.,* 1995).

In comparison to B cell epitopes, T cell epitopes are far more difficult to identify. Plasmodium protein antigens containing specific B cell epitopes can be identified systematically and with relative ease, yet this does not apply to T cell epitopes. A solution to this has been the development of protective T cell clones that are then used for identification of antigenic epitopes in fractionated proteins or DNA libraries. T cell clones were obtained from volunteers immunized with irradiated sporozoites, and were subsequently used to identify specific epitopes within antigens such the circumsporozoite (CS) protein (Calvo-Calle, 1997). T cell clones have been able to provide adoptive protection against malaria in mouse models (Brake *et al.*, 1986). One of these clones identified (CTR2.1), along with two subclones, was shown to effectively transfer protection to athymic nude mice infected with *P. chabaudi adami.* The clones were also shown to secrete IL-2 and interferon γ (IFN-γ), when stimulated with plasmodial antigen. This may suggest a role of these clones in the initiation of clonal expansion leading to further activation of host immune mechanisms, and the production of anti-parasitic factors. These studies indicated that it was possible that T cell epitopes were recognised on the surface of infected erythrocytes, causing the eventual immune response seen in malaria infection (Brake *et al.*, 1988). When athymic mice were infected with the CTR2.1 T cell clone, they produced a polyclonal IgG antibody response against a number of *Plasmodium chabaudi adami* antigens, as seen in euthymic mice. Due to the fact that the T cells were of a single clonal type, belonging to a specific subset, it may be that they provide help to a variety of B cell types, producing the array of antibodies seen in the sera. Due to the limited specificity of the T cell clones, it may be that there are several antigens present in the parasite containing the same or a universal T cell epitope, activating B cells to produce antibodies against these antigens (Goldring *et al.*, 1989).

Recent studies have shown a role for extrathymic T cells in protection against malaria in the *P. yoelii* mouse model. Although in the *P. chabaudi adami* model, in which athymic mice did not recover from challenge infection, in the *P. yoelii* model they did. It was proposed that the participation of extrathymic subsets of T cells played a role. These T cells participate in an innate-type immune system, associated with autoantibody producing B-1 cells (Kaiissar Mannoor *et al.*, 2002). This indicates how differently immune systems control the infection for different species of malaria and there are many factors that lead to cell mediated immunity. Infections with *Plasmodium vinckei petteri*, *P. chabaudi chabaudi,* and *Babesia microti*, in B cell deficient mice

appeared to exhibit a similar immune reaction as in normal mice. This further indicates the participation of the cell-mediated arm of immunity in resolving these species infections (Cavacini et al., 1990).

Current research on gamma delta T cells suggests the possibility of a protective role (Elloso *et al.*, 1994). These have been studied in murine models and have been found to function independently of antibodies and B cells (Van der Heyde *et al.*, 1995). The T cell receptor of gamma delta is expressed on a small proportion of peripheral lymphocytes and may participate as non-MHC restricted cytotoxic cells against intra-erythrocytic parasites (Ho *et al.*, 1990). During infections with *P. chabaudi* in mice, the γδ T cell populations increase in B cell deficient mice. These T cells were shown to produce IFN-γ when exposed to *P. chabaudi* parasites, indicating the possible involvement of macrophages. Recently γδ T cell deficient gene knockout mice were used to show the involvement of these cells in the pathology mouse of malaria infection (Seixas et al., 2002).

There have been several explanations as to how the participation of T cells in the immune system leads to the recovery from infection. Antibody-independent mechanisms may involve the role of activated macrophages. During erythrocytic cycles of growth, systemic release of pro-inflammatory cytokines is directed either by soluble malarial toxins or the stimulation of T cells by parasite antigens (Miller *et al.*, 1994). The specific non-antibody mediators released from host leukocytes include gamma interferon, tumour necrosis factor (TNF) and reactive forms of oxygen. TNF production by monocytes is facilitated by T cells and sensitizes neutrophils and macrophages to agents that cause them to secrete super-oxide compounds involved in the killing of pathogens. It is likely that malaria-dependent activation of T cells and phagocytes leads to both disease and killing of parasites (Miller *et al.*, 1994).

Reactive forms of oxygen and TNF are seen to play a related role in these mechanisms. Clark has shown that parasitaemia in *P. chabaudi adami* infected mice has been enhanced by ingestion, by the host, of free radical scavenger compounds such as butylated hydroxyanisole (Clark, 1987). Other researchers have reported that *in vivo* administration of TNF inhibits the survival of mouse malaria. The disadvantage in comparison to antibody mediated immunity is that overproduction of TNF by the host may be detrimental to survival and transmittance of the parasite due to the effects of TNF on the host. Symptoms associated with TNF match those seen during chronic infection of the disease and changes due to TNF toxicity in human malaria include symptoms such as headaches, fever, myalgias and nausea. This has also shown to occur in similar reactions seen in cancer patients receiving recombinant TNF infusions (Clark, 1987). Antibodies raised against TNF

have also been shown to inhibit the appearance of malaria infection symptoms in the pathology of *P. berghei* without inhibiting parasitic growth (Grau et al., 1987).

1.4 Vaccination against malaria

The current control of malaria lies mainly in the use of anti-vector approaches and antimalarial drugs. However both the *Anopheles* vector and the malaria parasite are becoming increasingly resistant towards these measures taken and this is seen as a major problem in regions of the world where malaria poses a problem This has shown the need for the development of a more long term control measure such as the development of a human malaria vaccine (Collins and Paskewitz, 1995). Several studies over the last 50 years have shown that the concept of vaccination against malaria may prove a useful tool in control of the disease (Holder, 1999). Humans that have been continuously exposed to malaria infection eventually develop immunity to the disease. It was found that immunization with antibodies from these individuals causes a dramatic reduction in blood stage parasitaemia (Cohen *et al.*, 1961). Furthermore, the inoculation of live attenuated parasites into humans and animals before challenge infection leads to protective roles in these models (Mitchell *et al.*, 1975). The development of a protective vaccine will involve the incorporation of the immune systems involved in acquiring natural immunity, as seen in exposed populations. Identification of the components of this protection, will allow insight into the development of a protective vaccine. Subunit and multi-component vaccine candidates have shown to partially protect against infection, yet with limited success. Another factor creating difficulty in vaccine development is the antigenic variation between different geographically located parasites causing different levels of protection to arise (Ranjit and Sharma, 1999).

1.4.1 Sporozoite stage vaccination

Protective immunity in malaria has been acquired by the immunization of the person with radiation-attenuated sporozoites in many animal and human models (Chen *et al.*, 1977). The protection is highly effective against large doses of infectious sporozoites and appears to be species and not strain specific. The introduction of these disabled whole parasites provides a complete range of natural antigens to the immune system. After the introduction of sporozoites from the mosquito liver cell invasion takes place. There are only two main proteins expressed at this stage involved in the invasion of liver cells. Antibodies specific against the surface proteins of the sporozoite can neutralise and prevent liver cell invasion (Holder, 1999). The infected liver cell has been identified in this way as a main target for sporozoite induced protection. This recognition

leads to cytotoxic T cells and killing of the infected hepatocyte. The T cells can also be stimulated to produce cytokines such as γ-interferon (Holder, 1999). It is known that this diversity shown in the T-cell dependent immune responses mediate this protection and $CD8^+$ T cell and IFN-γ-mediated immune responses should form the basis of vaccine design in the stages before erythrocyte invasion, which cause the main symptoms and pathogenesis of the disease (Doolan and Hoffman, 2000).

One of the main focuses of the development of a subunit synthetic peptide vaccine was based on reproducing similar immunity to that achieved using attenuated sporozoites, and has involved one of the expressed proteins at this stage, the circumsporozoite protein (CSP) (Targett, 1995). The protein was discovered in the rodent model of infection with *P.berghei*, and was soon cloned producing recombinant CSP (Dame *et al.*, 1984). This protein was previously shown to be a major component of the sporozoite surface membrane (Herrington *et al.*, 1991).

Cell mediated and humoral immune responses have been shown to be protective in rodents and nonhuman primates targeted to the CS protein (Calvo-Calle *et al.*, 1997). The success in CSP lies in the inclusion of relevant T cell and B cell epitopes in the development of synthetic vaccines. The repeat region of CSP has been shown to contain a B cell epitope $(NANP)_3$ that can elicit specific sporozoite neutralising polyclonal and monoclonal antibodies (Zavala *et al.*, 1985). This $(NANP)_3$ region was incorporated into a synthetic peptide, conjugated to tetanus toxoid as a protein carrier, and formed the first synthetic peptide malaria vaccine to undergo phase I, and phase II clinical trials (Herrington *et al.*, 1987). The parasite-specific antibody responses were low and the repeat sequence was unfortunately poorly recognised. In protective immunity, T cells expressing the $CD8^+$ antigen are required. This lead to investigations involving the transfer of T cells recognising specific epitopes contained in CSP, into mice. This passive transfer lead to a increased degree of protection against challenge, proving that $CD8^+$ T cells are directly involved in the response (Romero *et al.*, 1989).

It has since been realised that the use of irradiated attenuated sporozoites in providing long term protection is a direct result of the parasite producing infection in the hepatocytes. This induces a set of immune effector mechanisms mediated by cytokines and antibodies to various malaria antigens other than CSP (Targett, 1995). Gene expression studies of the malaria sporozoite stage revealed nearly two thousand expressed sequence tags involved in host cell invasion, providing a new handle for the development of pre-erythrocytic vaccines (Kappe *et al.*, 2001). The tandem repeat region of CSP has also been cloned into plasmid vectors and expressed as recombinant proteins in *Escherichia coli.* When injected into mice this protein induced high titres of antibodies

that react with the CSP protein on live sporozoties, blocking *in vitro* invasion of hepatoma cells. Immunization with vaccinia and influenza virus constructs expressing B cell or $CD8^+$ T cell epitopes of CSP or oral immunization with *P. falciparum* CSP co-expressed in *Salmonella typhi* have also been extensively investigated (Targett, 1995). One of the most interesting approaches included a DNA vaccine trial using DNA encoding the CSP, which caused induction of cytotoxic T cells and antibody responses providing good protection in monkeys (Sedegah *et al.*, 1994). The protozoan parasite *Toxoplasma gondii* causes strong cell-mediated immunity as well as non-specific resistance against other pathogens and tumors. For this reason the circumsporozoite protein of *Plasmodium yoelii* was engineered into *T. gondii* in order to create a stronger immune response against CSP. The CSP proteins were successfully expressed on the surface of *T. gondii* and high levels of anti-CSP antibodies were elicited. These results established a possible vaccine vehicle for the induction of $CD8^+$-dependent cell-mediated immunity (Charest *et al.*, 2000).

Hill and colleagues have pursued a different approach to induction of protective cytotoxic T cell responses by vaccination. They have shown that human leucocyte class I antigen (HLA-Bw53) and an HLA class II haplotype (DRB1*1302-DQB1*0501) is associated with protection from severe malaria. These haplotypes are common in humans throughout most of west African but rare in other racial groups. In this population they account for a reduction in disease incidence with similar proportions to the sickle cell variant. This data strongly supports the view that polymorphism of major histocompatibility complex (MHC) genes has evolved through time through natural selection by infectious agents (Plabanski and Hill, 2000). A search of pre-erythrocytic stage antigens for potential HLA-Bw53 epitopes revealed peptides in the liver stage antigen-1 (LSA-1). These epitopes were recognised by cytotoxic T cells from Gambian subjects with the identical class I antigen. The study involved 6 different haplotypes common in both African and Caucasian populations, revealing several epitopes in 4 pre-erythrocytic stage antigens ie. CSP, LSA-1, thrombospondin related anonymous protein (TRAP) and sporozoite threonine and asparagine rich protein (STARP) (Targett, 1995).

In a number of vaccine trials, it is often seen that the overall effectiveness of the vaccine is lowered by the genetic restrictions imposed by the various HLA types. Vaccines such as these should ultimately contain multiple protective $CD8^+$ T cell epitopes restricted by diverse MHC class I alleles, accounting for the genetic differences seen in populations in infectious areas. Another problem that reduces the effectiveness of vaccines is the genetic differences in the structure and antigenicity of the proteins used. The circumsporozoite protein showed heterogeneity in the repeat region and phenotypically different proteins were found in different strains and geographically located parasites (Rosenberg *et al.*, 1989). Doolan and colleagues (1996) found a

way to prevent this effect of genetic restriction of protection against malaria using DNA immunization with multi-gene representatives of the CSP.

Specific proteins have been shown to be required for sporozoite gliding motility and aspects of liver cell invasion and are part of a family of proteins found in the micronemes of invasive stages of apicomplexan parasites. These proteins are trans-membrane proteins and studies were performed investigating their role in cell invasion. Antibodies were effectively raised against these putative antigens, however they are not expressed in sufficient amounts on the cell surface for effective recognition and blocking by the immune system (Gantt *et al.*, 2000).

The advance in the technology of proteomics and genomics, along with more powerful methods of DNA sequencing, have allowed more detailed studies on the sporozoite stage, and the expression of proteins. cDNA Libraries were constructed, from the salivary gland sporozoites of the *P. yoelii yoelii* model (Kappe *et al.*, 2001). A total of 1972 expressed sequence tags (ESTs) were generated from the construction of the library, enabling all expressed proteins in this stage to be evaluated. Metabolic pathways and protein expression, thought not to occur during the sporozoite stages were identified, such as key enzymes in the shikimate pathway. Proteins were more effectively studied in the context of cell invasion. New insights into protein expression during the sporozoite stage were gained and various adhesion ligands were discovered which had potential for formulating a pre-erythrocytic based vaccine.

1.4.2 Merozoite stage vaccines

Most of the clinical symptoms of the disease are produced in the erythrocytic stage of the malaria life cycle. This shows that this stage should offer several vaccine candidates aimed at either clearing infection or reducing the detrimental effects of the disease. Red blood cells infected with *P. falciparum* have been shown to disappear from peripheral blood circulation as the intra-erythrocytic parasite develops into the schizont. This is due to the binding or adherence of the infected cells to microvascular endothelial cells (Miller *et al.*, 1994). This phenomenon is known as and often produces a lethal condition that in most cases leads to death. Cerebral malaria causes sequestration of red blood cells, which has been shown to be associated with the occurrence of knob formation on the surface of infected erythrocytes. A large antigenic variant surface protein called *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) was shown to represent specific ligands with host cells receptors CD36, thrombospondin (TSP) and intercellular adhesion molecule 1 (ICAM-1) (Baruch et al., 1995).

There are several proteins whose coding sequences belonging to the *var* gene family, including the gene for PfEMP1. This allows the *P. falciparum* parasite to express multiple forms of the protein, which contribute to the success of the parasite as a pathogen. These variant proteins have the ability to evade secondary disease, by the expression of different forms of proteins, avoiding detection by antibodies previously raised against proteins such as PfEMP1. This has been studied in several models including simian, murine and human malaria species indicating that antigenic variance in the surface of infected red blood cells has been linked to the lengthened persistence of infection, and the reestablishment of disease. Antibodies raised against specific forms of the PfEMP1 have shown to block adherence to CD36 in infection with the same isolates. This has paved the way for several important studies on vaccination against merozoite stage malaria (Baruch et al., 1995). These antibodies did not reduce the adherence to other receptors such as thrombospondin yet this study provided the molecular basis towards detailed studies on merozoite involvement in cerebral malaria.

The merozoite surface protein-1 (MSP-1) is a 180 to 210 kDa protein and has been found to elicit effective levels of immunity. This antigen is located on the surface of *P. falciparum* merozoites and a successful vaccine against asexual blood stage malaria may be based on this protein. Immunization with native or recombinant forms of MSP-1 has partially protected *Aotus* and *Saimiri* monkeys against *P. falciparum*. Polyclonal and monoclonal antibodies raised against MSP-1 have been shown to exhibit inhibitory effects on the growth of the parasites. Antibodies from human sera, affinity purified to MSP-1, inhibit parasite invasion *in vitro* and vaccinated mice elicits complete protection against a lethal challenge infection (Daly and Long, 1993). Vaccination of *Aotus* monkeys with two forms of MSP-1 provides protection against lethal challenge with *P. falciparum* malaria parasites (Egan *et al.*, 2000).

Protein P126 has been associated with the release of merozoites from mature schizonts, and has been identified as possessing immunogenic properties. The protein is expressed by the parasite in the 32^{nd} and 36^{th} hour of a 42-hour erythrocytic cycle before being stored inside a parasitophorous vacuole. This antigen could be involved in a successful subunit vaccine due to its relatedness to those antigens in several strains of *P. falciparum* and similarity in the amino acids in this region (Bhatia *et al.*, 1987). Monoclonal and polyclonal antibodies specific for p126 are effective in the inhibition of the growth of the parasite *in vitro* (Chulay *et al.*, 1987) and can induce partial protection against challenge in various species of monkeys (Banic *et al.*, 1998).

A 155-kDa protein was identified on the surfaces of infected read blood cells due to the recognition of specific modifications of the erythrocyte in comparison to uninfected cells (Perlman

et al., 1984). The protein was termed Pf155/RESA (ring-infected erythrocyte surface antigen) and is involved in merozoite invasion (Walin *et al.,* 1992). The sequence of the protein contains sets of blocks of tandemly repeated amino acid sequences mainly in the carboxy terminal (C-terminal) and central regions. These are repeats of the peptide (EENVEHDA) and (EENV) (Kabilan *et al.*, 1988). Antibodies raised against RESA have been shown to specifically inhibit the rate of parasite growth *in vitro*, showing the potential of this protein as a possible vaccine candidate (Collins *et al.*, 1986). Large proportions of human anti-Pf155 antibodies are directed against epitopes formed by these sequences, and have shown to be conserved between different isolates of *P. falciparum.* Intact Pf155 protein has lead to the *in vitro* proliferation, interleukin-2 (IL-2) release, and interferon-γ production in T cells specific to this antigen in infection. This antigen, indicating that the protein contains T-helper cell stimulating epitopes, induced T cell-dependent anti-Pf155 antibodies. The identification of these epitopes would contribute towards the development of a subunit vaccine against the merozoite blood stage (Kabilan *et al.*, 1988). Synthetic peptides corresponding to known epitopes of Pf155/RESA induce production of parasite-specific antibodies *in vitro* and were successful in verifying these epitopes (Chougnet *et al.*, 1991). This protein was at first believed to be associated with the micronemes and rhoptries of merozoites but has now been localised to dense granules in the apical end of merozoites by immunofluorescent techniques. During the process of erythrocyte invasion, the protein is expressed, released and transported to the red blood cell membrane due to interactions with the cytoskeleton structures, functioning in cell invasion by the parasite (Ruangjirachuporn, 1991).

A 96-kDa protein associated with the surface membrane of malaria infected mouse erythrocytes was identified to produce protective properties in *P. chabaudi.* Monoclonal antibodies against Pc96 identified cross-reacting antigens in *P. falciparum* (155 kDa), *P. vivax* (222kDa) and *P. cynomolgi* (200 kDa). A specific monoclonal antibody against the protein Pch105 recognises Pc96 (Wanidworanun *et al.,* 1989). Topographical location of Pc96 was further studied by immunofluorescent microscopy verifying that the protein is located in the surface membrane of infected erythrocytes. Further studies involved the elucidation of epitopes by mapping studies, identifying 3 distinct binding sites for the monoclonal antibodies. The fact that four different antigens in different species of malaria share the same, or similar topographical locations, along with at least one cross-reactive epitope indicates that the epitope is included in a conserved and functional region of the protein specific for the biology of the parasite. Pf155 and Pv222 share a cross-reacting epitope with Pc96, suggesting the potential of Pc96 as a vaccine candidate (Wanidworanun *et al.,* 1989). The cross reactivity with Pf155/RESA suggests that the *P. falciparum* protein may contain similar T cell epitopes to Pc96. In view of the protective T cell clones that Pc96 stimulated, producing antibody responses against multiple parasite antigens

elicited in athymic mice challenged with *P. chabaudi adami* (Goldring, 1989), an investigation into similar functions in *P. falciparum* would be worthwhile.

1.4.3 Sexual stage vaccination

The female *Anopheles* mosquito ingests gametocyte parasites in the blood, which develop infection in the midgut, with the formation of zygote, ookinete and oocyst stages (Figure 1.2). A few proteins studied on the surface of gametocytes, have been indicated in there role in blocking parasite development. This has opened up a relatively new area of vaccine research known as transmission blocking immunity (Holder, 1999). The effect of blocking transmission would prevent infection in the gut of the mosquito, and would prevent the spread of disease and drug resistant parasites through human populations (Stowers, 2000). Antibodies raised against antigens expressed during these stages, can block the development of the parasite. Pfs230 and Pfs48/45 are proteins identified to be expressed in the gametocyte stage of *P. falciparum*, and have been unsuccessful in the initiation of immune responses. The main objectives of a transmission blocking vaccine, involves an antigen that induces significant amount of antibodies in the host. Upon ingestion of the blood by the mosquito, the antibodies disrupt the development of the parasite in the gut, effectively preventing transmission. The fact that only a fraction of the thousands of parasites ingested must be eliminated, as apposed to the millions, present in the human host system as asexual erythrocytic-stages, allows the more efficient and rapid clearance of parasites (Duffy and Kaslow, 1997).

One of the more promising candidates for a transmission blocking vaccine is Pfs25, an antigen expressed on the surface of *P. falciparum* zygotes. This protein is a cysteine rich antigen of a 25 kDa mass which is made up of four domain-like tandem epidermal growth factors (EGF) which are anchored to the surface of the parasites (Stowers, 2000). Immune responses to zygote surface antigens such as Pfs25 are not induced in the host as they are only expressed in the mosquito vector. Recombinant forms of Pfs25 have been shown to induce strong transmission-blocking immunity in various models studies including rodents and monkeys (Targett, 1995). One such recombinant protein, TBV25H, which was secreted by *Saccharomyces cerevisae* can induce complete transmission-blocking antibodies (Stowers, 2000).

Pfs28 is a 28-kDa protein of the late-stage ookinetes and similarly to Pfs25, consists of EGF like domains. Pfs28 is also considered a major potential vaccine. Pgs28 is the avian parasite (*P. gallinaceum*) analog of Pfs28 and was used to show that polyclonal antibodies against Pgs28 block

the transmission of parasites by completely inhibiting the transformation from ookinetes to oocysts *in vitro*, in a similar mechanism to that of Pfs25. Due to the fact that the two antigens both block parasite development in the same stages, it was proposed that the two could act together as a transmission blocking vaccine. It was subsequently found that pooling the two antigens had little effect as a vaccine, but the fusion of the two to form a recombinant chimeric protein produced a potent transmission-blocking vaccine (Gozar *et al.*, 1998).

A different approach was taken whereby antigens from the mosquito midgut, involved in interactions with the parasite, were targeted. Polyclonal and monoclonal antibodies against the midgut inhibited the development of the parasite. *Plasmodium falciparum* and *Plasmodium vivax* were prevented from transport across the midgut reducing infection. The survival of the mosquito was also affected by a reduction in egg laying capacity, across a wide range of *Anopheles* species (Lal *et al.*, 2001).

1.4.4 Synthetic multivalent vaccines

The development of multivalent vaccines involves the hybridisation of different proteins from a number of different asexual or sexual blood stage antigens for the effective protection against infection. These antigens combine their immunogenic properties and each component induces a particular partial protective immune response, that when combined, protects a range of host populations (Holder, 1999). The antigen must contain properties that initiate effective immune response, taking into account factors such as antigenic variation. The delivery mechanisms used must also serve to present the antigens in a form that stimulates the required components of the immune system. As in the case of recombinant proteins, the development of better adjuvants for human use is a priority, and much success has been seen in the field of DNA vaccines (Holder, 1999). Of equal importance is much-needed research into the development of accurate assays for determining the nature of the protective immunity. A number of these have shown success in the evaluation of immune responses in humans exposed to natural infection such as the inhibition of sporozoite invasion assay. This has allowed the determination of the extent of hepatocyte invasion in the presence of blocking antibodies (Holder, 1999).

One of the most promising multivalent vaccines was the synthetic polypeptide SPf66 vaccine. This contained antigens based on pre-erythrocyte and the blood stages of *Plasmodium falciparum* This chimeric protein was one of the first multivalent vaccines of this type to be tested for it potential as a human vaccine. The monomer unit making up the vaccine is a chemically synthesised peptide of

45 amino acids containing sequences from three separately expressed asexual-blood-stage antigens of 83, 55 kDa and 35 kDa, linked by (PNANP) motifs derived from the circumsporozoite protein of *P. falciparum* (Alonso *et al.*, 1994).

Shi *et al.* (1999) have designed a recombinant protein coding for short peptide sequences from the sporozoite antigens CSP and SSP2, LSA-1, and the merozoite antigens MSP-1 and MSP2. RAP1 and AMA1 (rhoptry proteins), EBA175 (microneme protein) and a gametocyte protein (Pfg27) was also included into the structure of the vaccine. In total the construct is made up of twelve B cell epitopes, six T cell epitopes and three cytotoxic T lymphocyte epitopes derived from a total of nine *P. falciparum* antigens. The antibodies were raised against several constituent peptides in sporozoites, liver stage, blood stage and gametocytes, and native proteins in the blood stage, as determined by immunoelectron microscopy (Shi et al., 1999). The antibodies raised against this vaccine were successful in inhibiting sporozoite invasion and antibody-dependent cell inhibition assays, but on the whole were inactive in transmission blocking assays (Holder, 1999).

A potential universal T cell epitope for use in a malaria vaccine must ideally be capable of eliciting humoral and cell-mediated immunity in individuals of diverse HLA types, ie. a large amount, if not all, MHC class II molecules. A recent study involved the use of a synthetic polyoxime malaria vaccine containing *Plasmodium falciparum* B cell and universal T cell epitopes, eliciting immune responses throughout a diversity of HLA types (Nardin *et al.*, 2001). This vaccine, termed $(TIB)_9MAP$ contains the $(NANP)_3$ B cell epitope synthesised in tandem with a T cell epitope (T1) defined using CD4+ T cells of a *P. falciparum* sporozoite-immunised volunteer. The single tri-epitope peptide is linked at four oxime bonds to lysine residues forming the structure used in these experiments (Nardin *et al.*, 2001).

Through most of the multivalent vaccine candidates studied, very few have elicited powerful protection to humans against infection. Whether these recombinant multivalent approaches are the solution, or single antigen vaccines remains to be seen. Knowledge of the immune system involved in parasitic infection has increased significantly over the years and has shown that cell mediated responses play a role. Whether or not T cells are solely responsible for antibody responses to potential antigens, they are being more carefully considered for inclusion in vaccine constructs. The problem of genetic variance between the expressed antigens in heterogenous parasites and the diversity of HLA haplotypes in human populations requires the need for a universal epitope, common to many species, to be used in a vaccine.

1.4.5 Nucleic acid vaccines

DNA vaccines are seen as third generation vaccines, and are based purely on the genetic features of the parasite. Protective immunity resulting from DNA vaccination was first achieved in 1993, during studies involving the influenza virus, in particular the influenza A nucleoprotein which provided protection in mice (Ulmer et al., 1993). DNA vaccination involves the delivery of a plasmid bearing the pathogen DNA and a strong eukaryotic promoter into an organism used for immunization, followed by endogenous expression by the host and the induction of an immune response against the protein. The advantages to this system are that the plasmid DNA persists in the tissue for long periods of time, enabling the control of immunization times. The proteins expressed by the DNA vaccine plasmids, closely resemble eukaryotic structures, as apposed to recombinant proteins synthesised in prokaryotic expression systems such as *E. coli* (Alarcon et al., 1999).

The first experiments involving DNA vaccines against malaria were those conducted by Sedegah et al. (1994) where a plasmid encoding the circumsporozoite protein was used in attempts to immunise mice against lethal *P. yoelii* parasites. Although the sera showed limited inhibitory activity against the parasites, the titres of antibodies were increased after immunization. It was found that only a fraction of the antibodies raised against CSP were specific to previously defined antigens, accounting for the lack of sufficient antibodies capable of inhibiting the parasite. Despite this, there was evidence for a cell-mediated response, as a high percentage of mice were protected against infection, due to the action of the T cells on the infected liver cells (Sedegah *et al.*, 1994). This indicates that DNA vaccines can effectively be used for the generation of protective cell mediated responses, providing the DNA encodes previously characterised T cell epitopes such as the CSP. They can also be used in a multivalent format, expressing many antigens at the same time, and overcoming genetic restrictions imposed by the host.

A tool has been developed known as expression library immunization (ELI). This has been used in conjunction with DNA vaccine based methods of investigating protective features in malaria infection. It was shown that mice could be protected against *Mycoplasma pulmonis* infection with cloned genomic DNA, represented by pools of thousand of plasmids (Barry et al., 1995). This technique was adopted for studies on the mouse malaria model *Plasmodium chabaudi adami*, showing that 63% protection was achieved in challenged mice, previously vaccinated with plasmid pools. This allows the successful screening of the malaria genome, to identify possible sequences that contain protective features in infection (Smooker *et al.*, 2000).

1.5 Objectives of this study

Pc96, a protein associated with the membrane of *Plasmodium chabaudi* infected mice erythrocytes, showed several immunogenic and protective properties (Wanidworanun et al., 1989). Mice immunized with Pc96 were partially protected against further infection, indicating the potential of the antigen in the development of a vaccine. Monoclonal antibodies against Pc96 were used to identify cross-reacting antigens in other malaria species, identifying the Pf155 RESA antigen previously studied (Perlmann, 1984). However, the availability of sequence data representing both *P. yoelii* and *P. falciparum* genomes (PlasmoDB) provides a new method of identifying homologous proteins to Pc96 in these species, possibly containing similar epitopes. Pc96 was identified to contain certain T cell epitopes, using a specific protective T cell clone that previously was shown to provide adoptive protection in mice challenged with *P. chabaudi adami* (Brake et al., 1988). Antibodies raised against Pc96 identified proteins that induced the proliferation of the protective T cell clone.

The presence of homologous proteins in different malaria species was evaluated by the screening of protein databases with the protein sequence of Pc96 (Chapter 3). These searches were also used to determine the structure and possible function of the proteins. In this study a 178 kDa protein in *Plasmodium yoelii yoelii* (Pyy178) and a 403 kDa *Plasmodium falciparum* protein (Pf403) was identified using the sequence data from these genomes that contain regions of extensive similarity to the Pc96 antigen. The presence of potential T cell epitopes was evaluated, in order to determine possible involvement in similar protection to that provided by Pc96. The identification of these proteins and the relative sequence comparisons, are described in Chapter 3.

The molecular cloning of *Plasmodium falciparum* open reading frames has allowed malaria proteins to be studied effectively by the construction of recombinant proteins. In order to further study these proteins, their immunological significance, and respective functions in the malaria parasite, the proteins must be expressed, to allow further biochemical experimentation. The *E. coli* cloning systems pGEX and pMAL have shown success in several studies involving recombinant proteins. Due to the homology of this region in Pc96, Pyy178 and Pf403, as determined by multiple sequence alignments (ClustalW), it is highly likely that the region represents either a functional or structural domain, conserved between multiple species of malaria. The T cell epitopes identified in Chapter 3 to be highly conserved in these three proteins, indicate that further cloning and expression of this region, would be worthwhile in identifying the region and its possible involvement in protection against *Plasmodium falciparum* in humans, similarly to that seen with the Pc96 protein in *Plasmodium chabaudi adami* infection. Chapter 4 describes the

design of a strategy for cloning the region of Pf403 that shares this homology with the Pc96 antigen. Ultimately, a series of overlapping regions in the Pf403 protein in *Plasmodium falciparum* will be cloned and expressed in *E. coli* using these expression vectors. The first, and most relevant of these regions was expressed in *E. coli* using the pMAL expression vector (Chapter 5).

CHAPTER 2
MATERIALS AND METHODS

Chemicals used in this study were in most cases manufactured by Roche (Mannheim, Germany), Merck (Darmstadt, Germany) or Sigma (St Louis, USA). The nitrocellulose used for western blotting (Hybond™-C, 0.45 micron), and the T4 DNA Ligase, was obtained from Amersham Lifesciences (Cleveland, USA). The nylon membrane used for binding nucleic acids was from Osmonics Inc (Minnetonka, USA). Ampicillin antibiotic, *Taq* Polymerase was purchased from Roche (Mannheim, Germany). The Agarase I enzyme for degrading low-melting point agarose was from Sigma (St Louis, USA). Restriction enzymes used in this study were from Promega (Madison, USA) as well as the Klenow enzyme (DNA polymerase I large fragment). The pMOS*Blue* blunt ended cloning kit was purchased from Amersham Biosciences (Buckinghamshire, UK).

The Digoxigenin (DIG) Oligonucleotide 3'-end Labelling Kit with the terminal transferase enzyme, dATP's, the DIG-ddUTP molecule, and the anti-Digoxigenin antibody with alkaline phosphatase conjugate for detection, was from Roche (Mannheim, Germany). Purified plasmid DNA was isolated using the Nucleobond® AX kit. This was purchased from Macherey-Nagel (Duren, Germany).

The oligonucleotide primers for PCR were all synthesised at the University of Cape Town, by the Biochemistry Department synthetic laboratory. Isolated *Plasmodium falciparum* (3D7) genomic DNA was provided by Professor Christian Doerig (INSERM, Paris).

DNA sequencing was performed by Mariaan Ponsie in the Molecular Biology Unit, using the ABI PRISM® 3100 Genetic Analyser, with the BigDye® Terminator v3.1 Cycle Sequencing Kit.

Distilled water (dist.H_2O) was obtained from a Milli-RO® 15 purification system and deionised water was obtained from a Milli-Q plus Ultra-Pure Water System (Millipore, USA).

2.1 Bioinformatic methods

2.1.1 Database screening for proteins similar to Pc96

The nucleotide sequence of Pc96 was translated into six reading frames using the online Translate software (DNA to protein) at ExPaSy (http://us.expasy.org). The predicted amino acid sequence of Pc96 was used in a series of database searches with the online BLAST (Basic Local Alignment Search Tool) version 2.2.6. The database used for BLAST searches was on the NCBI (National Center for Biotechnology Information) website (http://www.ncbi.nlm.nih.gov/BLAST). This database is known as the nr database, containing links to GenBank and many other sources combining a huge range of different biological sequence data. This method is a highly sensitive database-searching tool, for the identification of proteins and analysis of sequence similarity. Comparisons of the query sequence and the sequences in the databases provide a measure of similarity or percent identity. BLAST reports results as an output of statistics and scores based on the matches it has found, involving algorithms responsible for locating similar strings within vast amounts of data at high speed negating the need to systematically compare sequences (Altschul *et al.,* 1997).

The E value depends on the score of the alignments based upon factors such as the lengths of database searched and properties of the query sequence. The E values for the hits allow interpretation of results. Scores below 0.01 occur by chance rarely, and so is a reasonable indication of homology. Even higher scores seen (1e-50) provide extremely high confidence that the query is comparible to that of selected targets in the database searched. PSI-BLAST (Position Specific Iterated - Basic Local Alignment Search Tool) was used with the Pc96 query sequence. This relies on gapped local alignments, building and iteratively refining patterns and profiles constructed from the alignments made. These profiles are used to re-search the database, with additional sequences identified in each cycle. This leads to the identification of sequences that may share homology with the query sequence but did not achieve a match using the original BLAST searching algorithm. Despite its increased sensitivity, care must be taken in using this algorithm for the annotation of

unknown protein sequences, as the profile generated can often mislead the alignments and provide false homologous hits (Baxevanis, 1998).

2.1.2 Prediction of T cell epitopes in Pc96

In order to evaluate the presence of any putative T cell epitopes on Pc96, online epitope prediction software was used. There are two programs available on the World Wide Web that allow unrestricted predictions, BIMAS (http://www-bimas.dcrt.nih.gov/molbio/hla_bind/) (Parker *et al.*, 1994) and SYFPEITHI (http://www.syfpeithi.de) (Rammensee *et al.*, 1999). These programs involved the recognition of T cell epitopes, due to chemically related amino acids in specific positions, corresponding to that of known MHC ligands. This allowed for the definition of a peptide 'motif' for every MHC allele (Falk *et al.*, 1991). SYFPEITHI uses motif matrices previously obtained and deduced from refined motifs based on single peptide analysis exclusively of natural ligands. Binders for various mouse, human and rat MHC class I molecules are compared according to the presence of primary and secondary anchor amino acids and other chemically similar residues. The predicted amino acid Pc96 sequence was used in the epitope prediction program SYFPEITHI, for naturally processed MHC class II epitopes, and I for a variety of HLA-types (Rammensee *et al.*, 1999).

2.1.3 Multiple sequence alignments of Pc96 and similar proteins using CLUSTALW

The CLUSTALW algorithm provided by the European Bioinformatics Institute (EBI) on the WWW has been widely used for the comparison of conserved amino acid sequences either between homologous proteins or domains, or between different species (Thompson *et al.*, 1996). The algorithm is based on the idea of progressive alignment, creating a series of pairwise alignments using the query sequence. A distance matrix is calculated based on these initial alignments, reflecting the relatedness of sequences. Pc96 and proteins matched using BLAST, were compared

in multiple and pairwise sequence alignments to investigate homology and identify regions within the sequences that have regions of similarity.

2.1.4 Comparison of structural features of Pc96, Pyy178, Pf403

The program PREDICT7 version 1.2 (Cármenes *et al.*, 1989) was used to determine structural properties of Pc96 and related proteins such as hydropathy, flexibility, surface probability, and antigenicity. Due to the fact that hydrophobic residues are located within the globular protein structure, and hydrophilic on the outside, exposed and interacting with water, an algorithm was designed by Hopp and Woods (1981) to represent this feature. Each amino acid is assigned a numerical value of hydrophilicity, which is averaged along the sequence of the protein, producing a plot of these trends. Flexibility, or sequence mobility, provides structural clues as to the nature and movement within the protein. Functional activity can often be assigned to these conformational variations. Regions of highest mobility exist often on the most highly accessible segments, usually on the surface of the molecule (Karplus and Schulz, 1985). Surface probability measures the proteins contact to solvents. Features of the protein such as buried residues can determine the surface regions of globular proteins (Janin, 1979). Antigenicity plots are based on the amino acid composition of regions of the protein and show a probability of the region being antigenic, due to previous characterisation of known antigens (Welling *et al.*, 1985). Due to the fact that these plots generate data specific to structural aspects of the protein, they were used primarily to compare homologous regions of interest in the identified protein homologues.

2.1.5 Identification of putative motifs and patterns from the PROSITE profile library and generation of similar 3D structures using 3DPSSM

The PROSITE analysis program was used with the query protein sequence to scan a profile library (Sigrist et al., 2002). The PROSITE profile library used for searching is an ExPASy database that is a collection of biologically significant motifs or sequence patterns. Due to the fact that some proteins are large and contain many domains and

functions, protein functional classifications are based mostly on domains rather than complete proteins. Pc96 and other similar proteins were used to screen the PROSITE database. Certain functional domains contain sets of conserved regions of amino acids, and the occurrence and positions of these regions along the length of a protein sequence, in relation to certain alignment features can be used to create a signature of the domain, allowing the identification of these patterns within query sequences. These signatures are a reflection of the 3-dimensional conformation of the domain, and can be used to assign function to regions of un-characterised proteins, providing the profile of that region is a significant match.

The 3-dimensional structure and function of protein sequences was also determined by analysis with the 3D-PSSM (Three Dimensional - Position-Specific Scoring Matrix) server. The server contains a database with known protein structures, which are compared to the query sequence and scored on a basis of compatibility. Factors such as secondary structure elements, and probability of occupying various states of hydrophilicity, in relation to overall shape are analysed (Kelley *et al.*, 2000). Entries in the PROSITE database make up the so-called BLOCKS database, used to identify families of proteins. This is not a comparison of the sequence itself, due to the fact that homologous proteins may not share sequence similarity. A block relates to motifs, or conserved stretches of amino acids conferring specific function to the structure of a protein. Individual proteins may contain several blocks in similar combination, corresponding to specific structure or function. The query sequence is aligned against all blocks in the database at all available positions. A score is derived from alignments using the position-specific scoring matrix (PSSM), taking into account matches at given positions and the probabilities of amino acids occupying specific positions in the block (Baxevanis, 1998).

2.2 Immunochemical techniques

2.2.1 Western blotting

This technique relied on the use of specific antibodies for the identification and characterisation of proteins separated by sodium dodecyl sulfate-polyacrylamide gel

electrophoresis (SDS-PAGE) and transferred electrophoretically to nitrocellulose. Nitrocellulose is a matrix with a high affinity for protein. The method used in this study was that of Towbin *et al.* (1979), with minor modifications. The transfer buffer included methanol to enhance the binding of protein-SDS complexes to the membrane (Goodenham, 1984). The unoccupied sites on the membrane were blocked with non-fat milk. Antigens transferred onto the nitrocellulose binds to primary antibody, which in turn binds to secondary antibody specific for the primary antibody.

The secondary antibody contains a detection system with labelled enzyme horseraddish peroxidase (HRPO). HRPO catalysed a reaction involving a chromogenic substrate, forming a visual, insoluble precipitate. These effectively labelled specific bands bound to the nitrocellulose.

2.2.1.1 Materials

10% (m/v) SDS. SDS (10 g) was dissolved in 100 ml dist. H2O with gentle heating if necessary.

Blotting buffer. Tris (27.23 g) and glycine (64.8 g) were dissolved in 3.5 litres of dist.H_20 and methanol (900 ml) was added. The volume was made up to 4.5 litres in a large beaker. Prior to use, 10% (m/v) SDS was added (4.5 ml).

Tris buffered saline (TBS; 20 mM Tris, 200 mM NaCl, pH 7.4). Tris (2.42g) and NaCl (11.69 g) were dissolved in 950 ml of dist.H_2O, adjusted to pH 7.4 with HCl, and made up to 1 litre.

1% (m/v) Ponceau S. Ponceau S (0.1 g) was dissolved in 1% (v/v) glacial acetic acid (100 ml).

Blocking solution [5% (m/v) non-fat milk powder]. Low fat milk powder (5 g) was dissolved in TBS (100 ml).

0.5% (m/v) BSA-TBS. Bovine serum albumin (0.1 g) was dissolved in TBS (20 ml).

4-chloro-1-naphthol substrate solution [0.06% (m/v) 4-chloro-1-naphthol, 0.0015% (v/v) H_2O_2]. 4-chloro-1-napthol (0.03 g) was dissolved in methanol (10 ml). 2 ml of this solution was diluted to 10 ml with TBS, with the addition of 30% hydrogen peroxide (4 µl).

2.2.1.2 Method

After SDS-PAGE had been performed on duplicate gels, one gel was used for staining as a reference for the blot, and the other was used for blotting onto nitrocellulose. The nitrocellulose was cut to a size similar to that of the SDS-PAGE gel and floated on blotting buffer to soak. Six pieces of Whatman No. 4 filter paper were cut slightly larger than the nitrocellulose. A sandwich was constructed (Figure 2.1) with blotting buffer to avoid the entrapment of air bubbles that can disturb the blotting pattern.

Two white Scotch-Brite pads, filter paper and the nitrocellulose and gel were positioned as seen in Figure 2.1. The sandwich was placed into a plastic support, into the western blotting apparatus and immersed in cold blotting buffer. The nitrocellulose was kept on the anodal side of the gel and blotting was achieved at 200 mA for 16 h. The blotting tank was placed in a larger container of ice. Blotting buffer was stirred by a magnetic stirrer throughout the procedure to keep the distribution of cooling even.

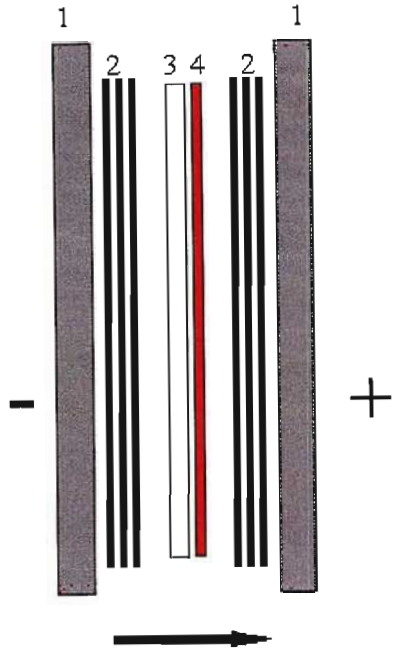


Figure 2.1 Arrangement of nitrocellulose sandwich prior to electro-transfer of protein. 1 - Scotch Brite pads; 2 - 3 x filter paper; 3 - SDS-PAGE gel; 4 - nitrocellulose membrane.

