CLONING OF THE PROMOTER REGIONS OF *TRYPANOSOMA BRUCEI* AND *TRYPANOSOMA CONGOLENSE* CYSTEINE PROTEASE GENES

by

THEMBILE LAWRENCE DALASILE

Submitted in fulfilment of the academic
requirements for the degree of
Master of Science
in the
Department of Biochemistry
University of Natal

Pietermaritzburg
1997

ABSTRACT

Trypanosoma brucei and *T. congolense* are protozoan parasites that infect humans, domestic livestock and wildlife in Africa. These parasites undergo complex morphological and biochemical changes, during the various stages of their life cycle. These changes correlate with alterations in the levels of trypanosomal lysosomal cysteine proteases, suggesting a role for transcriptional regulation of the cysteine protease in these parasites. The mechanism of this regulation is not yet understood nor have the promoter regions of the cloned trypanosome cysteine protease genes been investigated. This study involved an attempt to clone the *T. brucei* and *T. congolense* DNA fragments containing the promoter regions as the initial step in the investigation of the control elements of the cysteine protease gene.

Trypanosomes were isolated from infected rat blood employing a combination of the methods of isopicnic isolation on Percoll gradients and DEAE-cellulose anion exchange resin chromatography. Approximately 5 x 10^9 viable trypanosome cells were isolated from the infected rat blood and chromosomal DNA (approximately 500 µg) was extracted by alkaline-lysis method. Trypanosome genomic libraries were initially constructed in *Eschericia coli* HB101 employing the positive selection vector p*Eco*R251. The *Trypanosoma brucei* p*Eco*R251 library contained 6 000 recombinants and the *Trypanosoma congolense* library contained 15 000 recombinants. Plasmid DNA was then extracted from pools of recombinants, employing the alkaline-lysis method, digested with *Eco*R1 restriction endonuclease and resolved by agarose gel electrophoresis. After Southern hybridisation, the p*Eco*R251 libraries did not reveal any putative clones containing the fragment of interest when probed with both an oligonucleotide probe and the PCR generated dsDNA probe. Genomic libraries were then constructed in the phagemid pUC119. The *T. brucei* and *T. congolense* genomic libraries contained 33 000 and 27 000 recombinants respectively. Recombinants from the *T. brucei* and *T. congolense* libraries were pooled in lots of 400 and 300 respectively. Of the 80 *T. brucei* plasmid pools that were screened 30 pools contained fragments that hybridised with the probe whilst 12 pools from the 90 *T. congolense* library pools that were screened contained fragments that hybridised with the probe. Putative clones identified appeared to contain inserts, ranging between two and seven kb in size. A partial *T. congolense* library consisting of approximately 12 pools was screened by colony hybridisation for identification of individual clones and 76 putative clones were identified. After confirmation of these putative clones on a dot blot using a DIG-labelled dsDNA

probe, a selection of 30 putative clones were subjected to Southern hybridisation using a DIG-labelled DNA probe. Following Southern hybridisation 23 putative clones were identified to contain DNA inserts of interest in the range of two to seven kb. Five clones, designated pCPC1, pCPC2, pCPC3, pCPC4 and pCPC5 were then selected for further restriction mapping. Clone pCPC4 contains a seven kb fragment of *T. congolense* genomic DNA. A partial *T. brucei* library consisting of approximately 30 pools was screened by colony hybridisation for the identification of individual putative clones. Although plasmid pools containing putative clones were identified repeatedly by Southern blotting and DNA/DNA hybridisation, it was not possible to identify individual putative clones following transformation into *E. coli* MV1184 and colony hybridisation.

PREFACE

The experimental work described in this thesis was carried out in the Department of Biochemistry, University of Natal, Pietermaritzburg, from January 1994 to October 1996, under the supervision of Professor John D Lonsdale-Eccles and Dr Romilla Maharaj.

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

LIST OF CONTENTS

ACKNOWLEDGEMENTS

I would like to extend my sincere thanks to the following people and organisations for their valuable assistance, without which the completion of this thesis would not have been possible:

My supervisors and co-supervisors: Dr Romilla Maharaj, for her expert and sound guidance; Professor John D Lonsdale-Eccles, for his valuable suggestions and motivation during the course of this study, even after his relocation to the USA; and Professor Clive Dennison for his interest and willingness to supervise the compilation of the final document.

The University of Natal, for showing their faith in me and providing the financial opportunity to pursue these studies by appointing me as a Junior Lecturer. Dr Trevor Anderson, for taking the time to convince my family about the importance of this work. Mr Ron Berry, for his patience and willingness to help with the computers, especially saving files I thought were lost. Mr Omalokoho Tosomba, who taught me how to maximise parasite yield from initial procedures.

The Department of Biochemistry, for creating the ideal atmosphere to conduct research with everyone willing to help. The Department of Zoology and Entomology, for allowing me to use their thermocycler. The Department of Genetics, for giving me access to all of their equipment that I needed for my research.

The student "Interns": Senzo Makhathini, Sizwe Bhengu and Philisiwe Mazibuko for working with me in the laboratory, during vacations, as part of their training in the Internship Programme organised by the University.

My flatmates at King Edward Mansions from 1995 to 1997, for their encouraging moral support.

Finally, I have been blessed with an exceptional family: I thank my mother, Nomthandazo Dalasile for her unwavering moral support, encouragement, and family time sacrificed during this study and my brothers: Madoda for his encouraging support, Ludwala for doing the "cooking" while I was chasing after the results.

ABBREVIATIONS

AIDS	Acquired immunodeficiency syndrome
Ap	Ampicillin
Ap^r	Ampicillin resistant
BC	Basic copy
BHC	Benzene hexachloride
bp	Base pairs
BSA	Bovine serum albumin
Bz	Benzoyl
CIP	Calf intestine phosphatase
CP	Cysteine protease
DDT	Dichlorodiphenyl-trichloro-ethane
DEAE	Diethylaminoethyl
DFP	Diisopropylfluorophosphate
DIFP	Diisopropylfluorophosphate
DIG	Digoxigenin
dNTPs	Deoxynucleotide triphophates
DTT	Dithiothreitol
E-64	L-trans-epoxysuccinyl-leucylamido-(4-guanidino)-butane
EDTA	Ethyldiaminetetraacetic acid
ELC	Expression-linked copy
ESAG	Expression site associated gene
GPI	Glycosylphosphatidylinositol
GRESAG	Gene related to expression site associated gene
HDL	High density lipoprotein
HIV	Human immunodeficiency virus
IPTG	Isopropylthio-β-D-galactoside
LA	Luria agar
LB	Luria broth
MAP	Microtubule-associated repetitive protein
NMec	4-methyl-7-coumarylamide
NNap	2-naphthylamide
NBT	Nitroblue tetrazolium salt

PAG	Procyclin-associated genes
PARP	Procyclic acidic repetitive protein
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PMSF	Phenylmethylsulfonyl fluoride
PSG	Phosphate buffered saline glucose
SDS	Sodium dodecyl sulphate
SE	Sodium chloride EDTA
SET	Sodium chloride EDTA Tris-hydrochloric acid buffer
SSC	Sodium chloride sodium citrate
TAE	Tris-acetate EDTA buffer
TE	Tris EDTA
TF	Transferrin
VAT	Variable antigen type
VSG	Variable specific surface glycoprotein
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactoside
X-phosphatase	5-bromo-4-chloro-3-indolyl phosphate toluidium salt
YAC	Yeast artificial chromosomes
Z	Benzyloxycarbonyl

CHAPTER 1

GENERAL INTRODUCTION

Trypanosomatids are protozoan flagellates which appeared early in the evolution of the eukaryotes (Pays, 1993). Trypanosomes that infect mammals in Africa are responsible for major diseases, namely sleeping sickness in humans and trypanosomiasis or nagana in domestic livestock and wildlife. These parasites are transmitted to mammals following a bite by infected tsetse flies (*Glossina* species), which are the vectors. Trypanosomes that infect livestock and humans, also infect wildlife species, which serve as a reservoir of infection for flies which can then infect domestic animals and humans. Many wildlife species carry trypanosomes with no ill effects (ILRAD, 1990).

1.1 Trypanosome species and their infections

Trypanosome species that are of veterinary importance include *Trypanosoma* (*Duttonella*) *vivax, T.* (*Nannomonas*) *congolense,* and *T.* (*Trypanozoon*) *brucei* (subspecies *brucei*) (Table 1; Molyneux and Ashford, 1983; Stephen, 1986). Although both *T. vivax* and *T congolense* parasites cause nagana in cattle, they are morphologically and antigenically distinct from each other and from the *T. b. brucei* subspecies which also causes nagana in cattle (Hoare, 1972). The trypanosomes that cause sleeping sickness in humans are subspecies of *T. brucei*, namely *T. brucei gambiense* the causative agent of "Gambian" sleeping sickness and *T. brucei rhodesiense* the causative agent of "Rhodesian" sleeping sickness and chronic infection of animals (Hoare, 1970; Boothroyd, 1985). There are two other trypanosome species that are pathogenic to mammals, namely *T. evansi* the causative agent of surra in pigs and *T. equiperdum* the causative agent of dourine in horses. The latter two trypanosome species are not insect-borne but are closely related to other *brucei* subspecies. *T. evansi* is transmitted by tabanid flies through mechanical inoculation and *T. equiperdum* is transmitted by sexual contact between equines (Molyneux and Ashford, 1983). The distribution of animal trypanosomiasis due to infection by different species of trypanosomes is shown in Figure 1A and 1B.

1.2 Distribution of trypanosomiasis

The occurrence of trypanosomiasis in livestock, and sleeping sickness in humans, normally corresponds with that of the insect vectors which at present occur in about 37 countries in

Africa spanning an area of over 10 million square kilometres. There are three dispersal centres of *Glossina* in Africa. Firstly, the *fusca* group which established itself and spread out from a centre around the Gulf of Guinea which experiences rainy forest climatic conditions. Secondly, that of the *palpalis* group which developed in the same centre as the *fusca* group but moved westwards around the south of the Sahara, and thirdly, that of the *morsitans* group which arose in Eastern Africa (Ford, 1970). Isolated flybelts have also occurred due to the adaptation of parasites to new hosts and that of the insect vector to new climatic conditions.

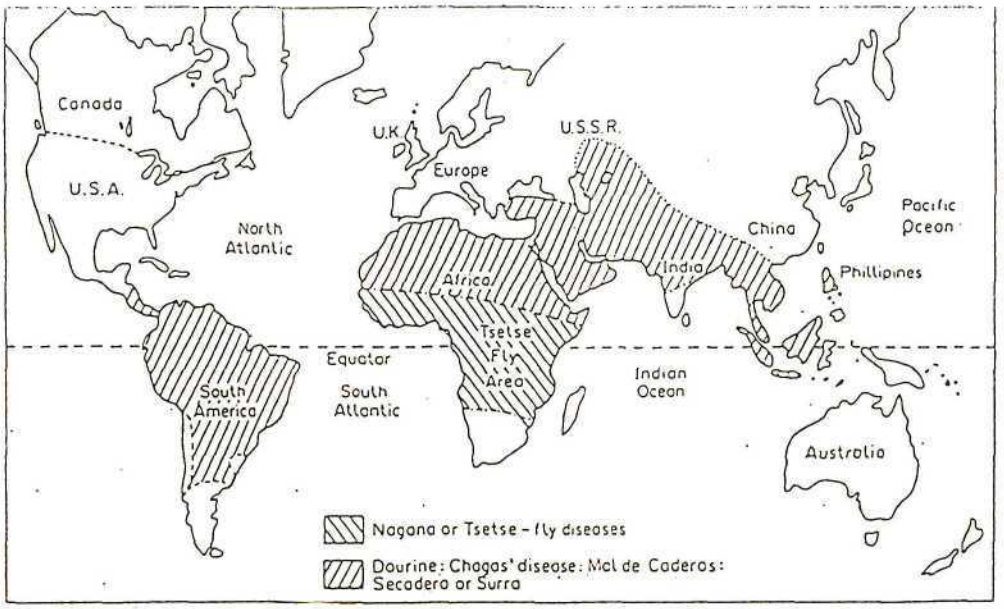
Table 1 Host, vector, disease, and mode of transmission of important salivarian trypanosomes (Molyneux and Ashford, 1983; Stephen, 1986).

Species	Host	Vector	Disease	Transmission
T. vivax	Ruminants and equines	Tsetse and tabanid flies	Animal trypanosomiasis	Cyclical and mechanical
T. congolense	Domestic and game animals	Tsetse fly	Animal trypanosomiasis	Cyclical
T. b. brucei	Domestic and game animals	Tsetse fly	Animal trypanosomiasis	Cyclical
T. b. gambiense	Man and game animals	Tsetse fly	Sleeping sickness	Cyclical
T. b. rhodesiense	Man and game animals	Tsetse fly	Sleeping sickness	Cyclical
T. evansi	Domestic and game animals	Bats, stomoxys and tabanid flies	Surra	Mechanical
T. equiperdum	Equines	None	Dourine	Coitus

Some of the insect flybelts have been reduced to a low level due to control methods that have been implemented, such as in Zululand in South Africa, while others have advanced to areas that were formerly free from infection (Pott, 1970). The insect vector *Glossina,* which inhabits over one third of the African continent, exposes about 30% of the estimated

160 million cattle, comparable numbers of small ruminants, and 50 million people to the risk of infection by trypanosomes (Figure 1; ILRAD, 1993/1994).

A



B

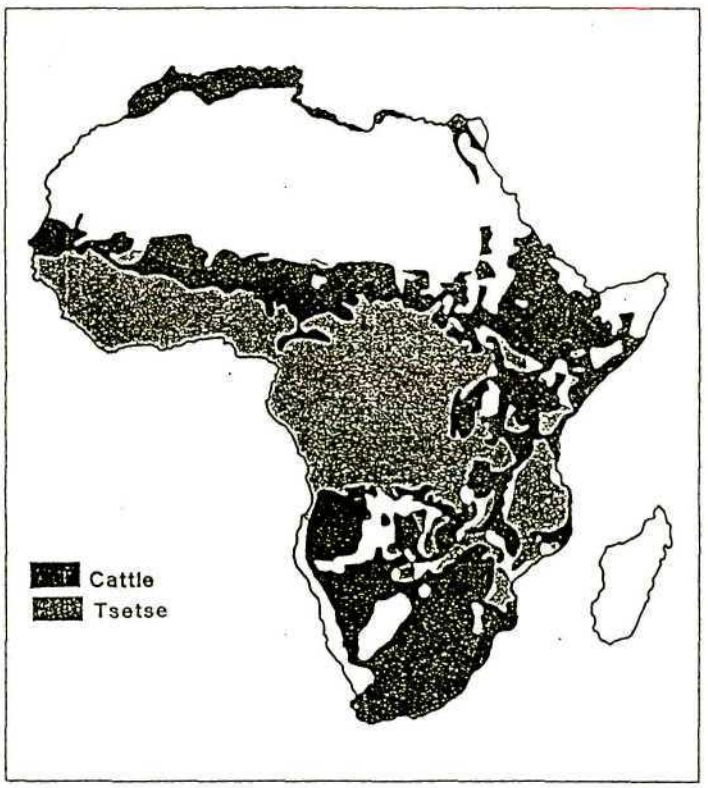


Figure 1 Geographical distribution of trypanosomiasis in the world (A) (Stephen 1986) and that of tsetse fly and cattle in Africa (B) (Molyneux and Ashford, 1983).

Bovine trypanosomiasis poses a major constraint on agricultural and economic development and the consequent improvement in human lives in many countries in tropical Africa.

Economical losses are estimated at 5 billion US dollars per year due to the loss in meat production, milk yield, tractive power, potential in livestock and crop production and to the cost of control programs against the disease (ILRAD, 1993/1994).

1.3 Control of trypanosomiasis

Methods of tsetse control can be grouped into two types, namely direct and indirect methods. Direct control methods include violent destruction of individual tsetse by catching and squashing, trapping or poisoning by insecticides and reducing a population by augmenting its burden of parasites (Hoare, 1970). Indirect methods often involve the use of destructive techniques against the tsetse food supply or their habitats. Control of tsetse by means of trapping has not been universally successful, but in Zululand, South Africa, this method was successful in eliminating most of the flies (Pott, 1970).

Three methods are currently used for controlling trypanosomiasis. These are the administration of drugs to treat or prevent the disease, the use of methods to reduce the population of flies that transmit the disease and the maintenance of a stock of indigenous livestock breeds that are resistant to the disease (Pott, 1970). In recent years, the most significant advance in the control of the tsetse population has been the use of insecticides which is now an established practice for the control of all the important species of tsetse. Some *Glossina* species are very susceptible to specific insecticides (Burnett, 1970) thus making this a preferred treatment. The first insecticides to be used against the tsetse fly, in practice and experimentally, were dichlorodiphenyl-trichloro-ethane (DDT) and benzene hexachloride (BHC). Pyrethrum, dieldrin, endosulfan, isobenzan, fenthion (organophosphate), aprocarb (carbamate), and bromocyclan were later used (Burnett, 1970). To date, each of these controlling methods, while remaining useful, have some drawbacks (ILRAD, 1993/1994).

In recent years, few new compounds have been developed for treating trypanosomiasis and those drugs that are available may cause unpleasant and fatal side effects. Furthermore, most trypanocidal compounds currently available have been widely used for many years, and have resulted in the development of drug resistant parasites (ILRAD, 1990). Consequently, it is hoped that a better understanding of the biochemistry of these parasites will help in the identification of unique and essential metabolic processes that may form the basis of new generations of drugs for the treatment of trypanosomiasis. Research towards

understanding of the biochemistry and molecular biology of these parasites was undertaken in search of the potential antiparasitic drugs that can be used in chemotherapy and chemoprophylaxis of this disease. Attention has focused on a variety of specific parasite molecules that play key roles in the parasite life cycle or in the pathogenesis of the disease caused (McKerrow *et al.*, 1993).