The gel was removed and stained with Ponsceau S (10 ml) in a glass petri dish to ensure that protein transfer had occurred. The nitrocellulose was removed and rinsed with dist H_2O until the dye is removed. The nitrocellulose was air-dried for 1.5 h. The membrane was then blocked for 1 h with 5% (m/v) low fat milk powder in TBS. It was then washed in TBS (3 x 5 min), followed by a 2 h incubation with primary antibody in 0.5% BSA-TBS. A washing step was repeated (3 x 5 min) and followed by incubation in secondary antibody with HRPO enzyme conjugate in 0.5% BSA-TBS for 1 h. 4-chloro-1-naphthol substrate solution (10 ml) was used as a substrate for the development of the chromogenic reaction.

2.2.2 Screening for recombinants using DIG nucleic acid hybridisation and detection

The Roche (Mannheim, Germany) Digoxigenin (DIG) Oligonucleotide 3'-end Labelling Kit was used to label oligonucleotide primers with the modified nucleotide analogue, digoxigenin-labelled dideoxyuridine-triphosphate (DIG-ddUTP), for the detection and screening of recombinant plasmid DNA. The enzyme terminal transferase is used to enzymatically label the 3'-OH termini of the oligonucleotide primer, incorporating a single DIG-ddUTP. The enzyme depends on Co^{2+} ions for its function and incorporates the DIG-ddUTP as a modified substrate. Only a single nucleotide is added to the 3' terminus of the oligonucleotide due to the chain-termination effect of this analogue.

The labelled oligonucleotides can then be detected by the specific high affinity binding of an antibody/alkaline phosphatase conjugate to the digoxigenin molecule. The alkaline phosphatase is then used to catalyse the indicator reaction with BCIP and NBT. The reaction products are insoluble purple-blue precipitates, which directly adhere to the nylon membrane. Figure 2.2 shows the structure of DIG-ddUTP (Schmitz *et al.*, 1991).

This technique can be applied to many fields of molecular biology, in particularly diagnostic and forensic applications. Labelled probes can be used to screen complex gene libraries, restriction site mapping of cloned DNA, gel shift assays and various other *in situ* hybridisation techniques. For the purposes of this study, the oligonucleotide primers used for polymerase chain reaction (PCR) were labelled and used to probe DNA for detection of the PCR product in the form of insert or recombinant DNA (Schmitz *et al.*, 1991).

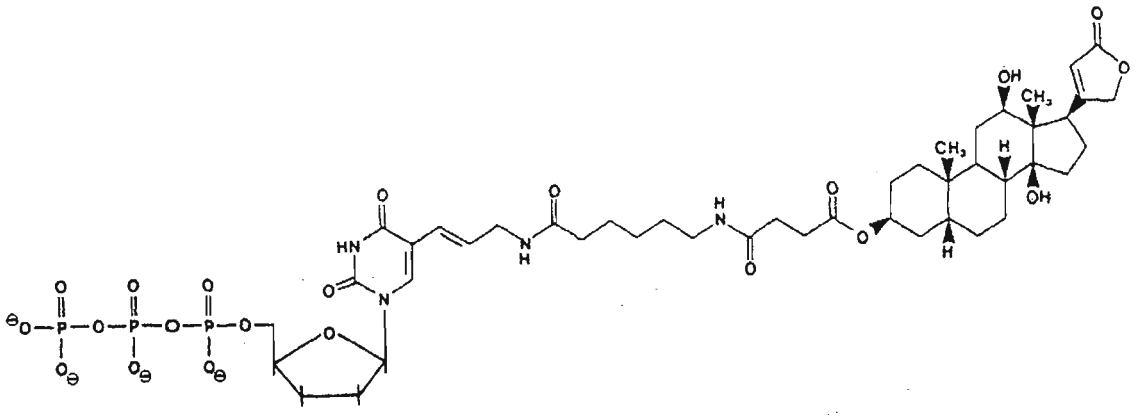


Figure 2.2 Structure of digoxigenin-labelled dideoxyuridine-triphosphate (DIG-ddUTP) (Schmitz *et al.*, 1991).

2.2.2.1 Materials

5 x Reaction buffer (1M Potassium cacodylate; 0.125 M Tris-Cl; 125 mg/ml Bovine Serum Albumin, pH6.6).

CoCl2 solution (25 mM $CoCl_2$).

DIG-ddUTP solution (1 mM DIG-dUTP).

dATP solution (10 mM dATP)

Terminal transferase (50 U/µl in 0.2 M Pottasium cacodylate, 1 m M EDTA, 200 mM KCl; 0.2 mg/ml Bovine serum albumin; 50% glycerol [v/v]; pH 6.5)

10% (m/v) SDS. SDS (10 g) was dissolved in 100 ml dist H2O with gentle heating if necessary

100 x Denhardt's solution (2% Ficoll; 2% polyvinylpyrrolidone; 2% Bovine serum albumin) Store at -20°C in 10 ml aliquots.

20 x SSC (3 M NaCl; 300 mM sodium citrate; pH 7). NaCl (175.3 g) and Na Citrate.$2H_2O$ (88.23 g) was dissolved in dist H2O (900 ml). The pH was titrated to 7.0 with HCl and the volume was made up to 1 litre. The solution was autoclaved to sterilise prior to use.

2 x SSC (0.1% SDS). 10 % SDS (1 ml) was combined with 20 x SSC (10 ml) and made up to 100 ml with dist.H2O.

0.2 x SSC (0.1% SDS) 10% SDS (1 ml) was combined with 20 x SSC (1 ml) and made up to 100 ml with dist.H2O.

Hybridisation solution (6 x SSC; 5 x Denhardt's solution; 0.5% SDS). 20 x SSC (30 ml), 100 x Denhardt's solution (5 ml) and10% SDS (5 ml) were diluted with dist H2O (60 ml).

Buffer 1 (0.1 M Tris-Cl; 0.15 M NaCl; pH 7.5). Tris (12.11 g) and NaCl (8.76 g) was dissolved in dist H2O (900 ml). The pH was adjusted to 7.5 with HCl and the made up to 1 litre with dist.H_2O.

Buffer 2 (1% Blocking solution). The blocking reagent (vial 6 of the DIG Nucleic Acid Detection Kit) was dissolved to a final concentration of 1 % (w/v) in buffer 1, with gentle heating if necessary.

Buffer 3 (0.1 M Tris-Cl; 0.1 M NaCl; 50 mM $MgCl_2$; pH 9.5).
Tris (12.11 g), NaCl (5.84 g), and $MgCl_2$ (4.76 g) were dissolved in dist H2O (900 ml). The pH was adjusted to 9.5 with HCl and made up to 1 litre with dist.H_2O.

Anti-Digoxigenin-AP, Fab fragments (Roche, Germany).

Colour substrate solution nitroblue tetrazolium salt (NBT) solution (45 µl) and 5-bromo-4-chloro-3-indolyl phosphate, toluidinium salt (BCIP) solution (35 µl) was combined with buffer 3 (10 ml).

2.2.2.2 Method

Crude preparations of DNA from each colony were obtained using the same method as per Section for colony PCR screening (Section 2.3.7). Colonies resulting from the transformation reactions were streaked onto sterile LB agar plates previously supplemented with ampicillin (100 µg/ml) and numbered. After sufficient overnight growth at 37°C, a portion of each colony was transferred to sterile dist H2O (50 µl) and vortexed to resuspend the colony. The samples were boiled for 5 min to lyse the cells and centrifuged briefly to remove contaminating cell debris. The supernatant, containing both chromosomal and plasmid DNA was transferred (30 µl) to a sterile microfuge tube.

Reaction components were combined in a sterile microfuge tube as described on Table 2.1, for the enzymatic transfer of the digoxigenin compound to the 3' terminus of the oligonucleotide.

Table 2.1 Reaction components for oligonucleotide tailing with DIG-ddUTP

Reaction component	Volume (µl)
5 x Reaction buffer	4
$CoCl_2$ solution	4
DIG-ddUTP solution	1
Terminal transferase	1
5 µM Oligonucleotide primer	2
dATP solution	1
Sterile dist.H_2O	7
Total	20

The above reactions (Table 2.1) were mixed and incubated at 37°C for 15-30 min and placed on ice. Hybridisation solution (100 µl) was added and the mixture was stored at -20°C. DNA samples were heat at 95-100°C for 15 min to denature the DNA. These were immediately dotted onto a nylon membrane onto pre-marked pencil dots. It was important not to let the sample cool before dotting, as they re-anneal and hybridisation is not possible. The primers used to make the probe were dotted as a positive control, and vector (pGEX, no insert) was dotted as a negative control, as the probe is not specific to vector DNA. The positive control for the detection procedure is DIG-labelled molecular weight markers. The membrane was then baked at 110°C for 15 min to bind the DNA to the nylon membrane.

Hybridisation solution (10 ml) was added as a pre-hybridisation step and the filter was incubated for 2 h at 37°C. The solution was removed and the probe (60 µl) and hybridisation solution (10 ml) was added and incubated at 37°C overnight. The filter was removed and washed at 37°C with 200 ml 2 x SSC, 0.1% SDS (2 x 15 min) then 200 ml 0.1 x SSC, 0.1% SDS (2 x 15 min). The filter was air-dried for 5 min and was washed for 1 min in buffer 1 (80 ml). The blocking step involved incubation at room temperature with shaking for 30 min in buffer 2 (100 ml). After this step the filter was washed briefly in buffer 1 (100 ml). The anti-digoxigenin-AP conjugate was diluted (1:5000) in buffer 2 (4 µl in 20 ml). The filter was incubated for 30 min with the diluted antibody conjugate (20 ml). Non-bound antibody conjugate was removed by washing (2 x 15 min) in buffer 1 (100 ml). The membrane was equilibrated for 2 min in buffer 3 (20 ml). The filter was then incubated with the colour substrate solution (10 ml) in the dark until the colour reaction developed.

2.3 Molecular biology techniques

2.3.1 Quantitation of DNA

Spectrophotometry was used for determination of the amount of DNA in solution. Readings were taken at 260 nm and 280nm. The reading at 260nm allows the concentration of the DNA to be calculated. An OD of 1 corresponds to50 µg/ml for

double stranded DNA. A ratio of the readings at 260nm and 280nm provides an estimate of the purity of the sample. If OD_{260}/OD_{280} values are significantly less that 1.8, quantitation of nucleic acid will be inaccurate (Sambrook *et al.*, 1989). DNA sample (5 µl) was added to sterile TE buffer (495 µl) in a 0.5 ml quartz cuvette. TE buffer (0.5 ml) was used as a blank and readings were taken at 260 nm and 280 nm.

DNA concentration (µg/ml) = A260 × 100 (dilution factor) × 50 µg/ml (OD of 1)

2.3.2 Agarose gel electrophoresis of DNA

Agarose gel electrophoresis of DNA is a technique commonly used in molecular biology for the study and analysis of nucleic acids. Agarose gels can be cast by melting agarose in a buffer and allowing the solution to cool causing the compound to form a matrix of uniform density. An electric field is applied across the gel and negatively charged DNA molecules migrate towards the anode. The gel matrix acts like a sieve, allowing fragments to migrate at a rate dependent on the length of the molecule. Large molecules move through the gel with less efficiency than small molecules due to the frictional hindrance provided by the agarose matrix (Sambrook *et al.*, 1989).

The fluorescent dye, ethidium bromide is used for visualisation of the DNA fragments. The dye intercalates between the nucleotide bases in DNA and displays an increased fluorescent activity compared to when in free solution. In this way, trace amounts of DNA can be detected by including ethidium bromide in the gel and electrophoresis buffer (Sambrook *et al.*, 1989).

2.3.2.1 Materials

50 x TAE buffer (40 mM Tris-acetate, 1 mM EDTA, pH 8.0). Tris was dissolved in 800 ml dist. H_2O. Glacial acetic acid (57.1 ml) was added, and made up to 900 mls with dist.H_2O. Ethylenediaminetetraacetic acid (EDTA) (18.612 g) was dissolved in 80 ml dist. H_2O, the pH was adjusted to 8.0 with glacial acetic acid and the solution

was made up to 100 ml with dist. H_2O. 0.5 M EDTA, pH 8.0 (100 ml) was combined with tris-acetate solution (900 ml) to make a 1 l solution.

1 x TAE working solution. 50 x TAE buffer (20 ml) was diluted to 1 l with dist. H_2O.

1% (m/v) agarose. Agarose (0.3 g) was placed in a Erlenmeyer flask with 1 x TAE buffer (30 ml). The solution was heated in a microwave oven, for about 1.5 min with occasional mixing until all the agarose had visibly melted. Different % gels can be used for different ranges of resolution (Table 2).

Gel-loading buffer (1 mM EDTA, 0.5% (m/v) bromophenol blue, 50 % (v/v) glycerol). EDTA (0.007 g) and bromophenol blue (0.1 g) were dissolved in sterile dist. H_2O (10 ml) and glycerol (10 ml) was added and thoroughly mixed.

1% (m/v) Ethidium bromide. Ethidium bromide (0.2 g) was dissolved in dist. H_2O (20 ml) and stored in a glass bottle wrapped in aluminium foil.

Tris-EDTA buffer (TE buffer; 10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Tris (0.121 g) and
EDTA (0.037 g) were dissolved in 90 ml dist. H_2O, adjusted to pH 8.0 with HCl and made up to 100 ml with dist. H_2O.

2.3.2.2 Method

A dry perspex-casting tray (100 x 65 mm) was cleaned, dried and masking tape was used to seal the open ends ensuring a closed mould for setting the agarose. A perspex 10 well comb was positioned in the casting tray. Once the molten agarose had cooled to approximately 55°C, ethidium bromide (1 µl) was added and thoroughly mixed in. The warm agarose solution was poured into the casting tray, making sure that no air bubbles were present. After approximately 20 min the masking tape and the perspex well comb were removed and the casting gel with the set agarose was positioned in an electrophoresis tank. 1 x TAE buffer was added to submerge the gel. Gel-loading

buffer (1µl / 5µl sample) was added to the DNA samples prior to loading. The glycerol in the buffer ensures that the DNA sample loaded is of a higher density than the 1 x TAE buffer and will sink to the bottom of the wells. The bromophenol blue is included so that the progress of the electrophoresis can be visualised.

Electrophoresis was carried out at 100 V until the bromophenol dye moved out of the wells and into the gel followed by 70-80V for 1.0 - 1.5 h, or until the dye front migrates sufficiently through the gel. The casting gel was removed from the electrophoresis tank and examined in an ultraviolet transilluminator. Photography was conducted through a red filter with a Polaroid® type 667 black and white instant films. Table 2.2 shows the various percentage agarose gels that can be used to resolve a wide range of sizes of nucleic acids (Sambrook *et al.*, 1989).

Table 2.2. Concentration of agarose for resolution of DNA sizes (Sambrook *et al.*,1989).

Agarose (%)	Effective range of resolution of DNA (Kb)
0.5	30 to 1
0.7	12 to 0.8
1	10 to 0.5
1.2	7 to 0.4
1.5	3 to 0.2

2.3.3 Restriction endonuclease digestion of DNA

Restriction enzymes are those occurring in bacteria species forming a defence mechanism against foreign DNA. 'Host-controlled restriction' was shown in the early 1950's when some strains of bacteria were immune to bacteriophage infection. Bacteria were shown to produce an enzyme that degrades the phage DNA before

replication and the synthesis of new phage particles occurs. The bacterium DNA is protected by additional methyl groups that block the enzyme action (Brown, 2001).

Purified restriction enzymes enable the molecular biologist to cut DNA molecules in a precise reproducible manner required for DNA analyses and gene cloning. Over 1200 different restriction enzymes have been characterized. There are three classes based on mode of action: type I, II and III. Type I and III have complex behaviour and have a limited role in their use in molecular biology, while Type II endonucleases play a vital role in gene cloning. The enzyme binds specifically to double-stranded DNA at specific sites or particular recognition sequences and cleaves the strand. Endonuclease enzymes are able to break internal phosphodiester bonds within a DNA molecule. Each restriction enzyme recognises particular unique hexanucleotide positions, which are usually palindromic (Sambrook *et al.*, 1989).

2.3.3.1 Method

The procedure is slightly different for each enzyme and was performed as per manufacturer's instructions. DNA (minimum 0.1 µg), 10 x restriction buffer (2 µl) and enzyme (1 U/µg DNA) were made up to 20 µl with sterile deionised H_2O in a sterile microfuge tube. The enzyme should be added last so that it is correctly buffered immediately. The volume of enzyme used cannot exceed 10% of the reaction volume due to presence of glycerol in the enzyme storage buffer which can decrease endonuclease activity. The sample was mixed by gentle finger tapping and pulse-spun if necessary. The tube was incubated at the recommended temperature (usually 37°C) for 2 hours to overnight. The reaction was stopped by heating at 65°C for 10 min to inactivate the restriction enzyme. An aliquot of the reaction can be run on an agarose gel (Section 2.3.2) to check the efficiency of the digestion.

2.3.4 Isolation of plasmid DNA (alkaline lysis)

The aim of this technique is the preparation of sufficient plasmid DNA for use in subsequent experiments. The presence of plasmid DNA in an *E. coli* cell is determined phenotypically by selectively isolating the plasmid DNA from cellular

DNA and RNA and additional cellular components such as protein, which can often contaminate and inhibit further downstream DNA manipulations. A number of different methods for isolating DNA have been developed, the most common of these being the alkaline lysis technique. The first step of plasmid isolation is lysis of the host cells carrying the plasmid. This method utilises sodium hydroxide and sodium dodecyl sulfate (SDS) for breaking open the bacterial cells. At this pH (12-12.6) the chromosomal and linear DNA are denatured. An acidic acetate solution is then used to neutralise the solution, and the chromosomal DNA forms an insoluble aggregate. Along with protein-SDS complexes, these contaminants can be removed by centrifugation. Due to its covalently closed circular nature, the plasmid DNA remains in solution, in the cleared lysate. Further treatment of the solution such as ethanol precipitation can be used to clean and concentrate the plasmid DNA for further experiments such as cloning or sequencing. A scaled down protocol using these techniques is known as the mini-prep technique (Sambrook *et al.*, 1989).

2.3.4.1 Materials

Ampicillin stock (5 mg/ml). Ampicillin (Roche) (50 mg) was dissolved in 10 ml sterile dist.H_2O. The solution was filter-sterilised and stored at 4°C. Dilute 2 ml per 100 ml LB medium (100 ug/ml final concentration).

LB (Luria Bertani) Medium. Bacto-tryptone (10 g), bacto yeast extract (5 g) and NaCl (10 g) was dissolved in dist H2O (1 l) with heating. Media was dispensed into conical flasks, stopped with cottonwool, and wrapped in tin foil. The media was sterilised by autoclaving (1 hour at 121°C).

GTE solution (50 mM glucose, 25 mM Tris.Cl pH 8.0, 10 mM EDTA). Glucose (4,5 g) was dissolved in sterile dist H2O (450 ml) along with Tris (1.51 g) and EDTA (1.86 g). The buffer was adjusted to pH 8.0 with HCl and made to 500 ml with dist.H_2O.

NaOH/SDS (0.2m NaOH, 1% SDS). NaOH (1.6 g) was dissolved in dist H2O (200 ml). 10% SDS (5 ml) was diluted with 0.2M NaOH (45 ml). This solution was prepared just before use.

3M Potassium acetate solution, pH 4.8. Anhydrous potassium acetate (58.89 g) was dissolved in dist H2O (50 ml) combined with glacial acetic acid (50 ml). The mixture was left to dissolve for 1.5 h. The solution was titrated to pH 4.8 with glacial acetic acid and adjusted to 200 ml with dist.H_2O.

2.3.4.2 Method

LB medium (10 ml) containing appropriate antibiotic (eg. 100 µg/ml ampicillin) was inoculated with a single bacterial colony of *E. coli* containing plasmid. The culture was grown overnight at 37°C with vigorous shaking. Cells were pelleted in sterile polypropylene centrifuge tubes, in the desktop Sigma 3K20 centrifuge (7250 x g; 5 min; 4°C) and the supernatant was poured off. The pellet was resuspended in GTE solution (200 µl) and left at room temperature for 5 min. The suspension was transferred to a sterile microfuge tube. RNase A (2 µl) was added. Freshly prepared NaOH/SDS (400 µl) was added to the sample and placed on ice for 5 min. Potassium acetate solution (300 µl) was added mixed thoroughly and the solution was placed on ice for 5 min. A thick white precipitate of cellular debris and chromosomal DNA forms on addition of this buffer. The precipitate was removed by centrifugation (12000 x g; 5 min; 4°C). The clear supernatant (800 µl) was carefully transferred to a fresh sterile microfuge tube. Chilled isopropanol (600 µl) was added to the sample and this was placed at -20°C for 1 h. Plasmid DNA was pelleted by centrifugation (12000 x g; 5 min; 4°C) and the supernatant was removed carefully from the pellet. The plasmid DNA was washed twice with ice cold 70% (v/v) ethanol (0.5 ml) and the pellet was eventually resuspended in sterile TE buffer (50 µl). This preparation is stable at -20°C.

2.3.5 Isolation of plasmid DNA (Nucleobond® AX kit)

Nucleobond® AX is a silica based anion exchange chromotography system which isolates plasmid DNA with extremely high levels of purity. The system is based on the principle behind the alkaline lysis method (Section 2.3.4). Bacterial cells are lysed by a NaOH/SDS solution, which denatures chromosomal DNA under the alkaline conditions. The addition of potassium acetate results in a precipitate of cellular compounds and chromosomal DNA. The plasmid DNA stays in solution and can be purified using the Nucleobond® AX cartridge. The ion-exchange silica is secured between inert filters in polypropylene cartridges allowing for the efficient separation of nucleic acids. Impurities and contaminants are completely removed which potentially inhibit further manipulations such as ligation or restriction digestion. The plasmid DNA binds to the matrix at a low pH and ionic strength, and is eluted when the pH is increased. The kit was used for preparative purposes, where the desired mini-prep needs to be highly pure.

2.3.5.1 Materials

The following materials were supplied with the Nucleobond® AX kit. The plasmid isolation was performed following manufacturers instructions supplied with the kit.

Buffer S1 (50 mM Tris/HCl, 10 mM EDTA, 100 µg RNase A/ml, pH 8.0). The lyophilised RNase supplied with the kit is dissolved in 1 ml buffer S1, then added to the bottle S1. This buffer was stored at 4°C.

Buffer S2 (200 mM NaOH, 1% SDS).

Buffer S3 (2.8 M Potassium acetate, pH 5.1)

Buffer N2 [100 mM Tris, 15% (v/v) ethanol, 900 mM KCl, 0.15% Triton X-100, pH 6.3]. The solution was adjusted with H_3PO_4 to pH 6.3.

Buffer N3 [100 mM Tris, 15% (v/v) ethanol, 1150 mM KCl, pH 6.3]. The solution was adjusted with H_3PO_4 to pH 6.3.

Buffer N5 [100 mM Tris, 15 % (v/v) ethanol, 1000 mM KCl, pH 8.5]. The solution was adjusted with H_3PO_4 to pH 8.5.

Nucleobond® AX polypropylene sterile cartridges.

Nucleobond® AX filters

2.3.5.2 Method

LB medium (30 ml) containing appropriate antibiotic (eg. 100 μg/ml ampicillin) was inoculated with a single bacterial colony of *E. coli* containing plasmid. The culture was grown overnight at 37°C with vigorous shaking. Cells were pelleted in sterile polypropylene centrifuge tubes, in the desktop Sigma 3K20 centrifuge (5000 x g; 10 min; 4°C) and the supernatant was poured off. The pellet was carefully resuspended in buffer S1 (4 ml). Buffer S2 (4 ml) was added and was mixed by inverting the sample 6 - 8 times. The mixture was incubated at room temperature for no longer than 5 min. Chromosomal DNA was precipitated by the addition of buffer S3 (4 ml) with gentle mixing by inversion and incubated on ice for 5 min. The white bacterial lysate precipitate can be removed either by centrifugation (12000 x g; 25 min; 4°C) or by filtration. The folded filter paper supplied with the kit was moistened with sterile dist H2O and the lysate was filtered, collecting the flow-through in a sterile 10 ml polypropylene tube. The cartridge was equilibrated by the addition of buffer N2 (2.5 ml). The filtered lysate was loaded onto the cartridge and allowed to flow through the macroporous silica by gravity. The cartridge was then washed twice with buffer N3 (2 x 5 ml). The plasmid DNA was eluted with buffer N5 (5 ml) and collected in a sterile polypropylene centrifuge tube. The plasmid DNA was precipitated with isopropanol (3.6 ml) and centrifuged immediately (15000 x g; 30 min; 4°C). Before centrifugation, the area where the plasmid DNA pellet is predicted to form was marked with a pencil and the centrifuge tube was placed accordingly in the rotor. This was to ensure that the position of the pellet is known and care can be taken to avoid losing the DNA. The DNA was washed with chilled 70 % (v/v) ethanol (5 ml) and

centrifuged (15000 x g; 10 min; 4°C). The ethanol was carefully removed and the pellet air dried for 10 min and resuspended in water or TE buffer (50 µl).

2.3.6 Purification and concentration of DNA from aqueous solutions

DNA can be effectively precipitated out of a solution using absolute ethanol, in the presence of monovalent cations such as sodium ions. At very low temperatures (-20°C), the DNA precipitate can be collected using centrifugation. The concentration of the DNA is effected when the pellet is resuspended in a volume less that the original sample (Brown, 2001).

2.3.6.1 Materials

3 M Sodium acetate, pH 5.5. Sodium acetate (40, 82 g) was dissolved in sterile dist H2O (20 ml) and glacial acetic acid (50 ml). The solution was titrated to pH 5.5 with glacial acetic acid and adjusted to 100 ml with sterile dist.H_2O.

2.3.6.2 Method

3 M sodium acetate was mixed in to 1/10th volume. 2.5 Volumes of ice-cold 100% ethanol was mixed in and left at -20°C for at least 1 h. The sample was centrifuged (14000 x g, 25 min, 4°C). The supernatant was removed carefully without disturbing the pellet. The pellet was washed with 70% (v/v) ethanol (0.5 ml) by centrifugation (12000 x g 5 min, 4°C). The ethanol is removed carefully and the pellet is left to air dry for 10 min. The dry pellet was dissolved in sterile H_2O if used in specific enzymatic reactions or in TE buffer if stored indefinitely.

2.3.7 Colony PCR

PCR is a powerful tool with a variety of applications in disciplines relating to molecular biology. The presence of an appropriate insert in a bacterial colony can be

verified using a PCR technique. Amplification of insert DNA cloned into a vector, using a crude preparation of the colony DNA as template, allows extremely quick detection and characterisation of the construct. In preference, universal primers to the vector are used for PCR, allowing the nature and size of the insert in the vector to be characterised. Primers specific to the insert, in this case primers originally used to amplify the insert can also be used to screen however care must be taken as false positives can be generated. This arises due to amplification of non-cloned DNA transformed into the competent cells, and is not a reflection of insert-containing vector DNA. Although the DNA cannot replicate within the host, the sensitivity of PCR is such that even trace amounts of insert DNA in a colony can be detected and lead to a false positive. Therefore, when using insert-specific primers, care must be taken in setting up the correct controls to safeguard against this occurrence (Alvarez *et al.*, 2002).

2.3.7.1 Materials

PCR reactions were performed as per Section 2.4.2

2.3.7.2 Method

The colonies tested were approximately 1 mm in diameter and were picked of the agar media using a sterile toothpick, or flame-sterilised loop, and placed in sterile dist H2O (50 µl). The colony was vortexed thoroughly to ensure dispersion of the cells in the water. The sample was boiled for 5 min to lyse the cells and denature the DNases, and then centrifuged (8000 x g, 2 min, room temperature) to remove insoluble contaminants that may interfere with the PCR reaction. The plasmid DNA present in the solution was removed (10 µl) for use in the PCR reaction. PCR was performed with the same reaction conditions and temperature cycles as described in section 2.4.2, for the amplification of the 817bp insert from *Plasmodium falciparum* DNA.

2.4 Amplification of a 817 bp region in the *Plasmodium falciparum* MAL3P6.11 gene by the Polymerase Chain Reaction

2.4.1 Design of oligonucleotides for PCR amplification of MAL3P6.11

In general, primers used in PCR are between 20 and 30 nucleotides in length, allowing a high annealing temperature to be used. There is an increase in specificity with longer oligonucleotides due to the occurrence of that specific sequence in the target DNA. The primer length and sequence are highly important in designing amplification. The melting temperature (T_m) of a nucleic acid hybrid increases with increasing (G+C) content, and length. The annealing temperature used in the PCR cycle depends on the composition and length of the primer. An annealing temperature (T_a) approximately 5°C below the lowest T_m of the pair of oligonucleotide primers should be used (Innis and Gelfand, 1990). The length of a primer depends on its (A+T) content. The higher the T_m of the primer, the more chance mismatch pairing and non-specific priming occurring. However the primer should be complex enough to decrease the likelihood of non-specific binding. Oligonucleotides were designed to amplify regions of the MAL3P6.11 gene in preparation for possible cloning into an expression vector (pMAL-C2x).

Due to the fact that specific regions of the gene are to be cloned, it was possible to introduce restriction site sequences onto the ends of the oligonucleotide primers. This is based on the method described by Scharf *et al.* (1986) in which through PCR, these sequences will become incorporated into the amplified product, which can then be digested with the restriction enzyme to produce blunt or sticky ends, as so required. Modifications at the 5' termini of oligonucleotides have little effect on the specificity of the amplification..

2.4.1.1 Materials

DNA sequence of MAL3P6.11 Obtained at PlasmoDB (The Plasmodium Genome Resource) (http://plasmodb.org/). The *Plasmodium falciparum* 3D7 database is provided by The Institute of Genomic Research (TIGR).

Primer 3

Online primer design software tool:

(http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi/primer3_www.cgi)

Amplify3 Primer design software tool, Dr Mitchell

2.4.1.2 Method

The online primer design software, Primer 3.0 was used to suggest possible primer binding sites within the MAL3P6.11 gene. The program analyses DNA sequences and selects oligonucleotide sequences suitable for PCR. The software takes into account factors such as melting temperature and primer dimer characteristics. If specific sequences are to be amplified, such as a region to be cloned, primers can be chosen manually taking into account several important rules (Innis and Gelfand, 1990):

1 Primers should be between 17-28 base pairs in length
2 The base composition (G+C) should be 50-60%
3 The 3' end of the primer should preferably end in at least one G or C, to prevent 'breathing' of ends, and increase the efficiency of the primer.
4 Melting temperatures between 55-80°C
5 The 3' ends of the primers should not be complementary. Avoid the possibility of causing primer dimers (hybridisation between oligonucleotide primers depletes the reaction of available primers for template DNA amplification).
6 Avoid primer self-complementarity. Sometimes self-complimentary sequences cause secondary structures such as hairpins.

Oligonucleotide primers were designed based on the above criteria, to amplify portions of the MAL3P6.11 gene. Due to the extremely low (G+C) content, suitable positions for primer binding were generally limited. The primer design software (Primer 3) provided possible binding sites for primers. This software predicts the most probable and safe primer binding sites within a given sequence based on melting temperature and primer dimer factors.

Due to the fact that some primers were altered by the addition of restriction sites, they were re-checked for primer-dimer formation with Amplify3.

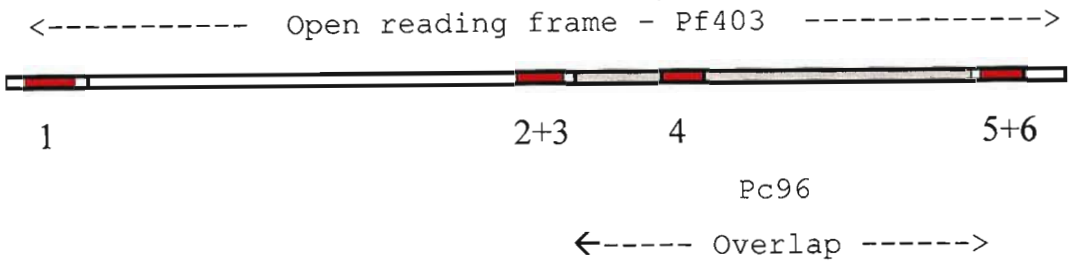


Figure 2.3. Diagram showing MAL3P6.11 gene with positions of primers designed: 1 - PfFOR1, 2 - PfFOR2, 3 - PfFOR2(*Bam*HI), 4 - PfREV1, 5 - PfREV2, 6 - PfREV2(*Sal*I). Grey region shows overlap with Pca96, respective primers binding sites is red.

The positions of the primers designed are shown in Figure 2.3. A set was designed to amplify the entire gene (primers 1 and 4), the Pc96 overlap region (primers 2 and 4), and the region of the overlap most similar to Pc96 (primers 2 and 3).

The restriction site chosen must not be present in the DNA target region, and so the sequence of MAL3P6.11 was used in an online restriction enzyme site map generator, Webcutter (http://www.firstmarket.com/cutter/cut2.html). This provided a list and position of all restriction endonuclease sites present in the gene. A *Bam*HI restriction site was designed into the sequence of the forward primer, PfFOR2(*Bam*HI) and a *Sal*I site into PfREVI(*Sal*I) (see Figure 2.3 and Table 2.3). *Bam*HI and *Sal*I sites do not occur in the target sequence and so can be used for PCR cloning. The *Bam*HI site was designed to keep the translation of sequence in frame with the expression vector protein. The *Sal*I site does not have to keep frame, being at the C-terminus of the protein.

Table 2.3. Oligonucleotide primer sequences for PCR of MAL3P6.11. Restriction sites *Bam*HI and *Sal*I are underlined.

Primer no.	5'-3' sequence	Primer name
1	tat gca tcc ctt tcg tgg tc	PfFOR1
2	tag ata tta aca act atc ttg	PfFOR2
3	gga tcc aac tat ctt gtt aat aat ctt c	PfFOR2(*Bam*HI)
4	gtt gta cgt ttt ccg tac cat c	PfREV1
5	tat tta aaa ata ttc aaa aag g	PfREV2
6	gtc gac tat tta aaa ata ttc aaa aag g	PfREV2(*Sal*I)

Two additional primers, used for the amplification of a 717 bp region of the ACC (acetyl-coA carboxylase) gene from *Plasmodium falciparum* DNA were used as an internal control for PCR. These are ACCXpf1 (GCT AAT AAT GGG ATG GCA GC) and ACCXpf2 (TTT GTG TAG TAC AAT CAC GAC).

2.4.2 Amplification of 817bp target region in the MAL3P6.11 gene from *Plasmodium falciparum* genomic DNA

2.4.2.1 Materials

10 x PCR reaction buffer with $MgCl_2$(100 mM Tris-HCl, 15 mM $MgCl_2$, 500 mM KCl, pH 8.3),

Roche (Mannheim, Germany).

dNTP mixture (2.5 mM of each dATP, dCTP, dGTP, dTTP). TaKara Biomedicals (Shiga, Japan).

Taq DNA Polymerase (1 U/µl), Roche (Mannheim, Germany).

Primer 1: 5 µM PfFOR2(*Bam*HI) (5' ggatccaactatcttgttaataatcttc 3')
Supplied at 2145.9 µM. 50µM primer stocks were made by diluting the primer stock (10 µl) into sterile 1 x TE buffer (419 µl). 5µM stocks were made by diluting the 50µM stock (10 µl) with sterile 1 x TE buffer (90 µl).

Primer 2 5 µM PfREV1 (5'gttgtacgttttccgaccatc 3')
Supplied at 2291.3 µM. 50µM primer stocks were made by diluting the primer stock (10 µl) into sterile 1 x TE buffer (448 µl). 5µM stocks were made by diluting the 50µM stock (10 µl) with sterile 1 x TE buffer (90 µl).

Plasmodium falciparum (3D7) genomic DNA. Isolated and provided by Prof. Christian Doerig

2.4.2.2 Method

Reagents were set up in sterile PCR tubes to a total volume of 50 µl, according to the amounts shown in Table 2.4. The PCR tubes were kept on ice at all times and the *Taq* polymerase enzyme was added last to the buffered reaction components.

Amplification was achieved using a Perkin Elmer GeneAmp 2400 Thermocycler according to the following temperatures and times (as previously optimised for the 717 bp ACC amplification from *Plasmodium falciparum* DNA): 94°C for 5 min, 25 cycles of the following: 94°C for 30 sec (denaturation); 55°C for 30 sec (annealing); 72°C for 1 min (extension), and 72°C for 7 min and finally to 4°C. PCR products were analysed on a 1% agarose gel with appropriate molecular mass markers.

Table 2.4. Components of PCR reaction used for amplification of 822bp region from isolated *Plasmodium falciparum* DNA.

Reaction component	Volume (µl)
Sterile distilled H_2O	35.5
10 x PCR reaction buffer with $MgCl_2$	5
dNTP mixture	4
Primer PfFOR2(*Bam*HI) stock (5µM)	2
Primer PfREV2 stock (5µM)	2
Taq DNA polymerase	1
Plasmodium falciparum genomic DNA	0.5
Total reaction volume	50

2.4.3 Optimisation of PCR reaction using a $MgCl_2$ profile

The $MgCl_2$ concentration in the PCR reaction mixture can be varied within a range of 0.5-5.0 mM in order to find the optimum. $Mg2^+$ ions form a soluble complex with the free dNTPs in the reaction mixture, which is vital for dNTP incorporation during strand synthesis by DNA polymerase. They also stimulate the polymerase activity and increase the T_m of the primer/template hybrid. The final concentration of $MgCl_2$ in the reaction can have a dramatic effect on the yield and specificity of the amplification. The optimum concentration is different for every PCR reaction, and so must be found for each reaction using a $MgCl_2$ profile. In general, low Mg^{2+} leads to low yields and an inefficient reaction, whereas high Mg^{2+} often results in the accumulation of non-specific products. Concentrations of 1.0 - 1.5 mM $MgCl_2$ are usually sufficient, and the profile is based around these values.

2.4.3.1 Materials

$MgCl_2$ stock solution (25mM $MgCl_2$) Roche (Mannheim, Germany).

10 x PCR reaction buffer without $MgCl_2$ (100mM Tris HCl, 500mM KCl, pH 8.3), Roche (Mannheim, Germany).

Other components as seen in Section 2.4.2

2.4.3.2 Method

The PCR reactions are set up as per section 2.4.2 (Table 2.4), but the 10 x PCR reaction buffer without $MgCl_2$ is used. This allows the $MgCl_2$ to be added at varying amounts. The amount of $MgCl_2$ added was taken into account when making the volume up to 50 µl with sterile dist.H_2O. Table 2.5 shows the amounts of $MgCl_2$ added for the profile. Products were analysed using agarose gel electrophoresis.

Table 2.5 Volume and final concentration of $MgCl_2$ added to the PCR reaction.

PCR reaction no.	1	2	3	4	5	6
[$MgCl_2$] (mM)	1	1.25	1.5	1.75	2	2.5
$MgCl_2$ (µl)	2	2.5	3	3.5	4	5

2.5 Directional cloning of 817 bp PCR product into pGEX-4T-3

2.5.1 Clean-up of 817 bp PCR product

2.5.1.1 Materials

Components for PCR are described in section 2.4.2.

Chloroform: isoamyl alcohol (CIAA)(24:1). Chloroform (240 µl) was combined with isoamyl alcohol (10 µl) and stored at room temperature.

2.5.1.2 Method

PCR amplification of the 817bp region in the MAL3P6.11 gene was conducted in a 100 µl reaction volume as described in section 2.4.2. An aliquot (5 µl) of the PCR reaction was evaluated for purity and size on a 1 % agarose gel. (90 µl) was added to the remaining PCR product. The mixture was transferred to a sterile microfuge tube and vortexed for 1 minute. The mixture was then centrifuged (12000 x g ; 1 minute) and 2 µl of the aqueous phase was transferred to a sterile microfuge tube. The PCR product was quantitated at 260 and 280 nm using spectrophotometry before and after the treatment to assess the clean up procedure. An aliquot (5 µl) was checked on an agarose gel for the efficiency of the CIAA treatment and yields obtained.

2.5.2 Klenow treatment and restriction digestion (*Bam*HI) of PCR product

2.5.2.1 Materials

DNA Polymerase I Large (Klenow) Fragment (5U/µl) (Promega, USA).
10 x BamHI restriction buffer (60 mM Tris-Cl; 1 M NaCl; 60 mM MgCl2; 10 mM DTT, pH7.5) (Promega, USA).

817 bp CIAA treated PCR product (500 µg/ml) As prepared in previous Section 2.5.1,

2.5.2.2 Method

The components described in Table 2.6 were added to a sterile microfuge tube and incubated for 15 min at 30°C. The reaction was stopped by heating for 10 min at 75°C. *Bam*HI (1 µl) was added to the reactions and incubated overnight at 37°C. This reaction was again stopped by heating at 65°C for 15 min.

Table 2.6 Reaction components for the blunt ending of PCR products using the Klenow enzyme.

Reaction component	Volume (µl)
CIAA treated PCR product	5
Klenow enzyme (5 U/µl)	1
10 x *Bam*HI restriction buffer	2
Sterile dist.H2O	12

2.5.3 Calculation of vector:insert ratios used for cloning 817 bp insert into pGEX-4T-3

The amount of the insert DNA prepared after the klenow treatment and restriction digestion (*Bam*HI) reactions must maintain an appropriate vector:insert ratio for the ligation reactions. A molar ratio of 1:5 (vector:insert) was chosen for the ligation reactions. The amount of insert (817 bp) DNA to use per 50 ng vector (4900 bp) for ligation was calculated according to the following equation:

$$\frac{\text{Insert length (bp)} \times \text{ng vector}}{\text{vector length bp}} \times \text{molar ratio} = \text{ng insert}$$

$$\frac{817 \text{ bp} \times 50}{4900 \text{ bp}} \times 5.0 = 41.68 \text{ ng insert}$$

Similarly, the amount of insert DNA to use in a ligation with a vector:insert ratio of 1:10 was calculated at 83.36 ng per 50 ng vector.

2.5.4 Ligation of 817 bp PCR product into pGEX-4T-3 vector

2.5.4.1 Materials

817 bp Insert DNA (125 µg/ml) As prepared in previous Section 2.5.2.

T4 DNA Ligase (Amersham Life Science, Inc, USA).
10 x ligation buffer (660 mM Tris-HCl, pH 7.6, 66 mM MgCl2, 100 mM DTT, 660 µM ATP) (Amersham Life Science, Inc, USA).

*Bam*HI/*Sma*I double digested pGEX-4T-3 DNA (50 µg/ml) (Amersham, Life Science, USA) pGEX-4T-3 plasmid DNA was isolated from *E. coli* culture using the (Nucleobond® AX kit) as described in Section 2.3.5. The purified DNA was digested with BamHI and SmaI restriction enzymes (Section 2.3.3) in preparation for ligation reactions.

2.5.4.2 Method

Ligation reactions were set up according to Table 2.7 in autoclaved microfuge tubes. Before the addition of the 10 x ligation buffer and the T4 DNA ligase, the reaction mixture was heated at 65°C for 5 min then cooled on ice. The reaction was incubated at 16°C overnight.

Table 2.7 Reaction components for the ligation of 817 bp PCR product into *Bam*HI/*Sma*I digested pGEX-4T-3.

	Volume (µl)					
Ligation reaction number	1	2	3	4	5	6
*Bam*HI/*Sma*I digested pGEX-4T-3 (50 µg/ml)	1	1	1	1	1	-
817 bp Insert DNA (125 µg/ml)	0.3	0.3	0.6	0.6	-	0.3
10 x ligation buffer	1	1	1	1	1	1
T4 DNA ligase	1	1	1	1	-	1
Sterile dist.H_2O	6.7	6.7	6.4	6.4	8	7.7
Total reaction volume	10	10	10	10	10	10

2.5.5 Preparation of competent *E. coli* DH5

The procedure developed by Cohen *et al.* in 1973, can yield competent cultures of *E. coli* strains for transformation protocols. Based on previous observations, bacteria treated with ice-cold solutions of $CaCl_2$, then heat shocked, can be transfected with bacteriophage λDNA. The treatment induces a state of 'competence' in the bacteria, in which they have the ability to take up DNA molecules derived from a variety of sources. How these divalent cations induce this state, and the mechanism by which plasmid DNA enters the cells is unknown. The method provides a means of obtaining a high frequency of transformation, used routinely and efficiently for cloning purposes (Sambrook, 1989). The protocol is essentially that of Cohen *et al.* (1973) with minor modifications.

2.5.5.1 Materials

E. coli (DH5α) *supE44 ΔlacU169 (φ80 lacZΔM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1*

Glycerol stocks were previously prepared by combining *E. coli* culture (0,625 ml) with sterile glycerol (0,375 ml). These were stored at -70°C.

LB agar medium. Bacto-tryptone (10 g), bacto yeast extract (5 g), NaCl (5 g) and agar (15 g) was dissolved in dist H2O (1 l) with heating. The media was sterilised by autoclaving (121°C for 1 h). Once the solution had cooled to about 50°C, the media was aseptically poured into sterile petri dishes.

LB broth media See Section 2.3.4.1

Sterile 50 ml polypropylene tubes. The tubes were cleaned, rinsed, and the lids and tubes were wrapped separately in tin foil. These were autoclaved at 121°C for 1 h.

0.1 M $CaCl_2$. $CaCl_2.2H_2O$ (1.47 g) was weighed and dissolved in dist H2O (100 ml). Solution was autoclaved at 121°C for 1 h.

2.5.5.2 Method

Glycerol stocks of *E. coli* strain DH5α, were streaked onto LB agar plates, and colonies were allowed to grow overnight at 37°C. A single colony was picked and transferred aseptically to sterile LB broth (50 ml) and cultured overnight at 37°C. 500 µl of this was inoculated into a further LB broth (100 ml) and grown at 37°C with vigorous shaking (300 cycles / min) until the OD_{600} is 0.3 - 0.4 (the number of viable cells per ml of culture should not exceed 10^8). The growth of the culture was monitored spectrophotometrically every 20-30 min. The cells (50 ml) were transferred to two sterile (autoclaved), ice-cold 50 ml polypropylene tubes. The cells were cooled by storage in the tubes, on ice for 10 min. The cells were recovered by centrifugation (4500 x g, 10 min, 4°C) and resuspended in ice cold sterile 0.1 M $CaCl_2$ (20 ml). The cells were again centrifuged (4500 x g, 10 min, 4°C) and carefully resuspended in 0.1 M $CaCl_2$ (1 ml). Cells can either be used for transformation at this point or stored at 4°C (no longer than 3 weeks). Glycerol storage of these cells allows for much longer storage, yet the transformation efficiencies decrease drastically after glycerol storage.

2.5.6 Transformation of competent *E. coli* DH5α cells

2.5.6.1 Materials

E. coli DH5α competent cells. As prepared in section 2.5.5.

SOC medium

pGEX-4T-3 test plasmid for transformation (0.2 ng/µl). Prepared by making a 0.2 µg/ml working stock of purified (Nucleobond® AX kit) pMAL-C2x plasmid DNA (Section 2.3.5).

2.5.6.2 Method

Competent DH5α *E. coli* cells (Section 2.5.5) were aliquoted (20 µl) into sterile microfuge tubes. The ligation reaction mixes (Table 2.2) (1 µl) were added directly to the cells (20 µl) and the sample was mixed gently. The test plasmid (pGEX-4T-3) for transformation (1 µl) was used as a transformation control to assess the efficiency of the DH5α competent cells. The tubes were left on ice for 30 min. SOC media was removed from -70°C and thawed. The cells were then heat-shocked for 90 seconds in a 42°C water bath. The samples were rapidly placed on ice for 2 min and SOC media (80 µl) was added and gently mixed with the cells. The tubes were incubated at 37°C for 1 h with rapid shaking (200-250 rpm) to allow the bacteria to recover and express the antibiotic resistant marker encoded by the plasmid DNA. A hockey stick for plating was made by heating and bending a Pasteur pipette, whose tip was sealed in a flame. The transformations (50 µl) were spread onto the surface of the L-agar plates aseptically, using 70%(v/v) ethanol to flame-sterilise the hockey stick. An additional control transformation (5 µl) was also spread. The plates were incubated overnight, inverted, at 37°C.

2.6 Cloning 817bp PCR product into pMOS*Blue* vector

The pMOS*Blue* blunt ended cloning kit (Amersham Biosciences, UK) was used to clone the 817 bp PCR product into the vector, pMOS*Blue*. Materials listed were partly provided by the kit. The procedure was followed according to the kit manual.

2.6.1 Phosphorylation and klenow treatment of DNA fragments

The large fragment of DNA Polymerase I (Klenow Fragment) is DNA polymerase I with its 5'-3' exonuclease activity removed, keeping the polymerase activity or the 3'-5'exonuclease activity intact. The 5'-3' exonuclease activity degrades 5' termini of primers bound to DNA templates and removes 5' phosphates that function in ligation reactions (Klenow and Henningsen, 1970). The polymerase activity can fill in recessed 3' termini, or degrade from free 3'-hydroxyl termini, polishing and blunt ending DNA fragments (Sambrook *et al.*, 1989).

Ligation reactions depend on the required 5'-phosphate residues at the termini of DNA fragments for the formation of new phosphodiester bonds between two DNA molecules. DNA molecules that lack the 5'-phosphate group can be prepared for ligation by phosphorylation with bacteriophage T4 polynucleotide kinase. This enzyme catalyzes the transfer of the γ-phosphate of ATP to the 5' terminus of DNA. The phosphorylation and klenow reactions were performed simultaneously in this experiment.

2.6.1.1 Materials

Positive control insert (4.5 ng/μl). (Amersham Biosciences, UK)

817 bp PCR product (500 μg/ml) Prepared by CIAA treatment (Section 2.5.1).

10 x PK buffer (Amersham Biosciences, UK)

PK Enzyme mix (Amersham Biosciences, UK)

100 mM dithiothreitol (DTT) (Amersham Biosciences, UK)

Sterile dist. H_2O

2.6.1.2 Method

PK reactions were set up according to Table 2.8 and were incubated at 22°C for 40 min. The samples were centrifuged briefly to collect the contents at the bottom of the tube. The reaction was heat inactivated at 75°C for 10 min. The reaction was cooled on ice for 2 min and centrifuged briefly to collect the condensate.

Table 2.8 Components of PK reactions for cloning 817bp insert into pMOS*Blue*

	Volume (µl)					
PK reaction number:	1	2	3	4	5	6
Control insert DNA	-	-	-	2	2	-
PCR aqueous phase	1	1	1	-	-	-
10 x Pk buffer	1	1	1	1	1	1
100 mM DTT	0.5	0.5	0.5	0.5	0.5	0.5
PK enzyme mix	1	1	1	1	-	1
sterile dist.H_2O	6.5	6.5	6.5	5.5	6.5	7.5
Final reaction volume	10	10	10	10	10	10

2.6.2 Blunt-end ligation of 817 bp insert to *Eco*RV digested pMOS*Blue* vector

2.6.2.1 Materials

T4 DNA ligase (4 units / µl)

*Eco*RV digested dephosphorylated pMOS*Blue* vector (50 ng / µl)

PK reactions from section 2.6.1

2.6.2.2 Method

Ligation reactions were set up as described on Table 2.9.

Table 2.9. Ligation reactions for cloning 817bp insert into 2887 bp pMOS*Blue* vector.

Ligation reaction	1	2	3	4	5	6
linearised vector (50 ng)	1	1	1	1	1	1
PK reaction (1-6)	10	10	10	10	10	10
T4 DNA ligase	1	1	1	1	1	1
Total reaction volume:	12	12	12	12	12	12

The PK reactions (10 µl) were combined with 1 µl *Eco*RV digested pMOS*Blue* vector according to the amounts indicated on Table 2.9. The T4 DNA ligase (1 µl) was added last to the two samples and gently stirred with the pipette tip. The ligation reactions (12 µl) were incubated at 22°C overnight.

2.6.3 Preparation of LB agar plates for spreading pMOS*Blue* transformants.

2.6.3.1 Materials

LB-agar (500 ml). Tryptone (5 g), yeast extract (2.5 g), NaCl (5 g) and agar (7.5 g) were dissolved in dist H2O (500 ml), in an autoclavable 1 l glass bottle with screw on

lid, with heating. The dissolved media, and a magnetic stirrer bar was sterilised by autoclaving (121°C, 1 h).

Tetracycline (12 mg/ml). Tetracycline (12 mg) was dissolved in 70%(v/v) ethanol (1 ml) and stored in the dark at -20°C. Dilute stock 0.1 ml per 100 ml L-broth (12 μg/ml final concentration).

5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) substrate (40 mg/ml). Dissolve X-gal (40 g) in dimethylformamide (DMF) (1 ml). Dilute 0.2 ml per 100 ml L-broth (80 μg/ml final concentration).

0.1M isopropyl B-D-thiogalactopyranoside (IPTG). IPTG (0.238 g) was dissolved in 10 ml dist.H_2O. The solution was filter sterilised and stored in a sterile 15 ml vial. This was added to 0.3 mM final concentration in the media.
Ampicillin (5 mg/ml). Ampicillin (50 mg) was dissolved in dist H2O (10 ml). The solution was filter sterilised and stored in a sterile 15 ml vial at -20°C. Dilute 1 ml per 100 ml L-broth (50 μg/ml final concentration).

2.6.3.2 Method

LB agar (493 ml) was prepared and autoclaved. The media was stirred using the magnetic stirrer bar until it cooled down to approximately 50°C. Tetracycline (0.5 ml), ampicillin (5 ml), X-gal (1 ml), and IPTG (0.5 ml) were added aseptically and mixed into the media. The LB media was poured into sterile 82mm plates until media covered the base. The plates were left to cool and set before refrigeration, to prevent condensation occurring inside the plates. Plates are stored inverted at 4°C.

2.6.4 Transformation of competent MOS*blue E. coli* cells

2.6.4.1 Materials

MOS*Blue* competent cells (component of pMOS*Blue* blunt ended cloning kit)

SOC media (component of pMOS*Blue* blunt ended cloning kit)

Test plasmid for transformation (0.2 ng/µl) (component of pMOS*Blue* blunt ended cloning kit)

2.6.4.2 Method

Sterile microfuge tubes were pre-chilled on ice. A sample of MOS*blue* competent cells was removed from -70°C storage and thawed on ice. The cells were dispensed into the sterile microfuge tubes (20 µl). The ligation reaction mixes (Table 2.9) (1 µl) was added directly to the cells (20 µl) and the sample was mixed gently. The test plasmid (pMOS*Blue*)for transformation (1 µl) was used as a transformation control to assess the efficiency of the MOS*blue* competent cells.

The tubes were left on ice for 30 min. SOC media was removed from -70°C and thawed to room temperature. The cells were then heat-shocked for 40 seconds in a 42°C water bath. The samples were placed on ice for 2 min and room temperature SOC media (80 µl) was added and gently mixed with the cells. The tubes were incubated at 37°C for 1 h with rapid shaking (200-250 rpm). A hockey stick was made, and the cells were spread onto the LB-agar plates as per section 2.5.6.

2.7 Subcloning 822 bp fragment from pMOS817 into expression vector pMAL-C2x

2.7.1 Purification of 822 bp insert from low-melting point agarose gel

Agarase I is isolated from *Psuedomonas atlantica* is an enzyme widely used in the field of molecular biology for isolating DNA fragments of many sizes from agarose gels (Burmeister and Lehrach, 1987). The enzyme is responsible for the degradation of agarose to oligosaccharides, and so disrupting the agarose gel matrix. Once treated and warmed, a gel will not reset upon cooling, allowing the DNA fragments to be free from the agarose gel. Agarase I and the oligosaccharides, generally do not interfere

with restriction enzyme digestion, and ligation. However the DNA can still be separated from these contaminants using standard procedures (phenol extraction or ethanol precipitation) (Burmeister and Lehrach, 1989).

Low melting point agarose contains hydroxyethyl groups that have been introduced into the polysaccharide chain. This substitution causes the gel to set at approximately 30°C and to melt at 65°C. Due to the low melting-temperature, the agarose remains fluid at 37°C, allowing various enzymatic manipulations to be carried out (ligation, synthesis of probes, digestion with restriction enzymes, etc.). The use of the enzyme agarase provides an extremely gentle method for the selective extraction and isolation of DNA bands from gels.

2.7.1.1 Materials

Agarase I enzyme (Sigma-Aldrich, Inc., USA). Stored in 40 mM Tris HCl, 50 mM NaCl, 50%(v/v) glycerol. One unit is the amount of enzyme that digests 100 µl of melted 1% low-melting agarose to agaro-oligosaccharides in 1 h at 45°C.

25X agarase I buffer (0.75 M Bis-Tris, 0.25 M EDTA, pH 7.1).

Low melting point agarose

3 M sodium acetate pH 5.2. Sodium acetate (40, 82 g) was dissolved in sterile dist H2O (20 ml) and glacial acetic acid (50 ml). The solution was titrated to pH 5.2 with glacial acetic acid and adjusted to 100 ml with sterile dist.H_2O.

2.7.1.2 Method

The band of interest, identified by comparison to molecular markers, was cut from the low melting point agarose gel and placed in a pre-weighed sterile microfuge tube. The tube was weighed to determine the weight of the gel slice (100 mg of gel yields 100 µl melted gel). 0.04 Volumes of 25X agarase I buffer was added and the gel was

melted at 65°C for 15 min. The sample was cooled to 45°C and the agarase I enzyme was added (1 U of agarase per 100 µl gel). This was incubated at 45°C for 1 h.

The DNA fragments were removed and concentrated from the sample by ethanol precipitation. This was performed as per Section 2.3.6 with minor modifications as follows. 0.1 Volume 3 M sodium acetate pH 5.2 was added to the melted agarose solution and was incubated for 15 min on ice. The sample was centrifuged (12000 x g; 15 min, 4°C) to remove the oligosaccharides, and the supernatant was transferred to a sterile microfuge tube. 3 Volumes of ice-cold 100%(v/v) ethanol was added and the tubes were incubated for 1 h on ice. The samples were centrifuged (14000 x g, 25 min, 4°C) to pellet the DNA, which was air dried for 5 min and dissolved in dist.H_2O. The DNA was quantitated using spectrophotometry (Section 2.3.1).

2.7.2 Ligation of 822bp fragment into pMAL-c2x

2.7.2.1 Materials

T4 DNA Ligase (Amersham Life Science, Inc, USA).

10 x ligation buffer (660 mM Tris-HCl, pH 7.6, 66 mM MgCl2, 100 mM DTT, 660 µM ATP) (Amersham Life Science, Inc, USA).

2.7.2.2 Method

pMAL-C2x Vector (1 µg) was used in a restriction digestion reaction (20 µl) using *Bam*HI, in preparation for cloning as per section 2.3.3. It was ensured that complete digestion had occurred, so as to minimise transformation of background colonies. The amount of insert DNA to use for a vector:insert molar ratio of 1:5, was insert (30.7 ng):vector (50 ng), as calculated in section 2.5.3. Table 2.10 shows the ligation reactions performed for this experiment. The insert DNA prepared in the previous section 2.6.1 was diluted to a 30.7 µg/ml working stock.

Table 2.10 Ligation reactions for cloning 822bp *Bam*HI insert into pMAL-C2x

Ligation reaction	1	2	3	4	5	6
1 µl linearised pMAL-C2x (50 µg/ml)	1	1	1	1	1	-
1 µl Insert DNA (30.7 µg/ml)	1	1	2	-	-	1
10 x ligation buffer	1	1	1	1	1	1
T4 DNA ligase	1	1	1	1	-	1
Sterile dist.H_2O	6	6	5	7	8	7
Total reaction volume	10	10	10	10	10	10

2.7.3 Preparation of competent *E. coli* JM103 cells

As performed in Section 2.5.5, but with the *E. coli* JM103 strain (*supE thi-1 endA1 hsdR4 sbcB15 strA* Δ(*proAB*) F'[*traD36 proAB*$^+$ *lacI*^{q} *lacZΔM15*).

2.7.4 Transformation of competent JM103 *E. coli* cells

The transformations were performed as described in Section 2.5.6, except the *E. coli* JM103 competent cells were used and the transformation efficiency was assessed using pMAL-C2x test plasmid for transformation (0.2 ng/µl), which was previously prepared by making a 0.2 µg/ml working stock of purified (Nucleobond® AX kit) pMAL-C2x plasmid DNA.

2.8 Protein techniques

2.8.1 Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was established in 1967 by Shapiro *et al.* The method is initially dependent on the interaction of the protein with SDS. SDS is a detergent with a 12 carbon hydrophobic tail and a hydrophilic, sulfonic acid head group. Proteins are first denatured in the presence of SDS by boiling. The SDS molecules interact with the proteins causing the formation of rod-like complexes with a consistent ration of *ca.* 1.4 mg of SDS per mg of protein (Pitt-Rivers and Impiombato, 1968). The negative charge of the SDS effectively masks that of the protein, causing all protein/SDS complexes to contain the same charge/mass ratio and an anodic migration (Weber and Osborn, 1969). While the charge differences between proteins is masked by SDS, separation of proteins is due to differences in size, therefore allowing effective separation based on size and estimation of molecular size.

SDS-PAGE is based on a discontinuous buffer system. This was developed where changes in the properties of the buffers and the gels in a vertical slab layout, would maximise the resolution and concentration of protein into defined visible bands. Two different gels are used, a stacking gel (large pore size) and a running gel (small pore size) (Ornstein, 1964). Both gels contain Tris-HCl buffer, but the buffer in the stacking gel is a lower pH than that in the running gel. In an electric field, a sharp interface develops between the high mobility 'leading ion' (Cl^-) and glycine, the less mobile 'trailing ion'. As the interface migrates downwards, protein molecules with mobilities intermediate between Cl^- and glycine become concentrated into a thin band. Protein molecules are kept in this sharp interface by voltage gradients behind or ahead of the interface (Jovin, 1973). As this interface reaches the change between the stacking and running gels the pH increases, resulting in an increased mobility of the glycine trailing ion (the anodic mobility of glycine increases with pH as the proportion of glycine ions increases). The interface overtakes the stack of proteins allowing separation to occur in a linear voltage gradient according to protein size. The running gel therefore forms a small-pore sieve effect. The smaller the protein, the more rapid migration towards the anode occurs (Dennison, 1999).

Polyacrylamide is a synthetic gel for electrophoresis that carries the advantage of being highly reproducible. The pore size is easily controlled by adjustments in the proportions of acrylamide and bisacrylamide, making up the synthetic polymer. Acrylamide is cross-linked by *N,N'*-methylenebisacrylamide, by the initiation of polymerisation with *N,N,N',N'*-tetramethylethylenediamene (TEMED). TEMED acts as a catalyst in the formation of free

radicals from ammonium persulfate, causing polymerisation (Hames and Rickwood, 1981).

The Laemmli (1970) method of polyacrylamide gel electrophoresis has become one of the most popular PAGE methods in use. However, for resolution of smaller proteins under 20 kDa, and alternative method is used described by Shagger and Von Jagow (1987), called the Tricine method. The estimation of molecular weight relies on reduction of the proteins as the disulfide bridges may interrupt the protein structure and prevent formation of protein/SDS complexes. Boiling in SDS not only denatures the protein but also dissociates the subunits of oligomeric proteins. The molecular weights obtained by SDS-PAGE are therefore representative of the subunits and not of the intact protein (Dennison, 1999).

2.8.1.1 Materials

Solution A: Monomer Solution [30% (m/v) acrylamide, 2.7% (m/v) Bis-acrylamide]. Acrylamide (73 g) and Bis-acrylamide (2 g) were dissolved and made up to 250 ml with dist H2O and stored in an amber coloured bottle at 4°C. Filter through Whatman No. 1 filter paper before use.

Solution B: 4 x Running Gel Buffer (1.5 M Tris-HCl, pH 8.8). Tris (45.37 g) was dissolved in approximately 200 ml of dist.H_2O, adjusted to pH 8.8 with HCl and made up to 250 ml. Filter through Whatman No. 1 filter paper before use.

Solution C: 4 x Stacking Gel Buffer (500 mM Tris-HCl, pH 6.8). Tris (3 g) was dissolved in 40 ml dist.H_2O, adjusted with HCl to pH 6.8 and made up to 50 ml. This buffer was made up weekly, because as a result of its poor buffering capacity at 2.1 pH units below its pKa at 4°C (Pharmacea products catalogue), pH drift led to anomalous running patterns in non-reducing SDS-PAGE. Filter through Whatman No. 1 filter paper before use.

Solution D: 10% (m/v) SDS. SDS (10 g) was dissolved in 100 ml dist H2O with gentle heating if necessary.

Solution E: Initiator [10% (m/v) ammonium persulfate]. Ammonium persulfate (0.2 g) was made up to 2 ml with dist H2O just before use.

Solution F: Tank buffer [250 mM Tris-HCl, 192 mM glycine, 0.1% (m/v) SDS, pH 8.3]. Tris (15 g) and glycine (72 g) were dissolved and made up to 5 litres with dist.H_2O. Prior to use, 2.5 ml SDS stock (solution E) was added to 250 ml for use in the Mighty Small II apparatus.

Solution G: Reducing Treatment Buffer [125 mM Tris-HCl, 4 % (m/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol, pH 6.8]. Buffer C (2.5 ml), 10% SDS (4 ml) (solution D), glycerol (2 ml) and 2-mercaptoethanol (1 ml) were made up to 10 ml with dist.H_2O.

Solution H: Non-reducing Treatment Buffer [125 mM Tris-HCl, 4 % (m/v) SDS, 20% (v/v) glycerol, pH 6.8]. Buffer C (2.5 ml), 10% SDS (4 ml) (solution D) and glycerol (2 ml) were made up to 10 ml with dist.H_2O.

Stain stock solution [1% (m/v) Coomasie blue R-250]. Coomasie blue R-250 (1 g) was dissolved in 100 ml of dist H2O by magnetic stirring for 1 h at room temperature. The solution was filtered through Whatman No. 1 filter paper.

Staining solution [0.125% (m/v) Coomasie blue R-250, 50% (v/v) methanol, 10% (v/v) acetic acid]. Stain stock (62.5 ml) was mixed with methanol (250 ml) and acetic acid (50 ml), and made up to 500 ml with dist.H_2O.

Destaining solution I [50% (v/v) methanol, 10% (v/v) acetic acid]. Methanol (500 ml) was mixed with acetic acid (100 ml) and made up to 1 litre with dist.H_2O.

Destaining solution II [7% (v/v) acetic acid, 5% (v/v) methanol]. Acetic acid (70 ml) was mixed with methanol (50 ml), and made up to 1 litre with dist.H_2O.

2.8.1.2 Method

The Bio-Rad Mini-PROTEAN II® vertical slab electrophoresis unit was used for SDS-PAGE, and was assembled as per manufacturer's manual. Glass plates, spacers, combs and casting stand gaskets were washed thoroughly with soap water and cleaned meticulously with 70% ethanol. The glass plates were clamped, separated by the 1 mm spacers and running gel was poured into the assembly using a 10 ml syringe with sterile needle. The gel was poured to a height leaving space for the stacking gel. Dist H2O was overlaid to prevent inhibition of polymerisation by oxygen. The gel was allowed to polymerise for 30 min to 1 h at room temperature. The gel had set when the interface was visible between the gel and the water. The water is poured off the running gel or carefully removed with a syringe. The stacking gel solution was poured in and the gel comb was inserted to form the sample application wells.

Table 2.11. Preparation of running and stacking gels of different acrylamide concentrations (Laemmli, 1970).

Reagent	Volume (ml)					
	Running gel (%)				Stacking gel (%)	
	15	12.5	7.5	5	4	3
A	7.5	6.25	3.75	2.5	0.94	0.71
B	3.75	3.75	3.75	3.75	0	0
C	0	0	0	0	1.75	1.75
D	0.15	0.15	0.15	0.15	0.07	0.07
E	0.075	0.075	0.075	0.075	0.035	0.035
dist.H_2O	3.5	4.75	7.25	8.5	4.3	4.53
TEMED	0.0075	0.0075	0.0075	0.0075	0.015	0.015

Once set (30 min to 1 h at room temperature), the comb was carefully removed and the wells rinsed with dist.H_2O. The clamps were placed onto the electrode apparatus and placed in an electrode tank. Tank buffer (Solution F, with SDS) (250 ml) was poured into the upper and lower electrode compartments.

The procedure for preparation of protein samples differs for different applications of SDS-PAGE. In general, for reducing SDS-PAGE, samples are combined with an equal volume of reducing treatment buffer (solution G) and boiled for 5 min. A dye, bromophenol blue tracker dye (5 µl) which migrates with the buffer front was added to each sample before loading into wells. Samples were loaded using gel-loading pipette tips into the set wells. After samples were loaded, the gel unit was connected to a power pack and run at 18 mA per gel until the bromophenol blue tracker dye in the sample is close to running off the gel. The power supply was disconnected, the clamps removed and the plates were separated using a spacer. The gel was removed using gloves and placed into staining solution for 4-to12 h. The staining solution was removed and the gels rinsed with destaining solution I, until gel background is clear. Destaining solution II is used to complete the stain and enlarge the gel. Gels were kept in plastic zip-seal backs and kept in destaining solution II until photographed or scanned.

2.8.2 Induction and expression of GST/MBP fusion proteins with IPTG

These protocols are similar to that of Sambrook *et al.*(1987) with minor modifications suited to optimisation of expression systems. The induction is based on the chemical activation of the *lacZ* gene with IPTG, during early exponential phase of *E. coli* culture. It is during this phase of the growth cycle that *E. coli* is expressing proteins at an optimal level, due the availability of growth nutrients, and the relative efficiency at which the cells divide. However these factors need to be optimised for each expression system, including the growth temperature at induction. An advantage of chemical induction with IPTG is the relatively high amounts of expression that can be obtained, even at low temperatures (20 - 30 °C) where in general, protein synthesis, rate of growth and cell division is low. SDS-PAGE analysis following induction, is an effective way of monitoring and detecting the levels of expression of a particular fusion protein, and provides a platform for further analysis such as western blotting.

2.8.2.1 Materials

0.1 M isopropylthio-β-D-galactoside (IPTG). IPTG (0.238 g) was dissolved in sterile dist H2O (10 ml) and filter sterilised.

1 x SDS gel loading buffer [125 mM Tris-HCl, 4 % (m/v) SDS, 20% (v/v) glycerol, 10% (v/v) 2-mercaptoethanol, pH 6.8].. Prepared as per Section 2.7.1, Solution G.

2.8.2.2 Method

E. coli cells containing the expression vector were initially streaked on an LB-agar plate (100 ug/ml Ampicillin) and grown overnight at 37°C in order to obtain single colony clones. Two cultures of a single colony was grown in separate LB medium (10 ml) containing ampicillin (100 ug/ml) overnight at 37°C with agitation. Further LB broth media (20 ml) was inoculated with the cultures (500 ul) and grown to an optical density (OD) of 0.6, as determined spectrophotometrically (600 nm). This was achieved by measuring the density every 30 min, against a sterile LB broth (1 ml) blank. At this point, one of the cultures was inoculated with 0.1M IPTG (3 ul per ml of culture) to 0.3 mM, taking into account the volume removed for spectrophotometric analysis. These cultures were grown for a further 4 h to ensure the induction of sufficient fusion protein for analysis using SDS-PAGE.

Cellular protein extracts were prepared by centrifuging (15000 x g) for 1 min at room temperature in a microfuge tube. The pellets were resuspended in 1 x SDS gel-loading buffer (100 ul). Samples were boiled for 3 min and cooled on ice, followed by loading (5 ul) the wells of a prepared SDS-PAGE gel for analysis of cellular protein. Cultured samples not induced were treated similarly and run simultaneously to the induced samples for direct comparison of protein proportions.

CHAPTER 3

IN SILICO DATABASE SCREENING WITH PC96 FOR CLASSIFICATION AND FUNCTIONAL ASSIGNMENT

3.1 Introduction

The alignment of novel DNA and amino acid sequences with previously characterised coding or amino acid sequences can provide important information about the identification of proteins. In recent years, the explosive growth in biotechnology and information technology, and efficient and automated sequencing methods have made it easier to characterise proteins based on their sequences. If significant sequence similarity exists between the query sequence and a previously characterised protein, much information can be acquired. The alignment of protein sequences can show common underlying structural roles for the protein. Sequence comparison procedures should be treated like any laboratory experiment taking care in the design of the particular method, and the correct evaluation of results gained (Brenner, 1998).

Pc96, a 96 kDa membrane associated antigen in *Plasmodium chabaudi adami,* has shown protective roles in infections of mice (Wanidworanun *et al.*, 1989). Monoclonal antibodies against Pc96 identified cross-reacting antigens in *P. falciparum* (155 kDa), *P. vivax* (222 kDa) and *P. cynomolgi* (200 kDa). Because Pf155 and Pv222 share a cross-reacting epitope with Pc96, this antigen was identified as a potential vaccine candidate. *P. chabaudi adami* infection in B-cell deficient mice, activates a T-cell dependent mechanism which terminates acute malaria in a similar manner to that seen in immunologically intact mice (Grun and Weidanz, 1981), indicating a significant role of cell mediated immunity. A protein was identified by antibodies raised against Pc96 that caused the proliferation of specific T cell clones, responsible for providing adoptive protection in a *Plasmodium chabaudi adami* infection. The identification of a similar protein in *Plasmodium falciparum* with similar T cell epitopes, and the identification of the T cell epitope in the sequence that may be responsible for a multiple antibody response would prove vital in the design of cell mediated protection against human malaria. The ring-infected erythrocyte surface antigen (Pf155/RESA) was thought to be the homologue of Pc96, based on cross-reactivity with antibodies in *P. falciparum.* In these experiments, protein sequence and structural analysis showed that Pf155/RESA is non-homologous to Pc96 and several other malaria proteins shows high homology in regions.

The sequence of the *Plasmodium falciparum* genome has recently been made availability at PlasmoDB (The *Plasmodium* Genome Resource) (Gardner *et al.*, 2002). This allows the database

to be screened for sequences similar to Pc96, thus providing an alternative approach to the identification of possible homologues. An open reading frame was identified in the falciparum PlasmoDB, coding for a 3394 amino acid protein predicted to be 403 kDa is size. BLAST screening with the Pc96 protein showed Pf403 to contain an extensive region of sequence similarity and similar secondary structure properties. Similar sequence similarities were found in a 178 kDa protein identified in the *Plasmodium yoelii yoelii* database, Pyy178. In an ideal situation, the BLAST search would have matched the Pc96 sequence to a full length, previously classified protein with known characteristics and functions. Proteins identified as showing high homology were previously not characterised, and were simply open reading frames within malaria genome DNA sequences, with unknown function.

The evaluation of protein sequences and subsequent identification of putative T cell epitopes from potential antigens has found many applications in the development of vaccines. Various strategies are available that select putative peptide candidates based on algorithms. Further experiments can be designed to verify the relevance of the epitopes (Schirle, *et al.*, 2001). Major histocompatibility complex (MHC) molecules are involved in the presentation of these epitopes to cells of the immune system. MHC class I molecules become associated with the peptides in the endoplasmic reticulum, after which is transported to the cell surface where it is exposed to the immune system for recognition by cytotoxic T cells. MHC class II ligands have more variety in their length (9-25 amino acids) and bind to MHC class II molecules in the late endosomal/lysosomal compartment. The complexes are exposed on the cell surface, for recognition by $CD4^+$ cells. A T cell response is mediated by the recognition of the peptide derived from protein antigen leading to T cell activation. This has since become the most successful method for the identification of T cell epitopes (Schirle, *et al.*, 2001).

The specificity of other cell components of cellular immune response (TAP and proteasome) have been applied as additional criteria to epitope prediction and the selection of appropriate T cell epitopes. T-cell mediated immunity is initiated by a series of cellular processes due to the recognition of MHC-peptide complexes by the T cell receptor.

In this Chapter, a bioinformatics approach was taken, utilising the various malaria databases in the search for proteins similar to the mouse malaria antigen, Pc96, and the possible identification of protein structure and function. A 403 kDa protein in *Plasmodium falciparum,* and a 178 kDa Pyy178 in *Plasmodium yoelii yoelii*, an alternative mouse model for malaria, showed extensive regions of similarity to Pc96 and contained putative conserved T cell epitopes similar to that found in Pc96. Features of these proteins such as hydropathy, flexibility and antigenicity were

evaluated, and motifs and signature sequences were identified using the PROSITE database. Further analysis was used to recognise the 3D structure of specific domain regions in Pf403 and homologous proteins. Features recognised in the Pf403 sequence include a unique leucine zipper patterns, along with the recognition of a Interferon-induced guanylate-binding protein 1 domain, provide evidence suggesting the involvement of the protein in DNA binding. Other domains highly similar to the *Clostridium botulinum* and *C. tetani* neurotoxins, and the Colicin Ia domain (Membrane and cell surface protein) suggest a role in membrane association. Three unique bipartite nuclear targeting sequence motifs indicate involvement of the protein in uptake and accumulation in the nucleus.

3.2 Results

3.2.1 Identification of proteins with sequence similarity to Pc96 amino acid sequence

The nucleotide and predicted amino acid sequence of Pc96 was used in a number of database searches using various BLAST algorithms designed to identify proteins with a high degree of sequence similarity. The results of a BLASTP search using Pc96 as the query are summarised in Table 3.1.

Table 3.1 Proteins with significant similarity to Pc96 identified using BLASTP

Significant hits with BLASTP	E value	GI accession number
hypothetical protein (*Plasmodium yoelii yoelii*)	2e-55	23481602
hypothetical protein C0760c - malaria parasite (*Plasmodium falciparum*)	1e-21	7494250

Table 3.1 shows the two protein sequences identified the from GenBank database at NCBI. Only proteins with sequence regions significantly homologous to Pc96 were chosen according to the statistical E value reported by the software. The E-values for these matches show a high significance, and an extremely low probability of these sequences being matched by chance. Hypothetical protein (un-characterised open reading frame) in the *Plasmodium yoelii yoelii* database (Pyy178) and hypothetical protein C0760c (*Plasmodium falciparum*) (Pf403) showed the highest similarity to Pc96.

Further analysis used the Pc96 sequence in the more sensitive and thorough algorithm, position-specific iterated-BLAST (PSI-BLAST). Proteins with similar regions can be identified using this search method, using profile data structures more sensitive to weak, yet significant matches. The profile constructed after the first search is used in a second pass search of the database, refining the profile. Table 3.2 shows ten database hits from the first PSI-BLAST iteration of the profile. The two most high-scoring hits were again Pyy178 and Pf403, with E values of 1e-115, and 3e-29 respectively. Using this algorithm, Pyy178 showed a higher probability of being homologous to Pc96 than Pf403.

Table 3.2 Proteins with significant similarity to Pc96 identified using PSI-BLAST

Significant hits with PSI-BLAST	E value	GI number
hypothetical protein (*Plasmodium yoelii yoelii*)	1.0e-115	23481602
hypothetical protein C0760c - malaria parasite (*Plasmodium falciparum*)	3.00e-29	7494250
hypothetical protein (*Plasmodium yoelii yoelii*)	0.0003	23482074
probable integral membrane protein PFB0710c - malaria parasite (*Plasmodium falciparum*)	0.0003	7494364
Plasmodium falciparum trophozoite antigen protein like family member (5U678) (*Caenorhabditis elegans*)	0.0005	17566482
hypothetical protein (*Plasmodium yoelii yoelii*)	0.0009	23478292
fork head domain protein, putative (*Plasmodium falciparum 3D7*)	0.001	23619015
myosin heavy chain (*Plasmodium yoelii yoelii*)	0.002	23485072
lipase precursor GehM (*Staphylococcus xylosus*)	0.003	11493973
protein conserved (4F151) (*Caenorhabditis elegans*)	0.003	25145616

A BLASTP search with the hypothetical protein C0760c (Pf403) was performed to identify similar proteins in GenBank. Only two hits were obtained, the hypothetical protein (*Plasmodium yoelii yoelii*) (Pyy178) and another hypothetical protein (*Plasmodium yoelii yoelii*) (Pyy84). As can be seen, there were no hits of proteins with characterised functional domains, allowing easy identification of the protein. Regions reported to be similar in these proteins showed a region of the Pf403 protein to be similar to the Pyy178 and Pc96 proteins, and a separate region to the Pyy84 protein, indicating the presence of two domains, yet the function or identity of these regions was not evident. The significance of these matches, were provided by E values for the alignments are seen on Table 3.3.

Table 3.3 Proteins with significant similarity to Pf403 identified using BLASTP

Significant hits with BLASTP	E value	GI accession number
hypothetical protein (*Plasmodium yoelii yoelii*)	9.00e-53	23481602
hypothetical protein (*Plasmodium yoelii yoelii*)	4.00e-39	23484590
caspase recruitment domain family, member 11 [*Mus musculus*]	0.095	28202025

In order to identify proteins similar to Pyy178, the sequence was used in a BLAST search, reflecting the similarity of Pyy178 to Pf403. The only significant hit was again hypothetical protein C0760c (Pf403), with an E value of 2e-72. No further proteins of any significance were detected, that may identify the protein and its homologues. The database was screened using the Pyy84 sequence, in order to further investigate the region of similarity, shared with Pf403. Due to its conserved nature in these proteins from these two species, it is likely that this region forms a particular function based on a domain or pattern. PSI-BLAST was used to identify other proteins containing similar regions to Pyy84 (Table 3.4).

Table 3.4 Proteins with significant similarity to Pyy84 identified using PSI-BLAST

Significant hits with PSI-BLAST	E value	GI accession number
hypothetical protein C0760c - malaria parasite (*Plasmodium falciparum*)	3.00e-51	7494250
Uncharacterized protein [*Methanopyrus kandleri* AV19]	0.001	20094208
Unknown (protein for IMAGE:3614358) [*Homo sapiens*]	0.001	18089283
enterophilin-2L [*Cavia porcellus*]	0.001	12718845
membrane protein p120 - *Mycoplasma hominis*	0.001	7481994
kinesin-like 7; kinesin-like protein 2 [*Homo sapiens*]	0.003	9910266

Table 3.5 shows a summary of the malaria proteins identified using BLAST showing sequence similarity.

Table 3.5. Features of malaria proteins showing regions similar to Pc96

Protein	GI accession	Size (amino acids)	Name	Molecular weight (Da)
hypothetical protein (*Plasmodium yoelii yoelii*)	23481602	1523	Pyy178	178052.17
hypothetical protein C0760c - malaria parasite (*Plasmodium falciparum*)	7494250	3394	Pyy403	403158.04
hypothetical protein (*Plasmodium yoelii yoelii*)	23484590	713	Pyy84	83779.93
Pc96 antigen (*Plasmodium chabaudi adami*)	Not deposited	942	Pc96	110054.92

The data obtained from the alignments of the BLAST hits, was used to construct a diagram, showing and comparing the sizes and relative positions of the proteins in relation to the *Plasmodium falciparum* protein Pf403 (Figure 3.1).

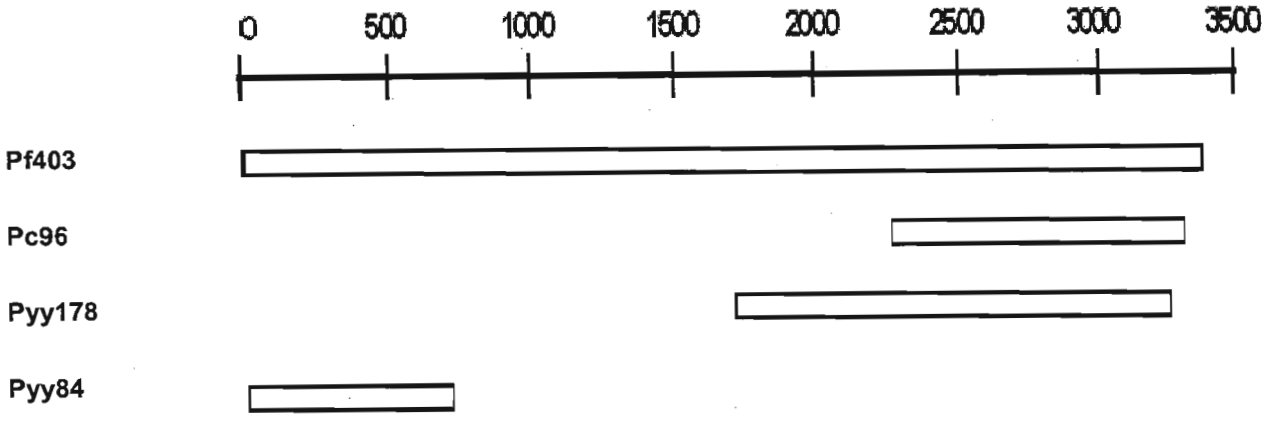


Figure 3.1 Positions of malaria proteins with homology to regions in Pf403. Pf403 (*Plasmodium falciparum*), Pc96 (*Plasmodium chabaudi adami*), Pyy178 (*Plasmodium yoelii yoelii*) and Pyy84 (*Plasmodium yoelii yoelii*) are shown in relation to amino acid number scale.

3.2.2 Prediction of HLA-type T-cell epitopes in Pc96

The SYFPEITHI program (Rammensee *et al.*, 1999) identified a number of putative T cell epitopes in the sequence of Pc96 (Figure 3.2). In total, 17 peptides were predicted. A total of 7 HLA-type epitopes were identified. Of these peptides (ILKMFKLGSCYLYII), (LIKYKNFI) and (LIKYKNFII) were shown to be conserved in Pf403 and Pyy178. As seen for Pc96 in mice infected with *P. chabaudi adami,* Pf403 may contain T cell epitopes capable of eliciting an immune response, directed against an array of *P. chabaudi adami* antigens. These epitopes

conserved in Pf403 are of high interest. The 17 peptides identified by SYFPEITHI were used in a BLAST search for short nearly exact matches of the *Plasmodium falciparum* malaria genome (PlasmoDB). This identified malaria proteins also containing these specific epitopes. Table 3.6 shows the predicted T cell epitopes for Pc96 and lists the malaria proteins found to contain the same epitope sequences. It was interesting to note that peptides (LIKYKNFI) and (LIKYKNFII), which are virtually identical in Pc96, Pf403 and Pyy178, exhibit exact similarity to those contained in a variety of other malaria proteins, as identified by the BLAST search for short nearly exact matches.

FSVSSDHITL IKCNILKMFK LGSCYLYIIN RNLKEIKILF DKINSLEENI QSLNLFINNL 60

KDQNKNNEVI KINNEEQILQ LKNSLQNNEN CINNLNDNLK QKDEMNNSNI KNLIKYKNFI 120

INLVHQTNVF LHIFKTMSTQ QVIQNSEYNQ LTQLLRKELD PYLNDSIMIS ELESKEKSDE 180

ANANNDLLNI PSTFSYENYE HIQIFTNKYN LIIERGQVSI FSDGKIKAPQ NEQTSTFFSN 240

ISSYFHNSNQ YNNTSNSKDA SENTPESPKD SQPQTYQEKE NEEESVRNDN GQMYSVEGNG 300

TDEEDYRDDL EEVEHDEDVE DEEREERDVR VEHDEDDDHD EHEERDEHGE LDEYDDHADH 360

VEHVESDEDD DHDDHDEHEE RDEHDELEEY DDHAGHIEHV EHDEDDEKSK ADGDTEVDED 420

ILSNNKEKEN QDVESYNEYN NDREEYENVS DSEKDEERSS IDNNYDEDHR DENETLQRYE 480

ENVIVENVDN LGEKELLPSN VESNVFENNY TSNDVNGNGI VPEVSEQNKV LDGDPISNIN 540

YENITTDINI NTSNMQNEEN ISPPFSNKLE NNNSTFVNGT TCSYMSEYEF NFFDKNDHPK 600

MKEAKKRKYH SDSEIVKQNN NNKTGYTKKK IRKIDHSYKF NNGDIPTEVV ENQIQQNINF 660

NNLDNNYTNQ LFNNYYEEQK NKYELLDNIN KINQTYNSNN LDAPNDERND NENGNNSLMK 720

SDKESQNENK DDKASDCYVI SSDDDAEVKD YDEEEDEAGE DDEDDDVKDY DEDESNEEID 780

EGEEEEDEAD YEQFDSKGEI NSSVADEGEE SNSDDPDKDM IDYDEMDNAN GMKIEMKRTA 840

KLNVTIVNLN VMKAMLIVIK VMLIVMKVMK IVMKVTLIAM KATFDVMKAT LIMMKVILVM 900

RRMTVVVIVT HILQATTQMK IIKIKNQLLV LQHQMMVRMM DQ 942

Figure 3.2 Distribution of 17 putative T cell epitopes predicted in the Pc96 amino acid sequence. (Epitopes are highlighted red and overlapping epitopes are highlighted in dark red).

Table 3.6. T-cell epitopes predicted by SYFPEITHI for Pc96. Shows the epitope sequence, HLA-type and other *P. falciparum* proteins containing the peptide.