One family of molecules that has attracted attention in recent years is the proteases produced by the parasite (McKerrow *et al.*, 1993a). Proteases have been shown to be key factors in the pathogenicity of many parasitic diseases, either by inducing tissue damage and facilitating invasion, or by enabling the parasites to salvage metabolites from host proteins (McKerrow, 1989). These enzymes are essential for the protozoa to break down host proteins to survive, reproduce, invade host tissues (Roose and Van Noorden, 1995), evade the host immune response and prevent blood coagulation (McKerrow, 1989).

Parasitic protozoa such as *T. brucei* and *T. congolense* are capable of undergoing complex morphological and biochemical changes during the various stages of their life cycle. These changes correlate with alterations in the levels of trypanosome lysosomal cysteine proteases (CP) (Lonsdale-Eccles and Mpimbaza, 1986; Mbawa *et al.*, 1991). Such alterations at different life cycle stages suggest a developmental regulation of the expression of the cysteine proteases (Pamer *et al.*, 1989; Fish *et al.*, 1995) and may be associated with developmental differentiation (Eakin *et al.*, 1990). It has been suggested that these alterations could be a result of transcriptional or posttranscriptional gene regulation of the cysteine protease gene(s) (Roose and Noorden, 1995). Cysteine proteases have been found to be highly active in the mammalian infective metacyclic and bloodstream forms of the parasite (Table 2) and northern blot analysis of different life cycle forms of *T. congolense* (Table 2) suggests a role for transcriptional regulation of the cysteine proteases in this parasite. However, the mechanism of this regulation is not yet understood.

Only three promoters (Figure 2), namely those of the genes coding for the variant-specific surface glycoproteins (VSGs), ribosomal proteins, and procyclin or procyclic acidic repetitive protein (PARP) have been characterised in trypanosomatids (Clayton *et al.*, 1990; Gottesdiener *et al.*, 1991). The transcription units in these parasites are arranged in a polycistronic manner. This is in contrast to other eukaryotes in which genes are monocistronic with each gene having its own promoter (Pays *et al.*, 1994). The expression

of a number of genes from a single promoter in trypanosomes suggests that trypanosomes do not control expression of genes in the same way as other eukaryotic organisms. This dissertation describes attempts made to clone the *T. brucei* and *T. congolense,* DNA fragments containing the promoter region, as the first step in the elucidation of the control elements of the cysteine protease gene.

Table 2 Comparison of the CP activity and the mRNA levels within and among the life cycle stages of *T. congolense* (Fish *et al.*, 1996) and *T. brucei* (Mbawa *et al.*, 1991; Pamer *et al.*, 1989)

Life Cycle Stage	***T. brucei***		***T. congolense***	
	CP activity	mRNA	CP activity	mRNA
Procyclic	±/-	nd		
Epimastigote	±/-	nd	±/-	±/-
Metacyclic	+	nd	±/-	±/-
Bloodstream (slender)	++	nd	+	++
(stumpy)	+++			

nd = not determined ;
- = not present
± = very low levels
+ = low levels
++ = high levels;
+++ = very high levels.

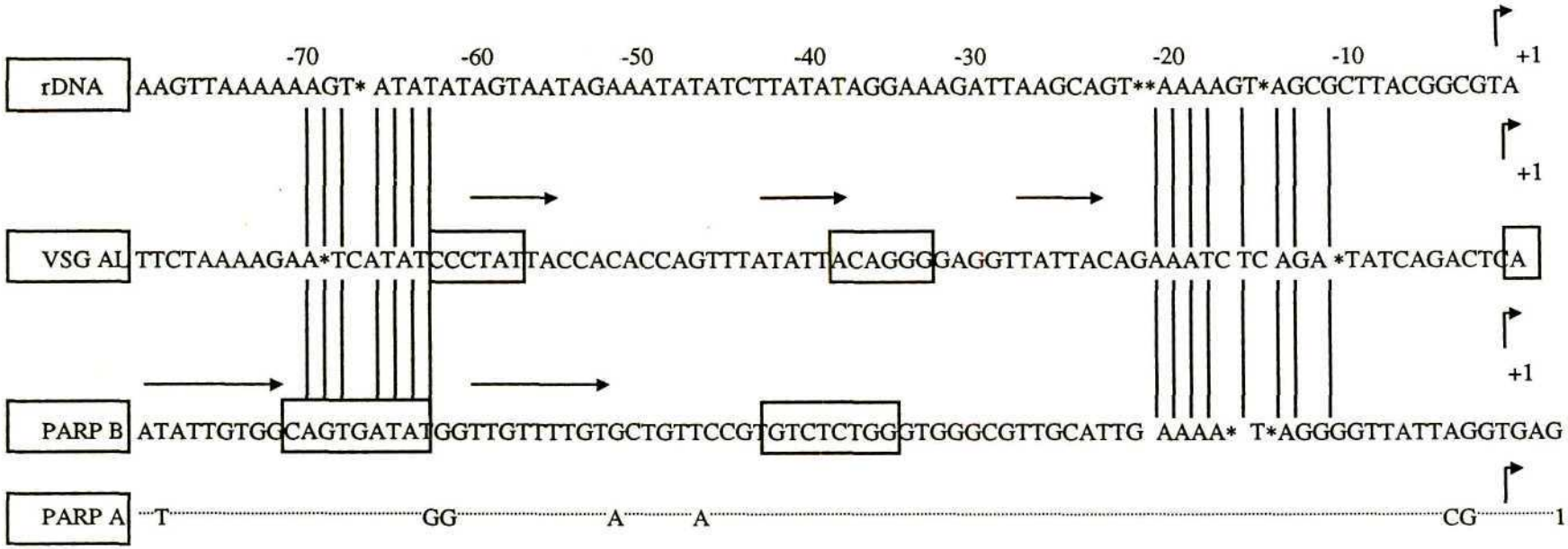


Figure 2 Comparison of the *T. brucei* rDNA, VSG and PARP promoters aligned for maximum homology. Arrows show direct repeats, boxes show essential regions as defined by mutagenesis and spaces between nucleotides are for alignment (Pays et al., 1994).

CHAPTER 2

TRYPANOSOMES

2.1 Classification

Mammalian trypanosomes have been classified in several ways. The systematic position of African trypanosomes among the protozoa is shown in Figure 3. Trypanosomes are currently grouped into two major sections, namely stercoraria and salivaria, based on their course of development in insect vectors. The section stercoraria comprises species that complete their development in the hind gut of the insect. Transmission of these parasites is mediated via parasite-containing faeces of the insects and the subsequent entry into the mammalian host system through exposed mucous membranes or injured skin (Molyneux and Ashford, 1983). Trypanosomes that complete their development in the anterior end of the insect vector, namely in the proboscis or salivary glands, fall under the section salivaria (Hoare, 1970). Transmission of this class of parasites is by inoculation with saliva when the insect feeds on the mammalian host blood. Trypanosomes that cause African sleeping sickness in humans and nagana in cattle belong to this section. Within the two major sections, trypanosomes are further subdivided into subgeneric groups and species based on the phylogenetic order as shown in Figure 3.

Classification of African trypanosomes is complicated as different species and genera are not always clearly distinguishable from each other (e.g. species from the subgenus trypanozoon). In some of these trypanosomes, the morphological characters separating the species are virtually indistinguishable and their cycles of development in *Glossina* appear to be similar. Their general biochemical and immunological properties overlap to a great extent and their response to chemotherapeutic agents also appears to be similar (Stephen, 1986). In others, only differential biological characters are present, namely host range, host parasite relations, geographical distribution and certain ecological features. All three trypanozoon species can infect a wide range of laboratory, domestic and wild animals although man has been shown to be refractory to infection by *T. b. brucei* (Hoare, 1970). *T. b. brucei* is noninfectious to humans as a result of its sensitivity to the cytolytic activity found in normal human serum, identified as due to a subclass of high-density lipoprotein (HDL). This sensitivity to HDL has been shown to be developmentally regulated and

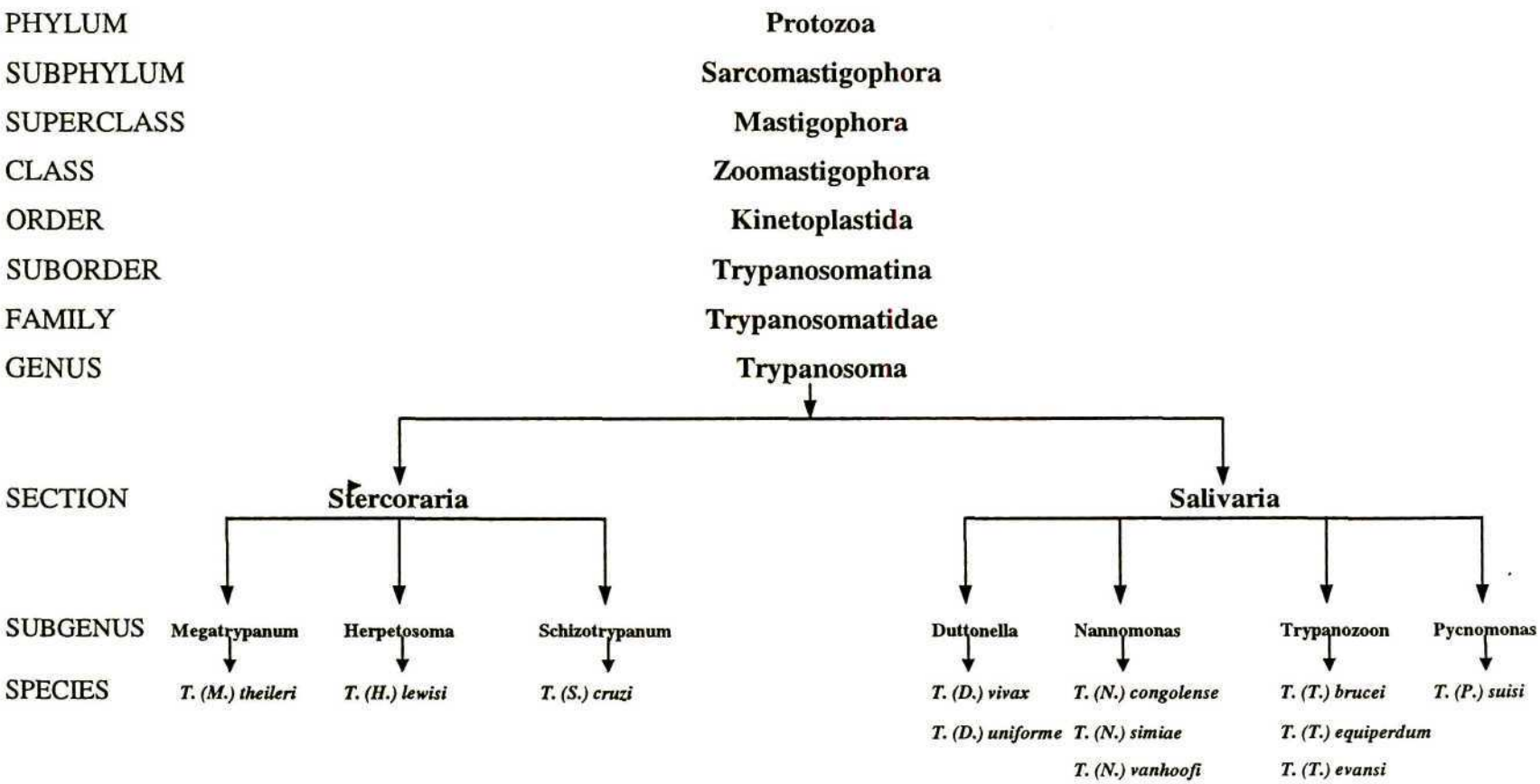


Figure 3 Schematic classification of trypanosomes within the phylum protozoa (Hoare, 1970; Stephen, 1986).

correlates with the life cycle specific changes in macromolecular uptake by the organism (Hajduk *et al.*, 1994; Smith *et al.*, 1995).

2.2 Anatomy of the trypanosome

The general cellular structure of the trypanosome cell is shown in Figure 4. Trypanosome cytoplasm is surrounded by a unit membrane that is composed of dense outer osmiophilic layers separated by a clear layer and is covered entirely by a coat of protein called the VSG, the site of the variable antigen in *T. brucei* (Molyneux and Ashford, 1983). The VSG is lost when the parasites are ingested by the tsetse fly (Vickerman, 1970) or grown under specific culture conditions. This process apparently involves proteolysis (Ziegelbauer *et al.*, 1993). The surface coat is reacquired as the epimastigotes transform into metacyclic trypanosomes (Hoare, 1970).

Pellicular microtubules are found beneath the entire cell membrane, except the flagella pocket, which is the point of attachment of the flagellar membrane to the body of the parasite. Microscopic examination of microtubules indicate that they extend from the pellicle to the posterior end of the trypanosome cell (ILRAD, 1990). The pellicular microtubules of trypanosomes have been suggested to serve an endoskeletal function (Kreier and Baker, 1987).

The undulating membrane is formed by the flagellum which runs along the pellicle and is attached to the trypanosome through the flagellar pocket at the origin of the flagellum. The flagellum begins as a structure called the basal body or kinetosome. The flagellar pocket, which represents 1 to 3% of the total cell surface is a specialised region which is involved in macromolecular uptake (Webster and Russell, 1993). The flagellar pocket differs from the rest of the cell membrane in that it is not associated with the subpellicular microtubules, but is covered by a thick layer of VSG molecules. Receptors for molecules such as transferrin have been identified in the flagellar pocket. When the flagellum reaches the anterior end of the body, it either becomes free, as seen in *T. brucei,* or terminates without a free portion as in the case of *T. congolense.* In the bloodstream form of trypanosomes found in the vertebrate host, the flagellum is responsible for movement while in the invertebrate it plays a dual role, that is one of movement and attachment.

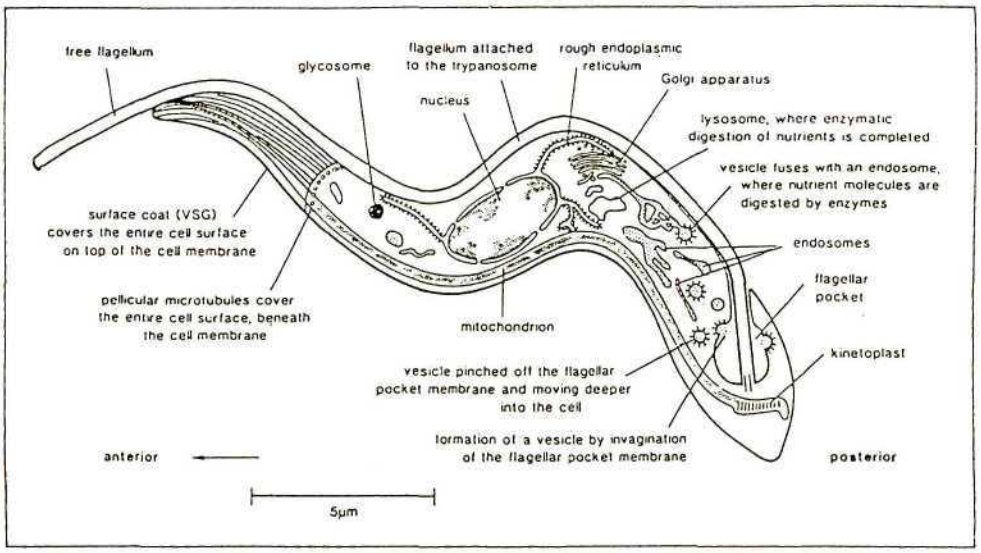


Figure 4 Schematic diagram showing the anatomy of *T. brucei* in its intermediate bloodstream form (ILRAD, 1990).

The basal body or kinetosome is found below the floor of the flagellar pocket at the posterior-end of the trypanosome. The kinetoplast is situated within close proximity of the basal body and is one of the most characteristic organelles of trypanosomes. It forms part of a single mitochondrion that extends from one end of the trypanosome to the other and contains deoxyribonucleic acid (kDNA). The mitochondrion differs in form from species to species and between different stages in the life cycle of the parasite. It is more developed in the tsetse fly stages and less developed in the blood stream forms, where energy is readily available in the form of glucose in the blood of the host (Vickerman, 1985).

The nucls, which is spherical or slightly ovoid, may be situated in various parts of the body of the parasite. In the bloodstream form it is normally situated in the centre or in the anterior half of the cell. As in all other eukaryotic cells the nucleus is bound by a double membrane punctured by pores. The outer membrane is continuous with the endoplasmic reticulum (ER) which is associated with the Golgi apparatus, subpellicular microtubules and the mitochondrial network. Chromatin is attached to the nuclear envelope and the nucleolus is centrally situated (Stephen, 1986). Other cell organelles present in the trypanosome include ribosomes, glycosomes, vesicles, endosomes, and lysosomes (Kreier and Baker, 1987; Webster, 1989; Webster and Fish, 1989; Somer and Wang, 1994). Glycosomes in trypanosomes mainly compartmentalise most glycolytic enzymes normally located in the cytosol in other eukaryotic cells. The Golgi apparatus is placed between the nucleus and the flagellar pocket and it produces surface coat proteins of trypanosomes and

enzymes which are stored in the lysosomes. The ER may be rough (rER), containing attached ribosomes, or smooth (sER) without the associated ribosomes. This organelle has been established to play a similar role to that in other eukaryotes, transporting proteins to the Golgi apparatus where additional molecules are added and packaged.

2.3 Life cycle

The life cycle of most trypanosomes is digenetic in that it involves alterations between two hosts, an invertebrate as the intermediate host (or vector) and a vertebrate. In cycling between the vector and host, the parasite has to cope with many alterations in its environment, including marked changes in temperature and nutrients (Koening-Martin *et al.*, 1992). Therefore, trypanosomes must adapt to extremely different environmental conditions. This adaptability could be due to the unique control of gene expression as demonstrated by the expression of different surface proteins when the parasites are in the vector midgut or in the vertebrate host (Hoare, 1972). In the vertebrate they develop in the blood and lymphatic system although some African trypanosomes are capable of invading the connective tissue, the brain and cerebrospinal fluid (Hoare, 1972). The parasite undergoes a number of transformations into different forms in both the tsetse fly vector and the mammalian host (Figure 5).

The tsetse fly ingests the bloodstream form of trypanosomes when it feeds on blood from an infected animal. In the case of *T. b. brucei*, these parasites may be either long slender forms that are rapidly dividing, or they may be nondividing, short stumpy forms. Once in the midgut of the fly, the short stumpy forms differentiate via several distinct forms (dividing promastigote and epimastigote forms) into the nondividing metacyclic forms, present in the salivary glands (ILRAD, 1993/1994). These metacyclic trypanosomes are able to reinfect a mammalian host when the infected tsetse fly next feeds on a mammalian host and the parasites are injected into the skin along with tsetse saliva. Once in the mammalian host, the metacyclic trypanosomes differentiate into bloodstream forms which are specifically adapted to live in blood. The bloodstream form of trypanosomes multiply rapidly by binary fission and enter the lymphatic system and blood circulation of the host. As tsetse flies feed on animals infected with the parasites, blood containing trypanosomes is taken up thus completing the life cycle.

In trypanosomes classified in section salivaria, cell division, which occurs in some parasite life cycle stages in both the insect and mammalian hosts, is by binary fission. This is initiated by bipartition of the kinetoplast, followed by the development of the new flagellum near the posterior kinetoplast. The new flagellum gradually grows in length and runs almost in parallel with the existing flagellum while the nucleus divides into two. When the two flagella are of the same length, the cytoplasm undergoes fission proceeding from the anterior to the posterior end of the body and the daughter cells separate.

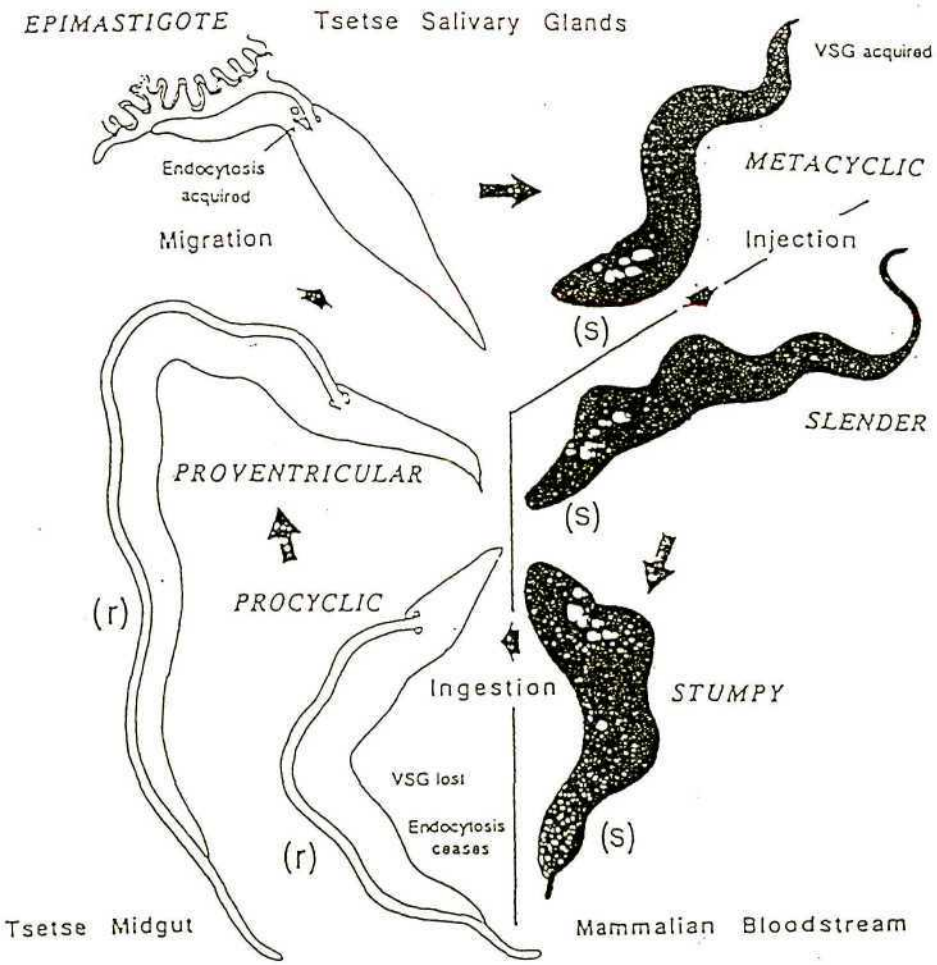


Figure 5 The life cycle stages of *T. brucei*: Developmental stages in the mammalian blood and the midgut and salivary glands of the insect. Black shading: VSG coat; Unshaded spots: endocytic vesicles; (s): sensitivity to human serum-mediated lysis; (r): resistant stages (Hajduk *et al.*, 1994).

Electron microscopy studies of the successive stages in the life cycle of trypanosomes show complex changes in the architecture of the parasite. The most striking changes take place in

the mitochondrial system and on the surface of the parasite. The former is as a result of the changes in the energy production system and the latter to assist evasion of the non-specific and specific defence mechanism of the mammalian host (Vickerman, 1985). In the bloodstream form trypanosomes, the mitochondrial functions are severely reduced with several metabolic pathways such as the tricarboxylic acid cycle and the cytochrome-dependent electron transport chain being completely suppressed. Thus, the parasite depends entirely on glycolysis and substrate level phosphorylation for its energy. In the glucose deficient environment of the insect midgut, generation of adenosine triphosphate (ATP) becomes the main source of energy. Synthesis of ATP is by oxidative phosphorylation in the well developed mitochondrion, in which the enzymes of the citric acid cycle and cytochrome systems are restored (Fairlamb, 1996).

2.4 Pathogenicity of trypanosomiasis

2.4.1 Trypanosomiasis in domestic animals

African trypanosomes infect all species of domestic animals. *T. congolense* is economically one of the most important parasites causing trypanosomiasis in African livestock. Symptoms of trypanosomiasis in cattle vary from area to area. This could be due to the different responses of genetically varying types of African cattle, different specific trypanosome strains, the nutritional status of the pastures or climatic variation. In addition, the relationship between the parasite and host is not static.

One or two weeks after infection, animals that are susceptible to the disease begin to show intermittent fever and anaemia. Anaemia is the most characteristic symptom of this disease and is due not only to the destruction of red blood cells in circulation, but also to inhibition of their synthesis in the haemopoietic tissues. There are two phases of anaemia, an active haemolytic anaemia associated with rising parasitaemia, and a chronic phase associated with declining levels of parasitaemia (Pays, 1985). Other important symptoms include oedema, and cachexia which leads to the emaciation of the animals. Under the stressful conditions which result from repeated bites by the tsetse fly carrying different types of trypanosomes and daily walks for forage and water, infected animals deteriorate over several months before dying (ILRAD, 1990). Different types of domestic animals show different degrees of tolerance to the disease, and the pathogenic rersponse of the different species and

subspecies of trypanosomes varies. Chronically infected animals normally show immunosuppression associated with depletion of lymphocytes (ILRAD, 1990).

2.4.2 Trypanosomiasis in humans

Different trypanosomes may infect a single host. There are two different types of human trypanosomiasis, the Gambian (West Africa) and Rhodesian (East Africa) sleeping sickness caused by *T. b. gambiense* and *T. b. rhodesiense,* respectively. The division into the two types is based on the clinical manifestations of the disease resulting from the strains of *T. brucei* that induce the disease (Hoare, 1972).

The Gambian sleeping sickness is caused by *T. b. gambiense* and transmitted by tsetse flies of the *palpalis* group. The incubation period after infection is from five to twenty days and is occasionally prolonged to months or years. The infected area, at the site of the tsetse fly bite, develops a local inflammatory reaction that produces a hard red nodule surrounded by a paler periphery forming a chancre. The chancre then disappears and the patient develops fever. Trypanosomes such as *T. vivax, T. congolense* and *T. brucei* have also been detected within the chancres of animals before they appear in the bloodstream (Gordon and Willet, 1958). Other symptoms of this disease are an increased heart beat, temporary rash, severe headaches, swollen and tender glands, loss of weight and itching. In addition, anaemia and oedema of the face may occur. Parasites may disappear from the glands which may shrink to normal size. Eventually, sleeping sickness conditions develop and the patient dies.

Rhodesian sleeping sickness is a disease that results from infection by *T. b. rhodesiense* which is transmitted by *G. morsitans* or *G. swynnertoni.* The Rhodesian infection develops in a similar way to the Gambian sleeping sickness, but is more acute. The incubation period of the disease is shorter and may begin with more pronounced symptoms. The central nervous system is affected soon after infection and death may occur well before symptoms of sleeping sickness conditions develop (Hoare, 1972).

2.5 The immunological responses to trypanosome infection

During the course of an infection trypanosomes multiply and as the parasitaemia increases, the patient suffers from fever followed by a decline in the number of trypanosomes (Hoare, 1972). This decline in parasitaemia is associated with the occurrence of antibodies in the blood that act against antigenic surface protein molecules. By binding to the parasites, the

antibodies initiate a series of events that eventually lead to the lysis and death of parasites. Parasites that have a different surface antigen are not killed and multiply to create a relapse population, against which the antibodies that overcame the initial population are ineffective. Antibodies which are later produced against the second population of parasites are then effective in bringing about a decline in the second population. Those parasites which escape the host immune response recover into a third population of parasites or antigenic form. These relapses recur until the host recovers completely, remains in a state of adjustment with the parasite, or dies. The most significant symptom in the development of the disease is the increase in B lymphocytes that could result in lymph node enlargement (Hoare, 1972).

As long ago as the beginning of the century it was shown that a single trypanosome could, in the course of its multiplication, give rise to at least twenty two different antigenic forms (Ritz, 1916). It was shown in *T. brucei* that its original antigenic properties are favoured when, at any stage, the trypanosomes undergo cyclical development in the tsetse fly regardless of the many different antigenic variants (Weitz, 1970). In addition, different variants of a strain have at least one antigen, known as the basic antigen which is common to all variants (Ford, 1971). However, Pays *et. al.* (1985) concluded that predominance is tightly linked to a particular variable antigen type (VAT) since a particular VAT may be predominant in one stock and occur late in another.

Defence mechanisms against mammalian trypanosomes include both cellular and humoral factors, with phagocytes playing a cellular role. Humans and wild and domestic animals produce antibodies in response to trypanosome infection, but to varying degrees. The types of antibodies produced vary considerably in different host species and at different times during infection, depending on the host's exposure to previous challenges or infections. A strong resistance to infection, referred to as an innate immunity may occur in some animals. In earlier studies, there was no clear evidence of acquired immunity to the pathogenic salivarian trypanosomes (Weitz, 1970). However, N'dama cattle, which are found in West Africa, are known to possess a degree of natural immunity or tolerance and develop low and transient parasitaemia and eventually recover from infection without manifesting any clinical symptoms. This breed of cattle, when born and bred in tsetse free localities are liable to contract a chronic disease and show an immune response comparable to that of susceptible cattle, which suggests that tolerance in N'dama cattle is due mainly to acquired immunity. Murray *et al.* (1981) showed that where tsetse challenge is high, trypanotolerant

breeds such as N'dama are as likely to be infected as trypanosusceptible breeds. In contrast, Zebu cattle are very susceptible to infection, producing very high levels of parasitaemia and highly marked clinical symptoms from which animals may die within three months.

The main immunoglobulins associated with trypanosomiasis are IgM and IgG. These two immunoglobulins may contain antibodies of the same specificity at the same time. In man, the increase in IgM protein and its presence in the cerebrospinal fluid of the infected patient is also indicative of an antibody response. The persistence of the antibodies in the blood after recovery, is variable, depending on the persistence of trypanosome antigens in the tissues of the host. The variable expression of antigens by trypanosomes during the infection promotes the induction of antibodies against each variant (Weitz, 1970).

2.6 Antigenic variation

Each trypanosome strain has an innate tendency to develop one particular antigenic form which is relatively stable, and is known as the "parent" antigenic type. Trypanosomes of one strain transmitted by different tsetse flies have been shown to possess antigens in common, and this antigen is known as the "basic" antigen. The basic type antigen will always produce the same sequence of antigenic types in primary and subsequent infections (Stephen, 1986). Certain other antigens which are different from the "basic" antigen in each strain of trypanosome tend to develop during the first few weeks after infection. Antigenic variation partially explains the production of different variants during the course of infection and the considerable differences between the variant specific antigens of different strains of the same species.

The metacyclic and bloodstream forms of African mammal-infecting trypanosomes are covered by VSG molecules. These molecules are tightly packed on the surface of the trypanosome to form an electron dense coat of about 12 to 15 nm thickness (Molyneux and Ashford, 1983) and are anchored into the surface membrane of the parasite by their C-terminal ends through the glycosylphosphatidylinositol structure (Cross, 1990; Dell, 1996). Carbohydrate and phospholipid are bonded to the polypeptide at the mature C-terminus of the protein. All *T. brucei* VSG molecules identified to date have molecular weights ranging between 60 and 65 kDa and are made up of a polypeptide chain of 450 to 480 residues, two or more carbohydrate side chains and phospholipid. A variable degree of amino acid sequence homology has been shown to exist between the different VSGs

(Boothryod, 1985) and this homology is higher in the C-terminal end of the protein. It has been established that, during the synthesis of VSG molecules, an amino terminal extension is synthesised on the nascent polypeptide as could be expected for a membrane protein that is exported across the endoplasmic reticulum (Boothroyd, 1985). A diagrammatic representation of VSG structure and synthesis is shown in Figure 6.

The biosynthesis of VSG molecules is not well understood. Synthesis begins with ribosomal translation, of a VSG mRNA, and the synthesis of the signal peptide. The signal peptide then attaches the complex to the membrane of the rER. With translation continuing, the nascent peptide is transported into the lumen of the rER. The leader sequence is removed and translation stops with the synthesis of the hydrophobic tail which then anchors the VSG to the membrane. Glycosylation occurs at the internal sites in the sER, followed by the replacement of the polypeptide tail with a glycolipid attached to the new C-terminus in the Golgi apparatus. The VSG molecule is represented as containing dimers, although this is only known to occur in VSGs in solution. The soluble form of VSG occurs as a result of phospholipase C activity, which leaves the lipid, presumably, still in the membrane and the phospho-sugar attached to the VSG. Current data suggest that the translation of mRNAs that code for VSGs is similar to that of eukaryotic surface proteins and occurs in membrane bound polysomes (Boothryod, 1985). However, the removal of the N-terminal extension (30-40 amino acids) occurs soon after translation is completed, but before the VSG molecule reaches the cell surface. The addition of the internal carbohydrate moieties is thought to occur in the endoplasmic reticulum and that of the phospholipid moiety in the Golgi apparatus (Grab *et al.*, 1984).

One, or more, of the carbohydrate chains is internally located within the VSG molecule and is of the high mannose type with a typical N-glycosidic linkage to the rest of the polypeptide via an asparagine residue. The function of this carbohydrate moiety is not known, nor is that of the other class of carbohydrate which is bonded through the α-carboxyl of the C-terminal residue, be it asparagine or serine. The C-terminal end carbohydrate moiety is a common antigenic determinant in all the released forms of VSGs in *T. brucei.* and in *T. congolense.* The phospholipid has been shown to be linked via a sn-3-phosphodiester bond to one of the sugars in the C-terminal carbohydrate group (Jackson *et al.*, 1985). The VSGs and the corresponding mRNAs coding for the proteins are extremely abundant during

the metacyclic and bloodstream form stages of the life cycle (Pays *et al.*, 1994), but are absent in the procyclic form when the trypanosomes are covered by a different protein called PARP. This is estimated at 10^7 identical VSG molecules per cell encoded by 5 to 10% of the mRNA.

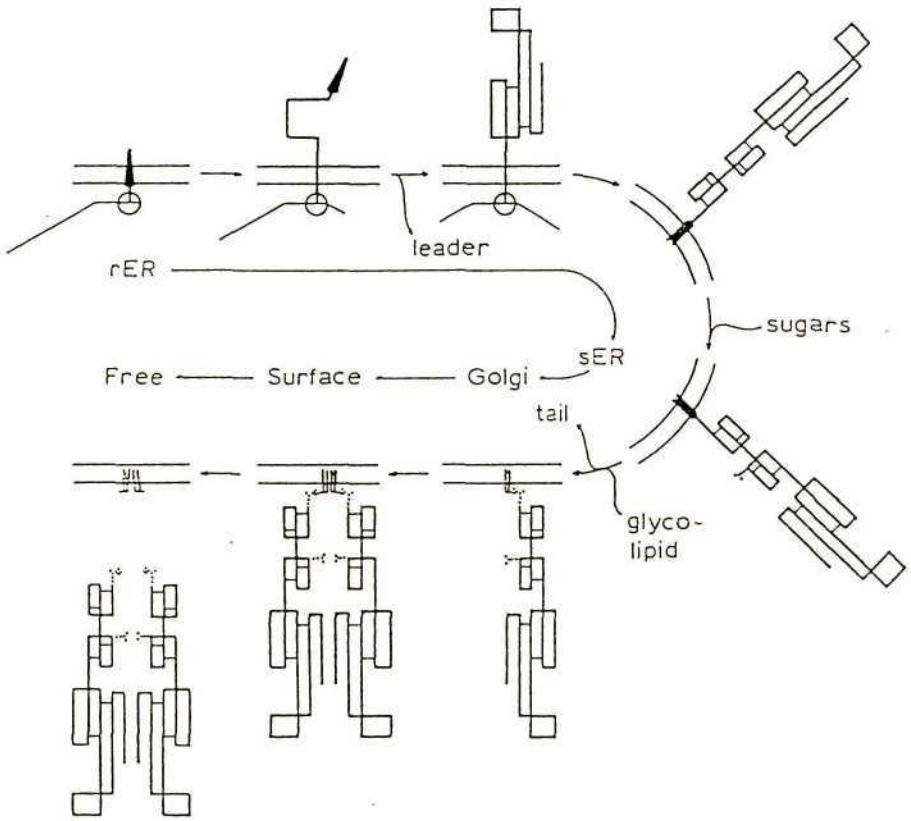


Figure 6 Schematic representation of the VSG structure and synthesis (Boothroyd, 1985). Open circle: ribosome; thin line: mRNA; arrow head: signal peptide; double line: membrane; arrow-tail: hydrophobic tail; dotted side chain: carbohydrate.