Epitope sequence	HLA-type	*P. falciparum* proteins containing epitope	gi accession number
ILKMFKLGSCYLYII	DRB1*0101	Hypothetical protein C0760c	7494250
LHIFKTMSTQQVIQN	DRB1*0101	-	-
IVNLNVMKAMLIVIK	DRB1*0101	-	-
KATFDVMKATLIMMK	DRB1*0101	-	-
TNVFLHIFKTMSTQQ	DRB1*1101	DNA replication licensing factor MCM2	11559510
EESVRNDNGQMYSVE	DRB1*0301	-	-
EIKILFDKINSLEEN	DRB1*0301	-	-
DGDPISNINY	A1	Trophozoite cysteine proteinase precursor (TCP)	118152
YYEEQKNKY	A1	Protein kinase cdc2-related 1 cdc2-related protein kinase 1 Erythrocyte membrane protein 1 Hypothetical protein C0820w Hypothetical protein PFMAL4P2 Putative dual-specificity protein phosphatase Hypothetical protein PFC0325c MAL1P4.01 PFC0005w	6.3047e+63
DVMKATLIMM	A26	-	-
ILFDKINSL	A*0201	GTP cyclohydrase	4377730
EQKNKYEL	B*08	Variant-specific surface protein 1 homolog PFB1055c Probable membrane associated protein PFB1055c Variant specific surface protein 1 PFMAL3P7 Hypothetical protein PFC1011c PFC0005w (MAL3P8.1)	7.4944e+41
EQKNKYELL	B*08	Variant-specific surface protein 1 homolog PFB1055c Probable membrane associated protein PFB1055c Variant specific surface protein 1 PFMAL3P7 Hypothetical protein PFB1045c	7.4944e+34
LIKYKFI	B*08	Hypothetical protein C0760c Hypothetical protein PFB0520w Hypothetical protein PFMAL4P2 Rhoptry-associated protein 1 Pr86 rhoptry precursor protein	7.4943e+33
LIKYKFNII	B*08	Ribosomal protein S11 Hypothetical protein C0760c Hypothetical protein PFB0250w Hypothetical protein PFMAL4P2 Hypothetical protein C0820w	7.5216e+34
IIKIKQNL	B*08	Novel serine/threonine-specific protein kinase PFB0520w Hypothetical protein MAL4P2.15 DNA-directed RNA polymerase III largest subunit DNA-directed RNA polymerase III largest chain	7.4280e+25
IIKIKNQLL	B*08	Hypthetical protein MAL4P2.15 Novel serine/threonine-specific protein kinase PFB0520w Hypothetical protein C0820w	6.5627e+20

3.2.3 Multiple sequence alignments using ClustalW of protein regions with high similarity to Pc96

Figure 3.3 shows the CLUSTALW alignment of Pc96 and Pf403 showing the high degree of similarity in these regions, including stretches of exact sequence identity.

```
Pc96            ----------------------------FSVSSDHITLIKCNILKMFKLGSCYLYIINRNL 33
Pf403           NNINNVNNINNVKNIVDINNYLVNNLQLNKDNDNIIIIKFNILKLFKLGSCYLYIINRNL 2400
                                           :. ..*:* :** ****:***************

Pc96            KEIKILFDKINSLEENIQSLNLFINNLKDQNKNNEVIKINNEEQILQLKNSLQNNENCIN 93
Pf403           KEIQMLKNQILSLEESIKSLNEFINNLKNENEKNELIKINNFEEILKLKNNLQDNESCIQ 2460
                ***::* ::* ****.*:*** ******::*::**:***** *:**:***.**:**.**:

Pc96            NLNDNLKQKDEMNNSNIKNLIKYKNFIINLVHQTNVFLHIFKTMSTQQVIQNSEYNQLTQ 153
Pf403           NLNNYLKKNEELNKINVKNIFKYKGYIIHLIQQSNVFCKIFKHFNENKIIDQSIINKL-L 2519
                ***: **:::*:*: *:**::***.:**:*::*:*** :*** :. :::*::*  *:*

Pc96            LLRKELDPYLNDSIMISELESKEKSDEANANNDLLNIPSTFSYENYEHIQIFTNKYNLII 213
Pf403           YLKKSFDFYMYDSVIQEIRE----------NKNIIINQDFLTDEYFKHIQTFTKTCNVLI 2569
                 *:*.:* *: **:: .  *          *::::   . :: * ::*** **:. *::*

Pc96            ERGQVSIFSDGK-IKAPQNEQTSTFFSNISSYFHNSNQYNNTSNSKDASENTPESPKDSQ 272
Pf403           QRGYLSILKDTNNDFFIQNKQSNQQGNQNGNHINMCNIYPDDEINVTADQQIFDGTENVQ 2629
                :** :**:.* :     **:*:.   .: ..::: .* * : . .  *.::  :..:: *

Pc96            PQTYQEKENEEESVRNDNGQMYSVEGNGTDEEDYRDDLEEVEHDEDVEDEEREERDVRVE 332
Pf403           QSLQNEEDYVNNEEMYTDKMDLDNNMNGDDDDDDDDDDDDDNNNNNNNNNNNNNNNMGDE 2689
                 .  :*::  ::.    :    . : ** *::*  ** :: ::::: ::::.::.::  *

Pc96            HDEDDDHDEHEERDEHGELDEYDDHADHVEHVESDEDDDHDDHDEHEERDEHDELEEYDD 392
Pf403           DNHLVNAFNNHNLLTNGNVKSDQINNETLERYEENIIQNIYTNDNVDNNQVIENINKILI 2749
                .:.  :  ::.:   :*::.. : : : :*: *.:  ::   :*: ::.:  :::::
```

Figure 3.3 Region of ClustalW multiple sequence alignment of Pc96 (Plasmodium chabaudi adami), Pf403 (Plasmodium falciparum). * Shows exact amino acid matches, : show conserved substitutions, and . denotes semi-conserved substitutions. Predicted T cell epitopes are highlighted on Pc96).

Pc96 aligned with Pf403, shows the first 250 amino acids to be homologous, producing a score of 973. To investigate and compare these sequences, pairwise alignments between Pyy178 and Pf403 was performed using CLUSTALW (Figure 3.4). The alignment of Pyy178 with Pf403 calculated an overall score of 2472. In this alignment it is evident that the first 700 amino acids in the Pyy178 sequence contain large regions of high similarity.

```
Pyy178          ----------------------------------------AQIIDNGKNDEMNMKKKEE 19
Pf403           IVVKDIKNNMRNEIDKLNNDINEKSYEIKLLKHENNNLINEMNILKNKETENMNIKQKEE 1740
                                                        :*:.* :.::**:*:***

Pyy178          EYIELLKKEKENVEKKFENTSEKYNEQINLNKKLTDDINLLISTHKEQVKVLNEQINILK 79
Pf403           DYIKLIKKDKTNIQNEYNDLLEKYNEVVVKNNMLYNDMNVLLKEHKEEIFLLKENIKILQ 1800
                :**:*:**:* *::::::::  ***** :  *: * :*:*:*:. ***:: :*:*:*:**:

Pyy178          KDNKYLSENLEKEVEISDNNLIIKNRLEQLVEINKDLYKEVQENYNNTEKMKFKIIELKE 139
Pf403           KDNTYLNDMFKNQINYVDNNLLKN-RLDQLFNINQDLQKHLDTNQKHLEQLKYDYIEIKE 1859
                ***.**.: :::::::  ****: : **:**.:**:** *.:: * :: *::*:. **:**

Pyy178          NIRVQKETHLQQQKCIIELR----------------------------TKLIDNNMASQ 170
Pf403           RLKIEKTKINKQEKYIIQLQKDNNLILNDFNSTTTTTNNNNNNDNNNDNNNDNNDNNNDT 1919
                .:::::* .  :*:* **:*:                                .:  ::*  .

Pyy178          KDEYVSNLKMNLESSRMQLKDLCDQLEKGNLNEKKCNMKIQMLETKLXKEEKERKRYQIE 230
Pf403           YQQFIHSLKANLENSRLELKELSNLNEKIQLSDEKNRMKITILEDKLFKNEKDKMKLQQI 1979
                 :::: .** ***.**::**:*.:  ** :*.::* .*** :** ** *:**:: : *

Pyy178          LANKNSTDSISYNKLKTQVEMVTEENKILLLRKECYEKEIEQLKRDSQFFNSTKNNDINI 290
Pf403           IDDNNKNYMIQYNKLKTNLDMLSEENRMLLLNKEEYEKQIEQLNHDHKLFISTKNNDIQI 2039
                : ::*..  *.******:::*::***::***.** ***:****::* ::* *******:*

Pyy178          IERDVLKKQVEEHITKINEKDKQIVNLNFEIKKLYNQLEEMKERMNRIETTTP------- 343
Pf403           IENEKLQEQVDQYITTINEKDKIIVHLNLQIKKLANQNEHMRSRCDIFNVAHSQDNIKND 2099
                **.: *::**:::**.****** **:**::**** ** *.*:.* : ::.: .

Pyy178          ------ILDGPTDEASTLNIDNSKTSQNNQ------------------------------ 367
Pf403           HMVVGEDIMGDTNHDVNKNIDQGTNQHINQGTNQHINQGTNQHDTCDGPNYNYVKVQNAT 2159
                       : * *:.  . ***:....: **

Pyy178          ----------LSLEIYKYINENIDLTAELENKNDVIEQMKEDXKNKNKEIAKLNKDVIN 416
Pf403           NREDNKNKERNLSQEIYKYINENIDLTSELEKKNDMLENYKNELKEKNEEIYKLNNDIDM 2219
                          ** *************:***:***::*: *:: *:**:** ***:*:

Pyy178          LSTNYDKLKESIYMMEKHKTNLNEYIKQKDEIISLLQQKGNNNNNSPTKFCNLVEN---N 473
Pf403           LSNNCKKLKESIMMMEKYKIIMNNNIQEKDEIIENLKNKYNNKLDDLINNYSVVDKSIVS 2279
                **.* .****** ****:*  :*: *::*****. *::* **: :.  :  .:*::   .

Pyy178          NTNKNGANEDDFDSPIKYNINLKNNNSKEIEEKIMGMEEIMNTSCEEVINTLKKIPQTNV 533
Pf403           CFEDSNIMSPSCNDILNVFNNLSKSNKKVCTNMDICNENMDSISSINNVNNINNVNNINN 2339
                  :...  . . :. ::   **.:.*.*   :  :  *:: . *. : :*.:::: : *

Pyy178          ENDDN--NNLNSSKNQFSVN------------SDHITLIKCNILKMFKLGSCYLYIINRN 579
Pf403           VNNINNVNNINNVKNIVDINNYLVNNLQLNKDNDNIIIIKFNILKLFKLGSCYLYIINRN 2399
                 *: *  **:*. ** ..:*            .*:* :** ****:**************

Pyy178          LKEIKILKDKIHSFEENIQSLNLFINNLKDQHKNNEMIKINNEEQIVELKNNLKNNENCI 639
Pf403           LKEIQMLKNQILSLEESIKSLNEFINNLKNENEKNELIKINNFEEILKLKNNLQDNESCI 2459
                ****::**::* *:**.*:*** ******:::::**:***** *:*::*****::**.**

Pyy178          NNLNDNLKQKDEMNNSNIQNLIKYKSFIINLVYQSNIFFHIFKNINKQKVIQNSIFNQLT 699
Pf403           QNLNNYLKKNEELNKINVKNIFKYKGYIIHLIQQSNVFCKIFKHFNENKIIDQSIINKLL 2519
                :***: **:::*:*: *::*::***.:**:*: ***:* :***::*::*:*::**:*:*
```

Figure 3.4 Region of ClustalW multiple sequence alignment of Pyy78 (*Plasmodium yoelii yoelii*), Pf403 (*Plasmodium falciparum*). * Shows exact amino acid matches,: show conserved substitutions, and . denotes semi-conserved substitutions.

```
Pyy178          LNSSKNQFSVNSDHITLIKCNILKMFKLGSCYLYIINRNLKEIKILKDKIHSFEENIQSL 600
pc96            -------FSVSSDHITLIKCNILKMFKLGSCYLYIINRNLKEIKILFDKINSLEENIQSL 53
                       ***.************************************ ***:*:*******

Pyy178          NLFINNLKDQHKNNEMIKINNEEQIVELKNNLKNNENCINNLNDNLKQKDEMNNSNIQNL 660
pc96            NLFINNLKDQNKNNEVIKINNEEQILQLKNSLQNNENCINNLNDNLKQKDEMNNSNIKNL 113
                **********:****:*********::***.*:************************:**

Pyy178          IKYKSFIINLVYQSNIFFHIFKNINKQKVIQNSIFNQLTQLR-KELDFYLNDSIMISELE 719
pc96            IKYKNFIINLVHQTNVFLHIFKTMSTQQVIQNSEYNQLTQLLRKELDPYLNDSIMISELE 173
                ****.******:*:*:*:****.:..*:***** :******  **** ************

Pyy178          NKEKSNLNNFENNLLNIVSTFSCENYEHIQIFTNKFNLIIERGQFSILADSKIQPSQNEK 779
pc96            SKEKSDEANANNDLLNIPSTFSYENYEHIQIFTNKYNLIIERGQVSIFSDGKIKAPQNEQ 233
                .****:  * :*:**** **** ************:********.**::*.**:..***:

Pyy178          QSENQNEKQNENLSENLSENLSEKQNENLSEKQNEKSNNFLSN-ISSYFQSSNKYNNAPT 838
pc96            TSTFFSNISSYFHNSNQYNNTSNSKDASENTPESPKDSQPQTYQEKENEEESVRNDNGQM 293
                 *   .: ..   ..*  :* *:.:: . .  :. *..:  :   ..  :.* : :*.

Pyy178          FKELNENTTEPQKDTQIQTVQISQEKQNEEENTKNENIIDTQVNNENYQIDALKNEEYND 898
pc96            YSVEGNGTDEEDYRDDLEEVEHDEDVEDEEREERDVRVEHDEDDDHDEHEERDEHGELDE 353
                :.  .:.* * :   ::: *: .:: ::**.: :: .: . : ::.: : :  :: * ::

Pyy178          DQMYSVEDDGSEEEEYRFNEEHAEVEIDGKQAEVETDGKHGEVEIDGEQAEVETDRKHGE 958
pc96            YDDHADHVEHVESDEDDDHDDHDEHEERDEHDELEEYDDHAGHIEHVEHDEDDEKSKADG 413
                 : :: . :  *.:*   :::* * *  .:: *:*  ..*.    . *: * : . * .

Pyy178          VGEHDEDMLNDSKEKESSNIESYNEYNSNREEYVDVSDNDEEKEGDEFNYSNKEKSSIDN 1018
pc96            DTEVDEDILSNNKEKENQDVESYNEYNNDREEYENVSDSEKD----------EERSSIDN 463
                  * ***:*.:.****..::*******.:**** :***.:::          :*:*****

Pyy178          NYDESNCDKSNCEENETLQRYEENVIVENGNNLGEKELLTSNVESNLFENDDNNNDDNND 1078
pc96            NYDEDHRD-----ENETLQRYEENVIVENVDNLGEKELLPSNVESNVFENNYTSNDVN-- 516
                ****.: *     ********************* :********.******:***: ..** *

Pyy178          DNNGDNGNNIVSEVGEENKLLGDNSVNHINYENIKTDVAVNVDIDANTSNLQNEENISPS 1138
pc96            ------GNGIVPEVSEQNKVLDGDPISNINYENITT------DININTSNMQNEENISP- 563
                      **.**.**.*:**:*..:.:.:******.*      **: ****:********

Pyy178          AFSNKLENQNNTFVNDSTRSYLSECEFNSFDKNDNLKIKESKKRKYHSDTEIVKQNNNNK 1198
pc96            PFSNKLENNNSTFVNGTTCSYMSEYEFNFFDKNDHPKMKEAKKRKYHSDSEIVKQNNNNK 623
                .*******:*.****.:* **:** *** *****: *:**:********:**********

Pyy178          VEYTKKKMRKMNNSHKFTSRGTSDVTVESPTPQKLSEKLSEKLSEKLNGNINGNINGDDN 1258
pc96            TGYTKKKIRKIDHSYKFNNGDIPTEVVEN-------------------QIQQNINFNNL 663
                . *****:**:::*:**.. . .  .**.                    :*: *** ::

Pyy178          ENEDDENYASESFDNYYDEKK-KYEILENISKLNQACNNNNLDATSDGNNYAKKNEHNFL 1317
pc96            DN----NYTNQLFNNYYEEQKNKYELLDNINKINQTYNSNNLDAPNDERN-DNENGNNSL 718
                :*    **:.: *:***:*:* ***:*:**.*:**: *.*****..* .*  ::* :* *

Pyy178          MKNDKESQNENAKE-SDSCYVISSDDDNRQIQDFEEDEEDEEDEEDEEDEEDEQDEADEY 1376
pc96            MKSDKESQNENKDDKASDCYVISSDDD-AEVKDYDE-EEDEAGEDDEDDDVKDYDEDESN 776
                **.******** .: :..*********  :::*::* **** .*:**:*: .: ** :.

Pyy178          DEYDDVENDDEYDEEYEEEEDDGEMD--EADDGEMDEADDGEMNEADDGEMDEADYDQFD 1434
pc96            EEIDEGE-EEEDEADYEQFDSKGEINSSVADEGEESNSDDPDKDMIDYDEMDNANGMKIE 835
                :* *: * ::* : :**: :..**::   **:** .::** : :  * .***:*:  :::

Pyy178          SK--GELNSSIDDDSNSDDADKDMIDYDDIDNINENQNEEDSEVENEDDNNFLDE----- 1487
pc96            MKRTAKLNVTIVN-LNVMKAMLIVIKVMLIVMKVMKIVMKVTLIAMKATFDVMKATLIMM 894
                 *  .:** :* :  *  .*   :*.   *     :   : : :  :   :.:.
```

Figure 3.5. Region of ClustalW pairwise sequence alignment of Pyy78 (*Plasmodium yoelii yoelii*), Pc96 (*Plasmodium chabaudi adami*). * Shows exact amino acid matches, : show conserved substitutions, and . denotes semi-conserved substitutions.

A score of 2435 resulted from the alignment of Pc96 with the *Plasmodium yoelii yoelii*, Pyy78 (Figure 3.6). It can be seen that the entire length of the Pc96 sequence is similar to a portion of the Pyy178 protein. This indicates a high probability of these sequences being homologous. The alignment shows several extended regions of sequence similarity. The full length of the Pc96 sequence shows high similarity to Pyy78 from residues 546 to 1523.

```
Pyy178      --VENDDNNNLNSSKNQEFSVNSDHITLIKCNILKMFKLGSCYLYIINRN  580
Pc96        ------------------FSVSSDHITLIKCNILKMFKLGSCYLYIINRN  32
Pf403       NNVKNIVDINNYLVNNLQLNKDNDNIIIIKFNILKLFKLGSCYLYIINRN  2399
                              :. ..*:* :** ****:**************

Pyy178      LKEIKILKDKIHSFEENIQSLNLFINNLKDQHKNNEMIKINNEEQIVELK  630
Pc96        LKEIKILFDKINSLEENIQSLNLFINNLKDQNKNNEVIKINNEEQILQLK  82
Pf403       LKEIQMLKNQILSLEESIKSLNEFINNLKNENEKNELIKINNFEEILKLK  2449
            ****::* ::* *:**.*:*** ******:::::**:***** *:*::**

Pyy178      NNLKNNENCINNLNDNLKQKDEMNNSNIQNLIKYKSFIINLVYQSNIFFH  680
Pc96        NSLQNNENCINNLNDNLKQKDEMNNSNIKNLIKYKNFIINLVHQTNVFLH  132
Pf403       NNLQDNESCIQNLNNYLKKNEELNKINVKNIFKYKGYIIHLIQQSNVFCK  2499
            *.*::**.**:***: **:::*:*: *::*::***.:**:*: *:*:* :

Pyy178      IFKNINKQKVIQNSIFNQLTQLR-KELDFYLNDSIMISELENKEKSNLNN  729
Pc96        IFKTMSTQQVIQNSEYNQLTQLLRKELDPYLNDSIMISELESKEKSDEAN  182
Pf403       IFKHFNENKIIDQSIINKLLYLK-KSFDFYMYDSVIQEIRENK-------  2542
            *** :. :::*::*  *:*  *  *.:* *: **:: .  *.*

Pyy178      FENNLLNIVSTFSCENYEHIQIFTNKFNLIIERGQFSILADSK-IQPSQN  778
Pc96        ANNDLLNIPSTFSYENYEHIQIFTNKYNLIIERGQVSIFSDGK-IKAPQN  231
Pf403       ---NIIINQDFLTDEYFKHIQTFTKTCNVLIQRGYLSILKDTNNDFFIQN  2589
               :::   . :: * ::*** **:. *::*:** .**: * :     **
```

Figure 3.6 ClustalW multiple sequence alignment of Pyy178 (*Plasmodium yoelii yoelii*), Pca96 (*Plasmodium chabaudi adami*) and Pf403 (*Plasmodium falciparum*). * Shows exact amino acid matches, : show conserved substitutions, and . denotes semi-conserved substitutions.

The alignment score for these three proteins was 5462. Pc96 shows homology from initial amino acids in the sequence, while the similar region in Pyy78 spans from amino acids 532-729. Pf403, being a much larger protein, is similar across a 221 amino acid region, from residues 2367 to 2589. From this multiple sequence alignment it can be seen that in these regions, there

are stretches of highly conserved amino acids. From these results it can be seen that Pf403 is more similar to Pyy78 than Pc96. The putative T cell epitopes, predicted in Section 2.1.2, are indicated in the alignments, along with the casein kinase II phosphorylation site, shared between, Pf403 and Pc96.

3.2.4 Secondary structure features of proteins similar to Pc96

The program PREDICT7 was used to calculate profiles based on secondary structure features such as antigenicity, flexibility, hydropathy and surface probability, to further illustrate and investigate the similarities between Pf403, Pc96 and Pyy178. Figure 3.7 shows these plots along with the predicted T cell epitopes.

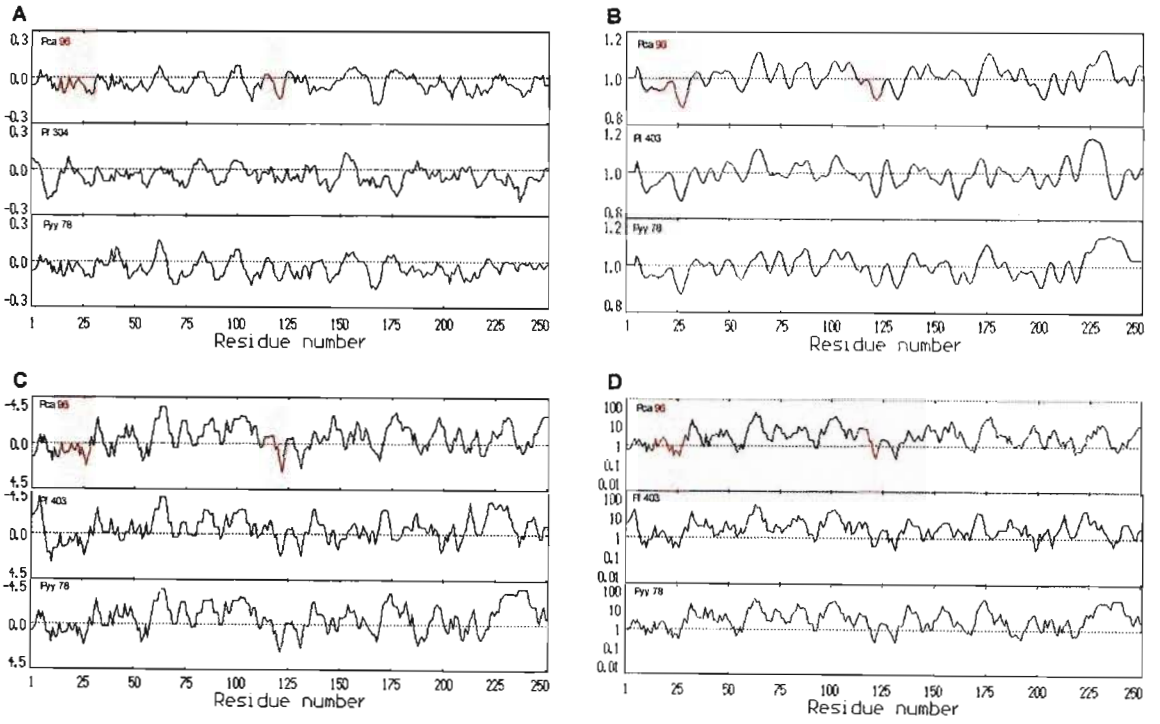


Figure 3.7 Secondary structure features predicted for 250 bp region in Pca96, Pf403 and Pyy178. A - Antigenicity plot (Welling et al., 1985); B - Flexibility plots (Karplus and Schultz, 1985), C - Hydropathy plots (Kyte and Doolittle, 1982); D - Surface probability plots (Janin, 1978). The position of predicted putative T cell epitopes are indicated in red.

As can be seen for all profiles, the pattern of the plot is similar for the three proteins, for the 250 bp region of high similarity. The plots generated over the T cell epitopes show the nature of the secondary structure, and indicate the similarities of the sequences containing the conserved epitopes on Pf403 and Pyy178.

3.2.5 Motif and pattern recognition using PROSITE and domain analysis using 3D-PSSM

Table 3.7 shows the identification of PROSITE profiles for the proteins used in this study. These proteins contain regions of sequence similarity to that of Pc96. The N-glycosylation sites, protein kinase C phosphorylation sites, casein kinase II phosphorylations sites and tyrosine kinase phosphorylation sites, are detected quite frequently within the sequences of these proteins. The Leucine zipper pattern detected in all proteins except Pyy78, has previously been shown to play a role in gene regulation. Although false hits have been identified in protein sequences known to lack this motif, verification is usually obtained by the identification of DNA binding motifs. In Pf403 there are three bipartite nuclear targeting sequences, increasing the likelihood of Pf403 playing a role in DNA binding or gene regulation. Although it may be difficult to draw any structural or functional conclusions about these proteins, these consensus sequences comprising of detected block motifs in the PROSITE database can be used for comparative analysis of 3D structure, and the possible assignment of function for certain domains.

Table 3.7. The presence and number of motif and sequence patterns identified using PROSITE for Pc96, Pf403, Pyy178 and Pyy84.

PROSITE pattern entry	Pc96	Pf403	Pyy78	Pyy84
N-glycosylation site	20	38	27	10
cAMP and cGMP-dependent protein kinase phosphorylation site	1	6	2	2
Protein kinase C phosphorylation site	9	16	26	6
Casein Kinase II phosphorylation site	24	49	35	14
Tyrosine kinase phosphorylation site	3	18	7	2
N-myristoylation site	2	6	8	8
Leucine Zipper pattern	2	1		1
Sulfonation	12	25	14	1
NUCLEAR Bipartite nuclear targeting sequence		3		

Tables 3.8 and 3.9 show the results of screening the 3DPSSM database with the four malaria proteins under study (Pf403, Pc96, Pyy178 and Pyy84). This particular program recognises certain combinations and positions of signature sequences and motif patterns from the sequence and compares this to a database with a collection of known 3D structures. The statistical relevance of the matches were reported as E values, reflecting the given probability of the domain recognition pattern occurring by chance, based upon the initial alignments in the PSSM algorithm. In general, E values below 0.1 are considered significant, and matches with values below 1 can still be investigated.

In an attempt to recognise specific structural and functional domains, overlapping regions of Pc96 and homologous proteins Pyy178 and Pyy84 were submitted for fold recognition at the 3DPSSM site. This provided clues as to the identity of these regions within the proteins. Analysis of Pc96 (Table 2.8) shows the identification of a region highly similar to the botulinum neutoxin, and the yeast t-snare protein (sso1), a membrane protein. The Pyy178 protein sequence was split into three regions, identifying the colicin Ia domain in the first, and the botulinum neurotoxin in the second and third. The hydrolase domain was again found, similar to that of phospholipase c beta and a region homologous to a hydrolase activator, proteasome component y7, was identified. Pyy84 showed homology to Botulinum neurotoxin and interferon-induced guanylate-binding protein 1, a DNA binding protein.

These features compare to that of Pf403, indicating a similarity in the 3D structure of Pc96 as well as the identified regions of sequence similarity. As can be seen on Table 3.9, a significant match with Colicin Ia was obtained located in the first 500 residues of Pf403. This indicates the possible role of Pf403 in biological membrane association as colicin Ia is involved in ion channel forming in membranes and cell surfaces (Stroud *et al.*, 1998). The interferon-induced guanylate-binding protein 1, may also indicate DNA-binding properties. The domain that showed the highest degree of structure recognition was the Botulinum neurotoxin catalytic domain. This protein is a zincin metalloprotease, and is also involved in membrane association. It is interesting to see the recognition of this domain, and that of colicin IA, in two specific regions of Pf403. The hydrolase domain found in angiostatin, was recognised in the last 500 residues, along with translation initiation factor eIF1a, a DNA-binding protein. Another putative hydrolase domain was detected in the first series of residues screened, similar to that of phospholipase c beta.

Table 3.8. 3D structure comparisons using 3DPSSM and identification of putative domains in Pc96, Pyy178 and Pyy84.

E value of Protein region	Protein Name	Description	Organism
Pc96(1-480)			
0.0259	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	Clostridium botulinum
0.12	yeast t-snare protein sso1	membrane protein	Yeast
Pc96(481-942)			
0.0254	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	Clostridium botulinum
0.0597	GST-alpha-Na,K-ATPase	crystal structure of the ankyrin binding domain of2 alpha-Na,K-atpase as a fusion protein with glutathione3 s-transferase - ankyrin binding.	
Pyy178(1-488)			
0.129	Colicin Ia	Membrane and cell surface protein - ion channel forming	Enteric bacteria
0.329	Interferon-induced guanylate-binding protein 1	c-terminal domain - binding of guanine nucleotides	Homo sapiens
0.489	phospholipase c beta	c-terminal - hydrolase enzyme	
Pyy178(488-977)			
0.000771	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	Clostridium botulinum
0.0487	yeast t-snare protein sso1	membrane protein	Yeast
Pyy178(977-1523)			
0.00312	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	*Clostridium botulinum*
0.052	proteasome component y7	hydrolase/hydrolase activator	
Pyy84 (1-713)			
0.171	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	*Clostridium botulinum*
0.642	Interferon-induced guanylate-binding protein 1	c-terminal domain - binding of guanine nucleotides - signalling protein	*Homo sapiens*

Table 3.9. 3D structure comparisons using 3DPSSM and identification of putative domains in Pf403 (*Plasmodium falciparum*).

E value (region of Pf403)	Protein name	Description	Organism
Pf403(1-780)			
0.042	Colicin Ia	Membrane and cell surface protein - ion channel forming	Enteric bacteria
0.326	Interferon-induced guanylate-binding protein 1	c-terminal domain - binding of guanine nucleotides - signalling protein	*Homo sapiens*
0.445	phospholipase c beta	c-terminal - hydrolase enzyme	
Pf403(781-1560)			
0.0000692	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	*Clostridium botulinum*
Pf403(1561-2340			
0.0605	Colicin Ia	Membrane and cell surface protein - ion channel forming	Enteric bacteria
0.544	Interferon-induced guanylate-binding protein 1	c-terminal domain - binding of guanine nucleotides	*Homo sapiens*
Pf403(2341-3120)			
0.0000688	Botulinum neurotoxin serotype A	Clostridium neurotoxins, catalytic domain - Zincin-like metalloprotease	*Clostridium botulinum*
Pf403(3121-3394)			
0.254	Angiostatin	hydrolase	Homo sapiens
0.265	Translation initiation factor eIF1a	Nucleic acid binding - Cold shock DNA-binding domain	Homo sapiens

These results together were used to construct a map of Pf403 and related proteins, according to regions where putative domains of known 3D structure were identified. Figure 2.8 shows Pf403, Pyy178, Pyy84 and Pc96, with the most significant matches obtained from the 3DPSSM analysis. Both Pyy84 and Pyy178, two separate *P. yoelii yoelii* proteins showed homology to two separate regions in Pf403, sharing similar comparisons to the 3D structures in these domains.

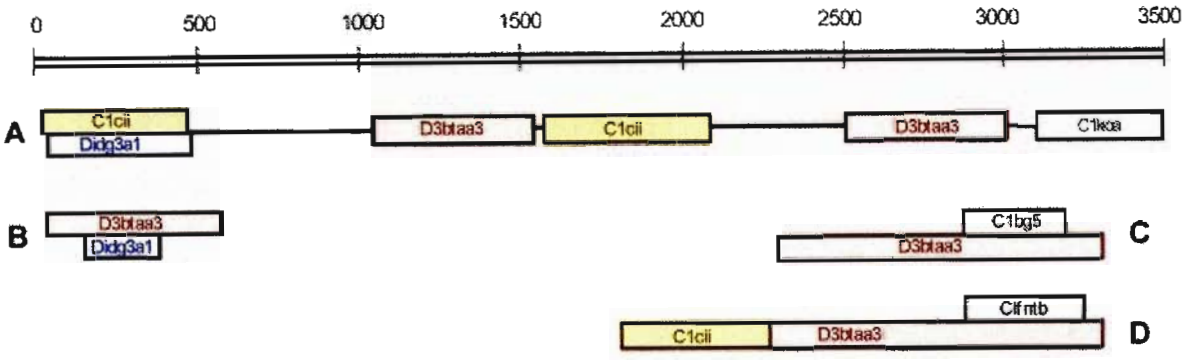


Figure 3.8 Diagram showing domain regions identified by 3DPSSM analysis. Proteins A - Pf403; B - Pyy84;C - Pc96; D - Pyy178. Structures identified by 3DPSSM shown on the diagram are: C1cii (Colicin Ia) - shaded orange; D3btaa3 (Botulinum neurotoxin serotype A) - shaded red; C14oa (Angiostatin hydrolase); C1bg5 (GST-alpha-Na,K-ATPase) and Clfntb (proteasome component y7 hydrolase activator); Didg3a1 (Interferon-induced guanylate-binding protein 1) - shaded blue.

Due to the fact that two regions in Pf403, and regions in Pc96, Pyy84 and Pyy178 showed a similar 3D structure to that of Botulinum neurotoxin serotype A, with comparatively high estimations of statistical similarities, the 3D structure viewing software Rasmol was used to generate structures of the protein (Figure 3.9). Both the backbone and ribbon structures were shown, to emphasise features of the shape, and the positioning of helices and sheets making up the protein.

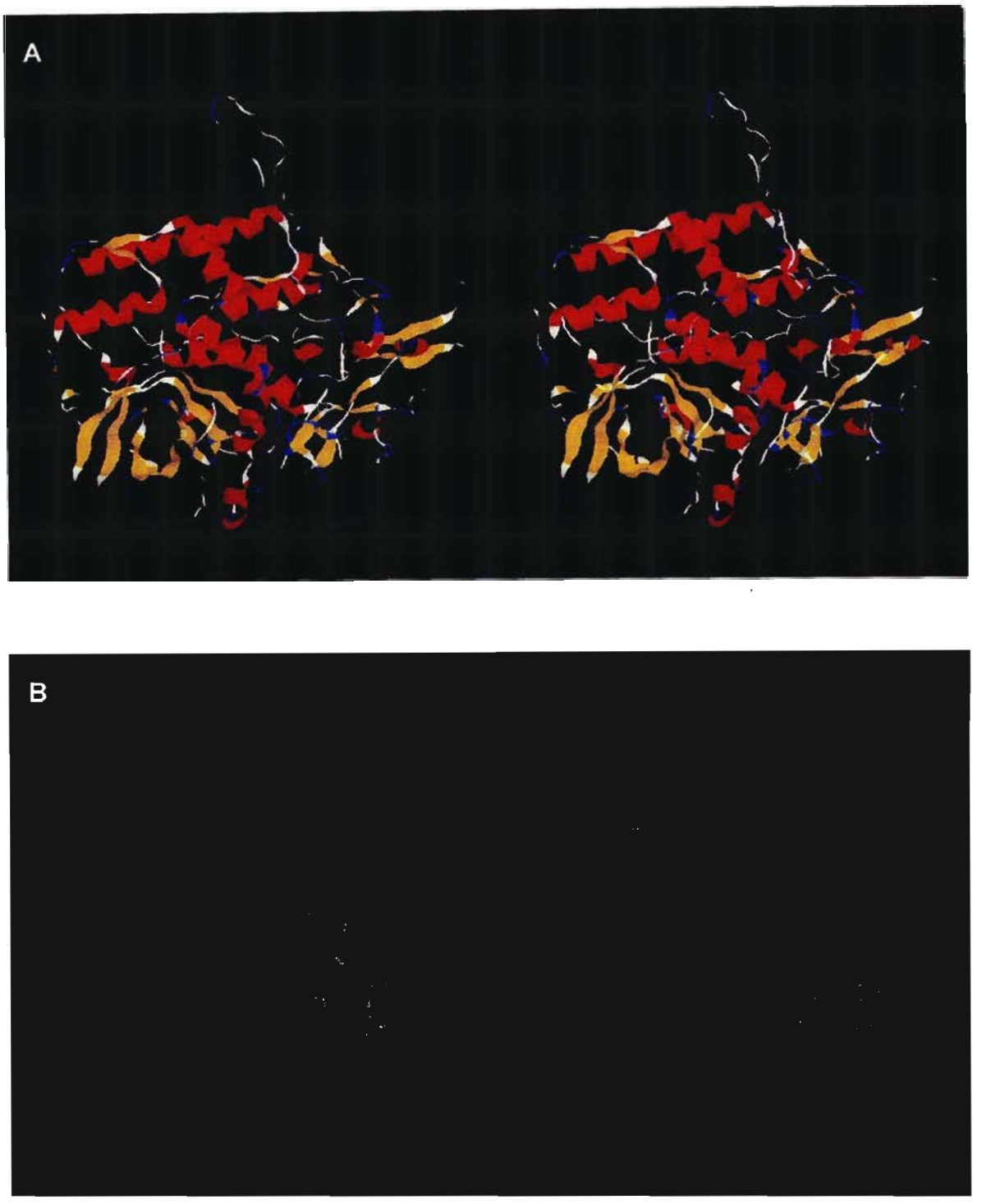


Figure 3.9. The 3 dimensional structures of the Botulinum neurotoxin domain, identified by 3DPSSM. Images are shown as stereograms for 3D visualisation. Panel A shows the ribbon structure and panel B, the backbone line structure.

3.3 Discussion and conclusion

The assignment of function to an unknown protein from its sequence relies on the positive identification of highly similar, previously characterised proteins. This technique has been previously used by Merckx et al (2003) for the identification of cyclin related proteins in the form of sequenced open reading frames in the *Plasmodium falciparum* sequence database (PlasmoDB) (Gardner *et al.*, 2002). Searches of sequence databases with the predicted amino acid sequence of Pc96, identified several proteins containing regions of extensive homology. Through variations of the BLAST algorithm, it was evident that all of the highest scoring hits were other plasmodial proteins. Pf403 (*Plasmodium falciparum*) and Pyy178 (*Plasmodium yoelii yoelii*) were hypothetical proteins found to contain regions homologous to the Pc96 sequence.

Hypothetical proteins are based on the detection of open reading frames (ORF) within the sequenced genomes, and properties of the proteins such as structure and function are unknown. The extent of similarity between the plasmodial proteins was further analysed by CLUSTALW pair-wise and multiple sequence alignment algorithms. Pyy178 showed the highest similarity to Pc96. Pf403 is a much larger protein but contains a region spanning several hundred amino acids with an extremely high degree of sequence similarity to that of Pc96 and Pyy178. Pyy84 was subsequently shown to be similar to a different region in Pf403.

The identification of T cells epitopes relied on the SYFPEITHI software for the selection of a number of putative peptides within the Pc96 sequence. *P. chabaudi adami* infection in B-cell deficient mice, activates a T-cell dependent mechanism which terminates acute malaria in a similar manner to that seen in immunologically intact mice (Grun and Weidanz, 1981). Due to the fact that the Pc96 protein was shown to contain T cell epitopes capable of eliciting a protective immune response in the mouse malaria model *Plasmoidum chabaudi adami* (Wanidworanun *et al.*, 1989), these epitopes were evaluated and compared to the Pf403 sequence. The identification of such epitopes in Pf403 is of interest, as there may be a similar epitope in falciparum malaria capable of generating a multiple antibody response. The fact that the epitopes were predicted using SYFPEITHI's algorithm of recognising putative peptides, indicates that results are not conclusive, and further research and evaluation is needed to verify the epitopes and their function within this antigen.

Three epitopes predicted in Pc96 (ILKMFKLGSCYLYII), (LIKYKNFI) and (LIKYKNFII) were shown to be conserved both in Pf403 and Pyy178 in a region of amino acids highly similar between the three molecules. Multiple sequence alignments of this region showed the similarity of the amino acids, and the secondary structure conformations reflecting the nature of solvent exposure in this region to be similar. The epitopes were used in a BLAST search of the *P. falciparum* genome database (PlasmoDB) in order to identify proteins containing the same peptide. Due to the fact that the T cell clone, in the protective chabaudi adami response provided help to a variety of B cell clonotypes, it is likely that the epitope was contained in several other chabaudi adami malaria proteins. The two conserved epitopes (LIKYKNFI) and (LIKYKNFII), and (YYEEQKNKY), were shown exist in a few other *Plasmodium falciparum* proteins. The correct identification of T cell epitopes from immunologically relevent antigens is a critical step in the development of vaccines, and in this chapter, the initial steps were taken for the selection of candidate peptides from the sequence. Subsequent verification of their *in vivo* relevance will be necessary.

Scanning the protein sequences against a PROSITE database, identified several unique features of the proteins similar to Pc96. N-glycosylation sites involve the attachment of carbohydrate chains to proteins by N-linkages with the nitrogen atom of asparagine residues. These proteins (glycoproteins) form integral components of biological membranes, lysosomal enzymes and the extracellular matrix (Karp, 1996). Potential N-glycosylation sites correspond to the consensus sequence Asn-Xaa-Ser/Thr. The folding of the protein plays an important role in the regulation of N-glycosylation. Several protein kinase phosphorylation sites were found. cAMP- And cGMP-dependent protein kinases share a preference for the phosphorylation of serine and threonine residues found close to at minimum two consecutive N-terminal basic residues. Protein kinases are enzymes that transfer phosphate groups to other proteins, involved in the regulation of a number of activities such as hormone activation, cell division and gene expression. They play an important role in cell signalling and the communication between cells and their environment.

The presence of a unique leucine zipper pattern in Pc96, Pyy403 and Pyy84 may indicate some role for the protein in gene regulation. This motif is present in many gene regulatory proteins such as cAMP response element (CRE) binding proteins, CCATT-box and enhancer binding proteins and various transcription factors. The motif consists of a

repetition of leucine residues at every seventh position over a region of the sequence including eight helical turns. Segments containing these periodic leucine residues are of an alpha-helical conformation. The leucine side chains in the coil, interact with those from a similar conformation on another polypeptide, causing dimerization and a coiled coil structure of these two regions. The motifs consensus pattern is L-x(6)-L-x(6)-L-x(6)-L (Landschulz et al., 1988).

The presence of three bipartite nuclear targeting sequences in Pf403 further suggests a DNA binding role. These patterns are present on proteins enabling the successful passage through the nuclear membrane. Uptake of the protein by the nucleus depends on this signal. Studies on the Large T antigen of SV40 in viruses have resulted in the identification of sequences required for nuclear translocation (Dingwall and Laskey, 1996). The bipartite nuclear targeting sequence is identified by two adjacent basic amino acids (Arg or Lys) followed by a extended spacer region (10 residues) and three basic residues (Arg or Lys) in the five following positions after the spacer. These motifs along with the leucine zipper pattern, suggest strongly that Pf403 is involved in gene regulation and possibly DNA binding.

3D structure analysis was performed using the 3DPSSM database for consecutive regions of Pc96, Pf403, Pyy78 and Pyy84 in an attempt to identify the structure and possible functions of the domain regions. The query sequences were deposited into the 3DPSSM software which screens a library of known structures based on the spacing of patterns and motifs. Two specific regions of Pf403 were identified to show high homology to that of a Zincin-like metalloproteases closely resembling the catalytic domain of *Clostridium tetani* and *Clostridium botulinum* neurotoxins. It is interesting to note that there are two of these domains present, along with a comparison to two colicin Ia domains. There are seven serologically different botulinum neurotoxins, involved in the inhibition of neurotransmitter release from peripheral cholinergic synapses in nerve cells. Two regions in Pf403 and other proteins studied contain 3D structure highly comparable to that of the catalytic domain of the neurotoxin A metalloprotease. Colicin Ia is a hydrolase enzyme involved ion channel formation in *E. coli* (Wiener *et al.*, 1997). This structure is associated with biological membranes.