The continuous expression of different trypanosome antigens (VSGs) in infected animals limits the progress of research into immunity against trypanosomiasis. Variation of the VSG in the blood allows parasites to escape the immune response of the host and the animal

develops a chronic infection (Pays *et al.*, 1994). Although the host may effect a strong antibody response to the VSG, antigenic variation ensures that only a portion of the parasites are destroyed by antibody mediated killing. A given strain of trypanosomes can produces similar antigens in different hosts, but the antigens often develop in a different sequence. A number of antigens may develop at the same time and the overall antigenic character of a variant strain is often a result of the combinations of the characteristics of several different antigens (Weitz, 1970). A clone derived from *T. brucei* has showed that specific variant antigen types (VATs) can reappear during the course of infection. It has also been shown that VATs found in other tissues of the host such as the brain and lymph nodes may differ from those present in the blood at any one time (Molyneux and Ashford, 1983).

2.7 Genome organisation

The size of the trypanosome genome is between 10^4 and 10^5 kb depending on the species (Gottesdiener *et al.*, 1990). To date, no introns have been found in the trypanosome genome and about 5% of the genes have been identified in *T. b. rhodesiense*. *T. b. rhodesiense* has been suggested to contain 800 genes on a 40 megabase pair (Mbp) genome (Fairlamb, 1996). Trypanosome genomes have been shown to be rich in repeated DNA sequences (Pays, 1993) and a large portion of the genome is devoted to the generation of antigenic variation (Pays, 1988). These parasites are characterised by the presence of a modified mitochondrion, the kinetoplast, hence the name Kinetoplastida. The kinetoplast is localised close to the base of the flagellum and its genome consists of a network of interlocking large and small DNA circles, the maxi and minicircles. There are normally 5 000 to 10 000 minicircles which vary in size from 0.5 to 2.5 kb and 50 maxicircles which range in size from 20 to 50 kb (Myler, 1993). Maxicircles are the functional equivalents of the mitochondrial DNA in other eukaryotes. They encode ribosomal RNA and several other proteins involved in the generation of energy in the mitochondria (Shapiro and Englund, 1995) and most of their transcripts undergo RNA editing, a process that involves addition or deletion of uridine residues at specific sites in the transcript. Minicircles make up over 90% of the kDNA and their main function is to form guide RNA molecules that control RNA editing specificity (Smith, 1996). The kinetoplast DNA is active during the insect phase of the life cycle of the parasite.

Although trypanosome genetic material has not been shown to condense at any stage during the cell cycle, nuclear chromosomes have been characterised by pulse field gradient gel-electrophoresis. The parasites normally have a large and variable number of chromosomes which are of widely differing sizes. *T. brucei* has been shown to contain approximately 100 minichromosomes ranging in size from 50 to 150 kb and at least 18 individual larger chromosomes up to 57 Mbp (Gottesdiener, 1990). Almost all minichromosomes have been suggested to carry VSG genes at their ends (Van der Ploeg *et al.,* 1984). Gottesdiener (1990) concluded that the genome of *T. brucei* contains homologous diploid chromosomes that vary in size although the minichromosomes were shown to be haploid. Minichromosomes are mainly composed of simple DNA sequences, with 90% of the chromosome length consisting of a tandem array of 177 bp repeats, in addition to the telomeric motif GGGTTA and subtelomeric motifs consisting of GG rich direct repeats separated by a bent helical AT-rich region (Pays, 1993). Minichromosomes do not appear to contain protein coding regions.

2.8 The genetic mechanisms of antigenic variation

Trypanosomes are parasites which have developed mechanisms that ensures a quick adaptability to a variety of very different environmental conditions. The genome of these parasites has been shown to contain many duplicated and amplified genes, e.g. adenylate cyclase genes which normally occur in four copies in *T. brucei.* The expression of genes, in these parasites, particularly rDNA, VSG, and PARP appear to be regulated by different levels of transcriptional and posttranscriptional regulation events in a light but reversible manner.

Trypanosomes are different from other eukaryotes in that they do not appear to primarily control the expression of their genes through a specific modulation of promoter activity. Only three trypanosome promoters, namely those of the VSG, procyclin, and ribosomal genes are known to date. These promoters show no significant homology between each other or with any other known eukaryotic promoter (Pays *et al.,* 1994). These promoters contain regions that have been identified as essential by site directed mutagenesis (Figure 2).

Apart from the short time period during switches of VSG expression, there appears to be selective expression of a single VSG gene at any one time from a repertoire of more than a thousand (Pays, 1994; Hajduk, 1994). Thus, a homogeneous VSG coat covering the

parasite is formed. Most VSG genes are present as silent copies in non-telomeric chromosomal regions although some silent copies are found in the telomeres (Beals, 1992). The only VSG genes that are expressed are found within the telomeric expression sites, where the chromatin is assumed to be in an open configuration (Smith, 1996) located at the ends of the chromosomes. Despite this, Muñnoz-Jordan *et al.* (1996) have been able to genetically manipulate *T. b. brucei* so that two VSGs are expressed simultaneously. Their observations provided no evidence for any alteration in infectivity or replication and they conclude that the normally observed singularity of expression of VSG is probably due to the level of expression determined at the site of transcription. Current data suggest that the majority of genes are organised in polycistronic transcription units rather than the monocistronic unit arrangement present in most eukaryotes. These genes are constitutively transcribed from RNA polymerase II promoters (Pays, 1994). In contrast, the VSGs and procyclin are transcribed by an RNA polymerase I-like enzyme that can be regulated at both initiation and elongation. The transcribed RNAs are subsequently processed by 5' transplicing, 3' polyadenylation, and cleavage to produce mature mRNA.

Pays (1988) showed that the expression sites found in telomeres that drive the expression of VSGs in the bloodstream form of trypanosomes are made up of a polycistronic unit (Figure 7). This region, which is located 45 to 60 kb downstream from the promoter, consists of a polycistronic unit that contains at least eight genes known as expression site associated genes (ESAGs). Some of the ESAGs encode surface proteins, such as the transferrin receptor and isoforms of adenylate cyclase. A large region made up of 70 bp repeats known as the "barren" region is devoid of restriction endonuclease cleavage sites and separates the ESAGs and the VSG at the telomere. There is in addition, an untranscribed 50 bp repeat region upstream of the VSG gene (Figure 7). Both repeats participate in the expression of the VSG molecules.

Metacyclic VSGs are composed of a limited subset of the total VSG repertoire, characteristic of a particular strain (Boothryod, 1985; Pays *et al.*, 1994). The original VAT when the parasite is ingested by the fly does not determine the metacyclic VAT subset. When the metacyclic trypanosomes are in the mammalian blood they rapidly switch antigenic types in favour of the expression of the gene that was active when the parasite was first ingested (Hajduk *et al.*, 1981). The metacyclic VSGs are found in telomeres of large chromosomes only and are devoid of the 70 bp repeats that are uniquely present in the

expression sites of bloodstream VSG. In addition, the metacyclic VSG genes do not contain all the ESAGs found in VSG genes of bloodstream form parasites. The procyclin genes are organised in tandems of two or three copies present in two nontelomeric and diploid loci (Pays *et al.*, 1994).

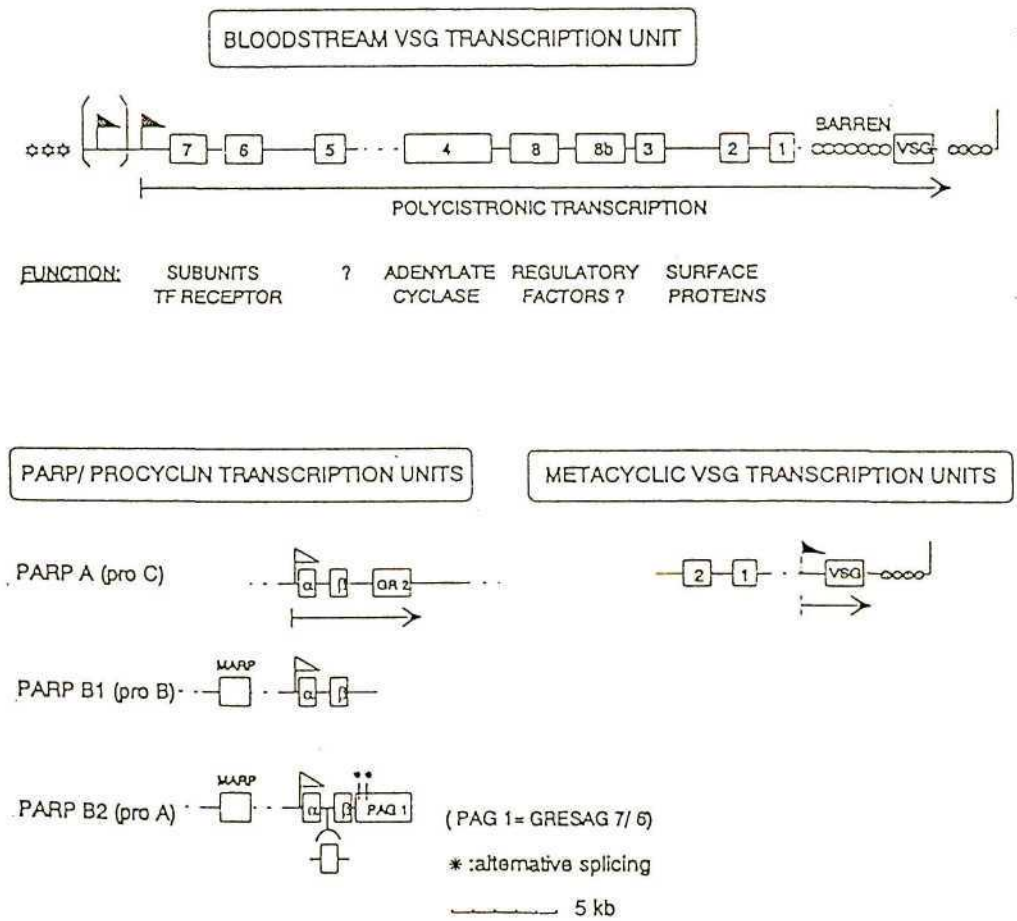


Figure 7 The expression sites for the major surface antigens of *T. brucei.* The star, open circle, and shaded circle represent different repeat regions. TF: transferrin; MAP: microtubule-associated repetitive protein; PAG: procyclin associated gene; GRESAG: gene related to ESAG (Pays *et al.*, 1994).

Switching of the expression of VSG genes is achieved by several different mechanisms that result in either the activation or deactivation of a different expression site or the alteration of the resident VSG gene in an already active site. Mechanisms of antigenic variations are

shown in Figure 8. Only one expression site, from a repertoire of 10 to 20 potential candidates is active at a given time. Three different mechanisms lead to the change of a VSG gene being expressed resulting in hybrid and mosaic genes as well as complete gene copies (Smith, 1996). Each VSG molecule is encoded by a specific gene known as the basic copy (BC). Firstly, in gene conversion (A in Figure 8) a variably sized gene known as the expression linked copy (ELC) replaces the previously expressed BC gene (X) by a copy of the ensuing gene, which may be telomeric (Y) or chromosome-internal.

The converted DNA sequence, ELC, is often about 3 kb in length. This DNA segment generally comprises the whole VSG gene which is approximately 1.5 kb and has a cotransposed sequence which is about the same size (Pays, 1985). Secondly, reciprocal recombination exchange of two telemetric genes may occur (B in Figure 8), leading to the expression of gene Y in the telomere of chromosome C1, along with the inactivation of gene X. Thirdly, the telomere of chromosome C1 may be deactivated and the resident antigen gene (X) silenced, while the telomere of chromosome C2 becomes the expression site of gene Y.

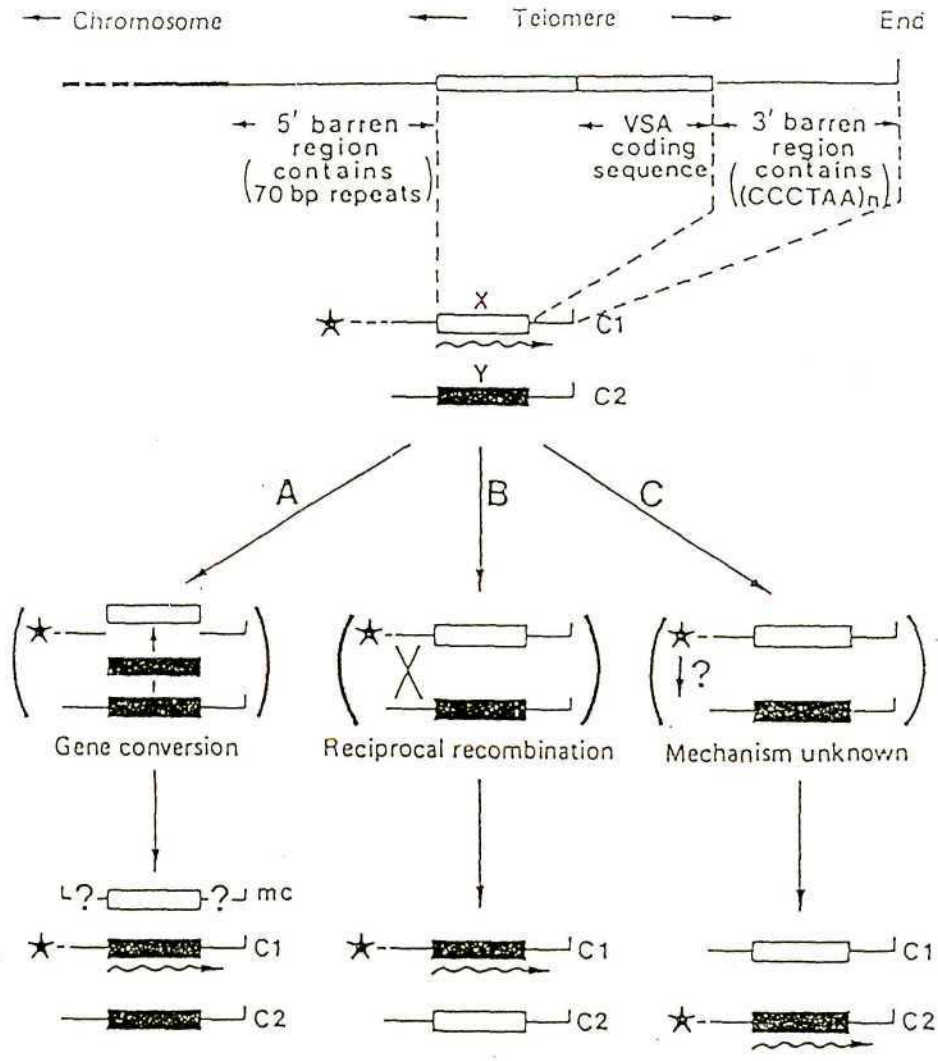


Figure 8 The different mechanisms of antigen switching in bloodstream form trypanosomes(Pays,1988).

CHAPTER 3

PROTEOLYTIC ENZYMES

3.1 Introduction

Proteolytic enzymes have been identified in animals, plants, and a wide variety of micro-organisms including African trypanosomes (Lonsdale-Eccles and Grab, 1987; McKerrow, 1993a; Turk and Bode, 1993). These enzymes, which catalyse the degradation of peptide bonds in both proteins and peptides to form peptides and/or amino acids, play important roles in many intracellular and extracellular processes of physiological importance. Some examples include, intracellular protein turnover, peptide hormone processing, blood clotting, and complement reactions. Proteolytic enzymes from microbial organisms have also been implicated in the development and progression of many human diseases, such as the penetration of host tissues and cells and in counteracting host defence mechanisms during infection (McKerrow, 1993a).

Proteolytic enzymes may be grouped into two broad categories on the basis of the position at which enzyme catalysis takes place. Enzymes that catalyse the cleavage of an internal peptide bond within a polypeptide are often called proteinases or endopeptidases while those that catalyse the cleavage of one or two amino acids from the amino or carboxy terminus of a polypeptide are referred to as exopeptidases or, more specifically aminopeptidases or carboxypeptidases (McKerrow, 1993a). However, the classification of proteolytic enzymes is usually based upon the amino acid or chemical residues found in the catalytic site of the enzyme.

3.2 Classification of proteolytic enzymes

Proteases have been classified in several ways. The procedure of classification that is normally used is the one proposed by Hartley (1960) which forms the basis of the system used by the Enzyme Commission. This method of classification categorises the proteases into four different classes of proteinases (Table 3), namely serine, cysteine, aspartic, and metalloproteinases. The different classes of proteolytic enzymes may be distinguished by their sensitivity to various types of inhibitors. This sensitivity is determined by their catalytic mechanisms and active site residues.

3.2.1 Serine proteinases

Serine proteinases form the largest family of proteolytic enzymes (McKerrow *et al.*, 1989) and they are considered to be some of the most important in nature (Sakanari, 1989). This group of enzymes play important roles in the physiology and metabolism of organisms and is also implicated in the pathogenesis of a number of infectious diseases. Most serine proteinases have been found to be most active at alkaline pH (North, 1982).

Table 3 Different classes of proteolytic enzymes and their properties (Barrett and Rawlings, 1993; Rawlings and Barrett, 1994).

Class	Representative proteinases	Essential active site residues	pH range for optimum activity	Typical inhibitor
Serine	Trypsin, Chymotrypsin	His-57, Asp-102, Ser-195	7-9	DIPF[a], PMSF[a]
Aspartic	Penicillopepsin cathepsin D.	Asp-33, Asp-213	2-7	Pepstatin A
Metallo	Thermolysin, A, collagenase,	Zn, Glu-143, His-231	7-9	EDTA[a], 1,10-phenanthroline
Cysteine	Papain, Cathepsin D	Asp-15, Cys-25, His-159	3-8	Iodoacetate, cystatin, E-64

[a] Abreviations defined in the text.

Typical inhibitors of the serine proteinases are diisopropylfluorophosphate (DIFP or DFP) and phenylmethylsulfonyl fluoride (PMSF) (North, 1982). The active site of these enzymes involves a serine amino acid, whose hydroxyl group catalyses hydrolysis. In addition to a serine residue at position 195 (chymotrypsin numbering), histidine and aspartic acid at positions 57 and 102, respectively are essential components of the active catalytic site (Barrett and Rawlings, 1993). These three amino acids were thought to form a "charge

relay" but this view has now been rejected because there is no evidence for the shuttling of a charge between these amino acids and they are now called the "catalytic triad".

Barrett and Rawlings (1993; 1995) have suggested that serine proteases evolved from different origins and proposed that they be subdivided into five or six clans, based on the way in which the catalytic residues are arranged in the linear sequence of the polypeptide chain. In the chymotrypsin clan, the three essential amino acids of the active site are arranged as His-57, Asp-102 and Ser-195. In the subtilisin and oligopeptidase clans the arrangement is Asp-32, His-64 and Ser-221 and Ser-554, Asp-529 and His-680, respectively. The serine type carboxypeptidases are from a family in which the order of the catalytic residues is Ser, Asp, and His and these enzymes are maximally active at approximately pH 5, a property that is different from the alkaline pH range in which serine proteinases are normally active.

3.2.2 Aspartic proteinases

Aspartic proteinases have been found in plants, animals (Barrett, 1986a) and fungi (North, 1982). They comprise a family of proteolytic enzymes that generally act in an acidic environment, primarily in the extra-cellular space (pepsin, penicillopepsin) and also in lysosomes (cathepsins D and E). Renin is the only aspartic protease that is known to be active at neutral pH (Scharpe *et al.*, 1991). A protease from the human immunodeficiency virus (HIV) was found to be optimally active in the pH range 4-6, in contrast to nearly all of the other eukaryotic aspartic proteases, which catalyse best in the pH range of 2-4 (Ido *et al.*, 1991; Wondrak *et al.*, 1993; Sugrue *et al.* 1994). This class of proteinases includes peptidases that depend upon the aspartic residues for their catalytic activity. North *et al.* (1982) reported aspartic proteinases to have molecular weight values in the range 30 kDa to 45 kDa and isoelectric points below pH 5.1. The most studied aspartic proteinase is pepsin, the proteinase that plays a role in the digestion of proteins in the stomach of higher animals. These enzymes have also been found to be less diverse compared to other proteases (Barrett and Rawlings, 1993) as demonstrated by their classification into only two families, which appear to share a common origin.

The enzymes in the pepsin family are composed of two homologous lobes that form the active site cleft between each lobe. The essential amino acid residues that form the active site are two aspartic residues at positions 32 and 215 (pig pepsin numbering) present in each lobe (McKerrow, 1993a). Amino acid sequence data and X-ray crystallography for pepsin

and other proteinases suggest that the enzymes evolved by genetic duplication (Tang, 1979). Specificity studies have shown that the size of the binding cleft is sufficient to accommodate seven to eight amino acid residues of a peptide substrate.

Specific inhibitors of this class of proteinases are pepstatin (although pepstatin does not inhibit fungal aspartic proteinases), acetyl pepstatin, diazoacetyl norleucine methyl ester, N-diazoacetyl-N'-2,4-dinitrophenylethylene diamine, and epoxy (p-nitrophenoxy) propane (North, 1982). Recently, inhibitors of the aspartyl proteinase of the HIV virus have been developed and used as a part of the treatment regimen for acquired immunodeficiency syndrome (AIDS) patients (Cohen, 1996).

3.2.3 Metalloproteinases

Metalloproteinases have been found in a wide range of living organisms such as bacteria, fungi and higher eukaryotes (Barrett, 1986a; Barrett and Rawlings, 1993). This group of enzymes has been considered to be proteases, enzymes that catalyse the cleavage of proteins, but not proteinases, enzymes which only catalyse the cleavage of internal peptide bonds within a polypeptide. The best characterised metalloproteinases are thermolysin (from *Clostridium histolyticum*) and the collagenases (from mammals). These enzymes normally show pH optima at pH 7 or greater. The active sites of natural forms of metalloproteinases usually require the metal ion zinc but some of these enzymes retain activity when this ion is replaced by another cation such as cobalt (Vallee *et al.*, 1990). Zinc has been shown to be essential for catalytic function of at least six classes of metalloproteinases established by the International Union of Biochemistry. The metal ion has been shown to be able to bind to three amino acid side chains of protein (two histidine side chains and one glutamic acid side chain) while the fourth ligand of the metal is a water molecule which helps initiate the attack on the peptide bond. In the case of structural zinc atoms, four protein ligands normally bind to the zinc. These zinc complexes are normally formed via interaction with the nitrogen, oxygen or sulphur atoms that are part of the respective amino acid residue side chain. The binding frequency is normally in the order His>>Glu>Asp=Cys (Vallee *et al.*, 1990).

Metalloproteinases include both enzymes that are very nonspecific such as thermolysin and others that have very strict substrate requirements such as mammalian collagenase. These enzymes will most often only cleave true peptide bonds formed between amino acid residues

and not peptidyl ester or peptidyl amide bonds. Frequently, the site of cleavage occurs at residues that have non-polar side chains and a bulky hydrophobic residue such as leucine or phenylalanine is preferred in position P1 (notation of Schechter and Berger, 1967; see "Figure 9", p32; Barrett, 1986b).

3.2.4 Cysteine proteinases

Cysteine proteinases, which are of particular relevance to the present study, are widely distributed in animals, plants and micro-organisms. This class of enzymes is involved in various intracellular and extracellular processes of physiological importance, namely proteolysis in protein regulation, normal tissue remodelling, and proteolysis involved in the morphological changes occurring during the complex life cycle of parasites (Turk and Bode, 1993; North, 1982). These enzymes have also been implicated to play a role in the development and progression of various human diseases such as arthritis, tumour invasion and metastasis. They are thought to be involved in the penetration of hosts by micro-organisms, countering host defence mechanisms, and nutrition of micro-organisms during infection (Roose and Noorden, 1995). Of the families of cysteine proteinases, the largest and best studied is the papain family (Barrett *et al.*, 1986b). This family also includes other cysteine endopeptidases of plants as well as lysosomal cathepsins (B, H, L, and S). Papain is a plant cysteine proteinase present in the latex of *Crania papaya* (Drenth *et al.*, 1970).

The cysteine proteinases all contain an essential cysteine residue in their active site but may differ in substrate specificities and other enzymatic properties due to evolutionary divergence (Wiederanders *et al.*, 1992). The thiol (-SH) group of the catalytic cysteine residue is involved in the binding and subsequent catalysis of the substrate (McKerrow, 1993a). The essential amino acids in the active site of cysteine proteinases are cysteine and histidine in positions 25 and 159, respectively, based on the numbering of the papain cysteine proteinase (Barrett and Rawlings, 1993). The active site of cysteine proteases is widely considered to consist of the zwitterionic triad $Asn^{\cdots}His^{+\cdots}Cys^{-}$ with the thiolate anion functioning as a nucleophile (Beveridge, 1996). Together with serine proteinases, the cysteine proteinases hydrolyse more peptide bonds than all other proteinases (Barrett and Kirschke, 1980).

Cysteine proteinases have been found to be generally most active at a slightly acidic pH and their activity is regulated by endogenous protein inhibitors called cystatins. Such inhibitors

are distributed throughout the plant and animal kingdoms (Turk and Bode, 1993). *In vitro* inhibitors of cysteine proteinases that are widely used include iodoacetate, 2,2'dipyridyl disulphide, L-trans-epoxysuccinyl-leucylamido-(4-guanidino)-butane (E-64) and mercuribenzoate, which also inhibits serine proteinases (Table 3). Cysteine proteinases are activated by the presence of cysteine, dithiothretol (DTT) and less frequently by ethyldiaminetetraacetic acid (EDTA) (North, 1982).

3.3 The activity of cysteine proteinases

Papain is a cysteine endopeptidase with broad specificity (Smith and Kimmel, 1960). It is widely considered as the model for cysteine proteases. Therefore, information obtained from this enzyme is frequently extrapolated to explain properties of the mammalian cysteine proteinases such as the cathepsins B, L, H and S. Indeed the enzymatic characteristics and amino acid sequences of cathepsin L and B, as well as the protozoan cysteine proteinases are similar to those of papain (North, 1982; Fish *et al.*, 1995).

Papain consists of a single polypeptide chain of 212 amino acids (Drenth *et al.*, 1970) folded into interacting domains delimiting a cleft on the surface of the enzyme where the substrate can bind. The catalytic residues Cys-25 and His-159 of papain (Menard, 1993), are located on the interface of the cleft on opposite domains of the molecule. These residues have been shown to be conserved in all members of the cysteine peptidase family. It has long been recognised that the enzyme acts by nucleophilic attack of the Cys-25 on an amide or ester carbonyl group (Lowe and Williams, 1965). The thiolate imidazolium ion pair between Cys-25 and His-159 (papain numbering) forms the reactive nucleophile (Barrett, 1986). An acyl enzyme intermediate is formed with the thiol of Cys-25 during the hydrolysis reaction. Other residues which are important for catalysis include Gln-19 which helps form the oxyanion hole and Asn-175 which orientates the imidazolium ring of His-159 (Menard, 1991).

Conditions required for the activity of cysteine proteinases vary widely with enzymes and substrates. The ionic interaction of cysteine and histidine in the catalytic site produces a structure with a pKa value of approximately 4 which is contrasted by a value close to pH 8 for a typical thiol group. This greatly increases the reactivity at acid pH. However, other additional factors such as the denaturation of some protein substrates at acid pH, can render the substrate more susceptible to hydrolysis. Although papain is known to have a broad pH

optimum, with activity extending well into the alkaline range, cathepsins L and B were thought to be active only at acidic pH because they are rapidly denatured at alkaline pH. However, Dehrmann *et al.* (1995; 1996) showed that the activity and stability of cathepsins B and L is affected by both pH and ionic strength, in a co-operative manner, which caused their pH optima, and stability at physiological pH, to be underestimated when buffers of constant molarity were used.

The binding site of papain is large and extends over 25 angstroms. It can be divided into 7 subsites (Schechter and Berger, 1967) as shown in Figure 9. Four of the subsites occur on the N-terminal side of the susceptible bond and are referred to as the S_1 to S_4, and three on the C-terminal side of the susceptible bond referred to as the S'_1 to S'_3 sites. The corresponding amino acid residues of the substrate that bind in these sites are normally called P_1 to P_4 and P'_1 to P'_3. Each subsite in the enzyme only accommodates one amino acid residue of the substrate. Therefore, the reactivity of a certain bond does not only depend on the two amino acid residues forming the susceptible bond (P_1 and P_1'), but also on the nature of the residues in its immediate neighbourhood.

3.4 The specificity of cysteine proteases

Synthetic substrates and inhibitors have been used extensively in the mapping of the specificity sites of papain. For example, papain exhibits broad specificity for N-acyl-L-amino acid derivatives in the S_1 subsite. Smith and Kimmel (1960) established that derivatives of L-arginine and L-lysine are hydrolysed more rapidly than uncharged amino acids, while those of L-aspartic acid and L-glutamic acid are hydrolysed more slowly than the uncharged amino acid derivatives.

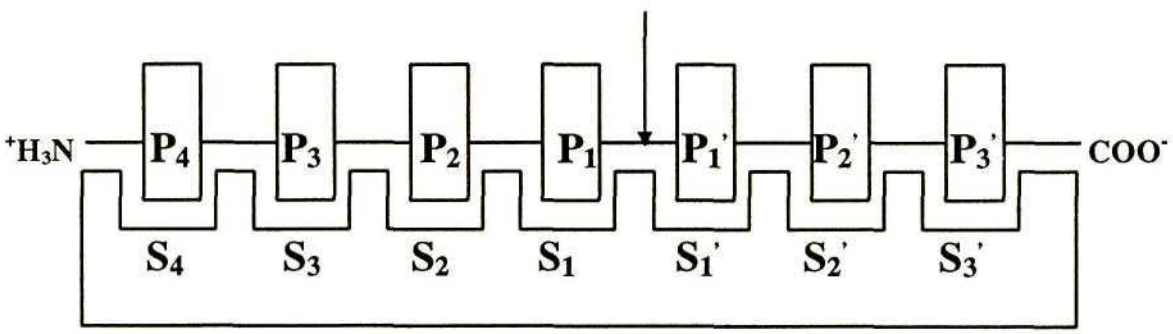


Figure 9 Schematic representation of the subsites of the proteinase active site (designated S) and the corresponding substrate amino acid residues (designated P). The arrow indicates the scissile bond (Schechter and Berger, 1967).

This is due to the fact that arginine and lysine are positively charged and as a result, are attracted to the active site, whereas aspartic acid and glutamic acid are negatively charged and are repelled from the S_1 subsite. Drenth *et al.* (1968) suggested that this could be due to the side chains of these substrates binding to the enzyme active site close to a carboxylic acid group. Thus, S_1 shows a preference for lysine over alanine and is stereo-specific for L-amino acid residues. In general, the cysteine proteases catalyse the hydrolysis of a variety of peptide, ester, and amide bonds of synthetic substrates.

Using X-ray crystallography Kamphius and Drenth (1984) established that subsite S_2, which binds a phenylalanine side chain very well, is a hydrophobic pocket in papain. Although there are some marked deviations from this generalisation, other enzymes of the papain family such as cathepsin L (Barrett, 1986b) and trypanopains (see later) tend to retain this characteristic of binding hydrophobic amino acids in their S_2-binding sites. The amino acid residue at the P_2 position is the predominant residue in defining the substrate specificity of papain and related cysteine proteases. Papain exhibits a strong preference for amino acids with bulky, non-polar side chains at that position (Schechter and Berger, 1968; Menard, 1993). Thus, the presence of the phenylalanine in P_2 enhances the susceptibility of a peptide to cleavage (Berger and Schechter, 1970; Drenth *et al.*, 1971). Thus, hexapeptide H_2N-Lys-Ala-Phe-Ala-Lys-Ala-COO^- will be cleaved at the peptide bond between alanine and lysine residues instead of between lysine and alanine preceding the C-terminal. Peptides that contain phenylalanine at the second residue from the C-terminal, H_2N-Asp Ala-Ala-Phe-Lys-COO^-, are strong inhibitors. Valine residues also show strong affinity for subsite S_2 (Berger and Schechter, 1970).

Like S_2 the S_1' subsite has a preference for hydrophobic residues, particularly tryptophan and leucine (Alecio *et al.*, 1974). The S_1' subsite could be contained in the hydrophobic region near Trp-177 and Ala-136. X-ray crystallographic and model binding studies have suggested that, in addition to the S_2-P_2 hydrophobic interaction in the binding pocket, the S_1-P_1 and S_2-P_2 hydrogen bonds present in the back-bone of the peptide chain are also critical for catalysis. It appears that the S_2-P_2 hydrogen bonding primarily accounts for the proper alignment of the substrate whereas S_2-P_2 hydrophobic interaction yields the major portion of S_2-P_2 binding energy (Asbóth, 1988). These two effects together confer the principal part of specificity in the catalysis by the proteases of the papain family.

3.4.1 Cathepsin B

Cathepsin B appears to act as an endopeptidase only in high enzyme to substrate ratios and contains at least six amino acid side chain-binding sites which are labelled S_1 to S_4 to the left of the sulfhydryl group and S_1' and S_2' to the right. Assays determining specificity of cathepsin B with synthetic substrates, show that P_2 may be the most important residue, just as it is for papain (Menard, 1993). However, the S_2 subsite specificity of cathepsin B is different from that shown by papain in that substrates with an arginine residue at P_2 are also hydrolysed at a significant rate while these substrates are only slowly hydrolysed by papain (Menard, 1993). Glutamic acid-245 which has a pKa of 5.1 in cathepsin B is responsible for the S_2-P_2 specificity for arginine containing substrates and controls the pH dependence of hydrolysis by the enzymes. In papain Val-133 and Val-157 at S_2 site are involved in the hydrophobic interactions. These residues are replaced by Ala-171 and Gly-195 in cathepsin B thus further reducing the hydrophobicity and bulkiness at the S_2 site of cathepsin B. Cathepsin B can accommodate a positively charged residue at the P_1 site, which suggest that the S_1 site of the enzyme has a negatively charged residue which interacts electrostatically with the P_1 site residue in the binding modes of the endopeptidase activities. S_2 and S_4 sites tend to prefer hydrophobic amino acids (Koga *et al.*, 1991).