There were several attributes of Pf403 that indicated its possible role in DNA-binding. The identification of a 3D structure similar to that of interferon-induced guanylate-binding protein 1 (GBP1) indicates that this protein may be characterised by its possbile ability to bind to guanine residues (GmP). GDP, and GTP). The presence of a unique leucine zipper motif in Pf403 is also indicative of DNA binding. Together, the DNA binding properties and membrane association comparisons, along with the high homology to the Zincin-like metalloprotease, provide the basis for more detailed analysis and identification of Pf403 and related malaria molecules. The fact that no full-length protein was found conclusively similar in primary structure and amino acid sequence indicates that the structure is unique and novel in its conformation, possibly functional in malaria parasites and no other organism. Further functional, structural and possibly enzymatic studies on these proteins will rely on the availability of large amounts of protein. Due to the fact that isolation of sufficient native protein from the malaria parasite is impractical and difficult, the genes can be cloned and the proteins recombinantly expressed, allowing relatively large amounts of the protein to be present for experimental evaluation. The subsequent chapters deal with the molecular cloning of a region of the Pf403 gene, most similar to the Pc96 protein, containing these putative T cell epitopes, and other features discussed.

CHAPTER 4
PCR-CLONING OF 817 BP REGION OF *PLASMODIUM FALCIPARUM* MAL3P6.11 GENE

4.1 Introduction

The polymerase chain reaction is a highly useful technique in the field of molecular biology for the *in vitro* amplification of DNA regions (Innis and Gelfand, 1990). Oligonucleotide primers complementary to the ends of a defined region of template DNA are used in an enzymic reaction involving DNA polymerase, deoxyribonucleotide triphosphates (dNTPs) and a specific set of reaction conditions. Oligonucleotide primers are short, single-stranded DNA molecules, also known as amplimers. The primers are extended on single-stranded denatured template DNA by the DNA polymerase enzyme. This results in double-stranded copies of the target region generated in a repeating cycle of heating and cooling. The template DNA is heat denatured in the range 93-100□C, forming single-stranded template DNA. The temperature is lowered to hybridisation temperature allowing the primers to anneal to their specific complementary regions on the template and then raised to extension temperature for synthesis by the polymerase enzyme in a 5' to 3' direction. This is normally repeated for at least 20 cycles copying the target DNA in an exponential manner (Mullis and Faloona, 1987).

The denaturation and annealing of oligonucleotide primers each cycle, requires polymerase enzymes capable of withstanding the high denaturation temperature. To avoid having to add fresh aliquots of polymerase each cycle, a thermostable polymerase was isolated from *Thermus aquaticus* (*Taq*). This has been a huge technological advancement, allowing routine use of PCR in clinical and research laboratories. The availability of temperature-cyclers has automated the reaction process eliminating the need for manual control of the reaction (Brown, 2001).

The aim of these PCR experiments is to amplify regions of the MAL3P6.11 gene coding for unique parts of the Pf403 protein, in preparation for further cloning into plasmid vectors. Cloning of PCR fragments can be performed in a number of ways, depending on the vector used and the cloning experiments to be performed. Blunt-end ligation is when PCR products are involved in a enzymatic reaction with the Klenow fragment of DNA

polymerase which fills in, or cleaves single-stranded overhangs at the termini of a DNA molecule, rendering the strand blunt-ended. The blunt fragments can then be ligated into vectors cut with restriction enzymes forming blunt ends (Innis and Gelfand, 1990). Double stranded PCR products are never blunt ended, and instead contain an additional adenine nucleotide at the 3' termini caused by *Taq* polymerase. Thymine-Adenine (TA) cloning takes advantage of this as the product is ligated into TA cloning vectors, which contain thymine residues at the 5' ends (Holton and Graham, 1990). The disadvantage of these methods is retaining an open reading frame if cloned into an expression vector. For this reason, restriction sites can be incorporated into the sequence of the primer (Kaufman and Evans, 1990). After PCR is performed, the product incorporates the restriction site into its sequence, allowing restriction digestion to create terminal sticky-ends, and controlled in-frame ligation with vector DNA. This method will be the primary approach for cloning experiments in this study.

An initial cloning strategy involved the direct cloning of the PCR products into the expression vector pGEX-4T-3. Due to several problems associated with direct cloning of PCR fragments, the PCR products were cloned into the initial cloning vector, pMOS*Blue*. The PCR product was cleaned and blunt-ended using the Klenow fragment of DNA polymerase. The termini of the fragments were phosphorylated using polynucleotide kinase and the modified strands were blunt end ligated into pMOS*Blue* using T4 DNA ligase. The resulting construct is highly useful for further sub-cloning experiments involving other vectors. The range of restriction endonuclease sites in the multiple cloning site of pMOS*Blue* is extensive (Figure 5.1). An insert cloned into the *Eco*RV site allows a variety of options for further cloning in many other vectors, including both prokaryotic and eukaryotic expression systems. The 817 bp region recombined with this vector allows further manipulations to be independent of PCR and growing up bacterial cultures containing the plasmid, allows batch amounts of the DNA to be obtained. PCR may introduce mutations and mistakes into the DNA. *In vivo* amplification of the DNA by replication inside bacterial cells provides a more effective way of obtaining large amounts of DNA. The synthesis of new strands is achieved using the full system of DNA replication provided by the host eg efficient proofreading activity.

Escherichia coli cloning vectors are highly useful for gene-cloning experiments. Bacterial plasmids are double-stranded, circular DNA molecules that carry a number of genes

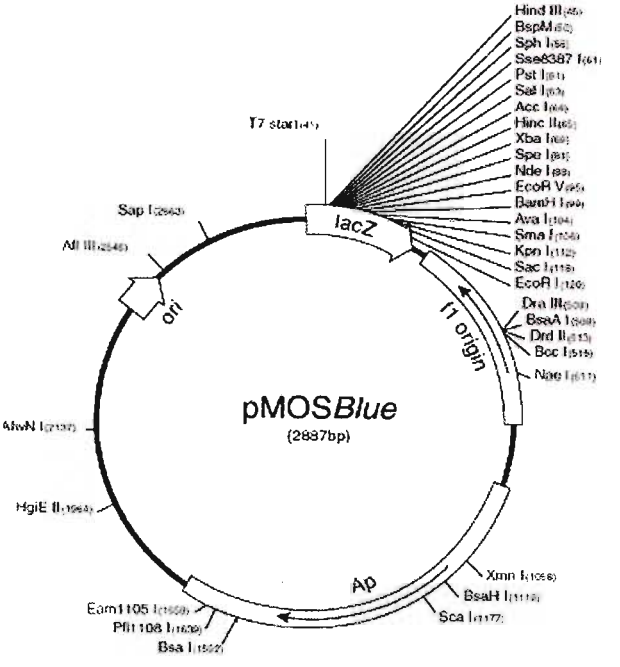
responsible for useful characteristics when introduced into the host bacterium. Plasmids are found in a variety of bacterial species, behaving as accessory genetic units, replicating independently of the host chromosome. Plasmids rely on the host for proteins and enzymes involved in replication and transcription. Advantageous or useful phenotypes produced in plasmid-bearing hosts include antibiotic resistance, degradation of organic compounds, and production of toxins and expression of cloned genes. The variety of cloning vectors available for researches is immense, in particular *E. coli* cloning vectors. The huge amount of information that exists concerning the biochemistry, microbiology and genetics of the *E. coli* host provides a valuable tool for the study of gene structure and function. Molecular biology research and gene cloning go hand in hand and provide an unlimited field and scope for biological research (Brown, 2001).

The most common selectable marker used in plasmid vectors is an ampicillin resistance gene (amp^R). This attribute allows the host bacteria to grow comfortably in the presence of the antibiotic ampicillin, allowing selection of only plasmid-containing hosts. Ampicillin binds to inhibit enzymes in the bacterial membrane involved in cell wall synthesis. The amp^R gene carried on the plasmid codes for an enzyme secreted into the periplasmic space of the bacterium, catalysing hydrolysis of the β-lactam ring, causing detoxification of the antibiotic (Sykes and Mathew, 1976).

The success in cloning experiments in this study was due to the particular system for the identification of recombinants known as lac selection or α-complementation. The plasmid contains a gene call *lacZ'*, which codes for a portion (the first 146 amino acids) of the enzyme β-galactosidase. The plasmid also contains the lac promoter, which is the sequence that controls transcription of the *lacZ* gene, and expression of β-galactosidase. The multiple cloning site of the vector is situated within the coding sequence for the *lacZ'* gene, and cloning with the vector causes insertional inactivation of the *lacZ'* gene. The strains of *E. coli* used in these experiments contain a modified *lacZ* gene that lacks the *lacZ'* portion, and codes for only the α-peptide portion of β-galactosidase (carboxy-terminal). These cells can only synthesise a functional β-galactosidase enzyme when they contain plasmid DNA coding for the rest of the enzyme, or *lacZ'* region. Cells that harbour plasmid DNA are rendered ampicillin resistant (amp^R). Cells containing recombinant plasmids DNA are β-galactosidase deficient ($\beta\text{-gal}^-$). A lactose analogue called 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-gal) can be broken down by β-galactosidase forming a deep blue product

(Horwitz *et al.*, 1967). If X-gal is included in the agar media, the bacteria containing plasmid DNA (β-gal$^+$) will form blue colonies due to the functional β-galactosidase enzyme. If the plasmid is recombinant the X-gal will not be broken down and the colonies will possess a white phenotype. This system is used to effectively identify recombinant colonies from a transformation plate without the need for time-consuming screening methods. This is especially helpful when the proportion of recombinant plasmid is low after ligation.

Vector map



Polylinker region

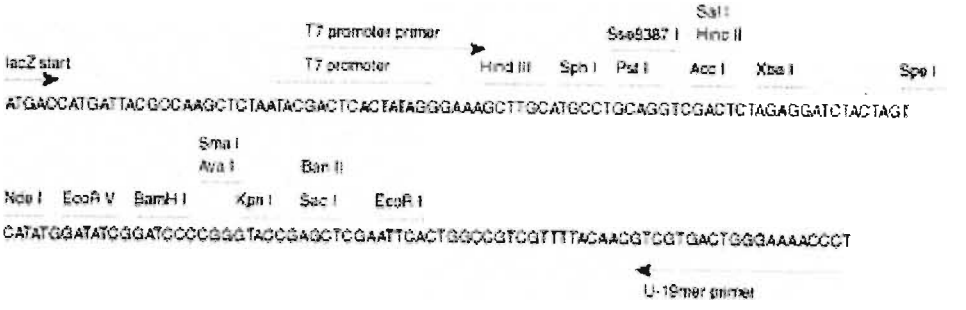


Figure 4.1 pMOS*Blue* vector map showing pMOS*Blue* polylinker region. (Taken from pMOS*Blue* blunt ended cloning kit instruction manual, Amersham Biosciences).

4.2 Results

4.2.1 Amplification of 757 bp target region in the *Plasmodium falciparum* acetyl-coA carboxylase gene

Oligonucleotide primers ACCXpf1 (GCT AAT AAT GGG ATG GCA GC) and ACCXpf2 (TTT GTG TAG TAC AAT CAC GAC) were used to amplify a region within the biotin carboxylase domain coding sequence of the ACC gene from the *Plasmodium falciparum* genomic DNA isolate used in this study. Figure 4.2 shows agarose gel electrophoresis of PCR samples obtained from this reaction and verification of product by restriction digestion analysis.

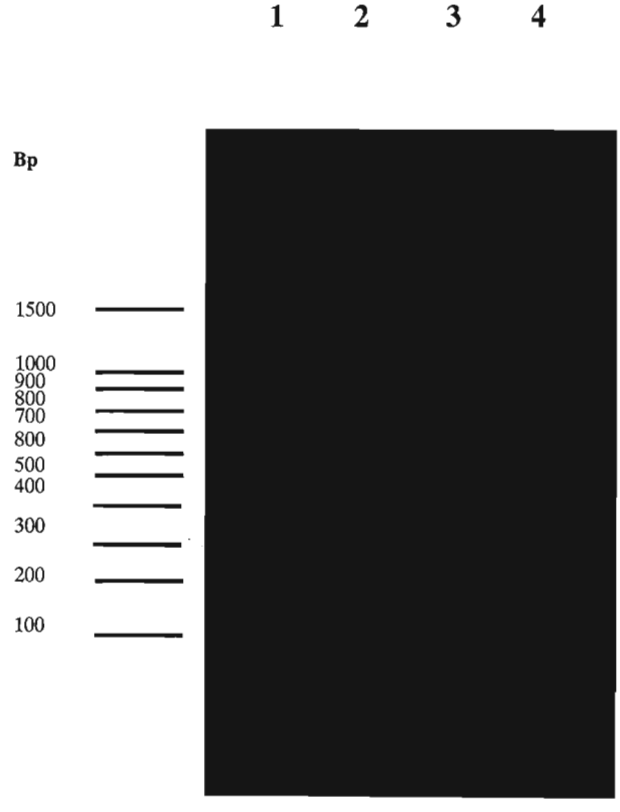


Figure 4.2 Amplification of ACC gene segment from *Plasmodium falciparum* genomic DNA. Lane 1 - 100bp DNA ladder; lane 2 - no product; lane 3 - 757 bp PCR product; lane 4 - *Hin*dIII digestion products of 757 bp PCR product.

These primers were previously designed to amplify this segment of DNA from isolated *P.*

falciparum DNA supplied by Professor P. Doerig (INSERM, Paris). The reaction was used as a control for the DNA used in this study. Lane 3 (Figure 4.2) shows a band estimated to lie within 700 - 800 bp size range, as estimated using the 100 bp markers in lane 1. This correlates to the expected size of 757 bp for this target region, as deduced from the genomic sequence available at PlasmoDB. The fragments contain an internal *Hin*dIII site, which was used to confirm the identity of this product. A restriction digestion reaction using the enzyme *Hin*dIII produces two bands estimated to be 350 bp and 400 bp is size, which compares to the sizes of 344bp and 413bp as predicted by sequence analysis.

These results show that the isolated *Plasmodium falciparum* genomic DNA as a good template for further reactions. This DNA sample, and the components of the PCR reactions used in this study, can effectively be used to amplify other target regions in the malaria genome.

4.2.2 Amplification of 817bp target region in the MAL3P6.11 gene from *Plasmodium falciparum* genomic DNA

A region of the MAL3P6.11 gene was amplified by PCR from *Plasmodium falciparum* genomic DNA using the primers PfFOR2(*Bam*HI) and PfREV1 (see Table 3.1). Gel electrophoresis of the PCR product showed an intense band, estimated using the molecular markers to be at roughly 800 bp. The amplification product was expected to be 817bp, based on the sequence of the target region. This shows that the DNA fragments synthesised during amplification correspond closely to that expected. As seen in lane 1 (Figure 4.3) there is an intense band at approximately 800 bp and what appears to be a non-specific PCR product at approximately 500bp. To verify that the PCR product obtained is the intended 817bp target sequence, restriction digestion analysis was performed. This made use of an *Eco*RI and *HindIII* restriction site present within the target sequence. The position of these sites is shown in Figure 3.2. The target sequence and position of the restriction sites in the sequence were used to calculate the sizes that would result from digestion with these enzymes. As seen in lane 2 (Figure 4.3), digestion with *Eco*RI produces fragments estimated at about190bp and 620bp. Digestion of the sequence with *Eco*RI was predicted to produces sizes of 193 bp and 624 bp respectively. The restriction digestion confirms the identity of the amplification product. Sequence identity was confirmed with *HindIII*, which digests once yielding fragments of 184bp and 633bp.

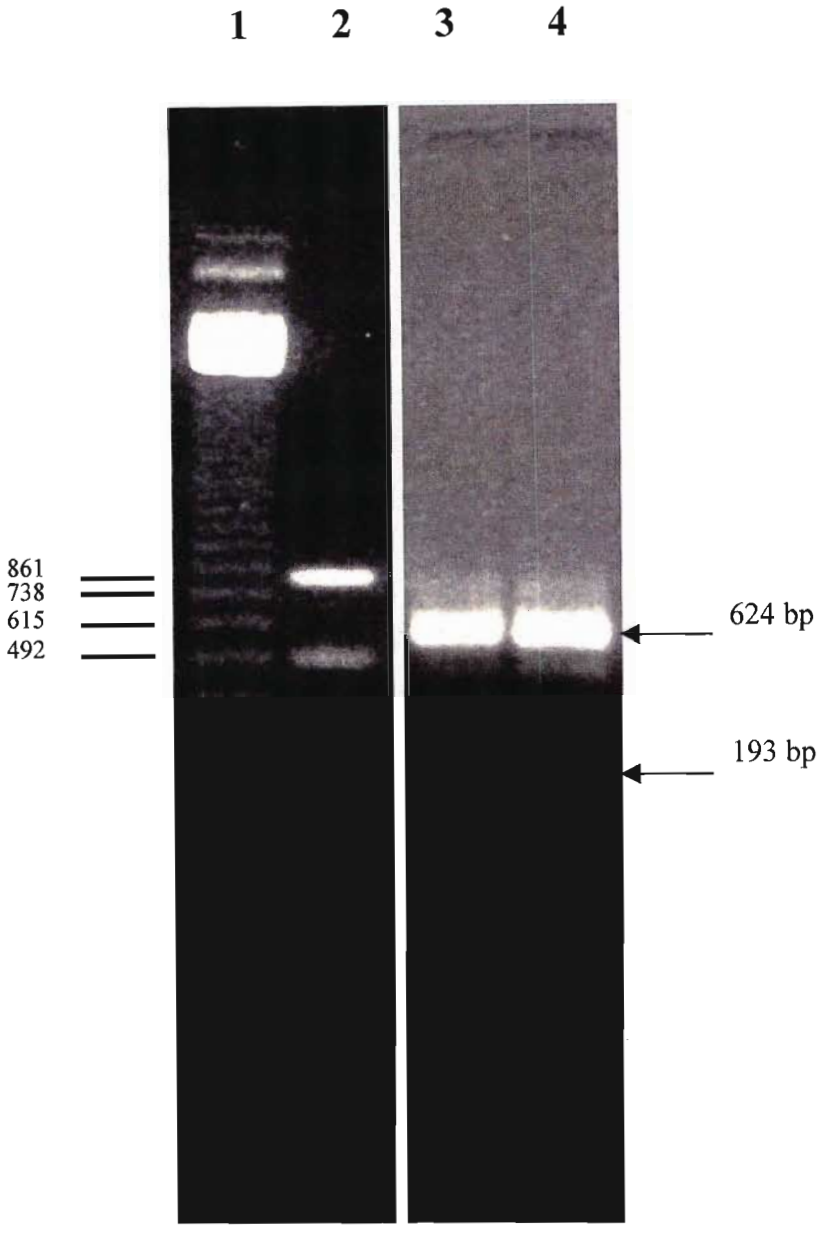


Figure 4.3 PCR amplification of 822bp fragment in MAL3P6.11 gene from *Plasmodium falciparum* genomic DNA. Agarose gel electrophoresis (1%). Lane 1 - 123bp Molecular markers; lane 2 - 822bp PCR product; lane 3 - Digested PCR product (*Eco*RI); lane 4 - Digested PCR product (*HindIII*) (arrows show restriction fragments).

```
  7051
   |                  5' GGATCCAACTATCTTGTTAATAATCTT  3'
5'aatgttaaaaacattgtagatattaacaactatcttgttaataatcttcaattaaataaagacaatgataatatt

  7126
   |
  attattattaaatttaatattttaaaactattcaaattaggttcatgctatttatatattattaatcgtaattta

  7201
   |                                                          HindIII   EcoRI
  aaagaaatccaaatgttgaaaaatcaaatcctttccttagaagaaagcattaaaagcttaaatgaattcattaat
  7276
   |
  aatctaaaaaacgaaaatgaaaaaaatgaattaattaaaataaataattttgaagaaatactcaaattaaaaaat

  7351
   |
  aatctacaagataatgaaagttgtatacaaaacttaaataattatttaaaaaaaaatgaagaattaaataaaatt
  7426
   |
  aatgtaaaaaatattttcaaatataaaggatatataattcatttaatacaacaaagtaatgtcttttgtaaaatt

  7501
   |
  tttaaacattttaatgaaaataaaattattgatcaaagtattataaacaaattactttatttaaaaaaatccttt

  7576
   |
  gatttttatatgtatgattcggttatacaagaaataagagaaaataaaaatataataataaatcaagatttttta

  7651
   |
  acagatgaatattttaaacatatacaaacctttaccaaaacatgtaatgtattaattcaaaggggatatctcagc
  7726
   |
  atcttaaaagatacaaacaatgatttctttatacaaaacaaacaaagtaatcaacaaggaaatcaaaatggtaac

  7801
   |
  catataaatatgtgtaacatatatccagatgatgaaatcaatgtaactgctgatcaacaaatttttgatggtacg
                                                                  3' CTACCATGC

  7876
   |
  gaaaacgtacaacagtccttacaaaatgaggaagattatgtaaataatgaggaaatgtatacggacaaaatggac
3'  CTTTTGCATGTTG 5'
```

Figure 4.4 DNA sequence of target region in MAL3P6.11 gene. The position of primers designed for amplifying 817bp are shown in bold along with the primers on the sequence. Restriction enzyme sites *Hin*dIII and *Eco*RI are highlighted.

4.2.3 Optimisation of PCR reaction (817bp) using different $MgCl_2$ concentrations

In an attempt to optimise the PCR reaction that amplifies the 817bp target in the MAL3P6.11 gene, a range of $MgCl_2$ concentration was evaluated. The non-specific reaction product estimated to be approximately 500 bp might be a non-specific reaction product caused by the amount of $MgCl_2$ in the reaction. A range of $MgCl_2$ concentration was set up from 1 - 2.5 µM. Figure 4.3 shows the agarose gel used to analyse the amplification products of the different PCR reactions.

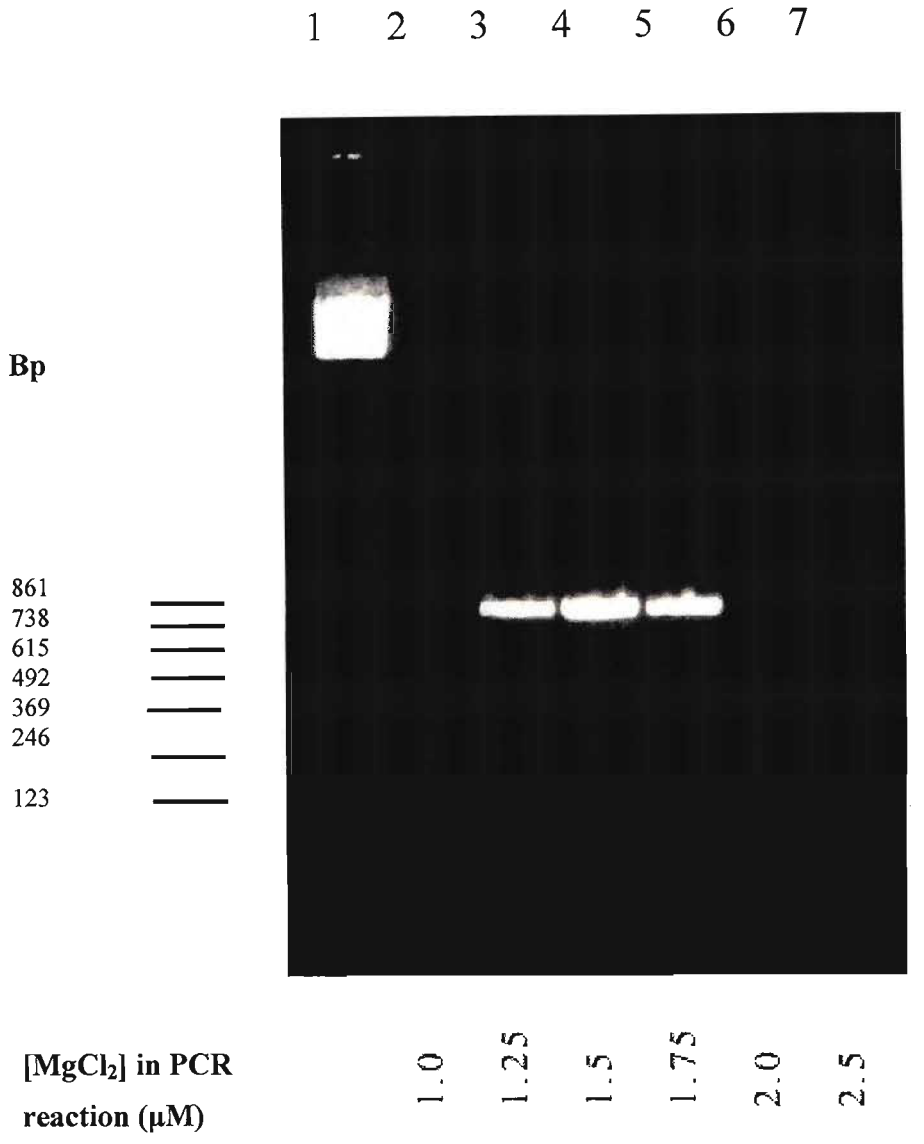


Figure 4.5. Optimisation of PCR reaction with range of $MgCl_2$ concentrations, for amplification of 817bp target in MAL3P6.11. Agarose gel electrophoresis (1%) showing the products of PCR reactions 1-6 in the presence of lane 2 - 1µM $MgCl_2$, lane 3 - 1.25µM $MgCl_2$, lane 4 - 1.5µM $MgCl_2$, lane 5 - 1.75µM $MgCl_2$, lane 6 - 2µM $MgCl_2$, lane 7 - 2.5µM $MgCl_2$. Lane 1 shows the 123 bp molecular weight markers used.

There were amplification products in PCR reactions 2-4 (1.25-1.75µM $MgCl_2$) and none in the other lanes. It appears that $MgCl_2$ concentrations below 1.25 µM and above 1.75 µM are unfavourable for successful amplification. Varying the $MgCl_2$ concentration did not affect the presence of the 500 bp non-specific reaction product, indicating that the non-specificity of the product is unrelated to $MgCl_2$ concentrations and primer annealing defects.

4.2.4 Further PCR reactions tested using various primers designed

Attempts to amplify larger regions of the MAL3P6.11 gene were unsuccessful (Section 2.4.1). This may be due to the length of the targets, and the difficulty in amplifying large regions. As can be seen on Table 4.1, the elongation time in the PCR temperature cycles were adjusted according to the target length.

Table 4.1 Alternative PCR reactions performed using designed primers, showing intended product length and elongation temperatures.

Primer pair used	Product size (bp)	Elongation duration (min)
PfFOR1 and PfREV2	10 532	10
PfFOR2 and PfREV2	3 202	3
PfFOR1 and PfREV1	8 146	8

Optimisation of each of these attempted reactions was performed using a range of $MgCl_2$ concentrations, as performed in section 4.5.2 for the successful reaction (817 bp), with no success in amplification due to a change in stringency conditions. The annealing temperature was also altered, to account for the low (G+C) content in the target sequences, and the cycle numbers for each reaction was increased to 40 compensating for possible late onset of amplification in the reactions. In a further attempt, aliquots of each reaction product were re-introduced as template into further reactions involving fresh *Taq* polymerase. These various optimisation attempts for these target regions did not succeed in amplification.

4.2.5 Preparation of insert DNA (817 bp PCR product) for ligation reactions

As can be seen in Figure 4.6, the PCR product shows a distinct clear band of the required size when visualised on an agarose gel. The PCR product was extracted with chloroform and isoamyl alcohol (24:1) (CIAA) to inactivate the polymerase. After extraction the aqueous phase was run alongside the untreated PCR product to assess the clean-up procedure. It was noted that the PCR product after treatment, was less concentrated than in the original PCR reaction due to loss of DNA in the CIAA extraction procedure. The product however, shows a clear distinct band on the gel, and a sufficient concentration for further enzymatic reactions (eg. ligation).

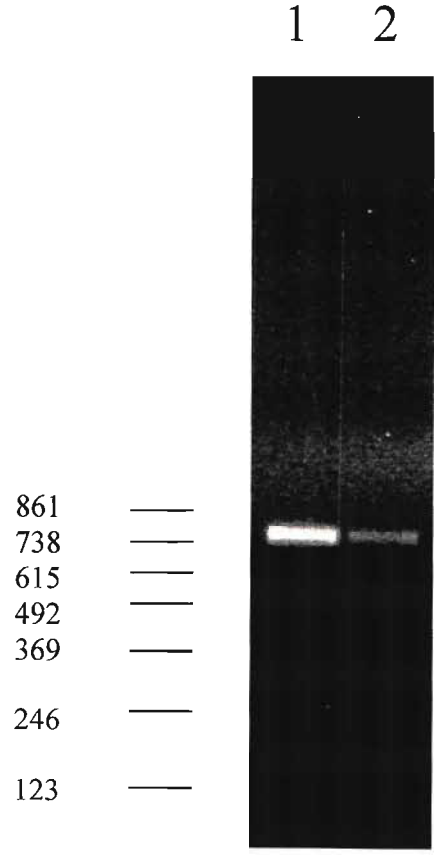


Figure 4.6 Clean-up of PCR product using chloroform / iso amyl alcohol (24:1) (CIAA) treatment. Agarose gel electrophoresis (1%). Lane 1 - Untreated PCR product (5 ul); lane 2 - CIAA treated PCR product (5 µl). DNA was stained with ethidium bromide and photographed under UV light.

The concentration of the PCR product before CIAA extraction, as determined spectrophotometrically (Section 2.3.1), was 760 µg/ml. After the clean up the concentration was 500 µg/ml indicating a 65.8 % yield. The purity of the samples (A_{260}/A_{280}) was measured and showed an increase in purity after the procedure (Table 4.2). Protein contamination in the original PCR reaction (*Taq* polymerase) was high and would interfere with further manipulations.

Table 4.2. Concentration and purity of PCR product, pre- and post-CIAA extraction.

	Before CIAA treatment	After CIAA treatment
Concentration (µg/ml)	760	500
Purity (A_{260}/A_{280})	1.64	1.95

4.2.6 Direct ligation of 817 bp insert DNA into pGEX-4T-3

4.2.6.1 Restriction digestion of pGEX-4T-3 with *Bam*HI and *Sma*I

Figure 4.7 shows the *Bam*HI and *Sma*I digestion of the pGEX-4T-3 vector in preparation for ligation reactions. Agarose gel electrophoresis of the digestion products, determines the extent and efficiency of these enzymes in the same buffer and the simultaneous digestion with *Bam*HI and *Sma*I. In comparison to the circular pGEX no-digested control, it is clear that the DNA was cut as both enzymes linearised the vector DNA in the specific restriction buffer used (Multi-core™ buffer, Promega), it was assumed that a double digest with these enzymes would result in vector fragments with terminal *Bam*HI and *Sma*I sites. Lane 4 in Figure 4.7 shows the results of a double digest reaction. Linear fragments of the plasmid were verified using molecular markers. Sizes of the linearised plasmid resulted in single bands were estimated to be approximately 4900 bp, correlating to the predicted size of the vector pGEX-4T-3 (4900 bp). These reaction products were used for directional cloning of the insert into the pGEX vector.

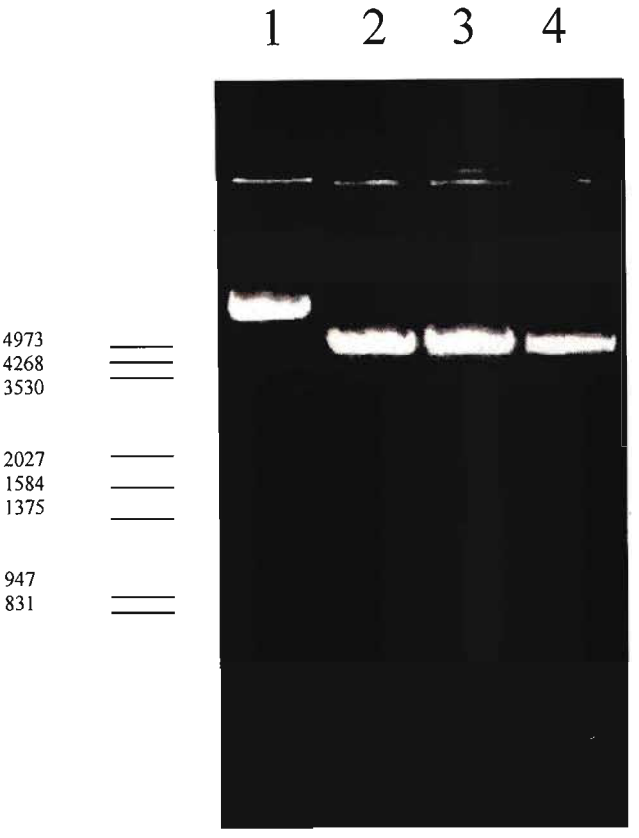


Figure 4.7 Restriction digestion of pGEX-4T-3 with *Bam*HI and *Sma*I. Agarose gel electrophoresis (1%) showing pGEX digestions: Lane 1 - Non-digested pGEX-4T-3; lane 2 - *Bam*HI digest; lane 3 - *Sma*I digest; lane 4 - *Bam*HI and *Sma*I double digest.

4.2.6.2 Transformation of competent *E. coli* DH5α cells

Competent *E. coli* DH5α cells were transformed with the test and control ligations as described in Section 2.5.4 and 2.5.5. The transformation resulted in over 300 colonies per 0.01 ng DNA, or 3.0 x 10^7 colony forming units (cfu) per μg DNA. This is comparable to the expected transformation efficiency for *E. coli* DH5α (4 x 10^7 cfu/μg test plasmid). This is a good indication that the cells were competent and transformation efficiencies for the ligation reactions were likely to be sufficient. Table 4.3 shows the number of colonies obtained in these transformation experiments.

Table 4.3. The number of transformed *E. coli* DH5α colonies per plate for test and control ligations with pGEX-4T-3.

Transformation:	1	2	3	4	5	6	7	8
no. colonies	17	32	29	22	31	0	>300	0

The ligation control (transformation 5) containing no insert DNA produced 31 colonies. These colonies may have resulted from re-ligation of cut vector DNA, or vector DNA not cut in the restriction digestion reactions, forming background colonies. The same approximate number of colonies is seen in the test ligations (with insert DNA). There are no colonies seen in the transformation of insert DNA only (no vector DNA) in the second ligation control (transformation 6) showing that all colonies seen in the test ligations are due to pGEX vectors providing ampicillin resistance to the host cells. There are unlikely to be contaminant colonies as there are none seen in the mock transformation control (transformation 8) where no DNA was transformed.

4.2.6.3 Screening for pGEX recombinants by the isolation of plasmid DNA

Figure 4.8 shows the miniprep plasmid isolates from colonies picked and grown up in batch culture from the transformation plate. Analysis of the plasmid isolations using agarose gel electrophoresis will indicate recombinants by the increased size in plasmid, and decreased rate of migration in an agarose gel. As can be seen, all the colonies appear to migrate at the same rate, and compared to the control plasmid (non-recombinant pGEX transformant) (lane 1), there are no recombinant plasmids. This method of screening was used for 30 colonies of the transformation plate.

This method was not used to screen all the colonies, as the technique is time consuming. In total, 40 colonies were screened in this manner, all non-recombinant. Due to the fact that the proportion of recombinant colonies from these ligations may be very low, and the amount of colonies to screen are too many, further more efficient methods must be used for screening, that encompass many more colonies at the same time.

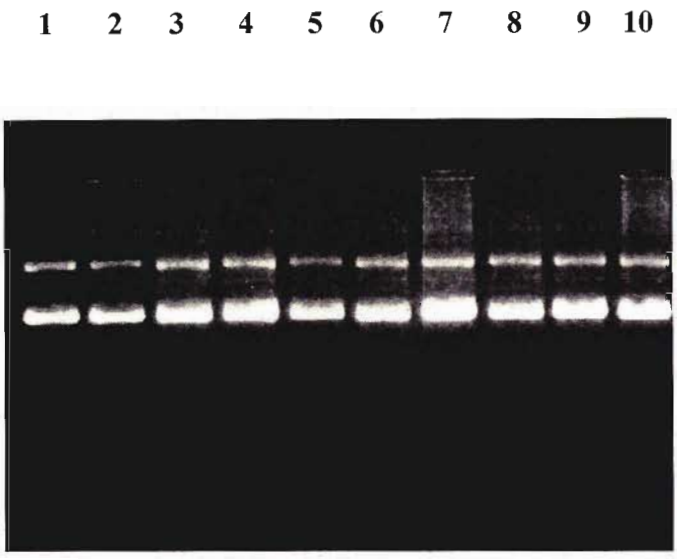


Figure 4.8 Screening for recombinant pGEX plasmids using size analysis of plasmid DNA. Lanes 1 to 10 shows the plasmid isolations from single colonies, lane 1 being the control non-recombinant pGEX picked of the transformation control plate.

4.2.6.4 Screening for recombinant pGEX plasmids by colony PCR

Colony PCR was performed using the method described in Section 2.3.7. Universal primers specific to the pGEX plasmid were used to screen for insert DNA. The presence of insert DNA was expected as part of the PCR product amplified, from the polylinker region in pGEX. The PCR products were analysed using agarose gel electrophoresis (Figure 4.9). Amplification of pGEX containing no insert using these universal primers results in a 173 bp product due to the positioning of the primers flanking the multiple cloning sites. This is evident in lane 2 by comparing the product size obtained to the molecular markers used in the experiment

The positive control involves pGEX containing an insert approximately 1100 bp in size resulting in a product less than 1300 bp using these primers. This was used to show that the insert DNA can effectively be amplified with these primers and any insert DNA that may be present in the form of recombinant pGEX, would be reflected in the same manner. Due to the fact that the 817bp insert is being screened, the predicted product size for a positive clone will be 990 bp in size. As can be seen on this gel in lanes 4 - 7, there were no inserts detected in these samples. Prepared DNA from 5 colonies was pooled in one PCR reaction to increase the amount of colonies to be screened, and four of these pools were evaluated.

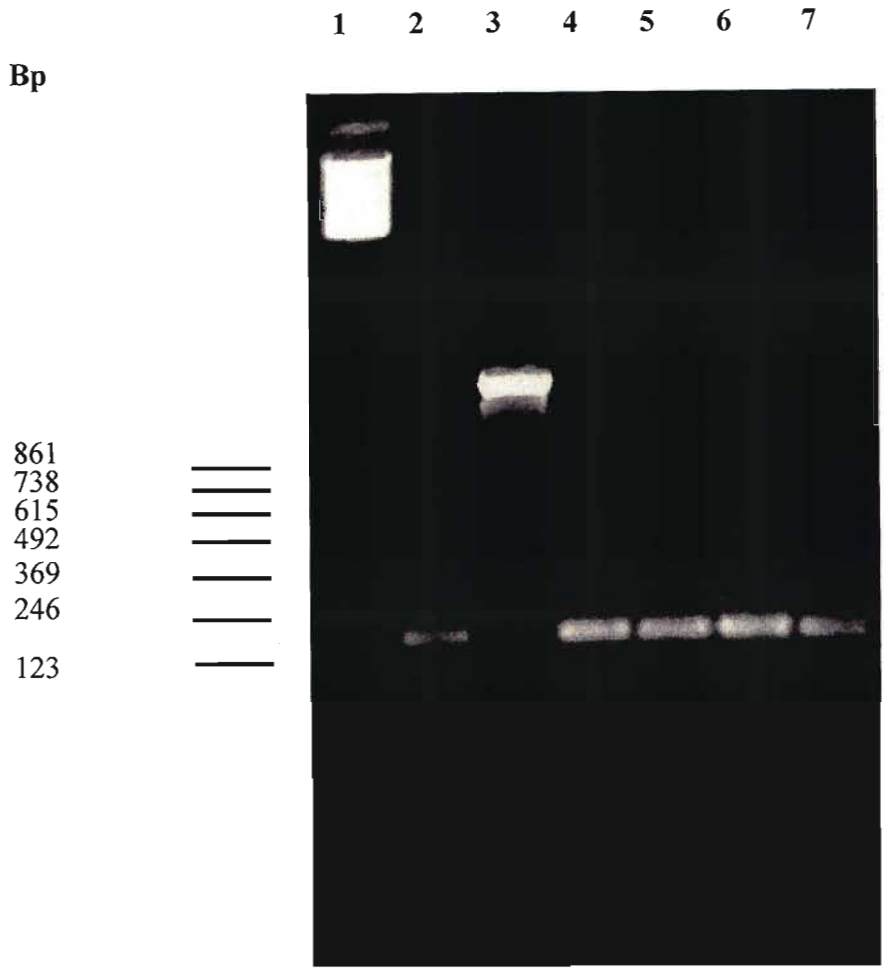


Figure 4.9 Screening of colonies using PCR with pGEX universal primers. Agarose gel electrophoresis (1%) Lane 1 - 123 bp molecular markers; lane 2 - PCR of non-recombinant pGEX; lane 3 - PCR of recombinant pGEX-GST-MAP2; lanes 4 to 8 are 20 colony DNA samples pooled into four PCR reactions.

4.2.6.5 Screening for recombinant pGEX plasmids using DIG nucleic acid detection

Insert DNA can be directly detected from DNA prepared from a single colony, and so allows for rapid screening of multiple colonies on a transformation plate. As a control for this screening technique, an oligonucleotide probe was constructed specific to pGEX vector DNA, independent of insert DNA. As described in Section 2.2.1, universal pGEX PCR and sequencing primer pG-Seq1 (GGG CTG GCA AGC CAC GTT TGG TG) was tailed at the 3' terminus with the DIG-ddUTP compound. Serial dilutions of purified pGEX-4T-3 plasmid

resulting from transformation with the pGEX (non-recombinant) vector. The membrane was probed using the DIG-labelled oligonucleotide, specific to pGEX vector DNA, to access the degree of detection and sensitivity of this technique. Figure 4.10 shows the nylon membrane with these samples, detected with anti-DIG antibody conjugated with alkaline phosphatase and, and detected with the substrates NBT and BCIP (Section 2.2.2).

Figure 4.10 DIG nucleic acid detection of DNA from colonies using pGEX specific probe. 1 to 3 - 5 ul DNA prepared from single colonies of pGEX; 4 to 10 - 5 ul of the serial dilution of purified vector DNA (1000 ng; 500 ng; 250 ng; 125 ng; 62.5 ng; 31.25 ng; 15.125 ng).

As can be seen in Figure 4.10, the colour reaction is more intense with a higher concentration of DNA and gradually diminishes as the concentration is decreased sequentially. The crude DNA preparation from the pGEX colonies showed positive colour reactions, and although not as intense, is still visible against the background as a positive reaction. In comparison to the serial dilution, the intensity of the colour reaction is similar to that of number 8, 9 and 10. This indicates that the plasmid DNA in each colony is present in the approximate range of 15,125 - 62.5 ng. This result also shows that the screening can be performed in the same manner, but replacing the vector specific probe with insert specific DIG-labelled oligonucleotide, for the detection of insert DNA.

The oligonucleotide primer PfFOR2 (*Bam*HI) used for the amplification of the 817 bp region (Section 4.2.3) was tailed with DIG, and used as a probe for the detection of insert DNA in colonies. A number of colonies were screened as shown in Figure 4.11. The control for the immunochemical detection of DIG and the alkaline phosphatase colour reaction was DIG-labelled molecular weight markers. The hybridisation reaction was checked using the purified

insert DNA as a positive control. As can be seen, both these samples showed a vigorous colour reaction, indicating that the hybridisation reaction was working efficiently, and the synthesised probe is specific to the 817 bp insert.