3.4.2 Cathepsin L

Cathepsin L has been postulated to have at least four side chain binding sites on either side of the cleavage point. This enzyme shows the importance of apolar residues in P_2 and partly in P_3 and, on the other hand, the relatively minor importance of P_1 and P_1'. The positive influence of apolar residues in P_2 and P_3 is generally accepted (Barrett and McDonald, 1980). In addition, a hydrophobic residue in P_1 can decrease the susceptibility of the peptide bonds for cathepsin L. In cathepsin L Val-133 and Val-157 which form part of the S_2 subsite of papain, are replaced by alanine and leucine, respectively. These substitutions are complementary as regards the net size of hydrophobic side chains. The decrease in size of the side chain at one site (Val-133), which is replaced by alanine is compensated for by the increase in that at the other site (Val-157), which is replaced by leucine (Koga *et al.*, 1990). The three synthetic substrates that are commonly used to assay for cathepsin L are Bz-Arg-NH_2, Z-Lys-$OPhN_2$ and Z-Phe-Arg-NMec. Generally, Z-Phe-Arg-NMec, which is a fluorogenic rather than a chromogenic substrate, is used most often because of its high sensitivity (Barrett and McDonald, 1980).

3.4.3 Cathepsin H

Cathepsin H differs from papain in that it is an exopeptidase (an aminopeptidase) rather than an endopeptidase (Takahashi *et al.*, 1988). The most sensitive synthetic substrates for cathepsin H are Arg-NNap, and Arg-NMec. Leu-NNap is less susceptible to hydrolysis by cathepsin H than are Arg-NNap or Arg-NMec (Barrett and McDonald, 1980). The specificity of cathepsin H is primarily dependent on S_1 side chain recognition (Takahashi *et al.*, 1988; Rothe and Dodt, 1992). It can also accommodate hydrophobic residues such as phenylalanine, tyrosine, tryptophan and leucine. Small or polar side chains are not favoured, and proline is prohibited. Hydrolysis cannot proceed when the S_1 pocket is occupied by proline (Koga *et al.*, 1990). Cathepsin H is, nevertheless, thought to play an important role in intracellular protein degradation as do other lysosomal proteinases.

3.4.4 Cysteine proteases from *T. brucei* and *T. congolense*

The cysteine proteinases of *T. brucei* and *T. congolense* show preference for substrates with bulky hydrophobic residues at the P_2 position indicating that they are cathepsin L-like rather than cathepsin B-like (Lonsdale-Eccles and Grab, 1987; Pamer *et al.* 1989; Mbawa *et al.*, 1992; Troeberg *et al.*, 1996). These proteinases hydrolyse only amide bonds with arginine or lysine at P_1, and at P_2 they require hydrophobic residues such as phenylalanine or leucine (Mbawa *et al.*, 1992). The cysteine protease from *T. brucei* is also reported to hydrolyse substrates with Arg-Arg in the P_2 and P_1 positions, cathepsin B like specificity (Pamer *et al.*,1989). However, Troeberg *et al.* (1996) dispute this claim. Similarly *T. congolense* cysteine proteinases also show preference for arg or lys in P_1 and hydrophobic amino acids at P_2 and P_3. Currently, the best synthetic substrate for the assay of these enzymes is the fluorogenic peptide Z-Phe-Arg-NHMec.

The structure, function and molecular evolution of trypanopains has been studied using molecular techniques. Mottram *et al.* (1989) cloned and sequenced a cDNA for a *T. brucei* cysteine protease gene. The deduced protein sequence, when compared to other cysteine proteases, was divided into four domains. The first three domains are the hydrophobic 20-residue pre-sequence, the hydrophilic 105-residue pro-sequence and a 217-residue central region. All these domains showed significant homology to cathepsin L with an overall homology of 45%. The residues which are involved in the catalytic activity of papain are also conserved in the central domain.

The cysteine protease from *T. brucei* was also found to contain a long C-terminal extension which distinguishes the trypanopains from all mammalian cysteine proteases identified to date (Mottram *et al.,* 1989). A similar C-terminal extension has been identified in trypanopains from *T. cruzi* (Åslund *et al.,* 1991), *T. b. rhodesiense* (Pamer *et al.,* 1990), *T. congolense* (Fish *et al.,* 1995) and *Leishmania* (Souza *et al.,* 1992; Traub-Cseko *et al.,* 1993). The extension begins, in trypanopains, with nine consecutive proline amino acid residues and continues with a further 100 amino acid residues including eight cysteine amino acid residues. The *T. b. rhodesiense* cysteine protease nucleotide sequence was found to be almost identical to that of *T. brucei.* The function of the C-terminal extension is not known, but is thought to be involved in the targeting of the enzyme within the cell (Mottram *et al.,* 1989). The cysteine protease from *T. b. rhodesiense* was found to be active in a truncated form, with the C-terminal extension excluded. Thus, the C-terminal amino acids are not necessary for proteolytic activity (Pamer *et al.,* 1991).

3.5 Role of proteolytic enzymes in trypanosomiasis

Inhibitors of cysteine proteases have been used extensively in establishing the roles played by the proteinases in cells. However, little work has been done in order to establish whether these inhibitors reach and inactivate their targets in cell culture or *in vivo* (Wilcox and Mason, 1992). It is suggested that proteolytic enzymes play an important role in the interaction between host and pathogenic parasites and to assist the pathogenic parasite in penetrating the host tissue, degrading connective tissue (McKerrow, 1993a) thereby enabling the parasite to utilise host proteins for their nutrition (North, 1982). They also play a major role in the evasion of the host immune response by degradation or inactivation of the immune molecules by the parasite and the prevention of blood coagulation (McKerrow, 1993b; Lonsdale-Eccles *et al.* 1993; Russo *et al.* 1994; Roose and Noorden, 1995).

The enzymes of the cysteine proteinase family have been especially implicated in metabolising host proteins for the development and growth of the parasite. These enzymes may also facilitate the transition between the morphologically different forms of a parasite that survives in extremely different environments. The cysteine proteinase in *T. cruzi* (etiological agent of American trypanosomiasis) has been found to play a major role in remodelling cytoskeletal or other structural components of the parasite (McKerrow, 1993a;

McKerrow, 1993b). Cysteine proteases have been found to be antigens that can elicit an immune response in some infected animals. This was established for *T. congolense* by Authié *et al.* (1993) and has also been demonstrated in *T. brucei* infections (Troeberg *et al.*; unpublished observations; personal communication). The enzymes are also thought to play a critical role in the formation of short stumpy form parasites from the long slender bloodstream form parasite in *T. brucei* (reviewed in McKerrow, 1993a; Roose and van Noorden, 1995). Since the short stumpy forms are infectious to the tsetse fly, inhibition of the proteinase could possibly interrupt the life cycle and stop the spread of the disease. Due to the major and diverse functions of proteases in the parasite life cycle and parasite infections, they have become an important target for drug and vaccine development.

CHAPTER 4

CLONING OF THE PROMOTER REGIONS OF *T. BRUCEI* AND *T. CONGOLENSE* CYSTEINE PROTEASE GENES IN *ESCHERICIA COLI*

4.1 Introduction

Advances in recombinant DNA technology have contributed to an explosion of biological knowledge through molecular analysis of genetic material from a variety of organisms. This has also made possible investigations relating to the nature and function of genes in both prokaryotes and eukaryotes. Gene cloning is a system that allows the isolation of the gene of interest and provides considerable information on genetic structure and function, including expression (Sreekrishna *et al.*, 1989). A wide range of cloning vehicles known as vectors, have been established. Examples of such cloning vectors are plasmids (Sambrook *et al.*, 1989), bacteriophage lambda, cosmid vectors (Collins and Hohn, 1978) and yeast artificial chromosomes (YAC) (Burke *et al.*, 1987). Recombinant DNA technology has also led to the construction of expression vectors which allow for the expression of a gene of interest under the control of a strong promoter (Cregg *et al.*, 1992).

While a number of cDNA sequences from trypanosomes have been cloned and sequenced, the promoter regions of these genes have not to date been isolated and characterised. In this investigation, genomic DNA libraries of *T. brucei* and *T. congolense* were constructed in p*EcoR*251 and pUC119 cloning vectors, in order to isolate the promoter region of the cysteine proteinase genes. Genomic libraries enable the identification of the control elements which are not present in the cDNAs. Construction and screening of genomic libraries and restriction endonuclease mapping of putative *T. congolense* clones are presented.

4.1.1 Vectors and cloning strategies

A wide variety of cloning vectors may be used with the bacterium *E. coli* as a host organism. *E. coli* and its vectors are used frequently as it is the most flexible type of host-vector system for DNA cloning, and is the most cost effective (Ratner, 1989). In addition, there exists an extensive understanding of its genetic, biochemical, and physiological systems. It is common practice for DNA fragments to be initially cloned and manipulated in

E. coli in order to understand the basic features, before the DNA is cloned and expressed in other organisms. However, the development of new and more sophisticated cloning vectors has been necessary and directed by research objectives (Ratner, 1989).

4.1.1.1 Bacterial plasmids

Bacterial plasmids are double-stranded closed circular DNA molecules that range in size from 1 to more than 200 kb. They exist as accessory genetic units that self replicate and are inherited independently of the bacterial chromosome (Sambrook *et al.*, 1989). Small bacterial plasmids are the simplest cloning vectors and the most widespread. Plasmid molecules also confer some phenotypic characteristics to a variety of bacterial cells. These include the degradation of complex organic molecules and the production of antibiotics, colicins, enterotoxins and restriction and modification enzymes. This cloning system combines the ease of purification with desirable properties such as high transformation efficiency, selectable markers for transformants and recombinants and the ability to clone reasonably large pieces of DNA.

The vector pBR322 (Bolivar *et al.*, 1977; Sutcliffe, 1979), one of the first vectors to be developed, is still frequently employed. Important features of this vector are its size, which minimises breakdown during purification, and a copy number of approximately 15 molecules per cell. *E. coli* plasmid vectors such as ColE1 (Bolivar *et al.*, 1977) and pMB1 (Hershfield *et al.*, 1974) are naturally occurring plasmids that have been modified to give rise to a series of vectors, including pBR322, which are suited for cloning DNA fragments ranging from a few hundred bp to 9 kb in size. The choice of a plasmid vector depends on the availability of restriction sites compatible with those present in the fragment to be cloned, the size of the fragment and the objectives of the study.

4.1.1.2 *E. coli* bacteriophage and its derived vectors

Extensive use of cloning vehicles specific to filamentous *E. coli* bacteriophage has also been made (Murray and Murray, 1974). Bacteriophages such as M13 have been used to construct various cloning vehicles ranging from M13mp1, M13mp2, to M13mp18 and M13mp19 (Messing, 1983; Norrander *et al.*, 1983; Yanisch-Perron *et al.*, 1985; Vieira and Messing, 1987). The mp series of bacteriophage M13 vectors have been constructed by the insertion of the *lac*Z fragments. Plasmid and phagemid vectors known as the pUC series have also been constructed with the insertion of the *lac*Z gene into a deleted version of

pBR322, with only the replication origin and the Ap^r gene remaining, thus creating a group of plasmid vectors that contain a variety of commonly used polycloning sites. The pUC series vectors have been further manipulated to produce plasmids pUC8 and pUC9 and phagemids pUC118 and pUC119. Phagemids such as pUC118 and pUC119 are able to produce single stranded versions of cloned DNA, in the presence of a helper phage, that may be used as template to determine the nucleotide sequence of the cloned DNA sequence. The cloning capacity of these vectors varies considerably with respect to the size of the DNA fragments that can be cloned in them. Their main disadvantage is the limit on the size of the fragment that can be cloned. The maximum capacity is generally accepted as 1.5 kb, although DNA fragments of 3 kb have been occasionally cloned. To increase the size of clonable DNA fragments in M13 derivatives, a number of novel vectors were developed by combining part of the bacteriophage with the pUC8 plasmid DNA. An example of such a hybrid plasmid-M13 vector is the phagemid pEMBL8 (Frischauf *et al.,* 1983). *E. coli* cells used as hosts for these vectors are subsequently infected with a helper phage, providing the necessary replicative enzymes and phage coat proteins. The phagemid pEMBL8 has been used to clone up to 20 kb genomic DNA fragments (Sambrook *et al.,* 1989). The pUC series cloning vectors have three main advantages that make them one of the frequently employed cloning vectors in *E. coli.* They have a high copy number (500-700 per cell) even before amplification. Secondly, identification of the recombinants can be achieved in a single step process by plating on a medium containing 5-bromo-4-chloro-3-indoyl-ϑ-D-galactoside (X-gal) in the presence of isopropylthio-ϑ-D-galactoside (IPTG) and ampicillin (Ap). Thirdly, they have a multiple cloning site which facilitates cloning of insert DNA molecules.

4.1.1.3 Lambda based cloning vectors

The need to clone large pieces of DNA, too large to be cloned into plasmid or M13 vectors, necessitated the development of lambda (λ) based cloning vectors. The two classes of λ based cloning vectors that were developed are λ-insertion and λ-replacement vectors. A λ-insertion vector contains at least one unique restriction site into which new DNA can be inserted. There are two popular insertion vectors, namely λgt10, which can carry up to 8 kb of new DNA (Huynh *et al.,* 1985). The other insertion vector is λZAPII which carries up to 10 kb of new DNA (Short *et al.,* 1988). In contrast, a λ-replacement vector contains

two restriction sites for the restriction endonuclease to be used for cloning. These vectors are normally designed to carry larger DNA fragments than λ-insertion vectors. The frequently used λ-replacement vectors include λWES.λB', in which two *Eco*RI sites flank the replacement DNA fragment and recombinant selection is solely on the basis of size. The other frequently used λ-replacement vector is λEMBL4, which carries up to 20 kb of inserted DNA (Frischauf *et al.*, 1983).

A more sophisticated λ based vector is the cosmid (Royal *et al.*, 1979). This type of vector is used to clone very large DNA fragments (35 to 45 kb) and has been constructed by combining a λ-phage DNA molecule and a bacterial plasmid. Recombinants are selected on the bases of size as non-recombinants will be too small to infect host cells. The important feature of all λ-based vectors is the ability to carry DNA fragments that are too large to be accommodated by plasmid or M13 and pUC series vectors. Vehicles for cloning larger DNA fragments are continuously being sought. Recent developments have centred around the bacteriophage P1 which can accommodate 110 kb of DNA into its capsid structure (Monaco and Larin, 1994). Another type of cloning vector, based on the F plasmids, can accommodate larger DNA fragments, up to 300 kb in size (Monaco and Larin, 1995).

Cloning vectors have also been developed for other species of bacteria, including *Streptomyces*, *Bacillus* and *Pseudomonas*. Some of these vectors are based on plasmids or bacteriophages specific to the host organism and some on broad host range plasmids able to replicate in a variety of bacterial hosts. Generally, these vectors are very similar to *E. coli* vehicles in structure and function. As mentioned above, *E. coli* as a host is particularly useful if the aim is to study basic features of molecular biology, namely gene structure and function. In areas such as in biotechnology where the aim is to use cloning in order to control or improve expression of a gene or to alter the phenotype of an organism, development of cloning vectors for other organisms becomes necessary.

4.1.1.4 Expression vectors

Production of large-scale quantities of recombinant DNA products in *E. coli* has met with various problems. Firstly, premature termination of transcription by sequences in the foreign gene may occur, which results in the loss of gene expression and lower yields (Cregg *et al.*, 1992). Secondly, bacteria cannot excise introns from primary mRNA transcripts which results in inefficient transcription and translation. Thirdly, these cells

cannot posttranslationally modify recombinant proteins efficiently. These problems led to the development of expression systems for higher organisms (Cregg *et al.*, 1993). Such expression systems include yeasts, baculoviruses and mammalian cells (Ratner, 1989). Yeasts have been used as host organisms for the production of important pharmaceuticals from cloned genes. The yeast, *Saccharomyces cerevisiae* is an important organisms in biotechnology, particularly in the brewing and breadmaking industries.

The development of cloning vectors for yeasts has been stimulated by the presence in these organisms of a 2 μm circle plasmid (Broach, 1982). Cloning vectors derived from the 2 μm yeast plasmid are called yeast episomal plasmids (YEps). Some YEps may contain the entire 2 μm yeast plasmid while others contain only the origin of replication. The vector YEp13 contains the 2 μm origin of replication, the entire pBR322 DNA and yeast chromosomal DNA. Thus, YEp13 is a shuttle vector which can replicate and be selected for in both yeast and *E. coli*. The other unique feature of YEps is that they can integrate into one of the yeast chromosomes. Other important yeast cloning vectors are yeast integrative plasmids (YIps), such as YIp5 and yeast replicative plasmids (YRps), such as YRp7. Three important factors that need to be considered when deciding on a suitable yeast vector are transformation frequency, copy number and stability of the plasmid in yeast cells.

Artificial chromosomes have now become a major research tool in both genome analysis and in the functional characterisation of genes. Yeast artificial chromosome vectors can accommodate DNA fragments of 1400 kb and greater in size (Monaco and Larin, 1994). However, these vectors are not always stably maintained and cloned DNA may become rearranged by intramolecular recombinations. Nevertheless, YACs are of immense value in accommodating large pieces of cloned DNA for large scale DNA sequencing programmes such as the human genome project. Recently, several alternative systems to YACs, based on bacterial host systems have been developed, namely P1 prophage clones, bacterial artificial chromosomes, P1-derived artificial chromosomes and mammalian artificial chromosomes. However, these have not as yet been characterised as thoroughly as YACs, for stability in particular. Some vectors have been derived for specific purposes, namely the production of fused polypeptides (Silhavy and Beckwith, 1985), extracellular secretion of cloned gene products (Chayeb, 1984), detecting promoters (Duvall *et al.*, 1984) and terminators (McKinney *et al.*, 1981). Efficient integrative vectors are now available for a

number of species including yeasts such as *Pichia pastoris* (Cregg *et al.,* 1993) and *Kluyveromyces lactis* (Das *et al.,* 1982) and the filamentous fungi *Aspergillus nidulans* and *Neurospora crassa.*

4.1.1.5 Gene cloning in higher eukaryotes

Vectors used for gene cloning, in plants and animals, are based on naturally occurring plasmids of *Agrobacterium* (Bevan, 1984). Examples of *Agribacterium* plasmids are Ti and Ri and are approximately 200 kb in size. Development of systems for cloning genes in animal cells has been researched extensively. Such vectors are needed for the synthesis of proteins from genes that do not function correctly on cloning into *E. coli* or yeast and in gene therapy. Direct gene transfer has been found to be less efficient and cloning experiments with animal cells depend on vectors based on viruses, namely simian virus 40 (SV40) (Sambrook *et al.,* 1986). Packaging constraints in SV40 limit the amount of insert DNA that can be cloned into the genome. The adenovirus (Graham, 1990), can accommodate larger sizes of DNA but is more difficult to work with because of the larger genome size. Papillomaviruses can accommodate larger DNA fragments and produce a stable transformed cell line. Baculoviruses enables large amounts of protein to be obtained from genes cloned in insect cells (Luckow and Summers, 1988).