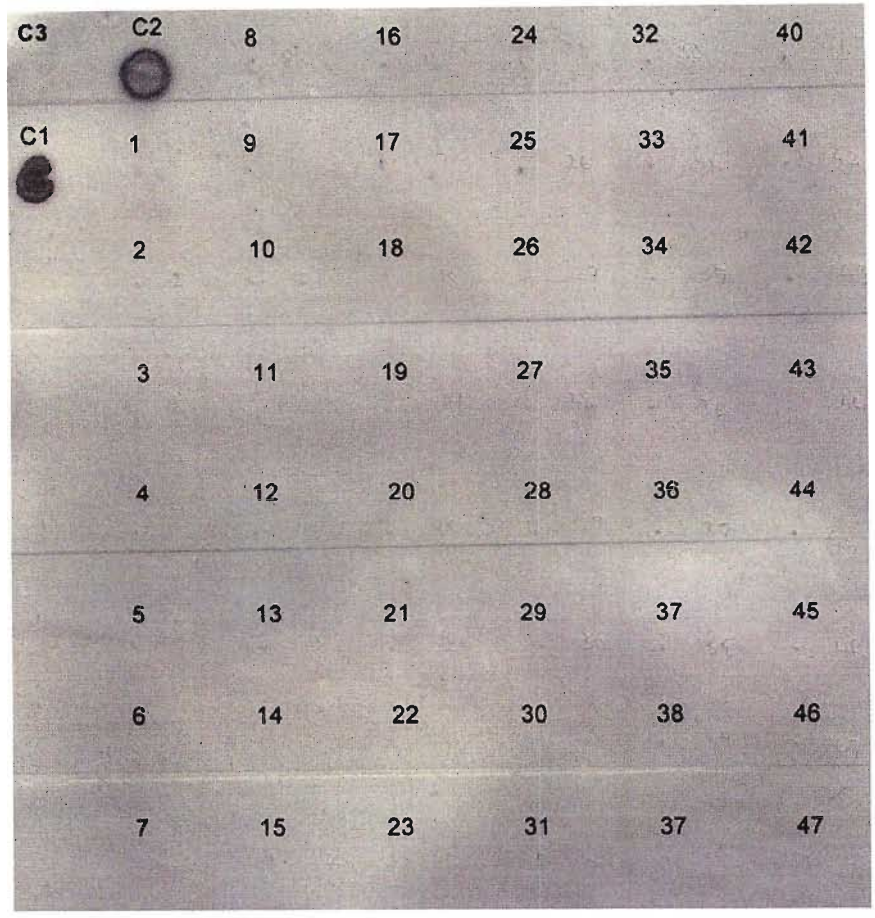


Figure 4.11 Screening by DIG nucleic acid detection using insert specific DIG-labelled oligonucleotide probe. C1 - DIG labelled molecular weight markers; C2 - 817 bp insert DNA; C3 - pGEX (non-recombinant) DNA; 1 to 47 - crude DNA preps from 47 colonies.

No colour reaction was detected for the pGEX control sample dotted, indicating that no false positives were likely to occur due to non-specific hybridisation. There were no positive colour reactions seen for any of the colonies screened. This experiment was repeated for a further 256 colonies, with no success in identifying a positive clone. Due to the fact that this method of screening, allowed a great number of clones to be screened, it was assumed that the ligation

reactions with pGEX did not result in recombinants. A more efficient system of cloning the 817 bp insert was attempted, using the pMOS*Blue* blunt-end cloning kit, for initially obtaining a vector containing the insert DNA attempted in these experiments.

4.2.7 Transformation of competent MOS*blue* cells

From this data (Table 4.4), the transformation efficiency of the competent MOS*Blue* cells can be assessed. The number of transformants per μg of DNA is the universal unit used for transformation efficiency. Considering that 0.2 ng DNA was used to transform 100 μl cells, plating 5 μl of this reaction, would be plating 0.01 ng. As seen in Table 4, the amount of colonies on the control transformation 8, was 390.

Table 4.4. The number and phenotype of pMOS*Blue* transformed *E. coli* colonies for test and control ligations.

DNA source:	Ligation reactions (1-6)						pMOS*Blue*		no DNA
Transformation number:	1	2	3	4	5	6	7	8	9
Volume spread (μl)	50	50	50	50	50	50	50	5	40
No. colonies (cfu)	47	45	42	53	23	0	>500	390	0
No. (β-gal$^-$) colonies	4	7	8	21	0	0	0	0	0

The transformation therefore produced 390 colonies per 0.01 ng DNA, or 3.9 x 10^7 colony forming units (cfu) per μg DNA. This is comparable to the expected transformation efficiency for MOS*blue* of 4 x 107 cfu/μg test plasmid.

The PK reaction, in which the cleaned PCR product was blunt ended (Klenow) and phosphorylated (T4 polynucleotide kinase), was evaluated using the control insert DNA ligated after PK treatment, and no PK treatment. Transformation 4 (Table 4) is the ligation reaction in which the controls insert DNA was PK treated, and transformation 5 the control insert was not treated. As can be seen there are a number of white (recombinant) colonies seen in no. 4, as a result of successful blunt-ending and phosphorylating of DNA termini, and successful ligation with vector DNA. Transformation 5, contains less white colonies, as a

result of the inability to ligate the insert with the vector, as the insert was not treated. This shows that the PK reaction is efficient in preparing the DNA for cloning into pMOS*Blue*. It also shows that the 23 colonies seen is the approximate amount of background colonies expected from these experiments. This also serves as a control for the ligation reaction.

There are no white colonies seen in transformation 6 (Table 5.2), as no insert was included in the ligation. This ensures that white colonies do not appear due to factors other than ligation such as contamination. The appearance of white colonies in the test and control ligations are entirely due to the addition of insert DNA to the ligation reactions. A mock transformation (transformation 9) was performed in which no DNA was transformed, to check for contaminants in the competent cells that may form colonies and false positives. The competent cells shouldn't form colonies, as they do not contain any plasmid DNA conferring ampicillin resistance.

4.2.8 Colony-PCR - screening for 817 bp insert in pMOS*Blue* plasmids.

PCR-screening was used to identify transformed *E. coli* colonies containing insert DNA. Primers PfFOR2(*Bam*HI) and PfREV1 were used to amplify the 817bp insert directly from crude preparations of bacterial DNA. After PCR, the samples were run an agarose gel. Figure 4.12 shows the PCR products from the colonies selected, lane 1 and 2 represent white colonies and a blue colony was picked as a negative control (lane 3). A distinct amplification product is seen for both white colonies screened, and there is no visible product for the blue colony. Lane 4 contains the PCR product of amplification using prepared insert DNA as the template. As can be seen in this lane, the band is intense, showing an efficient PCR reaction, serving as a positive control for the system. The size of the amplification products from PCR of both insert DNA and from white colony DNA were identical, due to the two bands similar migration rates. The size of the band was estimated to be approximately 800 bp, and due to the evaluation of this PCR reaction in Chapter 3, it can be assumed that the product is the predicted size of 817bp. This data confirms that the white colonies contain insert DNA serving as template for the PCR reaction.

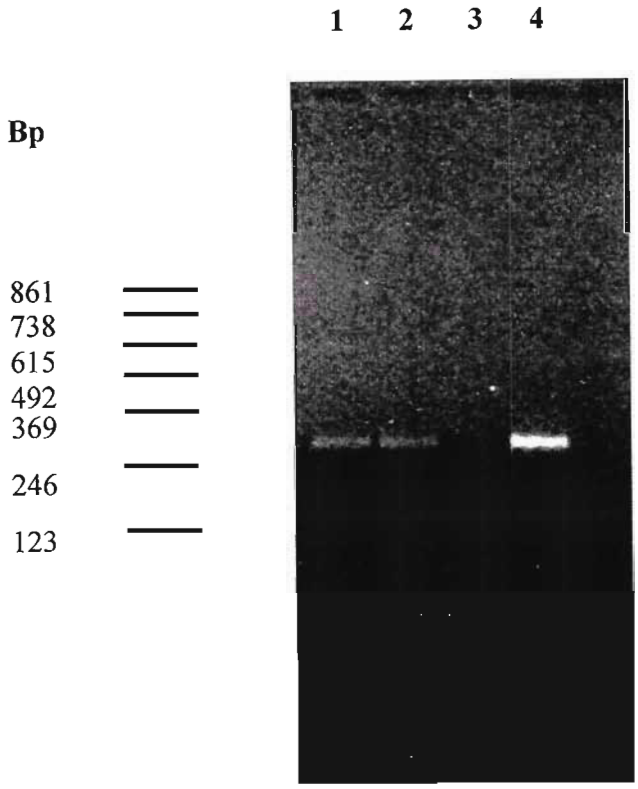


Figure 4.12 PCR based screening for recombinants. Lane 1 -PCR product from white colony, lane 2 - PCR product from white colony, lane 3 - blue colony (negative control - no insert); lane 4 - PCR product (822bp template). Molecular markers are shown for size analysis.

4.2.9 Preparation of plasmid DNA from transformant colonies for analysis of recombinants

A number of independently transformed bacterial colonies representing both white and blue phenotypes were picked and grown up in batch cultures. Plasmid DNA was isolated from these cultures using the Nucleobond® AX kit (see section 2.3.5). Based on the results obtained from colony PCR (Figure 4.12) the two white colonies identified as containing insert DNA

were picked, along with a blue colony as a control for non-recombinant plasmid. Aliquots of the isolated plasmid DNA were run on an agarose gel (Figure 4.13). The gel clearly shows characteristic plasmid DNA bands (super-coiled, closed circular, and nicked conformations). Molecular markers were not used in this gel as the different conformations of the plasmid produce an inaccurate size on an agarose gel.

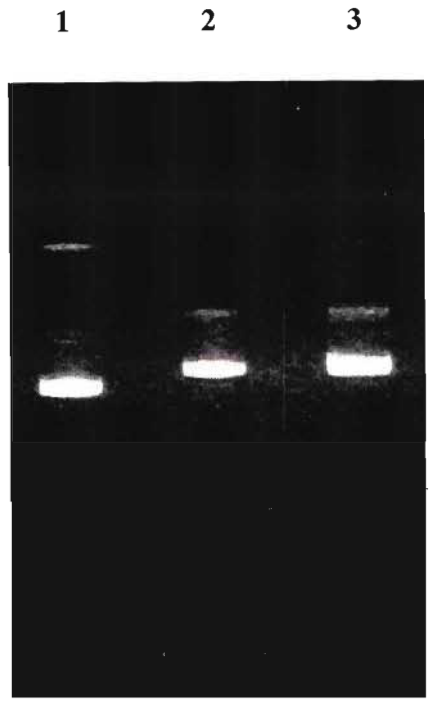


Figure 4.13 Isolation of pMOS*Blue* plasmid DNA for verification of recombinants. Agarose gel electrophoresis (1%) showing isolated plasmid DNA from: lane 1 - blue colony (non-recombinant control); lane 2 - white colony (recombinant); lane 3 - white colony. DNA was stained with ethidium bromide and photographed under UV light.

It can be seen that for the two recombinant plasmids (lanes 2 and 3), the migration through the gel has been slower, indicating that in comparison to the non-recombinant control in lane 1, the size is larger. This shows that the pMOS*Blue* plasmid DNA has incorporated insert DNA, and further verifies that the white colonies are indeed recombinant.

4.2.10 Restriction digestion analysis of recombinant pMOS*Blue* plasmid DNA.

Characteristics of the recombinant pMOS*Blue* plasmid were analysed in order to verify the effective ligation of the 817bp DNAinsert. Figure 4.14 shows agarose gel electrophoresis of the various restriction digestions performed on pMOS*Blue* and recombinant plasmid DNA. In order to estimate the size of plasmid DNA, the vector must be linearised by restriction digestion with an enzyme that contains one site in the sequence. Restriction digestion of non-recombinant pMOS*Blue* plasmid DNA with *Eco*RI (lane 2) yields a single band estimated to be approximately 2900 bp. This correlates to the known size of the vector, 2887 bp. For the recombinant pMOS*Blue* DNA, the cloned insert contains an *Eco*RI site, and so the enzyme *Sma*I was chosen for restriction digestion and size estimation. This digestion, shown in lane 3, produces a single band with an estimated size of 3600 bp. If the DNA insert (817 bp) has been ligated into pMOS*Blue* (2887), the expected size would be 3704 bp.

To verify that the insert DNA is 817 bp, the recombinant pMOS*Blue* DNA was digested with *Bam*HI, in an attempt to digest out a predicted 822 bp insert. The *Bam*HI site in the insert, and the *Bam*HI site present in the pMOS*Blue* multiple cloning region would cut out a fragment this size incorporating the cloned PCR product sequence. This assumes that the restriction site was successfully incorporated into the PCR product, and that the orientation of the insert DNA in the vector allows for the *Bam*HI sites to lie on either side of the insert. In lane 4 there are three bands seen, estimated to be 3600 bp, 2900 bp and 800 bp respectively. These products of the reaction result from partial digestion of the vector DNA. This is due to a proportion of the vector being digested at only one of the sites, at the same time as complete digestion yielding the 817 bp insert, and the pMOS*Blue* 2887 bp fragment. Vector that only cuts once is predicted to be at 3704 bp, which accounts for the top band seen in lane 4. This confirms that the insert digested out of pMOS*Blue*, is of approximate size to the insert used for the cloning experiment. To verify that the insert is the intended 817 bp PCR product, restriction digestion was performed with one of the internal restriction sites in the insert that is also in the vector. Based on the restriction sites of the insert and the vector, digestion with *Eco*RI would result in fragments of 625 bp and 3079 bp. Lane 5 shows this digestion with bands estimated to be approximately 650 bp and 3000 bp. The comparison of the estimated sizes of the restriction fragments seen in lane 5 with the expected sizes of an *Eco*RI digestion, confirms that the insert DNA in the recombinant pMOS*Blue* vector is the intended insert (817bp PCR product) used for cloning.

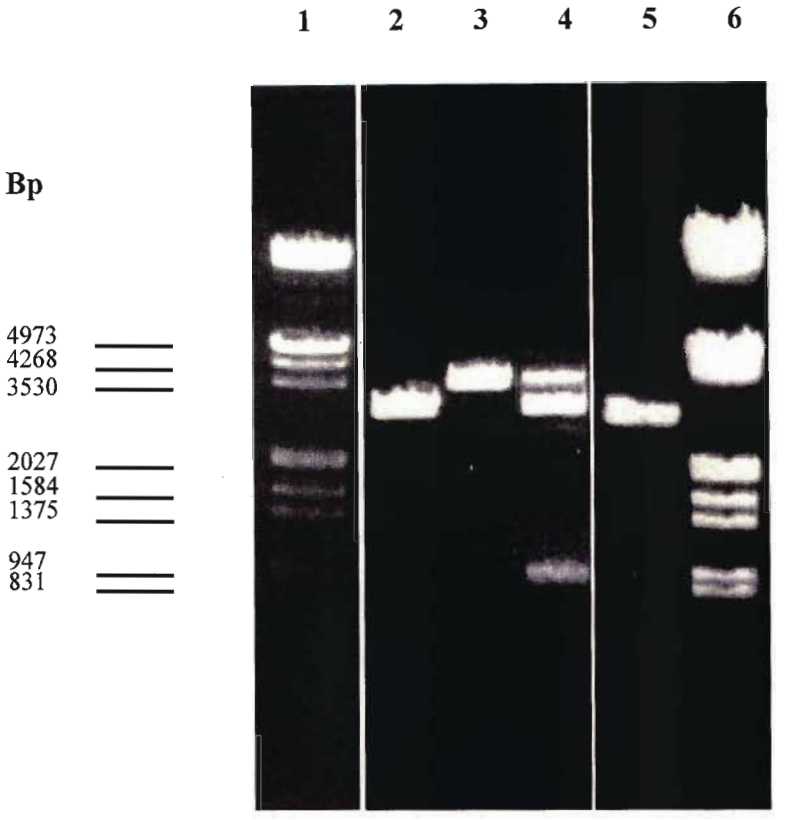


Figure 4.14 Restriction digestion analysis of recombinant pMOS*Blue* plasmid DNA. Agarose gel electrophoresis (1%). Lane 1 - Molecular markers (λ DNA digested with *Eco*RI and *Hin*dIII), lane 2 - Non-recombinant vector digested with *Eco*RI; lane 3 - recombinant vector digested with *Sma*I; lane 4 - recombinant vector digested with *Bam*HI; lane 5 - recombinant vector digested with *Eco*RI; lane 1 - molecular markers (λ DNA digested with *Eco*RI and *Hin*dIII).

4.3 Discussion and conclusion

A PCR reaction involving a 757 bp region of the acetyl-coA carboxylase gene in *Plasmodium falciparum*, was performed using the specific malaria genomic DNA isolate used in this study. The primers designed for this amplification was successful in amplifying a 757 bp band from isolated strain 3D7 DNA, and were used in this study to confirm the viability of the DNA template DNA. Any PCR reaction failures can therefore not be attributed to the template DNA, and are more likely to be the reaction itself. This amplification served to ensure the effectiveness of the *Taq* polymerase and other reaction components used in further experiments.

In these experiments oligonucleotide PCR primers were designed to amplify three main regions of the MAL3P6.11 gene (Figure 3.1) from isolated genomic *Plasmodium falciparum* DNA. The first set, PfFOR1 and PfREV2(*Sal*I) designed to flank the open reading frame (ORF) and amplify the whole gene, a 10532 bp region. The second set was designed to amplify a 3202 bp target using primers PfFOR2(*Bam*HI) and PfREV2(*Sal*I). Finally, primers PfFOR2(*Bam*HI) and PfREV1 were designed to amplify a 817 bp region. The 10532 bp and 3202 bp PCR reactions produced no amplification products that could be detected in an agarose gel. Due to the particularly long target length, a number of different reaction conditions were investigated. The elongation time was set at 10 min for the longer target, and 3 min for the 3202bp region. Increasing the elongation time, allows longer strands to be synthesised at that particular temperature in the cycle. Primer extension usually occurs up to 100 bases/sec (Innis and Gelfand, 1990), so a 3-minute elongation time would allow for this length. The cycle numbers were also increased for both reactions to 40, to allow more products to synthesise assuming that the reaction efficiency was low. The reactions were repeated using an aliquot of the first PCR mixture as template for a new reaction, in case the reaction components were depleted before sufficient amplification. The extension temperature was lowered from 72°C to 60°C to account for the high (A+T) content of the template DNA (Su *et al.*, 1996). Although there are polymerase enzymes more suited towards long targets such as *Vent* DNA Polymerase, these were not purchased due to financial constraints in this project and it was decided to adopt a 'staggered' cloning approach whereby several short targets would be amplified and studied, later to be cloned together using cloning systems more suited towards expressing large proteins.

PCR amplification of the 817bp target was successful. The region amplified encodes a length of amino acids whose sequence exhibits high homology to the *Plasmodium chabaudi adami* antigen, Pc96. The target also encodes features of the protein such as predicted T cell epitopes, a casein kinase II phosphorylation site, and various restriction sites, useful for further cloning strategies. The PCR reaction was performed and an aliquot of the reaction was evaluated on an agarose gel. An intense DNA band was seen (Figure 4.3), and in comparison to molecular markers, was estimated to be approximately the same size as the intended target region (817 bp). However, the size of the fragment does not confirm that the correct region has been amplified. Unique restriction sites within the sequence were used to verify the identity of the product. As Figure 4.4 shows, the presence of these restriction sites

at specific positions in the region being amplified, allows the synthesised DNA fragments to be cleaved producing two fragments of a defined size. The fact that there are two such sites (*Eco*RI and *Hin*dIII) increases the probability of verifying the expected product as being the correct target, as there is very little chance that a non-intended product has the same size and restriction sites. The results of these restriction digestion reactions on the PCR products (Figure 3.3) confirm the identity of the amplification product. A $MgCl_2$ profile was set up in an attempt to optimise the PCR reaction and remove the non-specific reaction product seen. The effect of altering the $MgCl_2$ concentration had little influence on the nature of the PCR product and amplification, except the fact that outside the range of 1.25 to 1.75 μM $MgCl_2$, the amplification did not work. Figure 4.15 shows the PCR product obtained and the incorporation of the *Bam*HI site into the amplification product.

pMAL c2x multiple cloning region

```
        Ile  Glu  Gly  Arg
5'   ATC GAG GGA AGG ATT TCA GAA TTC GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC TTG    3'
3'   TAG CTC CCT TCC TAA AGT CTT AAG CCT AGG AGA TCT CAG CTG GAC GTC CGT TCG AAC    5'
EcoRI     BamHI     XbaI      SalI      PstI       HindIII
```

Binding of forward primer PfFOR2(*Bam*HI) to MAL3P6.11 target sequence

```
                    BamHI
              5'    GGA TCC AAC TAT CTT GTT AAT AAT CTT C   3'
3' ... taa cat cta taa ttg ttg ata gaa caa tta tta gaa gtt aat tta ttt ctg ... 5'
              |                              |
             7074                           7104
```

817 bp Amplification product incorporates 5' *Bam*HI site

```
5'    GGA TCC AAC TAT CTT GTT AAT AAT CTT CAA TTA AAT AAA GAC ... 3'
3'    CCT AGG TTG ATA GAA CAA TTA TTA GAA GTT AAT TTA TTT CTG ... 5'
       BamHI
```

Figure 4.15 Diagram showing the incorporation of a *Bam*HI site into the 822 bp PCR product using primer PfFOR2(*Bam*HI).

For cloning experiments, the PCR product was treated with chloroform / iso amyl alcohol (24:1) to remove impurities and contaminants remaining from the amplification reactions (Taq polymerase, dNTP's etc.). Various attempts at cloning the PCR product directly into the expression vector, pGEX-4T-3 was unsuccessful. The PCR product was blunt ended, using the Klenow enzyme, and then digested with *Bam*HI to render the 5' terminus cohesive, for directional ligation into the pGEX vector. The plasmid was digested with *Bam*HI and *Sma*I (blunt-end restriction enzyme) to accommodate the insert. It was found that this strategy produced no recombinants. Hundreds of colonies were screened by colony PCR and miniprep plasmid DNA isolation and analysis using agarose gel electrophoresis.

A method using DIG nucleic acid detection was used for large-scale screening. Tailing the PCR oligonucleotide primer with digoxigenin made an insert specific probe. An initial experiment was performed to determine if DIG detection could be applied to crude preparations of colony DNA, by using a pGEX specific probe (DIG-labelled universal pGEX primer). There was significant detection seen, indicating that a positive clone, containing the intended PCR product insert DNA, would produce a positive detection colour reaction. A total of 256 colonies were screened in the insert specific clone, and no positive clones were detected.

The primary disadvantage of the pGEX system is the lack of α-complementation, which would allow the blue and white phenotype selection of recombinant colonies. In cases where the extent of correct ligation is low, this makes screening for recombinants difficult. Due to the amount of reactions with the insert DNA, the final product may not be clean enough for ligation (both Klenow reactions and restriction digestions were performed with the insert DNA). Although heat- inactivated, these enzymes may inhibit the ligation reactions. The problem may also be that the PCR product cannot be digested with *Bam*HI as the site lies directly at the termini of the fragments. The enzyme may not bind correctly to the termini as the mode of action for cleaving the strand at this site depends on flanking DNA sequences, stabilising the interaction (Kaufman and Evans, 1990). A solution to this problem is to first blunt-end the fragments using the Klenow enzyme, to compensate for additional adenosine overhangs resulting from the *Taq* polymerase activity, followed by blunt-end ligation into an appropriate vector (Innis and Gelfand, 1990). This ensures that the *Bam*HI site engineered into the termini of the PCR products can be efficiently digested, in preparation for further cloning. The pMOS*Blue* cloning vector was chosen for this strategy, due to unsuccessful

attempts at cloning the PCR product directly into pGEX. This system would ensure efficient digestion with *Bam*HI and would allow direct screening of recombinants using α-complementation. These experiments were shown to result in the successful blunt-end cloning of the 817 bp PCR product into the cloning vector pMOS*Blue*. The PCR product amplified from the MAL-3P6.11 gene in *Plasmodium falciparum*, was cleaned with chloroform / iso amyl alcohol (24:1) and cloned into pMOS*Blue* to form the construct, pMOS817. Figure 4.16 shows an outline and summary of the experiment detailing the ligation of the PCR product into the vector.

817 bp PCR product with 5' *Bam*HI site (Klenow treated; phosphorylated)

```
5'    GGA TCC AAC TAT CTT GTT AAT AAT CTT CAA TTA ... ... GAA AAC GTA CAA C    3'
3'    CCT AGG TTG ATA GAA CAA TTA TTA GAA GTT AAT ... ... CTT TTG CAT GTT G    5'
      BamHI
```

pMOS*Blue* multiple cloning region

```
5'...GGA TCT ACT AGT CAT ATG GAT ATC GGA TCC CCG GGT ACC GAG CTC GAA TTC ACT ...3'
3'...CCT AGA TGA TCA GTA TAC CTA TAG CCT AGG GGC CCA TGG CTC GAG CTT AAG TGA ...5'
          SpeI    NdeI    EcoRV   BamHI      KpnI     SacI    EcoRI
                                    SmaI
```

Restricted (*Eco*RV), and de-phosphorylated pMOS*Blue* vector with blunt ends

```
5'... GGA TCT ACT AGT CAT ATG GAT     ATC GGA TCC CCG GGT ACC GAG CTC GAA TTC...3'
3'... CCT AGA TGA TCA GTA TAC CTA     TAG CCT AGG GGC CCA TGG CTC GAG CTT AAG...5'
          SpeI    NdeI                  BamHI      KpnI     SacI    EcoRI
                                        SmaI
```

Blunt-end ligation of PCR product into restricted pMOS*Blue* vector

```
                  817 bp PCR product
5'.. ATG GAT gga tcc aac ... cta caa cAT CGG ATC CCC GGG TAC CGA GCT CGA ATT C.  .3'
3'.. TAC CTA cct agg ttg ... gat gtt gTA GCC TAG GGG CCC ATG GCT CGA GCT TAA G.  .5'
   NdeI      BamHI                      BamH       KpnI     SacI    EcoRI
                                        SmaI
```

Figure 4.16. Schematic diagram showing the construction of pMOS817 by blunt-end ligation of the 817bp PCR product into the *Eco*RV restriction site in pMOS*Blue*.

The identification and screening of recombinant pMOS*Blue* vector, was performed in a number of ways. Initially the α-complementation system provided by the *lacZ'* gene in pMOS*Blue* was used, and direct observation of colonies (blue and white phenotypes). Linearised (*Eco*RV digested) pMOS*Blue* vector that ligates with the PCR product, forms circular intact plasmid DNA, that when transformed into the MOS*Blue* cells, forms white colonies due to the truncated *lacZ'* gene. The system allows immediate indication of possible recombinants, and provides an easier method for screening many colonies. This method saves time, and further verification of recombinants allows rapid identification of positive clones.

Colony PCR screening, developed in initial experiments to detect a recombinant pGEX vector, was used to determine the presence of insert DNA in the colonies tested. Representative blue and white colonies were chosen and the primers used to amplify the PCR insert were used. Positive results were obtained for the white colonies tested and the blue colonies were negative, confirming that the product was insert-specific. The results were further verified as several studies have reported that false positives can arise using this technique, due to transformation of trace amounts of insert DNA in the cells (Alvarez et al., 2002). Primers specific to the *pMOSBlue* vector were unavailable, which would ensure that vector specific DNA was amplified. Other experiments were needed to characterise and identify the DNA in each of the white clones.

Further experiments required a large amount of purified plasmid DNA from each colony tested. Colonies were grown up in LB culture and isolated using the Nucleobond® AX kit. This was performed for three main reasons, to verify that the ampicillin resistant colonies picked contained plasmid DNA, to prove that the recombinant plasmids have actually incorporated extra DNA into its sequence by running the isolated preps on an agarose gel, and assessing the relative rates of migration of the DNA, and to provide a stock of DNA for further analysis and experiments. As can be seen in Figure 4.4, the white colonies plasmid DNA does migrate slightly slower through the agarose gel, due to its increase in size compared to the control (non-recombinant). Even though this still does not conclusively identify the inserts in these vectors to be the intended 817 bp region, two separate plasmid isolations of white colonies were shown to migrate identically indicating similar sizes.

The identity of the DNA insert sequence evaluated in the white colonies was verified using restriction enzyme analysis and visualisation on agarose gels (Figure 4.14). A theoretical map was constructed using the restriction enzyme sites in the insert and the polylinker region of the vector. The plasmid DNA was digested using specific restriction enzymes that only cut at one position in the vector, to form single linear fragments, which can then be compared to molecular mass markers. The sizes of the fragments seen on the gel corresponded to that of the actual sizes of pMOS*Blue* and recombinant pMOS817. The insert was also 'knocked out' of the vector using *Bam*HI which cuts on either side of the insert, producing two fragment sizes, a 822 bp region containing the insert, and 2882 bp vector fragment (Figure 4.14). This also verifies that the *Bam*HI site is now being cut successfully as may not have been the case during previous pGEX-4T-3 cloning attempts. The internal *Eco*RI site in the insert was used to analyse the size of the digestion fragments, further establishing the identity of the insert DNA.

These results show the positive identification of the insert DNA in the vector. The constructed plasmid pMOS817 can now be used for further sub-cloning experiments with expression vectors, and also provides a safe and easy method for the storage and amplification of the 817 bp malaria DNA region, without a need for further PCR and use of isolated *Plasmodium falciparum* genomic DNA.

CHAPTER 5

SUB-CLONING OF 822 BP INSERT FROM pMOS817 INTO pMAL-C2x AND EXPRESSION OF MBP -Pf32.5 FUSION PROTEIN

5.1 Introduction

There are many factors that may influence the level of gene expression. Expression of a protein relies on the presence of short sequences of nucleotides surrounding the gene that can be recognised by the host bacteria. These signals indicate the presence of a gene, providing instructions for the transcription and translation of the gene by the cell. Eukaryotic expression signals and their nucleotide sequences are not the same as those recognised in *E. coli*. The solution to this problem has been the design of expression vectors, which introduces the foreign gene in a manner where it is placed under control of *E. coli* expression signals. The promoter region is highly important in expression vectors, due to its direct involvement in gene expression and attachment to RNA polymerase. The promoter determines the rate at which mRNA is synthesized, and thus the yield of protein synthesised (Brown, 2001). The lac promoter is the sequence of DNA that controls the transcription of the *lacZ* gene coding for β-galactosidase in *E. coli* hosts. This promoter is induced chemically by isopropylthiogalactoside (IPTG) and addition of this chemical into the *E. coli* growth media, switches on transcription of the gene inserted downstream and in frame to that of the *lac* promoter. This promoter has been used in a wide variety of expression vectors due to its strength and efficiency in protein expression (Sambrook *et al.*, 1989).

With a fusion protein the gene is inserted into the polylinker, which interrupts a segment of a characterised *E. coli* gene in the plasmid. Insertion of the foreign gene into this position, conserving the correct reading frame, produces a hybrid gene that starts with the *E. coli* segment and joins into the foreign open reading frame without a break. The product of the expression is therefore a hybrid protein and has many advantages. Due to the fact that the attachment of the ribosome to its binding site depends largely on the nucleotide sequence at the start of the coding region, the use of an *E. coli* protein ensures efficient translation of the mRNA produced from the clone gene. Foreign proteins lack characteristics of bacterial proteins that prevent host degradation, and fusing the foreign protein to a host protein protects and stabilises the foreign protein.

pMAL™ vectors are highly useful for the expression and purification of a recombinant protein produced from a cloned gene or open reading frame. The cloned gene is ligated into the multiple cloning region downstream from the *malE* gene of *E. coli*, which encodes maltose-binding protein (MBP), resulting in the expression of a fusion protein (Guan *et al.*, 1987). The strong *tac* promoter (a minor variation of the *lac* promoter) and the *malE* translation initiation signals, provide a method of obtaining high-level expression of cloned sequences (Maina *et al.*, 1988). The system offers a one-step purification of the fusion protein using MBP's affinity for maltose (Kellerman and Ferenci, 1982). The restriction sites in the multiple cloning region are positioned in between the *malE* and the *lacZα* gene. In non-recombinant constructs the *lacZα* portion is expressed with the MBP protein, fulfilling the α-complementation role of the *lacZα* portion. Inserting a coding region of interest, into the multiple cloning site, inactivates the α-fragment activity of the *malE-lacZα* fusion, resulting in the formation of white colonies when transformed into an α-complementing host such as TB1 or JM103.

A signal peptide is present in some of the pMAL vectors (pMAL-P2x) directing fusion proteins to the periplasm. In case this case, folding and disulphide bond formation are allowed to take place in the periplasm of *E. coli*, allowing further purification from the periplasm. The pMAL-C2x vectors contain a deletion of the signal sequence, leading to cytoplasmic expression of the protein. Generally this vector produces more protein the pMAL-P2x, so the former is chosen for initial cloning experiments. If the fusion protein fails to fold correctly in the cytoplasm, or if it requires disulfide bonding to fold correctly, pMAL-P2x can be chosen for expression. The pMAL vectors also carry the origin of DNA replication of *E. coli* bacteriophage M13, allowing the production of single-stranded DNA, by infection of the host cells with a helper phage (Sambrook *et al.*, 1989).

These vectors also carry the *lacI*^{q} gene, which encodes for the *lac* repressor. This keeps expression from *Ptac* low when IPTG is not present, to minimise the effects of leaky expression. The vectors also contain a sequence coding for a recognition site of the specific protease factor Xa, located 5' to the polylinker mulitple cloning sites. This allows the MBP protein to be cleaved from the protein of interest after purification. Factor Xa is involved in cleaving a four amino acid recognition sequence located directly after the *malE* gene. This ensures that after cleavage, there are no remaining residues from the MBP protein present on

the cleaved fusion protein, that may interfere with the desired amino acid sequence of the cloned gene.

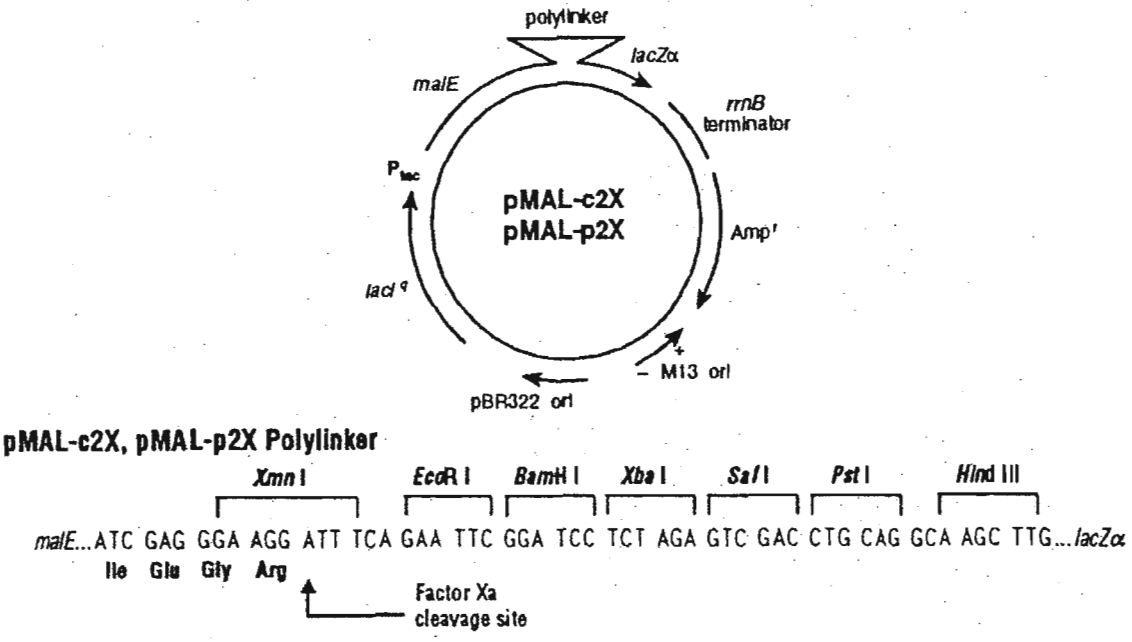


Figure 5.1 Map of pMAL-2x (6648 bp) vector showing the polylinker region and unique plasmid features. Arrows show the direction of transcription (taken from the pMAL™ Protein Fusion and Purification System instruction manual).

In order to produce a fusion protein using the pMAL vectors, the gene or open reading frame of interest can be inserted in the same translational reading frame as the vectors *malE* gene. As can be seen in Figure 5.1, the polylinker region is shown and the position of the *Bam*HI restriction site with regards to the translational reading frame of MBP. Due to the fact that a *Bam*HI site was incorporated into the PCR product and cloned into pMOS*Blue*, digestion of pMOS817 will yield fragments with terminal *Bam*HI cohesive ends. As the insert contains a cohesive *Bam*HI site at both ends of the molecule, the DNA can be inserted into pMAL in both orientations. The site at the 5' end of the product will have to be cloned in frame with MBP, in order to facilitate correct translation, and if ligated in the wrong way, the reading frame will be lost and translation will stop immediately downstream from the *malE* gene.

These experiments describe the sub-cloning of the 822 bp region of the malaria MAL3P6.11 gene (in pMOS*Blue*) into the pMAL expression vector, for the synthesis of a MBP fusion protein. The plasmid construct was sequenced to confirm the recombination with insert DNA, and preliminary induction and expression experiments were performed. Furthermore, two pGEX plasmids expressing malaria kinase fusion proteins (GST-MAP1 and GST-MAP2) were evaluated for their potential to be detected using antibodies raised against peptides designed from within their respective sequences.

5.2 Results

5.2.1 Preparation of insert DNA (822 bp PCR product)

In order to prepare the insert DNA for cloning and ligation into pMAL-C2x, the pMOS817 plasmid (Chapter 4) was digested with *Bam*HI, to excise an 822 bp insert. The restriction digestion reaction produced two products, the insert to be used for cloning, and the remainder of the pMOS*Blue* vector. The 822 bp fragments were separated from the vector DNA, so that the cut vector DNA did not interfere with cloning. For efficient ligation the insert was purified and cleaned yielding sufficient pure DNA for use in ligations. Figure 5.2 shows low melting point agarose gel electrophoresis of the digestion products. As much as 2 µg were digested (as in section 2.3.3) to ensure that enough insert DNA was present as a band at the required size of 822 bp. Four reaction products (lanes 1 to 4) were used to maximise the yield of DNA. As can be seen a single band representing the 822 bp product was isolated.

The bands seen at approximately 800 bp were taken as being the required 822 bp fragments due to previous experimental evidence. The four bands were excised from the low melting point gel and subjected to agarase treatment (section 2.7.1). The DNA was removed and concentrated from the sample by ethanol precipitation. The final DNA pellet was resuspended in 25 µl sterile dist H2O at a final concentration of 180 µg/ml, as determined spectrophotometrically (Section 2.3.1). An aliquot (1 µl) was run on an agarose gel (lane 5) and as can be seen in figure 5.2, the DNA was clean and only one product could be seen. There were sufficient amounts of this DNA for further experiments such as ligation.

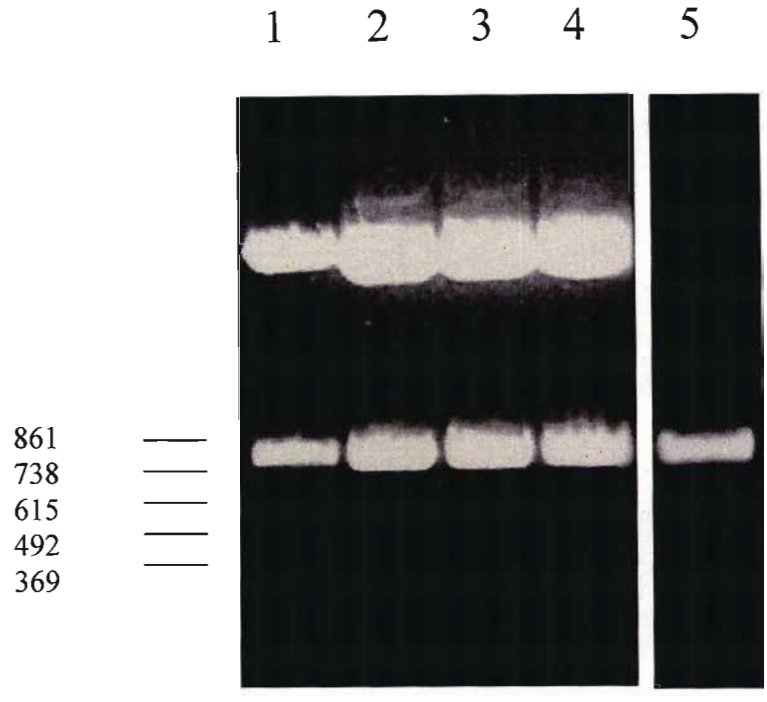


Figure 5.2. *Bam*HI restriction digestion of pMOS817, for the isolation of the 822 bp product from low melting point agarose gel. Lanes 1 - 4 - pMOS817 digested with *Bam*HI (used for excising 822 bp band), lane 5 - 822bp fragment after isolation.

5.2.2 Preparation of vector DNA (pMAL-C2x) for ligation of 822bp insert

pMAL-C2x DNA was digested with *Bam*HI (Section 2.3.3). Half of the reaction mixture was evaluated on an agarose gel to ensure complete digestion, and the other half used for ligation reactions. It is important that the vector DNA is completely linearised by *Bam*HI, to avoid background transformation colonies forming as a result of undigested plasmid. Figure 5.3 shows the results of the restriction enzyme digestion of pMAL-C2x with *Bam*HI the digestion of pMAL-C2x, and it can be seen that a single band is present, indicating complete linearisation of the plasmid due to efficient digestion. An undigested control was included (lane 1) for comparison of restriction digest products, and to ensure that the DNA had in fact been digested.

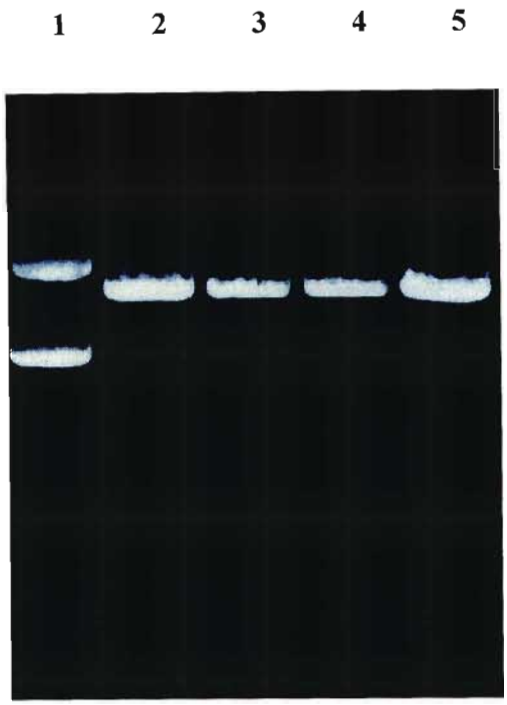


Figure 5.3. Restriction digestion of pMAL-C2x with *Bam*HI for ligation with 822 bp insert. Agarose gel electrophoresis (1%). Lane 1 - Non-digested pMAL-C2x DNA; lane 2 - 4 - pMALC2x digested with *Bam*HI.

5.2.3 Transformation of competent JM103 *E. coli* cells

Table 5.1 shows the results obtained after transformation with the test and control ligations as described in section 2.7.2.

Table 5.1. The number and phenotype of pMAL-C2x transformed *E. coli* colonies for test and control ligations.

DNA source:	Ligation reactions						pMAL-C2x		none
Transformation no:	1	2	3	4	5	6	7	8	9
Volume spread (µl)	50	50	50	50	50	50	50	5	50
No. colonies (cfu)	>400	>400	>400	>400	36	0	>500	260	0
No.(β-gal$^-$) colonies	8	5	22	0	0	0	0	0	0

The frequency of background (β-gal$^+$) colonies is very high in these transformations (Table 5.1). This is due to self-ligation of the *Bam*HI digested pMAL vector. Background colony numbers were expected to be this high, as the vector DNA was not dephosphorylated with calf intestinal alkaline phosphatase (CIAP). The vector is also only cut with one restriction enzyme (*Bam*HI), increasing the likelihood of religation. The extent of religation was tested using transformations 4 and 5. The *Bam*HI digested DNA, was included in a ligation reaction with no insert DNA, to assess the amount of religation, as reflected by ampicillin resistant transformant colonies. In comparison to the same reaction, with no DNA ligase enzyme included (transformation 5), the presence of ligase, results in over 400 colonies, indicating that in a ligation reaction, the extent of self-ligation is high.

The 'no-ligase' control (transformation 5) also served as a positive control for the ligation reaction, and comparing the amounts of colonies in transformations 4 and 5 (Table 5.1), it is clearly evident that the ligase is efficient under these reaction conditions. The 36 colonies formed in this control are also an indication of background colonies resulting from undigested DNA in the original vector *Bam*HI restriction digestion. Transformation 6 represents the ligation with no vector DNA, as a negative control. As can be seen there are no colonies transformed from this ligation.

The transformation efficiency was calculated for the competent JM103 *E. coli* cells used in this experiment. As much as 0.2 ng of DNA circular pMAL DNA was transformed as per Table 5.1 (transformation 7 and 8). When 5 µl of this transformation was plated, 260 colonies were counted, indicating that the transformation efficiency of these cells is 2.6×10^7 cfu/µg DNA. The competent cells did not contain any ampicillin resistant contaminants, as there were no colonies formed when no DNA was transformed (transformation 9). A total of 35 white colonies were counted on the test plates. There is a slight increase in the amount of white colonies when the molar ratio of vector:insert is 1:10 (transformation 3).

5.2.4 Preparation of plasmid DNA from transformed colonies for analysis of recombinants.

White and blue colonies were picked, and grown up in batch culture for the isolation of plasmid DNA using the Nucleobond® AX kit (section 2.3.5). Figure 5.4 shows plasmid DNA isolated from 3 white colonies and a blue colony (control), run on an agarose gel.