4.1.2 DNA library construction

The identification and isolation of specific DNA sequences from a complex genome is preceded by cloning random fragments of genomic DNA to produce a genomic library. Random cleavage of total cellular DNA decreases the probability of systematic exclusion of sequences when constructing a library. The number of recombinant plasmids required to obtain a representative genomic library for a given organism may be calculated from the formula:

$$N = \frac{\ln(1-P)}{\ln(1-\frac{a}{b})}$$

Where N = the number of clones required,

P = the desired probability,

a = the size of the DNA insert

b = the total size of the genome

(Clarke and Carbon, 1976).

The number of clones required for a complete genomic library in bacteria, yeast and fungi is determined by the size of the genome and the capacity of the vector. The availability of vectors able to accommodate large DNA fragments has significantly reduced the labour involved in the construction, propagation, and screening of genomic DNA libraries of organisms with large genomes. Copy DNA (cDNA) libraries can be established by the process of enzymatic conversion of mRNA to double stranded DNA, followed by the insertion of this DNA into cloning vectors (Grubler and Hoffman, 1983). The cDNA libraries are important in establishing information relating to the protein coding region of a gene as the cDNA lacks introns.

Cloning by the polymerase chain reaction (PCR) allows for the construction of libraries from small amounts of DNA or RNA. The basic method of PCR cloning has been reviewed by Arnheim and Erlich (1992). There are three PCR cloning strategies. In the first method of cloning using PCR, cloning onto a suitable vector is done by means of blunt-end ligation. However, non-template directed nucleotide addition by the *Taq* DNA polymerase enzyme can lead to lower efficiency of ligation due to the difficulties associated with "nibbling" off of these additional nucleotides to create blunt ends prior to ligation (Arnheim and Erlich, 1992). In the second method, the PCR cloning strategy is performed such that it introduces "sticky ends" into the PCR product and the cloning vector by appropriately designed primers. The second strategy involves the use of PCR cloning vectors that contain 3' terminal thymidine molecules at both ends. Thus, the 3' A overhang of the PCR product, added by the *Taq* DNA polymerase, can ligate to the 3' T overhang of the vector which greatly improves the efficiency of ligation of PCR products. PCR cloning vectors are available commercially. Examples include pGEM-T, pCR-Script Amp SK (+), pCR-Script Cam SK (+), pCR-Script Direct SK (+) and pMOS Blue (Boehringer Mannheim, Germany). In the third method, appropriate restriction endonuclease sites are incorporated into the 5' ends of the custom made primers (Scharf *et al.*, 1996). Prior to ligation, the product is

digested to yield sticky ends which can be ligated to an appropriately digested cloning vector by directed cloning (Arnheim *et al.*, 1992). This strategy requires proper cutting of the restriction endonucleases or the cloning process becomes inefficient.

4.1.3 Identification of the clone of interest

The identification of the clone of interest from a gene library may be facilitated by an appropriate cloning strategy. Procedures available for the identification of the gene of interest are based on two strategies; direct selection for the gene of interest and the identification of the clone from a gene library. The technique known as marker rescue, makes use of mutant strains as the hosts for transformation. Transformants containing a wild-type gene will transform mutant cells to a wild-type phenotype. Marker rescue and other direct selection techniques do not work in all situations and there are many genes that cannot be selected by this method. In addition, these techniques are not of much use in providing bacterial clones of genes from higher organisms, as some of the enzymes may not be functional in bacterial cells. A reliable and frequently used method in identifying a desired clone is that of nucleic acid hybridisation (Southern, 1975). Heterologous probing based on hybridisation between related sequences for clone identification can also be employed. Probes may also be designed based on the amino acid sequence of the protein related to the protein of interest by preparing degenerate oligonucleotides.

Immunological screening (Young and Davis, 1983) has also been employed as an alternative to nucleic acid hybridisation. This strategy requires that the gene of interest be expressed and the resultant product should not normally be present in the host cells. Expression vectors promote the expression of the cloned gene in a specific host such as λgt11, λgt18-23, λZAP and λORF8 (Huynh *et al.*, 1985; Meissner *et al.*, 1987) and are therefore critical in constructing libraries to be screened with antibodies.

For cDNA library screening, the abundancy probing method may be used. In this method, a number of individual cDNA clones from the library are used to probe all the other members of the library. The abundant clone is considered the desired recombinant on the basis that cDNA libraries are normally formed in order to obtain a clone of a gene being expressed at a relatively high level in a particular cell type.

4.1.4 Cloning of the cysteine proteinase gene from trypanosomes

Gene cloning has been used extensively in determining the nucleotide sequences of many genes in trypanosomes, including the VSG and cysteine proteinase genes. Degenerate DNA oligonucleotides were used by Mottram *et al.* (1989) to identify a cDNA cysteine protease gene from a *T. brucei* λgt11 cDNA library. These oligonucleotides were designed on the basis of two conserved amino acid sequence regions of eukaryotic cysteine proteinases and the known partial amino acid sequence of a *T. cruzi* cysteine proteinase. An equivalent enzyme was identified from a *T. b. rhodesiense* λgt11 cDNA library by screening with a 495 bp DNA molecule encoding the highly conserved active site of the *T. brucei* cysteine protease (Eakin *et al.,* 1990; Pamer *et al.,* 1990). The cysteine proteinase gene fragments have also been amplified from a variety of other organisms, namely *T. cruzi, T. brucei* and *Entamoeba histolytica* by PCR cloning. The PCR primers were based on the consensus sequences flanking the active site Cys-25 and the active site Asp-175 of a cysteine proteinase of the slime mould *Dictyostelium discoideum* (Williams *et al.,* 1985; Eakin *et al.,* 1990). Åslund *et al.* (1991) used PCR to amplify part of the cDNA cysteine proteainase gene that includes the C-terminal extension from a λgt11 cDNA library constructed from epimastigotes forms of *T. cruzi,* for sequencing. The primers used corresponded to a sequence in the middle region of the cysteine proteinase gene and one complementary to the λgt11 vector. A cysteine proteinase cDNA gene from *T. congolense* has been cloned and sequenced (Fish *et al.,* 1995), using a *Pst*I fragment, containing much of the *T. b. brucei* cysteine proteinase open reading frame as a probe for screening the metacyclic λgt11 cDNA library.

4.2 Materials and methods

4.2.1 Trypanosomes

Trypanosome species used in this investigation were *T. congolense* IL3000 (Bienen *et al.,* 1991) and *T. brucei* ILTAT1.1. These parasites were obtained from the International Laboratory for Research on Animal Diseases (ILRAD), in Nairobi, Kenya.

4.2.2 Bacteria and plasmids

Bacterial strains used were *E. coli* MV1184 (Vieira and Messing, 1987) and *E. coli* HB101 (Bolivar and Backman, 1979). Vectors used were p*Eco*R251 (Remaut *et al.*, 1983) and the phagemid pUC119 (Messing *et al.*, 1987).

4.2.3 Propagation of trypanosomes, bacteria and plasmids

All media and solutions used are listed in Appendix B. Trypanosome parasites were propagated in rats (Appendix C) and the *E. coli* strains were grown in Luria medium (Miller, 1978). Plasmid transformed bacteria were selected for in Luria agar (LA) containing Ap (50 Tg/ml).

4.2.4 Chromosomal DNA extraction

Chromosomal DNA was isolated from trypanosomes employing a modification of the methods of Blin and Stafford (1976) and Ozaki and Cseko (1984). A pellet containing approximately 5 x 10^9 cells was resuspended in 10 ml cold SE buffer (Appendix B) containing freshly added proteinase K (Appendix B) to a concentration of 100 µg/ml. Sarkosyl (Appendix B) was added to a final concentration of 0.5% (w/v), the solution mixed gently and incubated at 50°C for 30 min, swirling periodically to effect cell lysis. After cooling the solution to room temperature, an equal vol. of phenol (Appendix B) was added and the two phases were mixed gently by turning the tube end over end, for 1 h using a roller apparatus. The two phases were separated by centrifugation at (5000 x g; 15 min; room temperature (RT)). Using a wide bore pipette the viscous aqueous phase was transferred to a clean centrifuge tube and extraction was repeated once more. Ammonium acetate (0.2 vol.; 10 M) was added and followed by the addition of 2 vol. cold isopropanol (4°C) to the aqueous phase in a clean centrifuge tube and the mixture was mixed thoroughly by swirling. Precipitated genomic DNA was spooled out of solution and resuspended in TE buffer (Appendix B).

4.2.5 Plasmid DNA extraction

Plasmid DNA from *E. coli* cells was prepared by the alkali-lysis method of Ish-Horowicz and Burke (1981) as modified by Sambrook *et al.* (1989) and described in Appendix C.2.

4.2.6 Restriction enzyme digestion of DNA

Standard techniques for the analysis of DNA were adopted from Sambrook *et al.* (1989). All restriction endonucleases together with the respective buffers were commercially obtained and used in accordance with the manufacturer's specifications (Appendix C.5).

4.2.7 Amplification of DNA by PCR

Probes for DNA/DNA hybridisation were synthesised employing the PCR method (Appendix C.3) and custom made primers. The custom made primers were synthesised in the Department of Virology at the University of Natal (Medical School). The DNA fragment amplified was defined by a fragment spanning the cDNA cysteine protease sequence from base pair 293 to 1045 of the *T. brucei* cDNA (Mottram *et al.,* 1989) and 317 to 1069 of the *T. congolense* cDNA sequence (Fish *et al.,* 1995). The amplified DNA was purified by transfer onto a DEAE membrane (Appendix C.4). The PCR amplified *T. congolense* DNA fragment was digested with *Alu*I and *Hinc*II. The PCR amplified *T. brucei* DNA fragment was singularly digested with *Pst*I and *Hinc*II. Diagnostic digests together with the controls, namely uncut PCR fragment, PCR reactions with everything except the enzyme, or template, or primer, were analysed by agarose gel to determine whether the correct regions have been amplified.

4.2.8 Construction of *T. brucei* and *T. congolense* genomic libraries

To construct genomic libraries in p*Eco*R251 insert DNA was ligated into the *Bgl*II restriction endonuclease site of the vector. Genomic DNA from the isolated parasites (4.2.4) was partially digested with *Sau*3A restriction endonuclease and size fractionated using a sucrose density gradient (Appendix C.7). Fragments of genomic DNA (4-10 kb in size) were ligated (Appendix C.9) with linearised vector in a 1:5 molar ratio. The ligation mixture was transformed into competent *E. coli* HB101 cells as described in Appendix C.11. Approximately 6 000 *T. brucei* and 15 000 *T. congolense* Ap^r transformants were obtained. Transformants constituting these libraries were pooled in lots of approximately 300 colonies. Plasmid DNA was extracted from each pool and stored at 4°C until further analysis.

Phagemid library construction involved linearising pUC119 with *Bam*H1 restriction endonuclease, dephosphorylating the vector using calf intestine alkaline phosphatase (CIP;

Sambrook *et al.,* 1989) and ligation with size fractionated DNA fragments (4-10 kb) in a vector to insert molar ratio of approximately 1:5. The ligation mixture was transformed into competent *E. coli* HB101 cells and the mixture was plated onto LA plates containing Ap in the presence of the substrate X-gal and the inducer IPTG (refer to C.11 for details). After incubation at 37°C overnight, a total of 33 000 *T. brucei* and 27 000 *T. congolense* recombinants were obtained. Recombinants from the *T. brucei* and *T. congolense* libraries were pooled in lots of approximately 400 and 300 colonies, respectively. Plasmid DNA was extracted from each pool by "midiprep" preparation (Appendix C.2.2) and stored at 4°C until further analysis.

4.2.9 DNA hybridisation

The techniques employed in this section are described in Appendix C. Single digests of chromosomal DNA from the parasites were made with *Eco*R1, *Hind*III, and *Bam*H1 restriction endonucleases. The digested genomic DNA, *Bam*HI linearised vector and the PCR synthesised DNA fragment were resolved by electrophoresis in a 0.8% (w/v) agarose gel and transferred onto a Magna charge nylon membrane (ICN Pharmaceuticals Inc.). Pools of recombinant plasmids were double digested with *Eco*R1 and *Hind*III restriction endonucleases, in order to release the inserted DNA fragment and the digests were resolved in a 1% (w/v) agarose gel, alongside an *Eco*R1 and *Hind*III linearised vector and the PCR synthesised DNA fragment, and transferred onto a nylon membrane.

Colony hybridisation, a method of Grunstein and Hogness (1975) and modified by Mason and Williams (1985) (Appendix C.13), was used to identify putative clones contained in the plasmid pools. In this method pools of DNA containing recombinant plasmids were grown on LA, replica plated onto Magna lift membrane disc filters (Micron Separations Inc.), colonies lysed *in situ* and the released DNA hybridised with an appropriate radiolabelled probe DNA fragment probe (Appendix C.14).

The PCR synthesised DNA probe was labelled employing the digoxigenin (DIG) labelling system (Boehringer Mannheim, Germany) (Appendix C.15) for probing of genomic blots and by nick translation (Amersham) (Rigby *et al.,* 1977) with [alpha-^{32}P]-deoxycytidine-5' monophosphate ([α-^{32}P]-dCTP) (Appendix C.14) for colony hybridisation and for probing Southern blots.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Isolation of parasites

Parasites were isolated from infected rat blood employing a combination of the methods of isopycnic isolation on a Percoll gradient (Grab and Bwayo, 1982) and an anion exchange chromatography (Lanham, 1968; Lanham and Godfrey, 1970). Approximately, 5×10^9 *T. brucei* cells were isolated from approximately 30 ml of blood (1.66×10^8 parasite cells per ml of blood). Similarly, approximately 5×10^9 *T. congolense* cells were obtained from approximately 50 ml of blood (1×10^8 parasite cells per ml of blood). This result indicates that the *T. brucei* parasites reached higher levels of parasitaemia in rats than did *T. congolense* parasites in these experiments. This was also demonstrated by the larger number of *T. brucei* parasites per microscopic field at peak parasitaemia.

Earlier methods used to isolate trypanosome parasites from infected rat blood, namely hemagglutination and differential centrifugation (Simmons *et al.,* 1964) and sucrose gradient centrifugation (Williamson and Cover, 1966) produced low yields. In order to improve the yield and resolution and to preserve cell viability the isolation methods used in this study were developed. The Percoll gradient method exploits the self generating gradient property of Percoll for the isolation of the parasites. In addition, contamination by platelets is reduced to negligible levels, with blood highly infected with trypanosomes. However, contamination by platelets may be significant with low parasitaemia. After centrifugation trypanosomes float as one band near the top of the gradient, well separated from the red blood cells which sediment to the bottom. Anion exchange chromatography exploits the fact that the ionic charges on the surface coat of salivarian trypanosomes differ significantly from the surface charges of mammalian blood cell components. Blood cells are preferentially retained on the cellulose resin while the trypanosomes pass through. In the present study, equivalent yields of parasites were obtained by anion-exchange without the application of the Percoll gradient step.

Isolation of *T. congolense* parasites was not always successful using this method as the trypanosome cells were sometimes trapped in the resin. This could be due to the resin particles being too fine for parasites to elute. However, the parasites eluted again after a few more attempts, under similar conditions. The Percoll density gradient method coupled

with anion exchange chromatography has proved to be relatively reproducible and easy to perform. The yield of the parasites was comparable to that in the recommended protocol for the isolation of genomic DNA, 1 x 10^{10} (Ozaki and Cseko, 1984). The eluant was cloudy with no blood cells as these were trapped in the resin and the trypanosome cells appeared to be highly viable as demonstrated by their motility under the microscope.

5.2 Extraction of chromosomal DNA

Genomic DNA was extracted from approximately 5 x 10^9 cells (500 Tg) employing the modification of the methods of Blin and Stafford (1976) and Ozaki and Cseko (1984). The genomic DNA was resolved in a 0.8% agarose gel electrophoresis and was found to be of high molecular mass. Contamination with RNA was low in extracts prepared without the addition of RNAse and absent when extracts were treated with RNase during the extraction procedure. The genomic DNA yield (approximately 500 Tg) and purity were fairly constant in subsequent extractions, and in both parasite species. Restriction endonuclease digestion and DNA amplification of the genomic DNA proved it to be sufficiently pure.

5.3 PCR synthesis of DNA probes

In order to facilitate probing of the genomic libraries, dsDNA probes were synthesised employing the PCR technique. PCR amplification of *T. brucei* genomic DNA, employing custom made primers corresponding to bp 293 to 313 and 1015 to 1045 on the cDNA sequence (Mottram *et al.*, 1989) produced a single band of approximately 752 bp as judged by agarose gel electrophoresis (Figure 10). The negative controls which lacked either the template, enzyme or primer did not result in amplification of the template DNA. This is indicative of the specificity of the amplification reaction. Pilot experiments were performed to determine optimal conditions and a temperature combination of 94°C, 50°C and 72°C was found to produce higher yields with all other variables as recommended by Sambrook *et al.* (1989). The concentration of the amplified fragment, as determined by spectrophotometry, from each reaction mixture was approximately 2 µg per 100 µl reaction mixture and this yield was relatively constant in all subsequent amplification reactions. The specificity of the amplified fragment was confirmed by diagnostic digests using restriction endonucleases for internal restriction sites established from the cDNA sequences. Digestion of the *T. brucei* PCR DNA fragment with *Pst*1 and *Hinc*II restriction endonucleases produced two fragments of 144 and 608 bp and three smaller fragments of 69, 200 and 483

bp, respectively (Figure 10). However, only two fragments of 200 and 483 bp could be visualised from the *Hinc*II digest. The resultant smaller fragments, as judged from agarose gel electrophoresis, confirm the specificity of the amplification reaction. In order to improve the resolution of the smaller fragments polyacrylamide gel electrophoresis was used (results not shown).