Figure 5.4 Analysis of isolated of pMAL-C2x plasmid DNA for verification of recombinants. Agarose gel electrophoresis (1%) showing Lane's 1, 3 and 4 - white (β-gal$^-$) colonies; lane 2 - blue colony (β-gal$^+$) (non-recombinant control).

These results show the slower migration of the plasmid DNA in lane's 1, 3 and 4 indicating that the circular DNA is a larger molecule than the original pMAL-C2x non-recombinant DNA. This correlates to the fact that these plasmids were isolated from colonies containing the β-gal$^-$ or white phenotype (indicative of recombination). This indicates that the white colonies are recombinant and have probably incorporated the 822 bp insert DNA used for cloning. Further analysis of the plasmid DNA will confirm the identity of the ligated insert. The size was not compared to molecular markers as the circular nature of the plasmid DNA results in different conformations, that cause a migration through the agarose gel, disproportional to that of the linear marker bands.

5.2.5 Restriction digestion analysis of recombinant pMAL plasmid DNA.

The plasmid DNA that was isolated in the previous experiment (Figure 5.4) was used in *Bam*HI digestions to confirm the size and of the insert DNA. Figure 5.5 shows an agarose gel electrophoresis of the digestion products. Lanes 3,5 and 6 each show a band of a similar size to the control (lane 2) (resulting from *Bam*HI digestion). If this was the insert used for cloning then the fragment of DNA 'knocked out' of the recombinant vector should be the same size. This can be seen on Figure 5.5 for the three recombinant colonies, in comparison to the control. The vector DNA fragment after digestion was estimated to be roughly 6600 bp, which correlates to the actual size of the vector, 6648 bp. As can be seen (lane 4), plasmid DNA isolated from a non-recombinant colony does not contain any insert DNA and when digested with *Bam*HI, only one size is seen, that of linearised pMAL vector DNA. This negative control validates the use of *Bam*HI to screen and verify recombinant colonies. Together, this data suggests that the white colonies do contain insert the same size as the intended 822 bp insert, and that the insert was successfully ligated into the *Bam*HI site in the pMAL polylinker region.

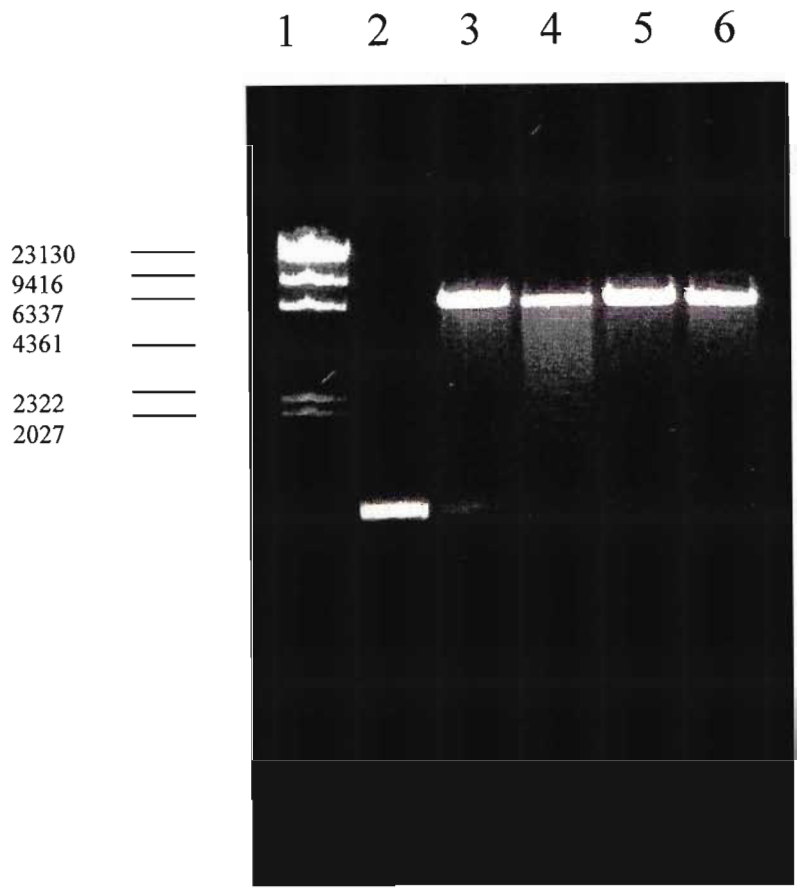


Figure 5.5 Analysis of *Bam*HI digestion of isolated pMAL DNA showing 822bp insert. Agarose gel electrophoresis (1%) showing lane 1 - Molecular markers II; lane 2 - positive control 822 bp DNA, lane 3, 5 and 6 - *Bam*HI digested recombinant (β-gal$^-$) vector; lane 4 - *Bam*HI digested non-recombinant vector (β-gal$^+$).

To confirm the identity and orientation of the insert cloned in pMAL-C2x, the isolated plasmid DNA from two separate white colonies were analysed. Figure 5.6 shows restriction digestion analysis of two recombinants. As can be seen in lanes 2 and 3, the plasmid DNA from the two colonies chosen contains a 800 bp band corresponding to the 822 bp insert. These colonies were both digested with *Hin*dIII (lanes 4 and 5). *Hin*dIII occurs naturally in the sequence of the insert (Chapter 4, Figure 4.4) originally cloned using PCR. Due to the fact the *Hin*dIII is also present in the polylinker, downstream from the *Bam*HI site, digestion of recombinant DNA with this enzyme, determines the identity and orientation of the insert DNA.

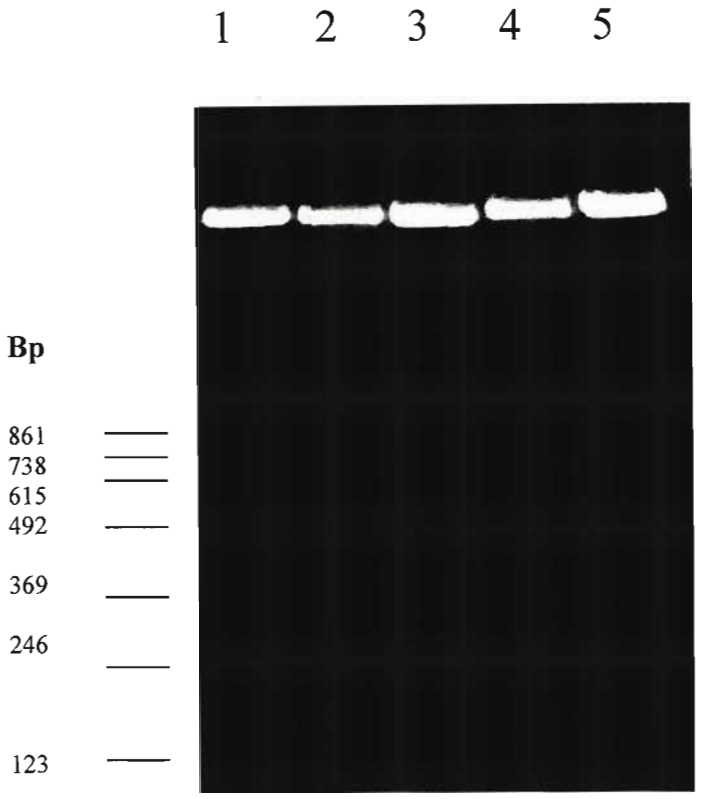


Figure 5.6 Analysis of recombinant pMAL clones using restriction enzyme digestion.. Agarose gel electrophoresis (1%). Lane 1 - *Bam*HI digested non-recombinant pMAL; lane 2 - *Bam*HI digestion of recombinant pMAL (white colony 1); lane 3 - *Bam*HI digestion of recombinant pMAL (white colony 2); lane 4 - *Hin*dIII digested pMAL (white colony 1); lane 5 - *Hin*dIII digested pMAL (white colony 2).

According to the sequence of the 822 bp insert cloned into pMAL, digesting with *Hin*dIII should produce fragments of 629 bp and 7277 bp if the insert is in the desired orientation. If the insert has been ligated in the opposite direction, the sizes were predicted to be 193 bp and 6841 bp. Colony 1 plasmid DNA digested with *Hin*dIII (lane 4) produces two bands. The size of the smaller of these bands was estimated using the molecular markers, to be 615 bp. This correlates to the sizes predicted for this digestion, and confirms the orientation of insert DNA. The insert DNA can be verified as being the correct 822 bp insert, sub-cloned from pMOS817. Lane 5 shows colony 2 plasmid DNA resulting in a band estimated to be 200 bp, from a similar digestion with *Hin*dIII. This compares to the expected size when the insert is in an incorrect orientation. The sizes of the high molecular weight bands could not be effectively resolved at this gel percentage, although it can be seen that the degree of migration is slower for vector with smaller fragments cleaved out, indicating an expected correlation between the two reaction product sizes.

5.2.6 Nucleotide sequencing of polylinker region and insert DNA in recombinant pMAL plasmid

The insert in the pTS822 plasmid was sequenced using universal primers pMAL(F) (GGT CGT CAG ACT GTC GAT GAA GCC) and pMAL(R) (CGC CAG GGT TTT CCC AGT CAC GAC). The universal primer used for sequencing pMAL vectors (New England Biolabs) initiates sequencing near the 3' end of the *malE* gene, roughly 78-81 bases upstream of the polylinker site. Figure 5.7 shows the nucleotide sequence generated from the forward and reverse cycle sequencing reactions of the insert in the pMAL vector.

GCGCGCAGAC TAATTCGAGC TCGAACAACA ACAACAATAA CAATAACAAC AACCTCGGGA
TCGAGGGAAG GATTTCAGAA TTCGGATCCA ACTATCTTGT TAATAATCTT CAATTAAATA
AAGACAATGA TAATATTATT ATTATTAAAT TTAATATTTT AAAACTATTC AAATTAGGTT
CATGCTATTT ATATATTATT AATCGTAATT TAAAAGAAAT CCAAATGTTG AAAAATCAAA
TCCTTTCCTT AGAAGAAAGC ATTAAAAGCT TAAATGAATT CATTAATAAT CTAAAAAACG
AAAATGAAAA AAATGAATTA ATTAAAATAA ATAATTTTGA AGAAATACTC AAATTAAAAA
ATAATCTACA AGATAATGAA AGTTGTATAC AAAACTTAAA TAATTATTTA AAAAAAAATG
AAGAATTAAA TAAAATTAAT GTAAAAAATA TTTTCAAATA TAAAGGATAT ATAATTCATT
TAATACAACA AAGTAATGTC TTTTGTAAAA TTTTTAAACA TTTTAATGAA AATAAAATTA
TTGATCAAAG TATTATAAAC AAATTACTTT ATTTAAAAAA ATCCTTTGAT TTTTATATGT
ATGATTCGGT TATACAAGAA ATAAGAGAAA ATAAAAATAT AATAATAAAT CAAGATTTTT
TAACAGATGA ATATTTTAAA CATATACAAA CCTTTACCAA AACATGTAAT GTATTAATTC
AAAGGGGATA TCTCAGCATC TTAAAAGATA CAAACAATGA TTTCTTTATA CAAAACAAAC
AAAGTAATCA ACAAGGAAAT CAAAATGGTA ACCATATAAA TATGTGTAAC ATATATCCAG
ATGATGAAAT CAATGTAACT GCTGATCAAC AAATTTTTGA TGGGACGGAA AACGTACCAC
ATCGGATCC**T CTAGAGTCGA CCTGCAGGCA AGCTGGGCAC TGGCCGTCGT TTTACAACGT**
CGTGACTGGG AAAACCCTGG CGTTACCCAA CTTAATCGCC TTGCAGCACA TCCCCCTTTC
GCCGCTGG

Figure 5.7. DNA sequence obtained from sequencing reaction of insert region in pMAL-C2x. Vector specific DNA (polylinker region) is indicated in bold and the *Bam*HI site used for cloning is underlined.

Figure 5.8 show this sequence translated into amino acids to verify the successful incorporation of the insert DNA.

```
          A   Q   T   N   S   S   S   N   N   N   N   N   N   N
5'  GGC GCG CAG ACT AAT TCG AGC TCG AAC AAC AAC AAC AAT AAC AAT
     N   N   N   L   G   I   E   G   R   I   S   E   F   G   S
    AAC AAC AAC CTC GGG ATC GAG GGA AGG ATT TCA GAA TTC GGA TCC
     N   Y   L   V   N   N   L   Q   L   N   K   D   N   D   N
    AAC TAT CTT GTT AAT AAT CTT CAA TTA AAT AAA GAC AAT GAT AAT
     I   I   I   I   K   F   N   I   L   K   L   F   K   L   G
    ATT ATT ATT ATT AAA TTT AAT ATT TTA AAA CTA TTC AAA TTA GGT
     S   C   Y   L   Y   I   I   N   R   N   L   K   E   I   Q
    TCA TGC TAT TTA TAT ATT ATT AAT CGT AAT TTA AAA GAA ATC CAA
     M   L   K   N   Q   I   L   S   L   E   E   S   I   K   S
    ATG TTG AAA AAT CAA ATC CTT TCC TTA GAA GAA AGC ATT AAA AGC
     L   N   E   F   I   N   N   L   K   N   E   N   E   K   N
    TTA AAT GAA TTC ATT AAT AAT CTA AAA AAC GAA AAT GAA AAA AAT
     E   L   I   K   I   N   N   F   E   E   I   L   K   L   K
    GAA TTA ATT AAA ATA AAT AAT TTT GAA GAA ATA CTC AAA TTA AAA
     N   N   L   Q   D   N   E   S   C   I   Q   N   L   N   N
    AAT AAT CTA CAA GAT AAT GAA AGT TGT ATA CAA AAC TTA AAT AAT
     Y   L   K   K   N   E   E   L   N   K   I   N   V   K   N
    TAT TTA AAA AAA AAT GAA GAA TTA AAT AAA ATT AAT GTA AAA AAT
     I   F   K   Y   K   G   Y   I   I   H   L   I   Q   Q   S
    ATT TTC AAA TAT AAA GGA TAT ATA ATT CAT TTA ATA CAA CAA AGT
     N   V   F   C   K   I   F   K   H   F   N   E   N   K   I
    AAT GTC TTT TGT AAA ATT TTT AAA CAT TTT AAT GAA AAT AAA ATT
     I   D   Q   S   I   I   N   K   L   L   Y   L   K   K   S
    ATT GAT CAA AGT ATT ATA AAC AAA TTA CTT TAT TTA AAA AAA TCC
     F   D   F   Y   M   Y   D   S   V   I   Q   E   I   R   E
    TTT GAT TTT TAT ATG TAT GAT TCG GTT ATA CAA GAA ATA AGA GAA
     N   K   N   I   I   I   N   Q   D   F   L   T   D   E   Y
    AAT AAA AAT ATA ATA ATA AAT CAA GAT TTT TTA ACA GAT GAA TAT
     F   K   H   I   Q   T   F   T   K   T   C   N   V   L   I
    TTT AAA CAT ATA CAA ACC TTT ACC AAA ACA TGT AAT GTA TTA ATT
     Q   R   G   Y   L   S   I   L   K   D   T   N   N   D   F
    CAA AGG GGA TAT CTC AGC ATC TTA AAA GAT ACA AAC AAT GAT TTC
     F   I   Q   N   K   Q   S   N   Q   Q   G   N   Q   N   G
    TTT ATA CAA AAC AAA CAA AGT AAT CAA CAA GGA AAT CAA AAT GGT
     N   H   I   N   M   C   N   I   Y   P   D   D   E   I   N
    AAC CAT ATA AAT ATG TGT AAC ATA TAT CCA GAT GAT GAA ATC AAT
     V   T   A   D   Q   Q   I   F   D   G   T   E   N   V   P
    GTA ACT GCT GAT CAA CAA ATT TTT GAT GGG ACG GAA AAC GTA CCA
     H   R   I   L   *
    CAT CGG ATC CTC TAG AGT CGA CCT GCA GGC AAG CTG GGC ACT GGC

    CGT CGT TTT ACA ACG TCG TGA CTG GGA AAA CCC TGG CGT TAC CCA

    ACT TAA TCG CCT TGC AGC ACA TCC CCC TTT CGC CGC TGG    3'
```

Figure 5.8 Translation of DNA sequence obtained from sequencing reaction. The insert DNA and protein is shown in bold, the *Bam*HI sites are underlined, the T cell epitopes are shown in red and the casein kinase phosphorylation site is shown in blue. * Denotes a stop codon.

As can be seen (Figure 5.8) the reading frame is conserved as the cloned 822 bp open reading frame is continuous with the 3' end of the *malE* gene. The introduction of the *Bam*HI site in the 5' end of the insert causes a glycine and a serine to replace the isoleucine and asparagine residues in the orignal sequence. There are no other mutations of the DNA sequence region to be expressed as a fusion protein, as checked by an alignment using this amino acid sequence and that of the original sequence from the database.

5.2.7 Induction and expression of MBP-Pf32.5 fusion protein

The presence of expressed fusion protein was evaluated using SDS-PAGE of *E. coli* total cell extract (Figure 5.9). Cells containing the pMAL-C2x non-recombinant plasmid, were induced with 0.3mM IPTG (Section 2.8.1). A fusion protein of approximately 50 kDa is seen in lane 3 and the presence of a large dark band indicates the induction of the *malE* gene product to be efficient, producing high yields of the MBP-β-gal-α fusion protein (50.8 kDa). Three recombinant colonies were chosen. pTS822A contains the 822 bp insert in the reverse orientation (Figure 5.12), therefore disrupting the open reading frame upstream from the *lacZα* region. This plasmid was used as a control in this experiment. pTS822B and pTS822C contain insert DNA in the correct orientation and reading frame, as confirmed by restriction analysis and DNA sequencing.

Lanes 4 and 5 are the non-induced and induced culture fractions respectively of clones containing the insert in the wrong orientation. The MBP is expressed without the *lacZα* fusion, as the open reading frame is disrupted. The MBP protein in this state is known as MBP2 and has a molecular weight of 42,5 kDa. This is seen in lane 5 where upon induction with IPTG there appears to be an expressed protein of approximately 40 kDa. In contrast the constructs (lane 6 - 9) with in-frame inserts, do not have expression protein bands at this size. There appears to be an increase in the density of the protein band estimated to be 75 kDa, which can be seen more markedly in lanes 7 and 9 (see arrow, Figure 5.9), in comparison to that of the control lanes. The expected size of the MBP-Pf32.5 fusion expressed from this plasmid is 75 kDa, as the cloned open reading frame codes for a 32,5 kDa region and the MBP protein is 42,5 kDa. It can also be seen that some of the bands around 40 kDa appear to be more intense, possibly due to degradation of the fusion protein. The relative size of the band at approximately 75 kDa is much larger and more evident in the correct clones than in the incorrect plasmid. The 50 kDa, and at 42 kDa bands are absent from these two lanes further indicating the identity of the 75 kDa protein.

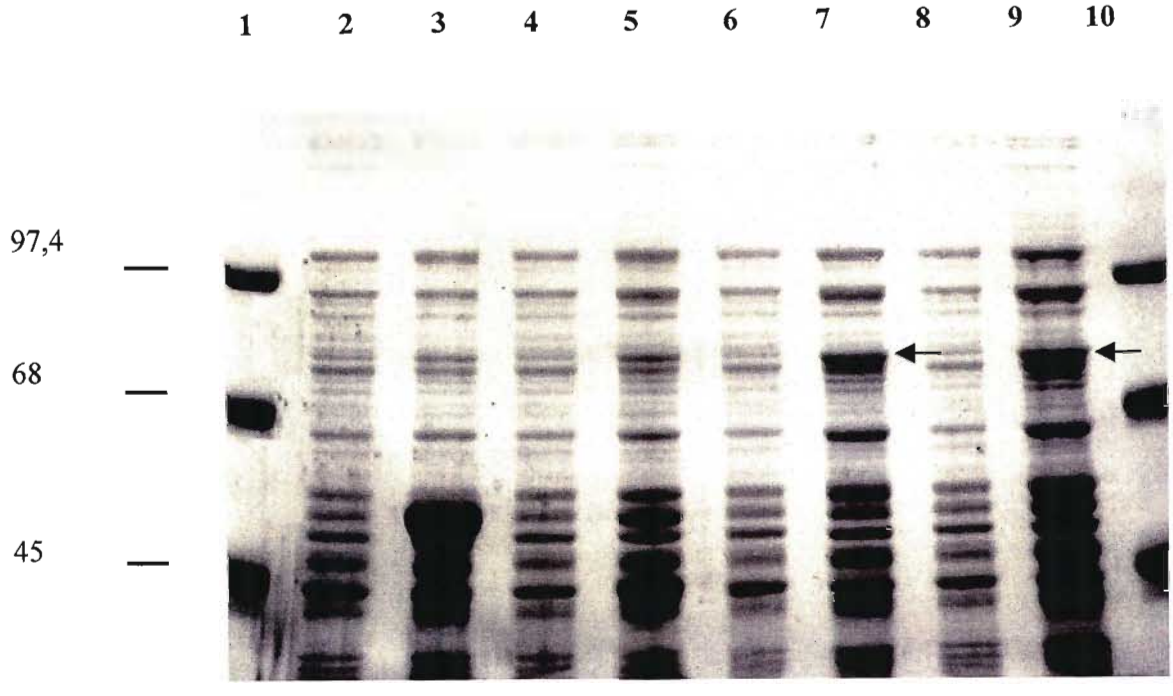


Figure 5.9 SDS-PAGE analysis of induction and expression of MBP-Pf32.5 in *E. coli* JM103. (7.5% running gel; 4% stacking gel). Lane 1 - Low weight molecular markers (Phosphorylase b, 97.4 kDa; Bovine serum albumin, 68 kDa; Ovalbumin, 45 kDa), lane 2 - Non-induced MBP control, lane 3 - induced MBP control, lane 4 - non-induced pTS822A, lane 5 - induced pTS822A, lane 6 - non-induced pTS822B, lane 7 - induced pTS822B, lane 8 - non-induced pTS822C, lane 9 - induced pTS822C, lane 10 - Low weight molecular markers. The arrows indicate the band thought to represent the MBP-Pf32,5 fusion protein.

5.2.8 Detection of fusion protein by western blot analysis using anti-MBP antibodies

The visualisation of expressed protein bands, in conditions of inefficient induction, is often ambiguous due to the presence of many other *E. coli* proteins in the fractionate. One of the methods for identifying which of the bands may be the expressed fusion protein is western blotting, and immunochemical detection of the MBP protein using anti-MBP antibodies. The antibodies can be used to label specific bands from fractions run on an SDS-PAGE gel. Figure 5.10 shows the detection of the MBP by western blotting using anti-MBP antibodies (as per Section 2.2.1).

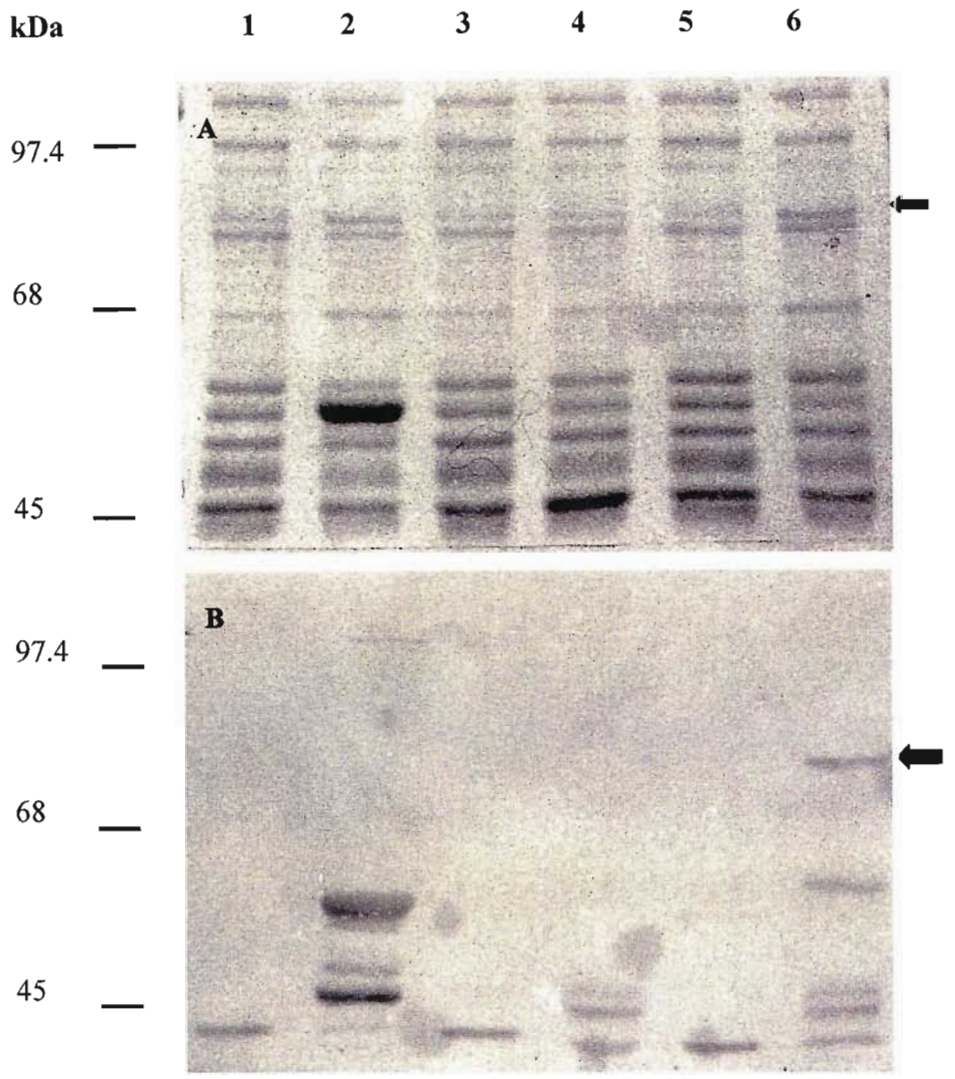


Figure 5.10 Western blot analysis of MBP and MBP-Pf32.5 expression using anti-MBP anitbodies. (A). SDS-PAGE (7.5%; 4%) Lane 1 - non-induced MBP control, lane 2 - induced MBP control, lane 3 - non-induced pTS822A, lane 4 - induced pTS822A, lane 5 - non-induced pTS822B, lane 6 - induced pTS822B. (B) Western blot analysis of SDS-PAGE gel in (A), using rabbit anti-MBP antibodies and goat anti-rabbit antibodies with HRPO conjugate for substrate based detection. The arrow indicates the fusion protein MBP-Pf32.5.

As can be seen in the control pMAL (no insert) lane 2, a large band was detected estimated to be approximately 50 kDa (Figure 5.10B). A band is also seen just below the 45 kDa marker. Lanes 3 and 4 are the recombinant pMAL plasmid with the insert in the opposite orientation. The expressed protein should be MBP2 (42,7 kDa) as the insert disrupts the reading frame and the full MBP *lacZα* fusion (50,8 kDa) is not synthesised. A band in lane 4 of Figure 5.10(A) corresponding to the size of MBP2 was also detected by anti-MBP antibodies in the western blot (Figure 6.10B). The band thought to represent the recombinant fusion protein, induced with IPTG (estimated to be approximately 75-80 kDa), was detected with anti-MBP antibodies in the western blot. This confirms that this band is an MBP fusion protein, and due to the size analysis, appears to be expressed at the expected size of 74.99 kDa.

There also appears to be a protein band detected at about 60 kDa and minor bands lower down in the gel. Due to the fact that the host *E. coli* strain is a cloning strain and not an expression strain (protease deficient), this may represent degradation products of the expressed protein due to host proteases. This is also seen in the other lanes where there are several other sizes detected. These lower molecular weight bands detected in the recombinant fractionate are present in the control lanes (lane 4), and so are not as a result of the open reading frame, Pf32.5. The lowest molecular weight band seen detected in the western blot appears to be present in the non-induced lanes, and so may be a non-specific detection.

5.2.9 Induction and expression of GST-fusion proteins MAP1 and MAP2 and detection using anti-peptide antibodies

Figure 5.11 shows analysis of fusion protein expression using SDS-PAGE, for the protein kinases, MAP1 and MAP2. These are GST fusions, previously cloned into pGEX-4T-3. GST-MAP1 is approximately 76 kDa, and GST-MAP2, 69 kDa. The GST protein is not shown in this figure as it is approximately 27 kDa and is poorly resolved at a SDS-PAGE gel percentage of 7.5%. The fusion proteins are seen to be expressed in the induced sample lanes, at sizes estimated using the molecular markers to be 75 kDa and 70 kDa, for GST-MAP1 and GST-MAP2 respectively. Due to these estimated sizes corresponding to the predicted sizes for these fusions, these expressed bands are suggested as being the fusion proteins of interest.

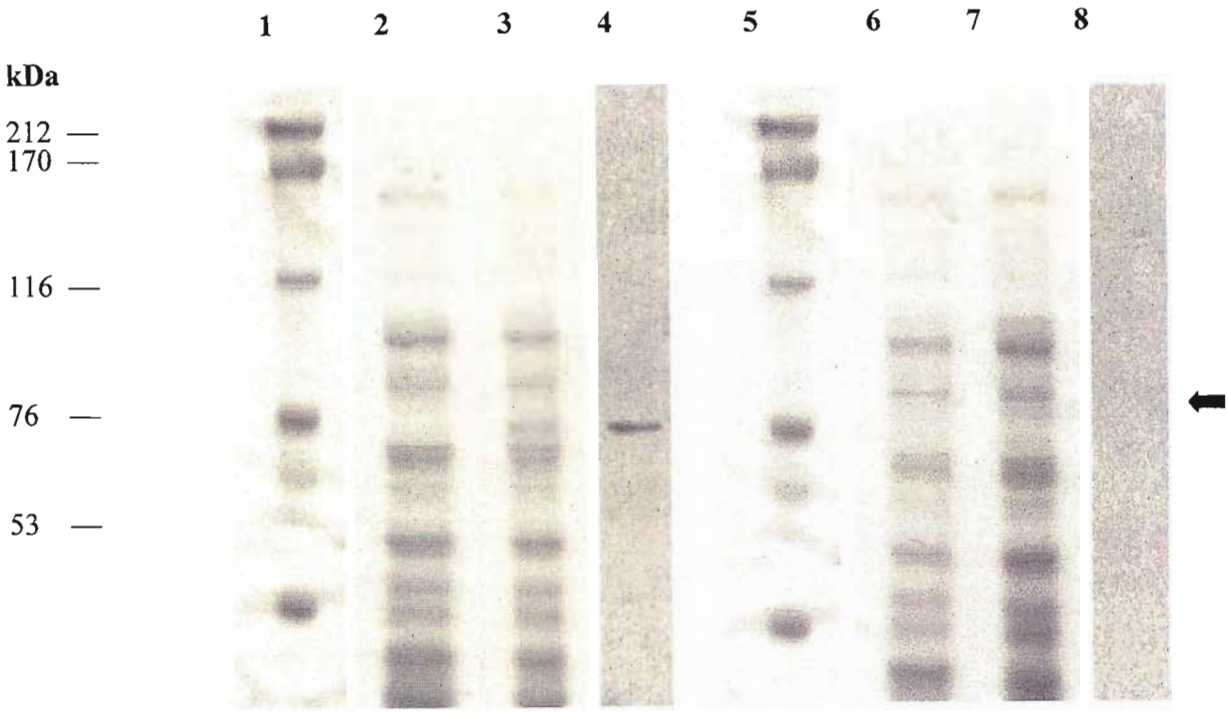


Figure 5.11. SDS-PAGE analysis of expressed MAP1 and MAP2 GST-fusion proteins, and detection of MAP2 on a Western blot. (7.5% running gel; 4% stacking gel). Lanes 1 and 5 - High weight molecular markers (Myosin. 212 kDa; α2-macroglobulin, 170 kDa; β-Galactosidase, 116 kDa; Transferrin, 76 kDa; Glutamic dehydrogenase, 53 kDa); lane 2 - non-induced GST-MAP2; lane 3 - induced GST-MAP2; lane 4 - western blot analysis using anti-map-2 antibody; lane 6 - non-induced GST-MAP1; lane 7 - induced GST-MAP1; lane 8 - western blot analysis using anti-map-1 antibody. The arrow shows the position where GST-MAP1 should be expressed.

Antibodies raised in chickens against peptides contained within the MAP1 and MAP2 proteins (peptides VHE and CPQ) were used in a western blot experiment as the primary antibody to detect the expressed proteins. As can be seen GST-MAP2 (lane 4) was detected, verifying the expression seen for this construct. Though the MAP1 protein is visibly expressed (lane 7) the antibodies used did not give any positive detection band.

5.3 Discussion and conclusion

These experiments describe the sub-cloning of the 822bp malaria insert from the pMOS817 plasmid into the expression vector pMAL-C2x, to form pTS822. The white colonies obtained in the transformations were screened and analysed, to identify the size and verify the sequence of the insert DNA. Due to the fact that the insert DNA prepared for cloning contains *Bam*HI sites on either end of the fragments, the insert can be cloned into the *Bam*HI site of pMAL-C2x in either orientation. Also, because only one restriction site is used in the polylinker, there is a high degree of re-circularisation of vector DNA without the incorporation of insert. To reduce this, the vector DNA can be dephosphorylated using calf intestinal alkaline phosphatase (CIAP), which removes the 5' phosphates from the termini of DNA fragments, preventing self-ligation. Increasing the amount of insert used in the reaction also increases the chances for recombination. As can be seen from these experiments, the presence of several white colonies (β-gal$^-$) in the test transformations, and the identification of a positive clone, negates the need to retry other methods to increase the recombination efficiency.

It is interesting to note that in the 'no-ligase' control, there are still a number of colonies seen, resulting from plasmid that were not digested in the initial *Bam*HI digestion of vector DNA. Even though there was no evidence of circular vector when the reaction was visualised on an agarose gel, these background colonies result. The appearance of hundreds of colonies when ligase was included in this reaction indicates the effectiveness of the enzyme in ligating DNA termini. The transformation efficiency (2.6×10^7 cfu/μg DNA) was high enough to conduct these experiments for the competent cells prepared by the $CaCl_2$ technique. Although this efficiency was lower than that of the MOS*Blue* cells used for blunt-cloning the initial PCR product, this was expected as the transformation efficiency decreases with an increase in the size of the plasmid DNA. The pMAL vectors are 6648 bp, compared to the 2887 bp pMOS*Blue* vector. Due to the size of the vector it is more difficult for the plasmid to be introduced into the cell, and smaller particles are generally transformed more efficiently.

The method used to determine the orientation of the insert DNA cloned into the vector, relied on restriction digestion analysis. This took advantage of the natural internal *Hin*dIII site in the insert sequence. Figure 5.12 shows the basis for these analyses, and the sizes of

the restriction fragments expected. As can be seen, digestion with *Hin*dIII will cleave at the position in the multiple cloning site and the site within the 822 bp insert. This allows the orientation of the insert to be determined. Digestion of the in-frame construct will result in a 629 bp fragment. This analysis (section 5.2.5) allowed a clone with the desired insert orientation to be identified and picked. The plasmid (pTS822) was grown up and stock cultures made, to allow for further experimentation with the plasmid.

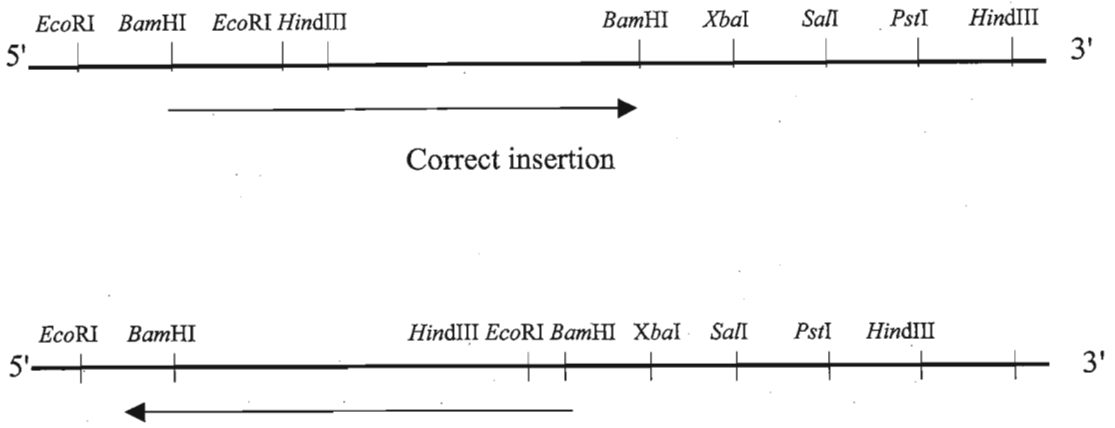


Figure 5.12 Restriction maps of pMAL ligated with 822 bp insert in both orientations, showing restriction sites used for analysis.

Sequencing the insert in pTS822 using universal pMAL sequencing primers verified the successful ligation of the insert, in frame with MBP. This also confirmed the sequence of the insert and served to check for any mutations in the sequence introduced during the original PCR reaction where the *Taq* polymerase may have caused mistakes in the synthesis. It is also interesting to note that the DNA sequence cloned, is identical to that obtained originally off the PlasmoDB (*Plasmodium falciparum,* 3D7) database. This result allows further protein induction and expression experiments to be conducted using this plasmid. The plasmid-containing insert DNA in the reverse orientation was also sequenced. It was predicted from this sequence data that the incorrect reading frame introduces a stop codon shortly after the start of the insert. IPTG was used to induce the bacteria to express the MBP fusion protein. This was analysed using SDS-PAGE where the protein bands between non-induced and induced cultures were compared (Figure 5.9).

MBP by itself (non-recombinant plasmid) showed efficient expression of the MBP protein and served as a control for the induction. Western blotting using antibodies raised against MBP were used to detect the protein. The position of the detection band in relation to the SDS-PAGE gel allows the identification of appropriate MBP fusion proteins. There was a minimal amount of expression seen for the recombinant fusion protein, MBP-Pf32.5.

In these experiments it was noted that some lower weight bands were thought to represent degradation products. The *E. coli* strain JM103 is primarily used for cloning experiments while other strains are more suited towards expression, and are protease deficient, reducing the degradation of fusion proteins. As levels of fusion protein expressed for the recombinant plasmid were relatively low, further experimentation can be performed to optimise the levels of expression. This plasmid, pTS822 can be re-transformed into a protease deficient host such as the BL21 strain. Optimisation of growth and induction temperatures and times, will determine and maximise protein expression. This degree of degradation may also contribute towards reduced protein expression levels, and re-transformation into an appropriate expression host, may increase the amount of fusion protein synthesised.

In a separate experiment, recombinant malaria protein kinases map-1 and map-2, were used to futher evaluate recombinant gene expression in *E. coli*, involving cloned malaria open reading frames. Two pGEX expression vectors were evaluated for their expression, GST-MAP1 and GST-MAP2. SDS-PAGE analysis was used to visualise the induction and expression of these proteins. Affinity purified antibodies raised in chickens against peptides from immunodominant regions contained in the sequences of these proteins were used to detect the fusion protein by western blot analysis. The anti-map-2 antibodies were successful in the detection of MAP2 whilst under the same conditions MAP1 was not detected. These results show how antibodies raised against peptides designed from the proteins sequence, can be used to positively identify recombinant expression. Where MBP fusion proteins were detected using antibodies specific to MBP, this method shows the specific detection of the malaria protein cloned.

pMOS817 construct - 817bp PCR product cloned into pMOS*Blue* vector

817 bp PCR product

```
5' ATG GAT gga tcc aac aac tat ctt ... aac cta caa cAT CGG ATC CCC GGG TAC CGA 3'
3' TAC CTA cct agg ttg aac tat ctt ... ttg gat gtt gTA GCC TAG GGG CCC ATG GCT 5'
    NdeI        BamHI                                    BamH        KpnI
                                                          SmaI
```

Restriction digestion of pMOS817 with *Bam*HI, producing insert for pMAL-C2x cloning

```
5'    GA TCC AAC AAC TAT CTT ... AAC CTA CAA CAT CG          3'
3'       G TTG TGG ATA GAA ... TTG CAT GTT GTA GCC TAG       5'
```

pMAL c2x multiple cloning region

```
          Ile  Glu  Gly  Arg
5' ATC GAG GGA AGG ATT TCA GAA TTC GGA TCC TCT AGA GTC GAC CTG CAG GCA AGC TTG 3'
3' TAG CTC CCT TCC TAA AGT CTT AAG CCT AGG AGA TCT CAG CTG GAC GTC CGT TCG AAC 5'

                              EcoRI  BamHI  XbaI  SalI  PstI  HindIII
```

Restriction digestion of pMAL c2x with *Bam*HI

```
     Ile  Glu  Gly  Arg  Ile  Ser  Glu
5'   C GAG GGA AGG ATT TCA GAA TTC G          GA TCC TCT AGA GTC GAC CTG CAG GCA C 3'
3'   G CTC CCT TCC TAA AGT CTT AAG CCT AG          G AGA TCT CAG CTG GAC GTC CGT T 5'
                               EcoRI            XbaI  SalI  PstI  HindIII
```

Cohesive-end ligation of insert DNA into restricted pMAL-c2x vector in correct orientation

```
     Arg  Ile  Ser  Glu  Phe  Gly  Ser  Asn  Asn  Tyr  Leu
5' ATT TCA GAA TTC Gga tcc aac aac tat ctt ... aac cta caa cat cgG ATC CTC TAG  3'
3' TAA AGT CTT AAG CCT Agg ttg ttg ata gaa ... ttg gat gtt gta gcc tag GAG ATC  5'
          EcoRI  BamHI                                                XbaI
```

Figure 5.13. Schematic diagram showing the construction of pTS822 by cohesive-end ligation of sub-cloned insert (pMOS817) into the pMAL-C2x expression vector.

The detection of expressed fusion protein MBP-Pf32.5, combined with the results obtained from sequencing, verify the successful synthesis of recombinant protein coded by the 822 bp insert (276 amino acids) cloned in these experiments. Figure 5.13 shows an overall summary of the cloning scheme used for these experiments, based on the insert obtained from the pMOS817 plasmid. The diagram also indicates how the correct translational reading frame is preserved in the ligation of the foreign open reading frame into pMAL-C2x. This plasmid can now be used as a tool for investigating properties of this region of Pf403 (amino acids 2355 - 2631). Analysis on the various features within the protein sequence (T cell epitopes, phosphorylation sites etc.) can be performed. The recombinant protein can be used to raise antibodies, for further immunological studies. These in turn can be used to identify the presence and localisation of the protein using immunofluorescent microscopy. This clone serves as the first of several overlapping regions to be expressed for studies on antigenic properties. It codes for the region of Pf403 that is most similar to the protective mouse malaria antigen, Pc96.

CHAPTER 6
GENERAL DISCUSSION

Over four thousand humans die from malaria every day in sub-Saharan Africa, mostly due to the spread of drug resistance in *P. falciparum*, the parasite responsible for the majority of deaths, particularly in infants (Merckx *et al.*, 2003). To reduce the number of fatalities, there is an urgent need for control of the disease and the development of new effective chemotherapeutic agents and vaccines (Ridley, 2002). The availability of databases containing the sequenced genomes of *P. falciparum* and the rodent malaria *P. yoelii yoelii* has revolutionised the study of malaria. A wealth of molecular biological information can be gained from utilisation of this sequence data, providing insights into the genetic and biochemical mechanisms within the parasite. The DNA sequencing of both the human pathogen *P. falciparum* and the rodent model *P. yoelii yoelii*, has allowed research into possible drug and vaccine targets. Homologues of novel proteins of interest can be identified, cloned and expressed for further characterization.

Efforts based on the development of a human malaria vaccine have largely been based on the assumption that antibodies mediate protection. However it is evident that cell-mediated mechanisms play a significant role in resistance to malaria (Grun and Weidanz, 1981). T cell epitopes on the parasite antigens, providing protective cellular immune responses have been considered as potential immunizing agents. The purpose of this study was the identification of *P. falciparum* malaria proteins that may share homology to Pc96, due to the unique immunological properties of this mouse antigen (Wanidworanun *et al.*, 1989). The protein has been shown to be associated with the outer membrane of mouse erythrocytes, using antibodies raised against Pc96. A T cell clone that provided protection when grafted into nude mice challenged with *P. chabaudi adami* (Brake *et al.*, 1988) proliferated in the presence of recombinant and native Pc96. The mice produced antibodies against an array of parasite antigens, similar to those seen in infected euthymic mice. This was unexpected due to the restricted ability of the clone to recognise diverse antigenic epitopes. This may suggest that this T cell clone was capable of providing help to a variety of B cell clonotypes, specific to a range of malaria antigens (Goldring *et al.*, 1989). It was therefore of interest to identify the *P. falciparum* homolog and to obtain recombinant protein material to begin characterization of the protein, and the role that the protein may have in promoting a human immune response.

In the experiments described in Chapter 3, open reading frames containing overlapping identity to the amino acid sequence of Pc96 (*P. chabaudi adami*) were identified in *P. falciparum* and *P. yoelii yoelii.* The sequences were obtained from the PlasmoDB database and relate to Pc96 at the primary and secondary structure levels, with the identification of possible tertiary structure and function. Pyy84 and Pyy178 were shown to resemble regions of Pf403, the *P. falciparum* homologue of Pc96. Pc96 was evaluated for the presence of T cell epitopes using the online T cell epitope recognition software, SYFPEITHI. Three of the peptides identified were shown to be conserved in the homologous regions of Pf403 and Pyy178 identified in *Plasmodium falciparum* and *Plasmodium yoelii yoelii* respectively, and may play a role in cell-mediated immunity. To evaluate the possible presence of these epitopes in other malaria proteins, the *P. falciparum* database (PlasmoDB) was screened using the epitope sequences in a BLAST search for short nearly exact matches. In particular one epitope (LIKYKFNII) conserved in Pc96, Pf403 and Pyy178 was aligned with several other malaria proteins.

In an effort to elucidate features of structural homology in these proteins, the sequences were screened against a database (3DPSSM) of known tertiary structure profiles. Two regions of Pf403 were found to resemble the catalytic domain of the *Botulinum* neurotoxin, a zinc metalloprotease, and several other features were found to indicate membrane association and DNA binding. No other homologues were identified in other organisms providing functional information.

In order to further study Pf403 to establish a possible function for the protein and its interaction with the human host immune system, a series of PCR reactions were designed. A range of oligonucleotide primers were designed for the *in vitro* amplification of the gene using PCR, from purified *P. falciparum* DNA, in preparation for cloning and recombinant expression. The strategy for cloning in these studies was based on PCR-cloning, in which restriction enzyme sites are included into the primer design for digestion and ligation into the vectors (Kaufman and Evans, 1990). A number of open reading frames were successfully cloned in this way. Cloning and recombinant expression of a variety of protein kinases involved in the regulation of parasite development and programming of the cell cycle, was achieved. PCR-cloning strategies were primarily used, resulting in the production and isolation of fusion proteins in *E. coli* expression systems. Two mitogen-activated protein kinase homologues, Pfmap-1 and Pfmap-2, were cloned into the pGEX expression vector

(GST fusion proteins) and were shown to be activated by phosphorylation. This indicates that the open reading frame cloned, results in the correct folding and synthesis of the protein in an active form. Methods used for the induction and expression of these two fusion proteins were used in this study as a control for expression of malaria proteins in *E. coli* expression systems. Recently, three novel cyclin-related proteins from *P. falciparum* were identified and characterised by molecular cloning and bacterial expression (Merckx *et al.*, 2003). In this study, these pGEX plasmids, expressing pfmap-1 and pfmap-2 were used as controls for protein expression and evaluation on western blots with specific antibodies.

Recombinant malaria proteins can be used to determine the interactions of the proteins with components of the immune system. In Chapter 5, the pfmap-1 and pfmap-2 GST fusion proteins were detected in a western blot system. Antibodies were previously raised in chickens against peptides contained within the amino acid sequence of the proteins. The successful detection of pfmap-2 using anti-peptide antibodies, allows the antibodies to be used in further experiments such as localisation of the native proteins using immunofluorescent microscopy. The antibodies can be used to screen expression libraries, identify the cross-reactivity of the antibodies, and for the isolation of native protein using affinity chromatography. Further experiments will involve the evaluation of antibodies isolated from the sera of humans infected with *P. falciparum*, with regard to the recognition of the recombinant proteins. The extent of recognition of proteins by the human immune system can be evaluated using western-blotting and enzyme-linked immunosorbent assay (ELISA)-based techniques. This approach will form a basis for the evaluation of recombinant proteins related to the Pc96 antigen, and their natural effect on infected populations of humans.

Due to the unknown function of the Pc96 protein, and the difficulty in identifying domains and regions of specific 3D conformation, the region of Pf403 corresponding to, or overlapping Pc96 and Pyy178 was initially chosen for cloning. This region includes the two conserved putative predicted T cell epitopes along with other features such as the leucine zipper conformation, and a casein kinase phosphorylation site, in the hopes that the conserved nature and extended region of similarity to Pc96 would render a region of the protein with a specific structure, conformation and function. Studies on this expressed protein will provide useful information about Pf403, and the Pyy178 and Pc96 proteins. Certain immunological experimentation is not possible using human malaria, due to the difficulty in obtaining human

material from malaria patients and complex regulations involving human studies. With Pyy178 in *Plasmodium yoelii yoelii*, studies using the rodent model of infection are possible.

An initial study of the amplification of target malaria genes was performed using PCR amplification of a 757 bp target from *Plasmodium falciparum*. The primers were designed to amplify this region of the acetyl-coA carboxylase gene. Previously isolated genomic *P. falciparum* (3D7) DNA was used as a template for the reaction, and resulted in the amplification of the 757 bp sequence, as verified by agarose gel electrophoresis. The amplicons identity was confirmed using restriction digestion analysis. This PCR reaction was used as a control for many of the reactions attempted for the amplification of target regions in MAL3P6.11, encoding Pf403. One pair of oligonucleotides resulted in the amplification of a 817 bp target region. A *Bam*HI restriction site was included into the design of the forward primer, to facilitate directional cloning into plasmid expression vectors. The PCR product was identified by restriction digestion analysis, using restriction sites based within the sequence.

Attempts at directly cloning the 817 bp PCR product into pGEX were unsuccessful (Chapter 4). The strategy involved the blunt ending of the PCR product using the Klenow enzyme, followed by restriction digestion with *Bam*HI to generate a 5' cohesive overhang. These fragments would have contained a cohesive and a blunt end for directional ligation into the *Bam*HI and *Sma*I sites in the poly-cloning region. It may have been that the *Bam*HI site was not digested, preventing ligation with pGEX, or that the extent of ligation was too low to identify recombinants with the screening methods used. The use of the pMOS*Blue*-cloning vector, introduced the α-complementation system of screening where colonies of the white phenotype are selected as recombinants. This is observed directly off the transformation plate allowing immediate selection of putative positive clones, in a system where the extent of recombination may be too low to detect using conventional screening procedures. The incorporation of the insert in the cloning vector allowed the *Bam*HI sites to become internalised, allowing complete and efficient digestion with restriction enzymes, as analysed by agarose gel electrophoresis. The introduction of this step allowed the sub-cloning of the insert DNA into the pMAL expression vector.

The pMAL expression vector (New England Biolabs) was chosen in this study due to its success in previous cloning experiments. These involved PCR amplification of open reading frames from *P. falciparum*, in particularly several protein kinases, by Doerig's group over the

last number of years (Merckx *et al.*, 2003). The vector offers the blue/white screening capabilities, simplifying the process of the identification and screening of recombinants, and enabling the isolation of fusion proteins using affinity chromatography (Guan *et al.,* 1987). The clones obtained were analysed using various restriction enzyme cleavage sites contained within the vector and the insert DNA. The insert was sub-cloned from pMOS, containing *Bam*HI sites on either termini, one within the forward primer sequence used for the original amplification, and the other derived from the pMOS polylinker region. The designed *Bam*HI site in the insert served to clone the fragment in frame to the *lacZ* coding sequence of pMAL, forming a fusion protein with MBP. Successful in frame ligation depended on the inserts orientation, which could have been in either direction due to the *Bam*HI sites at both termini (Figure 5.12, Chapter 5). The polylinker region in the recombinant pMAL plasmid was sequenced to confirm that the insert had ligated in frame with the MBP coding sequence. The recombinant pMAL vector was named pTS822 and was used in a series of studies on protein expression.

The fusion protein encoded by pTS822 includes a 32,5 kDa region of the *Plasmodium falciparum* open reading frame, Pf403 and a 42,5 kDa MBP protein. The theoretical mass of the MBP fusion protein is 75 kDa. In initial experiments, using standard growth and induction conditions, and SDS-PAGE, it appears as if there is low expression of a fusion protein estimated to be at a similar size to that expected. Anti-MBP antibodies raised against the MBP protein were used to detect fusion proteins more sensitively by western blotting. This was performed directly from the SDS-PAGE gels used to resolve the *E. coli* cellular proteins. The antibodies detected the control and non-recombinant pMAL MBP protein, as well as the fusion protein seen at approximately 75 kDa. Although several products of other sizes were detected, these are a result of protein degradation effects, and are seen in the controls. These degradation products can be removed by further optimisation of expression with this plasmid. Isolated pTS822 DNA can be used to re-transform an *E. coli* host strain more suited to protein expression than cloning experiments. The strain used for these expression experiments was JM103, a strain more genetically suited to cloning experiments (Chapter 5). Strains such as Bl21 and DH5α may be more suited towards expression of the fusion protein, due to the lack of host protease enzymes. The induction experiments can also be optimised for temperature and time and duration of induction. The amounts of expression in relation to other *E. coli* proteins will obviously influence the amount that can be isolated from an affinity column using MBP's affinity for maltose. As the PCR product was initially cloned into the pMOS*Blue*

vector (Chapter 3), and the insert cloned further into the pMAL vector (Chapter 4), it should be possible to sub clone the insert into the pGEX vector using a similar strategy. Although MBP has been succesful in the expression of foreign proteins, fusion with an alternative protein such as GST might improve the expression of the recombinant protein.

The cloned region (homologous to Pc96) of the gene encoding Pf403 in *Plasmodium falciparum* (Pf32.5) will be used in further studies to validate certain features previously investigated in the protein sequence (Chapter 3) that may be used to study the nature of the recombinant protein and its possible involvement in an immune response in humans. T cell epitopes predicted within the sequence can be studied further, by synthesising peptides and studying MHC binding (Vogt *et al.,* 1994). A further approach can be the identification of T cells that may recognised the antigen, once processed naturally. Several assays are available that measure the response of T cells to antigens, such as the detection of cytokines and proliferation studies. Antibodies raised against IL-2 and IFN-γ, can be used in an ELISA system to measure T cell responses (Troye-Blomberg *et al.*, 1985). The MHC heavy chain has been engineered to allow the biotinylation of the MHC-peptide complex and subsequently detected by dye-bound avidin (Crawford *et al.*, 1998). This indicates the proportion of T cells specific to the antigen or T cell epitope studied. Another method for detecting cytokine production is the amplification of induced cytokine mRNA using quantitative real time reverse transcriptase polymerase chain reaction (RT-PCR). This has allowed the relative expression of cytokines to be quantitated in a reproducible and sensitive fashion in disease. This is also highly useful during vaccination studies with the peptides or antigen, as a determination of the recognition and activation of cytokine producing T cells (Schirle *et al.*, 2001).

The region of the Pf403 protein was chosen due to its similarity to the Pc96 antigen. Regions of the Pyy178 protein overlapping with Pc96 can be similarly cloned and expressed. This will provide further material to study the Pf32.5 recombinant protein, and any further regions of Pf403 expressed due to the availability of a mouse model for parallel *in vivo* experiments on the *P. yoelii yoelii* homolog. The establishment of mouse malaria with *P. yoelii yoelii*, will also provide opportunities to further study the function of Pyy178 in malaria possibly indicating a role for Pf403 in *P. falciparum*. The 817 bp PCR product amplified in this study from *P. falciparum* DNA, or similar targets amplified from Pyy178, can be used to screen various stages of parasite development, by the evaluation of mRNA in each stage. cDNA libraries can be generated allowing further detection and cloning studies to be performed.

To further study the Pf403 protein, B cell epitopes can be selected within the amino acid sequence and synthesised as peptides, which can be used to raise antibodies. This has been used successfully by Merckx *et al* (2003) in studies involving malaria protein kinases. Peptides were chosen within the sequences of recombinant Pfcyc-1, Pfcyc-2, Pfcyc-3 and Pfcyc-4, and synthesised, and used to raise antibodies in chickens. IgY was isolated from the chicken egg yolks according the method used by Polson *et al.* (1985) using polyethylene glycol (PEG) precipitation. The antibodies were affinity purified using the peptides, and used in various pull-down experiments involving malaria kinase proteins in lysed malaria material. Parasite extracts were exposed to the anti-peptide antibodies and native malaria proteins were immunoprecipitated, allowing an evaluation of kinase activity within the extract. The anti-peptide antibodies can be used in similar pull-down experiments, involving the isolation of native Pf403 protein, and comparisons to the recombinant Pf32.5. These antibodies can be further used in experiments such as immunofluorescent microscopy, for the detection of the protein in various organelles. As the Pc96 protein was identified in the surface membrane of red blood cells, this technique can be used to determine whether Pf403 is distributed similarly.

The expression of recombinant malaria proteins has played a vital role in understanding the biology, metabolism and immunology of malaria. Due to the relative difficulty in the isolation of pure native proteins from malaria extracts, this technique has been widely used for the production and isolation of recombinant proteins. The recombinant forms of these proteins have been shown to exhibit functional activity in many studies discussed. Recombinant malaria proteins show potential for use in vaccine research, as particular regions or features of proteins can be chosen according to their particular immunogenic characterisitics. Genetic engineering of recombinant antigens has allowed the synthesis of different components, hoping to achieve an effective multi-component response.

In these experiments a region of a *P. falciparum* protein Pf403, which was shown to exhibit close homology to the *P. chabaudi adami* antigen Pc96, was cloned and expressed in *E. coli.* The pMAL expression vector was used to clone and express a MBP fusion protein (MBP-Pf32.5). This recombinant protein could be used as a tool for evaluating its role in the *P. falciparum* parasite, and its possible function in eliciting a protective immune response in humans.

REFERENCES

Alarcon, J.B., G.W. Waine, and D.P. McManus. 1999. DNA vaccines: technology and application as anti-parasite and anti-microbial agents. In Advances in Parasitology. Vol. 42. J.R. Baker and D. Rollinson, editors. Academic Press Limited. 344-390.

Allison, A.C., and E.M. Eugui. 1983. The role of cell-mediated immune responses in resistance to malaria, with special reference to oxidant stress. *Annual Review of Immunology.* **1**:361-388.

Alonso, P.L., T. Smith, J.R.M. Amrstrong Schellenberg, H. Masanja, S. Mwankusye, H. Urassa, I. Bastos de Azevedo, J. Chongela, S. Kobero, C. Menendez, N. Hurt, M.C. Thomas, E. Lyimo, N.A. Weiss, R. Hayes, A.Y. Kitua, M.C. Lopez, W.L. Kilama, T. Teuscher, and M. Tanner. 1994. Randomised trial of efficacy of Spf66 vaccine against *Plasmodium falciparum* malaria in children in southern Tanzania. *The Lancet.* **344**:1175-1181.

Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D.L. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research.* **25**:3389-3402.

Alvarez, N., V.M. Gonzalez, and A. Jimenez-Ruiz. 2002. Be careful with the primers when screening your clones by polymerase chain reaction. *Analytical Biochemistry.* **308**:189-191.

Aucan, C., Y. Traore, F. Tall, B. Nacro, T. Traore-Leroux, F. Fumoux, and P. Rihet. 2000. High Immunoglobulin G2 (IgG2) and low IgG4 levels are associated with human resistance to *Plasmodium falciparum* malaria. *Infection and Immunity.* **68**:1252-1258.

Baird, J.K., I. Wiady, D.J. Fryauff, M.A. Sutanihardja, B. Leksana, H. Widjaya, T. Kysdarmanto, and B. Subianto. 1997. *In vivo* resistance to chloroquine by *Plasmodium vivax* and *Plasmodium falciparum* at Nabire, Irian Jaya, Indonesia. *American Journal of Tropical Medicine and Hygiene.* **56**:627-631.

Banic, D.M., J. de Oliveira-Ferreira, L.R. Pratt-Riccio, V. Conseil, D. Goncalves, R.R. Fialho, H. Gras-Masse, C.T. Daniel-Ribeiro, and D. Camus. 1998. Immune response and lack of immune response to *Plasmodium falciparum* P126 antigen and its amino-terminal repeat in malaria-infected humans. *American Journal of Tropical Medicine and Hygiene.* **58**:768-774.

Barry, M.A., W.C. Lai, and S.A. Johnston. 1995. Protection against mycoplasma infection using expression-library immunization. *Nature.* **377**:632-635.

Baruch, D.I., B. Gamain, J.W. Barnwell, J.S. Sullivan, A. Stowers, G. Galland, L.H. Miller, and W.E. Collins. 2002. Immunization of *Aotus* monkeys with a functional domain of the *Plasmodium falciparum* variant antigen induces protection against a lethal parasite line. *Proceedings of the National Academy of Sciences, USA.* **99**:3860-3865.

Baxevanis, A.D. 1998. Practical aspects of multiple sequence alignment. In Bioinformatics: a practical guide to the analysis of genes and proteins. A.D. Baxevanis and B.F.F. Ouellette, editors. Wiley, New York. 172-188.

Bhatia, A., P. Delplace, B. Fortier, J. F. Dubremetz, and A. Vernes. 1987. Immunochemical analysis of a major antigen of *Plasmodium falciparum* (P126) among ten geographic isolates. *American Journal of Tropical Medicine and Hygiene.* **36**:15-19.

Billingsley, F.F., and R.E. Sinden. 1997. Determinants of malaria-mosquito specificity. *Parasitology Today.* **13**:297-301.

Brake, D.A., C.A. Long, and W.P. Weidanz. 1988. Adoptive protection against *Plasmodium chabaudi adami* malaria in athymic nude mice by a cloned T cell line. *Journal of Immunology.* **140**:1989-1993.

Brake, D.A., and W.P. Weidanz. 1986. Antigen specific, interleukin-2-propagated T lymphocytes confer resistance to a murine malarial parasite. *Journal of Immunology.* **137**:347-350.

Bredenkamp, B., B.L. Sharp, S.D. Mthembu, D.N. Durrheim, and K.I. Barnes. 2001. Faliure of sulphadoxine-pyrimethamine in treating *Plasmodium falciparum* malaria in KwaZulu-Natal. *South African Medical Journal.* **91**:970-972.

Brenner, S.E. 1998. Practical database searching. In *Trends Guide to Bioinformatics*. M. Patterson and M. Handel, editors. Elsevier Trends Journals, Cambridge. 9-11.

Brock, T.D., M.T. Madigan, J.M. Martinko, and J. Parker. 1994. Biology of Microorganisms (7th edition). Prentice-Hall International Inc, New Jersey. 425-426.

Brown, T.A. 2001. Gene cloning and DNA analysis - an introduction. Blackwell Science, Oxford. 35-65.

Burmeister, M., and H. Lehrach. 1989. Isolation of large DNA fragments from agarose gels using agarase. *Trends in Genetics* **5**, 1132-1135.

Butler, D. 1997. Vaccines: A roller-coaster of hopes. *Nature*. **396:**537-538.

Calvo-Calle, J.M., J. Hammer, F. Sinigaglia, P. Clavijo, Z.R. Moya-Castro, and E.H. Nardin. 1997. Binding of malaria T cell epitopes to DR and DQ molecules in vitro correlates with immunogenicity *in vivo*: identification of a universal T cell epitope in the *Plasmodium falciparum* circumsporozoite protein. *Journal of Immunology*. **159**:1362-1373.

Carmenes, R.S., J.P. Freije, M.M. Molina, and J.M. Martin. 1989. Predict7, a program for protein structure prediction. *Biochemical and Biophysical Research Communications*. **159**:687-691.

Cavacini, L.A., L.A. Parke, and W.P. Weidanz. 1990. Resolution of acute malarial infections by T cell-dependent non-antibody-mediated mechanisms of immunity. *Infection and Immunity:* **58.** 2946-2950.

Charest, H., M. Sedegah, G.S. Yap, R.T. Gazzinelli, P. Caspar, S.L. Hoffman, and A. Sher. 2000. Recombinant attenuated *Toxoplasma gondii* expressing the *Plasmodium yoelii* circumsporozoite protein provides highly effective priming for $CD8^+$ T cell-dependent protective immunity against malaria. *Journal of Immunology*. **165**:2084-2092.

Chen, D.H., R.E. Tigelaar, and F.I. Weinbaum. 1977. Immunity to sporozoite-induced malaria infection in mice. I the effect of immunization of T and B cell deficient mice. *Journal of Immunology*. **118:**1322-1327.

Chougnet, C., M. Troye-Blomberg, P. Deloron, L. Kabilan, J.P. Lepers, J. Savel, and P. Perlmann. 1991. Human immune response in *Plasmodium falciparum* malaria. Synthetic

peptides corresponding to known epitopes of the Pf155/RESA antigen induce production of parasite-specific antibodies *in vitro*. *Journal of Immunology*. **147**:2295-2301.

Chulay, J.D., J.A. Lyon, J.D. Haynes, A.I. Meierovics, C.T. Atkinson, and M. Aikawa. 1987. Monoclonal antibody characterization of *Plasmodium falciparum* antigens in immune complexes formed when schizonts rupture in the presence of immune serum.. *Journal of Immunology*. **139**:2768-2774.

Clark, I.A. 1987. Cell-mediated immunity in protection and pathology of malaria. *Parasitology Today*. **3**:300-305.

Cohen, S., A. McGregor, and S. Carrington. 1961. Gamma globulin and acquired immunity to human malaria. *Nature*. **192**:733-737.

Cohen, S.N., A.C.Y. Chang, H.W. Boyer, and R.B. Helling. 1973. Construction of biologically functional bacterial plasmids *in vitro*. *Proceedings of the National Academy of Sciences, USA*. **70**:3240-3244.

Collins, F.H., and S.M. Paskewitz. 1995. Malaria: current and future prospects for control. *Annual Review of Entomology*. **40**:195-219.

Collins, W.E., R.F. Anders, M. Pappaionou, G.H. Campbell, G.V. Brown, D.J. Kemp, R.L. Coppel, J.C. Skinner, P.M. Andrysiak, J.M. Favaloro, L.M. Corcoran, J.R. Broderson, G.F. Mitchell, and C.C. Campbell. 1986. Immunization of *Aotus* monkeys with recombinant proteins of an erythrocyte surface antigen of *Plasmodium falciparum*. *Nature*. **323**:259-262.

Crawford, F., H. Kozono, J. White, P. Marrack, and J. Kappler. 1998. Detection of antigen-specific T cells with multivalent soluble class II MHC covalent peptide complexes. *Immunity*. **8**:675-680.

Daly, T.M., and C.A. Long. 1993. A recombinant 15-kilodalton carboxyl-terminal fragment of *Plasmodium yoelii yoelii* 17XL merozoite surface protein 1 induces a protective immune response in mice. *Infection and Immunity.* **61**:2462-2467.

Dame, J.B., J.L. Williams, T.F. McCutchan, J.L. Weber, R.A. Wirtz, W.T. Hockmeyer, W.L. Maloy, J.D. Haynes, I. Schneider, and D. Roberts. 1984. Structure of the gene encoding the immunodominant surface antigen on the sporozoite of the human malaria parasite *Plasmodium falciparum. Science.* **225**:593-599.

Dennison, C. 1999. A guide to protein isolation. Kluwer Academic Publishers, The Netherlands. 35-100.

Dingwall, C., and R.A. Lasky. 1991. Nuclear targetting sequences-a consensus? *Trends in Biochemical Sciences.* **16**:478-481.

Doolan, D.L., and S.L. Hoffman. 2000. The Complexity of protective immunity against liver-stage malaria. *The Journal of Immunology.* **165:**1453-1462.

Duffy, P.E., and D.C. Kaslow. 1997. A novel malaria protein, Pfs28, and Pfs25 are genetically linked and synergistic as falciparum malaria transmission-blocking vaccines. *Infection and Immunity.* **65**:1109-1113.

Egan, A.F., M.J. Blackman, and D.C. Kaslow. 2000. Vaccine efficacy of recombinant *Plasmodium falciparum* merozoite surface protein 1 in malaria-naive, -exposed, and/or -rechallenged *Aotus vociferans* monkeys. *Infection and Immunity.* **68**:1418-1427.

Elloso, M.M., H.C. Van der Heyde, J.A. Vande Waa, D.D. Manning, and W.P. Weidanz. 1994. Inhibition of *Plasmodium falciparum in vitro* by human gamma delta T cells. *Journal of Immunology.* **153**:1187-1194.

Falk, K., O. Rotzschke, S. Stevanovic, G. Jung, and H. Rammensee. 1991. Allele-specific motifs revealed by sequencing of self peptides eluted from MHC molecules. *Nature.* **351.** 52-58

Gantt, S., C. Persson, K. Rose, A.J. Birkett, R. Abagyan, and V. Nussenzweig. 2000. Antibodies against thrombospondin-related anonymous protein do not inhibit *Plasmodium* sporozoite infectivity *In Vivo*. *Infection and Immunity*. **68**:3667-3673.

Gardner, M.J., N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, J.M. Carlton, A. Pain, K.E. Nelson, S. Bowman, I.T. Paulsen, K.James, J.A. Eisen, K. Rutherford, S.L. Salzberg, A. Craig, S. Kyes, M. Chan, V. Nene, S.J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M.W. Mather, A.B. Vaidya, D.M.A. Martin, A.H. Fairlamb, M.J. Fraunholz, D.S. Roos, S.A. Ralph, G.I. McFadden, L.M. Cummings, G.M. Subramanian, C. Mungall, J.C. Venter, D.J. Carucci, S.L. Hoffman, C. Newbold, R.W. Davis, C.M. Fraser, and B.Barrell. 2002. Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature*. **415**:686-693.

Goldring, J.P.D., D.A. Brake, Cavacini. L.A, C.A. Long, and W.P. Weidanz. 1989. Cloned T cells provide help for malaria-specific polyclonal antibody responses. *European Journal of Immunology*. **19**:559-561.

Goodenham, K. 1984. Methods in Molecular Biology Vol. 1. Hymana, New Jersey. 165-178 pp.

Gozar, M.M.G., V.L. Price, and D.C. Kaslow. 1998. *Saccharomyces cerevisiae* - secreted fusion proteins Pfs25 and Pfs28 elicit potent *Plasmodium falciparum* transmission-blocking antibodies in mice. *Infection and Immunity*. **66**:59-64.

Grau, G.E., L.F. Fajardo, P.F. Piguet, B. Allet, P.H. Lambert, and P. Vassalli. 1987. Tumor necrosis factor (cachectin) as an essential mediator in murine cerebral malaria. *Science*. **237:**1210-1212.

Greenwood, B., K. Marsh, and R.W. Snow. 1991. Why do some children develop severe malaria? *Parasitology Today*. **7**:277-281.

Grun, J.L., and W.P. Weidanz. 1981. Immunity to *Plasmodium chabaudi adami* in the B-cell-deficient mouse. *Nature*. **290**:143-145.

Guan, C., P. Li, P.D. Riggs, and H. Inouye. 1987. Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose binding protein. *Gene.* **67**:21-30.

Hames, B.D., and D. Rickwood. 1981. Gel electrophoresis of proteins: a practical approach. IRL Press, Oxford. 32-40.

Herrington, D., D.F. Clyde, G. Losonsky, M. Cortesia, J.R. Murphy, J. Davis, S. Baqar, A.M. Felix, E.P. Heimer, and D. Gillessen. 1987. Safety and immunogenicity in man of a synthetic peptide malaria vaccine against *Plasmodium falciparum* sporozoites. *Nature.* **328**:257-259.

Herrington, D., J. Davis, E. Nardin, M. Beier, J. Cortese, H. Eddy, G. Losonsky, M. Hollingdale, M. Sztein, and M. Levine. 1991. Successful immunization of humans with irradiated malaria sporozoites: humoral and cellular responses of the protected individuals. *American Journal of Tropical Medicine and Hygiene.* **45**:539-547.

Ho, M., H.K. Webster, P. Tongtawe, K. Pattanapanyasat, and W.P. Weidanz. 1990. Increased gamma delta T cells in acute *Plasmodium falciparum* malaria. *Immunology Letters.* **25**:139-141.

Holder, A.A. 1999. Malaria vaccines. *Proceedings of the National Academy of Sciences, USA.* **96**:1167-1169.

Holton, T.A., and M.W. Graham. 1990. A simple and efficient method for direct clinging of PCR products using ddT- tailed vectors. *Nucleic Acids Research.* **19**:1156-1160.

Hopp, T.P., and W. K.R. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences, USA.* **78**:3824-3828.

Horwitz, J.P., J. Chua, R.J. Curby, A.J Tomson, M.A Da Rooge, B.E Fisher, J. Mauricio, and I. Klundt.1964. Substrates for cytochemical demonstration of enzyme activity. I. Some substituted 3-indolyl-_-D-glycopyranosides. *Journal of Medical Chemistry.* **9**, 447-448

Innis, M.A., and D.H. Gelfand. 1990. Optimization of PCRs. In PCR protocols. Innis, Gelfand, Sninsky, and White, editors. Academic Press, New York.

Janin, J. 1979. Surface and inside volumes in globular proteins. *Nature*. **277**:491-492.

Jayawardena, A.N., D.B. Murphy, C.A. Janeway, and R.K. Gershon. 1982. T cell-mediated immunity in malaria. I. The *Ly* phenotype of T cells mediating resistance to *Plasmodium yoelii*. *Journal of Immunology*. **129**:377-381.

Jovin, T.M. 1973. Multiphasic zone electrophoresis. I. Steady-state moving boundary systems formed by different electrolyte combinations. *Biochemistry*. **12**:871-879.

Kabilan, L., M. Troye-Blomberg, H. Perlmann, G. Andersson, B. Hogh, E. Peterson, A. Bjorkman, and P. Perlmann. 1988. T cell epitopes in Pf155/RESA, a major candidate for a *Plasmodium falciparum* malaria vaccine. *Proceedings of the National Academy of Sciences, USA*. **85**:5659-5663.

Kaiissar Mannoor, M., R. C. Halder, S.R.M Morshed, A Ariyasinghe, H.Y. Bakir, H. Kawamura, H. Watanabe, H. Sekikawa, and T. Abo. 2002. Essential role of extrathymic T cells in protection against malaria. *Journal of Immunology*. **169**:301-306.

Kappe, S.H.I., M.J. Gardner, S.M. Brown, J. Ross, K. Matuschewski, J.M. Ribeiro, J.H. Adams, J. Quackenbush, J. Cho, D.L. Carucci, S.L. Hoffman, and N. V. 2001. Exploring the transcriptome of the malaria sporozoite stage. *Proceedings of the National Academy of Sciences, USA*. **98**:9895-9900.

Karp, G. 1996. Cell and molecular biology. John Wiley & Sons, Inc, New York.

Karplus, P.A., and G.E. Schulz. 1985. Prediction of chain flexibility in proteins. *Naturwissemschaften*. **72**:221-228.

Kaufman, D.L., and G.A. Evans. 1990. Restriction endonuclease cleavage at the termini of PCR products. *Biotechniques*. **9**:304-306.

Kellerman, O.K., and T. Ferenci. 1982. Maltose binding protein from *E. coli*. *Methods in Enzymology*. **90**:459-463.

Kelley, L.A., R.M. MacCallum, and M.J.E. Sternberg. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *Journal of Molecular Biology.* **299**:499-520.

Klenow, H., and I. Henningsen. 1970. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proceedings of the National Academy of Sciences, USA.* **65**:168-175.

Kyte, J., and R.F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. *Journal of Molecular Biology.* **157**:105-132.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T. *Nature.* **277**:680-685.

Lal, A.A., P.S. Patterson, J.B. Sacci, J. A Vaughan, C. Paul, W.E. Collins., R.A. Wirtz, and A.F. Azad. 2001. Anti-mosquito midgut antibodies block development of *Plasmodium falciparum* and *Plasmodium vivax* in multiple species of *Anopheles* mosquitoes and reduce vector fecundity and survivorship. *Proceedings of the National Academy of Sciences, USA.* **98**:5228-5233.

Landschultz, W.H., P.F. Johnson, and S.L. McKnight. 1988. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science.* **240**:1759-1764.

Maina, C.V., P.D. Riggs, A.G. Grandea, B.E. Slatko, L.S. Moran, J.A. Tagliamonte, L.A. McReynolds, and C. Guan. 1988. A Vector to express and purify foreign proteins in *Escherichia coli* by fusion to, and separation from, maltose binding protein. *Gene.* **74**:365-373.

Merckx, A., K. Le Roch, M.P. Nivez, D. Dorin, P. Alano, G.J. Guiterrez, A.R. Nebreda, D. Goldring, C. Whittle, S. Patterson, D. Chakrabarti, and C. Doerig. 2003. Identification and initial characterization of three novel cyclin-related proteins of the human malaria parasite *Plasmodium falciparum. Journal of Biological Chemistry.* **10**:1074-1086.

Miller, L.H., M.F. Good, and G. Milon. 1994. Malaria pathogenesis. *Science.* **264**:1878-1883.

Mitchell, G.H., G.A. Butcher, and S. Cohen. 1975. Merozoite vaccination against *Plasmodium knowlesi* malaria. *Immunology*. **29**:397-407.

Mullis, K.B., and F.A. Faloona. 1987. Specific synthesis of DNA *in Vitro* via a polymerase - catalyzed chain reaction. *Methods in Enzymology*. **155**:335-350.

Nardin, E.H., J.M. Calvo-Calle, G.A. Oliveira, R.S. Nussenzweig, M. Schneider, J.M. Tiercy, L. Loutan, D. Hochstrasser, and K. Rose. 2001. A totally synthetic polyoxime malaria vaccine containing *Plasmodium falciparum* B cell and universal T cell epitopes elicits immune responses in volunteers of diverse HLA types. *Journal of Immunology*. **166**:481-489.

Oaks, S.C., V.S. Mitchell, G.W. Pearson, and C.C.J. Pearson. 1991. Malaria: obstacles and oppurtunitites. National Academy Press, Washington DC.

Ornstein, L. 1964. Disc electrophoresis - I - background and theory. *Annals of the New York Academy of Science*. 121:321-349.

Parker, K.C., M.A. Bednarek, and J.E. Coligan. 1994. Scheme for ranking potential HLA-A2 binding peptides based on independent binding of individual peptide side-chains. *Journal of Immunology*. **152**:163-169.

Perlmann, H., M. Berzins, M. Wahlgren, J. Carlsson, A. Bjorkman, M.E. Patarroyo, and P. Perlmann. 1984. Antibodies in malaria sera to parasite antigens in the membrane of erythrocytes infected with early asexual stages of *Plasmodium falciparum*. *Journal of Experimental Medicine*. **159**:1686-1704.

Phillips, R.S. 2001. Current status of malaria and potential for control. *Clinical Microbiology Reviews*. **14**:208-226.

Pitt-Rivers, R., and F.S.A. Impiombato. 1968. The binding of sodium dodecyl sulphate to various proteins. *Biochemical Journal*. **109**:825-830.

Plebanski, M., and A.V.S. Hill. 2000. The immunology of malaria infection. *Current Opinion in Immunology*. **12**:437-441.

Polson, A., T.H.T. Coetzer, J. Kruger, E. von Maltzahn, and K.J. Van der Merve. 1985. Improvements in the isolation of IgY from the yolks of eggs laid by immunised hens. *Immunological Investigations*. **14**:323-327.

Rammensee, H., J. Bachmann, N.P. Emmerich, O.A. Bachor, and S. Stevanovic. 1999. SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*. **50**.213-217.

Ranjit, M.R., and Y.D. Sharma. 1999. Genetic polymorphism of falciparum malaria vaccine candidate antigen genes among field isolates in India. *American Journal of Tropical Medicine and Hygiene*. **61**:103-108.

Ridley, R. 2002. Malaria: To kill a parasite. *Nature*. **415**:686-693.

Roberts, M.R., and W.P. Weidanz. 1983. T-cell immunity to malaria in the B-cell deficient mouse. *American Journal of Tropical Medicine and Hygiene*. **28**:1-3.

Romero, P., J.L. Maryanski, G. Corradin, R.S. Nussenzweig, V. Nussenzweig, and F. Zavala. 1988. Cloned cytotoxic T cells recognise an epitope in the circumsporozoite protein and protect against malaria. *Nature*. **341**:323-326.

Rosenberg, R., R.A. Wirtz, D.E. Lanar, J. Sattabongkot, T. Hall, A.P. Waters, and C. Prasittisuk. 1989. Circumsporozoite protein heterogeneity in the human malaria parasite *Plasmodium vivax*. *Science*. **245**:973-976.

Ruangjirachuporn, W., R. Udomsangpetch, J. Carlsson, D. Drenckhahn, P. Perlmann, and K. Berzins. 1991. *Plasmodium falciparum*: analysis of the interaction of antigen Pf155/RESA with the erythrocyte membrane. *Experimental Parasitology*. **73**:62-72.

Sambrook, J., E.F. Fritch, and T. Maniatis. 1989. Molecular Cloning: A laboratory manual. Cold Spring Harbor Laboratory Press, New York.

Schirle, M., T. Weinschenk, and S. Stevanovic. 2001. Combining computer algorithms with experimental approaches permits the rapid and accurate identification of T cell epitopes from defined antigens. *Journal of Immunological Methods*. **257**:1-16.

Schmitz, G.G., T. Walter, R. Seibl, and C. Kessler. 1991. Nonradioactive labelling of oligonucleotides in vitro with the hapten digoxigenin by tailing with terminal transferase. *Analytical Biochemistry*. **192**:222-231.

Schofield, L., J. Villaquiran, A. Ferreira, H. Schellekens, R.S. Nussenzweig, and N. V. 1987. Gamma interferon, $CD8^+$ T cells and antibodies required for immunity to malaria sporozoites. *Nature*. **330**:664-666.

Sedegah, M., R. Hedstrom, P. Hobart, and S.L. Hoffman. 1994. Protection against malaria by immunization with plasmid DNA encoding circumsporozoite protein. *Proceedings of the National Academy of Sciences, USA*. **91**:9866-9870.

Seixas, E., L. Fonseca, and J. Langhorne. 2002. The influence of T cells on the $CD4^+$ T cell and antibody response during a primary *Plasmodium chabudi chabudi* infection in mice. *Parasite Immunology*. **24**:131-140.

Shagger, H., and G. von Jagow. 1987. Tricine-sodium dodecyl-sulphate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1-100kDa. *Analytical Biochemistry*. **166**:368-379.

Shapiro, A.L., E. Vinuela, and J.V.J. Maizel. 1967. Molecular weight estimation of polypeptide chains by electrophoresis in SDS-polyacrylamide gels. *Biochemical and Biophysical Research Communications*. **28**:815-820.

Sharp, B.L., and D. Le Sueur. 1996. Malaria in South Africa - the past, the present and selected implications for the future. *South African Medical Journal*. **86**:83-89.

Shi, Y.P., S.E. Hasnain, J.B. Sacci, B.P. Holloway, H. Fujioka, N. Kumar, R. Wohlhueter, S.L. Hoffman, W.E. Collins, and A.A. Lal. 1999. Immunogenicity and *in vitro* protective efficacy of a recombinant multistage *Plasmodium falciparum* candidate vaccine. *Proceedings of the National Academy of Sciences, USA*. **96**:1615-1620.

Sigrist, C.J., L. Cerutti, N. Hulo, A. Gattiker, L. Falquet, M. Pagni, A. Bairoch, and P. Bucher. 2002. PROSITE: a documented database using patterns and profiles as motif descriptors. *Brief Bioinformatics*. **3**:265-274.

Smooker, P.M., Y.Y. Setiady, A. Rainczuk, and T.W. Spithill. 2000. Expression library immunization protects mice against a challenge with virulent rodent malaria. *Vaccine.* **18**:2533-2540.

Stanley, S.L. 1998. Malaria vaccines: are seven antigens better than one? *The Lancet.* **352**:1163-1164.

Stowers, A.W., D.B. Keister, O. Muratova, and D.C. Kaslow. 2000. A region of *Plasmodium falciparum* antigen Pfs25 that is the target of highly potent transmission-blocking antibodies. *Infection and Immunity.* **68**:5530-5538.

Stroud, R., K. Reiling, M. Wiener, and D. Freymann. 1998. Ion-channel-forming colicins. *Current Opinion in Structural Biology.* **8**:525-533.

Su, X., Y. Wu, C.D. Sifri, and T.E. Wellems. 1996. Reduced extension temperatures required for PCR amplification of extremely A+T-rich DNA. *Nucleic Acids Research.* **24**:1574-1575.

Sykes, R.B., and M. Matthew. 1976. The beta-lactamases of gram-negative bacteria and their role in resistance to beta-lactam antibiotics. *Journal of Antimicrobial Chemotherapy.* **2**:115-157.

Targett, G.A. 1995. Malaria vaccines - now and the future. *Translations of the Royal Society of Tropical Medicine and Hygiene.* **89**:585-587.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1996. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research.* **22**:4673-4680.

Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitro-cellulose sheets: procedure and some applications. *Proceedings of the National Academy of Sciences, USA.* **76**:4350-4354.

Troye-Blomberg, M., G. Andersson, M. Stoczkowska, R. Shabo, P. Romero, M.E. Patarroyo, W. H, and P. Perlmann. 1985. Production of IL 2 and IFN-gamma by T cells from malaria patients in response to *Plasmodium falciparum* or erythrocyte antigens in vitro. *Journal of Immunology.* **135**:3498-3503.

Ulmer, J.B., J.J. Donnelly, S.E. Parker, G.H. Rhodes, P.L. Felgner, V.J. Dwarki, S.H. Gromkowski, R.R. Deck, C.M. De Witt, A. Friedman, L.A. Hawe, K.R. Leander, D. Martinez, H.C. Perry, J.W. Shiver, D.L. Montgomery, and M.A. Liu. 1993. Heterologous protection against influenza by injection of DNA encoding a viral protein. *Science.* **259**:1745-1749.

Van der Heyde, H.C., M.M. Elloso, W.L. Chang, M. Kaplan, D.D. Manning, and W.P. Weidanz. 1995. Gamma delta T cells function in cell-mediated immunity to acute blood- stage *Plasmodium chabaudi adami* malaria. *Journal of Immunology.* **154**:3985-3990.

Van Regenmortel, M.H.V., J.P. Briand, S. Muller, and S. Plaue. 1988. Synthetic polypeptides as antigens. Elsevier, New York. 2-41 pp.

Vogt, A.B., H. Kropshofer, H. Kalbacher, M. Kalbus, H.G. Rammensee, H.G. Coligan, and R. Martin. 1994. Ligand motifs of HLA-DRB5*0101 and DRB1*1501 molecules delineated from self-peptides. *The Journal of Immunology.* **153**:1665-1670.

Vonderheide, R.H., W.C. Hahn, J.L. Schultze, and L.M. Nadler. 1999. The telomerase catalytic subunit is a widely expressed tumor-associated antigen recognized by cytotoxic T lymphocytes. *Immunity.* **10**:673.

Walin, B., A. Sjolander, N. Ahlborg, R. Udomsangpetch, A. Scherf, D. Mattei, K. Berzins, and P. Perlmann. 1992. Involvement of Pf155/RESA and cross-reacting antigens in *Plasmodium falciparum* merozoite invasion in vitro. *Infection and Immunity.* **60**:443-449.

Wang, R., Y. Charoenvit, G. Corradin, R. Porrozzi, R.L. Hunter, G. Glenn, C.R. Alving, P. Church, and S.L. Hoffman. 1995. Induction of protective polyclonal antibodies by immunization with a *Plasmodium yoelii* circumsporozoite protein multiple antigen peptide vaccine. *Journal of Immunology.* **154**:2784-2793.

Wanidworanun, C., J.W. Barnwell, H. Nagasawa, M. Aikawa, and H.L. Shear. 1989. Cross-reacting antigens to Pc96, a protective antigen of *Plasmodium chabaudi*, in *P. falciparum, P. vivax* and *P. cynomolgi. American Journal of Tropical Medicine and Hygiene.* **40**:579-584.

Weber, K., and M. Osborn. 1969. The reliability of molecular weight determinations by dodecyl-sulphate-polyacrylamide gel electrophoresis. *Journal of Biological Chemistry.* **244**:4406-4412.

Welling, G.W., and W.J. Weijer. 1985. Prediction of sequential antigenic regions in proteins. *FEBS Letters.* **188**:215-218.

Wiener, M., D. Freymann, P. Ghosh, and R. Stroud. 1997. Crystal structure of colicin Ia. *Nature.* **385**:461-464.

Zavala, F., J.P. Tam, M.R. Hollingdale, A.H. Cochrane, I. Quakyi, R.S. Nussenzweig, and V. Nussenzweig. 1985. Rationale for development of a synthetic vaccine against *Plasmodium falciparum* malaria. *Science.* **228**:1436-40.