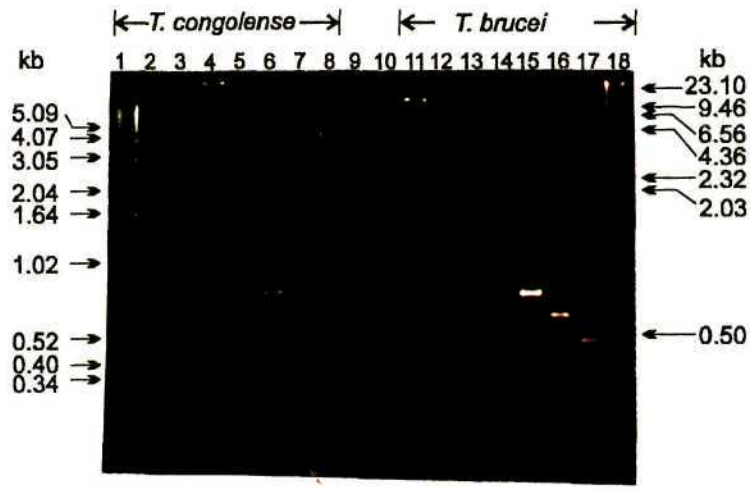


Figure 10 Agarose gel electrophoresis of PCR mixtures and restriction endonuclease digests of amplified DNA fragments.
Lane 1: kb DNA ladder marker; lanes 2 and 11: genomic DNA; lanes 3 and 12: no template DNA controls; lanes 4 and 13: no primer controls; lanes 5 and 14: no enzyme controls; lanes 6 and 15: test DNAs; lane 7: *Alu*I digest; lanes 8 and 17: *Hinc*II digests; lane 16: *Pst*I digest; lane 18: DNA markers.

The PCR amplification of *T. congolense* genomic DNA, employing custom made primers corresponding to bp 317 to 336 and 1039 to 1069 on the cDNA sequence (Fish *et al.*, 1995), similarly produced a single band of approximately 752 bp as judged by agarose gel electrophoresis (Figure 10). Optimal reaction conditions were determined as for *T. brucei* and similar optimal conditions were obtained. Digestion of the *T. congolense* PCR DNA fragment with *Alu*1 and *Hinc*II restriction endonucleases produced two fragments of 267 and 485 bp and four smaller fragments of 38, 150, 162 and 402 bp, respectively (Figure 10). However, only two DNA fragments, a doublet of 150 and 162 bp and a fragment of 402 bp could be visualised from the *Hinc*II digest. In Figure 10 it is shown that the expected smaller DNA fragments were obtained, thus initially confirming the amplification of the desired region of the cysteine protease gene from the genomic DNA of the parasites. Southern hybridisation was employed to confirm that the *T. brucei* and *T. congolense*

genomic DNA was indeed amplified. The *T. brucei* probe hybridised to two *Hind*III fragments of 1.85 and 3.95 kb (Figure 11).

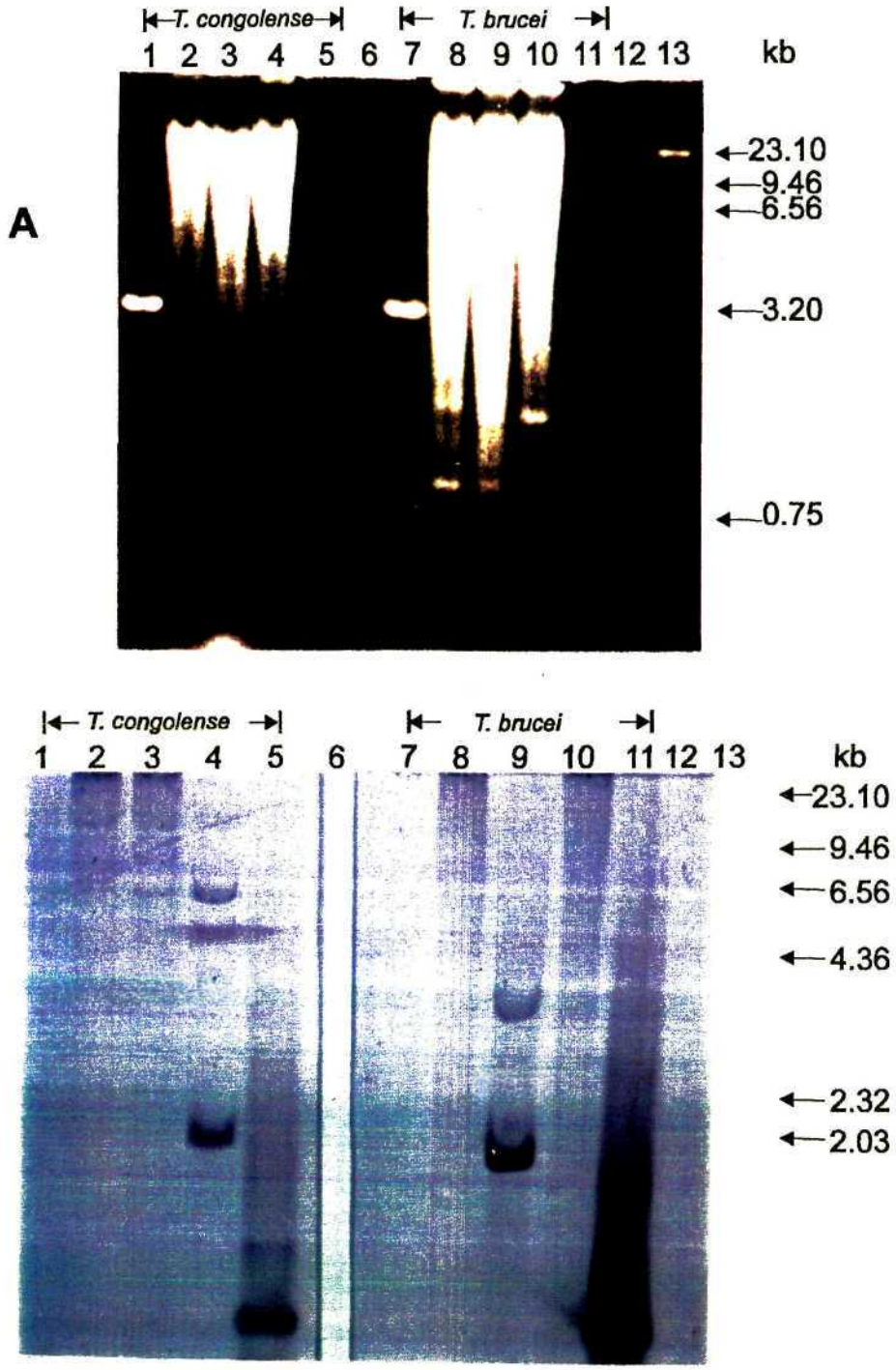


Figure 11 Ethidium bromide stained agarose gel electrophoresis (A) and the corresponding Southern blot of genomic DNA (B).
The genomic blots were hybridised with DIG-labelled homologous DNA probes. Lanes 1 and 7: linearised pUC119; lanes 2 and 8: *Eco*R1 digests of genomic DNA; lanes 3 and 9: *Hind*III digests of genomic DNA; lanes 4 and 10: *Bam*HI digests of genomic DNA; lane 5: unlabelled *T. congolense* probe; lane 11: unlabelled *T. brucei* probe; lane 13: DNA markers.

The *T. brucei* probe did not hybridise to the *Eco*R1 and *Bam*H1 genomic digests of *T. brucei* nor to the lambda DNA marker. The probe hybridised to the unlabelled PCR DNA fragment and produced a stronger signal, confirming the specificity of the probe. Thus, the probe hybridised specifically to the trypanosome genomic DNA and to the unlabelled DNA probe. Mottram *et al.* (1989) used an oligonucleotide probe based on the cDNA sequence of *T. brucei* to probe various complete and partial *T. brucei* genomic digests. A complete *Hind*III digest produced fragments of 1.75 and 3.75 kb that hybridised to the probe. The differences in the size of the DNA fragments between 1.75 and 3.75 and 1.85 and 3.95 is within experimental error. Identification of two bands, also in *T. congolense* (Figure 11), is thought to suggest that the cysteine protease gene could either be present in the parasite genome in tandem repeats, or in two copies.

The *T. congolense* probe hybridised to two *Bam*HI fragments of 1.98 and 4.5 kb (Figure 11). The *T. congolense* probe did not bind to the *T. congolense Eco*RI and *Hind*III genomic DNA digests nor to the lambda DNA marker. Hybridisation occurred with the unlabelled probe also, confirming the specificity of the probe. Thus, the probe hybridised specifically to the trypanosome genomic DNA and to the unlabelled probe.Hybridisation of the *T. brucei* genomic digests using the *T. congolense* probe gave the same hybridisation pattern as for the *T. brucei* probe, but with a low signal and *vice versa.* Therefore, it would not be advisable to use a probe from the *T. brucei* species to probe the *T. congolense* library as the intensity of the signal is critical in the screening of the genomic libraries. Both PCR fragments hybridised specifically to the trypanosome genomic digests indicating their suitability for use as probes for screening the genomic libraries.

5.4 Preparation of insert DNA

Sequences located in regions in which there are fewer restriction endonuclease sites are likely not to be present in libraries constructed from restriction endonuclease digests. On the other hand those sequences that are present in regions having numerous restriction sites for the specific restriction endonuclease have a greater probability of being cloned and the library frequency may be higher. Hence *Sau*3A restriction endonuclease, a frequent base cutter, was used to partially digest the genomic DNA.

Genomic DNA partially digested with *Sau*3A restriction endonuclease was size fractionated by sucrose density gradient centrifugation. Sucrose density gradient fractions were then

resolved in a 0.8% (w/v) agarose gel (Figure 12). Fractions containing 4-10 kb sized DNA fragments (Figure 12, well numbers 13, 14 and 15) were pooled, and used in ligation reactions. The concentration of pooled DNA fragments was 600 ng/Tl as determined by spectrophotometry. The precipitated DNA was sufficiently pure as judged by agarose gel electrophoresis for ligation reactions.

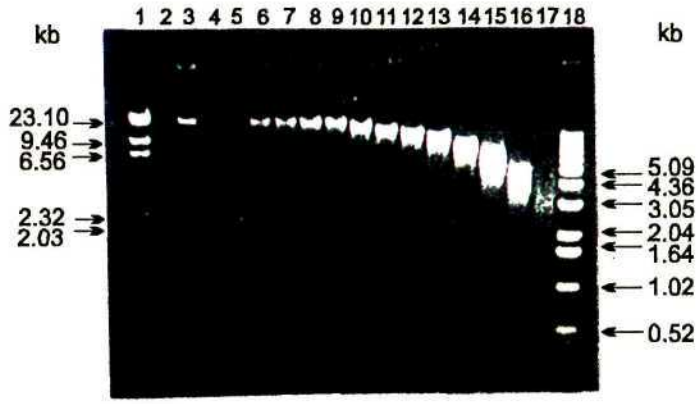


Figure 12 Agarose gel electrophoresis of sucrose density gradient fractions of *T. congolense* genomic DNA.
Genomic DNA was partially digested with Sau3A. Lane 1: *Hind*III lambda DNA marker; lanes 2-17: DNA fractions; lane 18: kb DNA ladder marker.

5.5 Construction and initial probing of *T. brucei* and *T. congolense* genomic libraries

Initially, an attempt to construct genomic libraries, from both trypanosome parasite species, in the *E. coli* suicide vector p*Eco*R251 was made. In this vector only recombinants are expected to grow due to the insertional inactivation of the *Eco*RI gene under the lambda rightward promoter or regulation by plasmid pc1857 which contains a temperature sensitive lambda repressor gene (Remaunt *et al.*, 1983). The *T. brucei* p*Eco*R251 library contained 6 000 recombinants and the *T. congolense* library contained 15 000 recombinants. Recombinant plasmids from each library were pooled in lots of 300 colonies, digested with *Eco*R1 restriction endonuclease and resolved by agarose gel electrophoresis. The recombinants appeared to contain inserts as demonstrated by the multiple bands present in the agarose gel lanes. The size of the released fragments ranged from approximately 500 bp to 10 kb.

Pools of recombinant plasmids were digested with *Eco*RI restriction endonuclease, Southern blotted and probed with an oligonucleotide probe spanning bp 293 to 313 on the

T. brucei cDNA sequence and 5' end-labelled with [γ-^{32}P]-ATP and a PCR synthesised dsDNA fragment that was labelled with [α-^{32}P]-dCTP. The p*Eco*R251 libraries did not reveal any putative clones containing the fragment of interest when probed with both an end-labelled oligonucleotide probe and a PCR generated dsDNA probe. This suggests that the libraries were not fully representative of all the gene sequences present in the trypanosome genome or that the gene of interest was producing a lethal product in the *E. coli* MV1184 cells.

Clarke and Carbon (1976) established that for a 99% probability of isolating a specific gene from a trypanosome genomic library with an average insert size of 7 kb, 20 000 recombinant plasmids would have to be obtained. Therefore, the number of Ap^r p*Eco*R251 transformants in *T. brucei* and *T. congolense* libraries were probably insufficient for a full representation of the respective trypanosome genomes. However, the formulae used to calculate the number of colonies that are required for a full representation of all the DNA sequences present in the genome have their own limitations. One of the major factors that plays an important role in determining which sequences or fragments are ligated amongst DNA fragments of differing sizes is differential kinetics of ligation. This factor was demonstrated by the variable number of transformants and percentage of recombinants obtained from different ligation reactions. As a result of these limitations it is a matter of chance whether or not the gene of interest could be in one of the recombinant plasmids.

Genomic libraries were then constructed in the phagemid pUC119. The ability to select for recombinants in this vector has allowed maximisation of the number of recombinant clones for screening, as this method makes it possible for the parentals to be eliminated. The *T. brucei* and *T. congolense* libraries contained 33 000 and 27 000 recombinants, respectively. The number of recombinants in each library was higher than the calculated minimum number of recombinants required to ensure full representation of the genome of the trypanosome parasites. This is expected to increase the chances of the clone(s) of interest being present in the libraries. Of the 80 plasmid pools from the *T brucei* library that were screened, 30 pools contained fragments that hybridised with the probe whilst 12 of the 90 plasmid pools constituting the *T. congolense* library contained fragments that hybridised with the probe. A majority of the positive clones identified appeared to contain inserts of similar sizes ranging from two to seven kb. This reflects one of the limitations of the

cloning strategy employed, namely internal duplication of clones. Representative results from the screening of the libraries are shown in Figure 13 for *T. brucei* and Figure 14 for *T. congolense.*

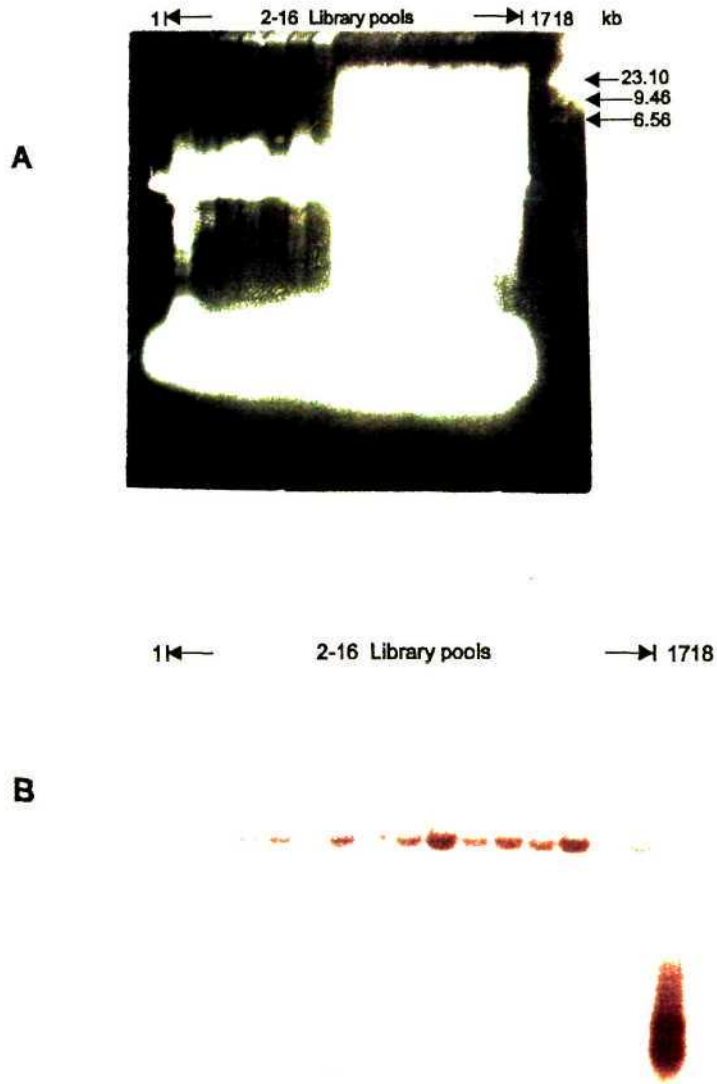


Figure 13 Screening of the *T. brucei* genomic library constructed in pUC119. Agarose gel electrophoresis of the pools of digested plasmid DNA (A) and the corresponding autoradiogram (B) following hybridisation with an [α-^{32}P]-dCTP-labelled DNA probe from *T. brucei.* Lane 1: linearised pUC119; lanes 2-16: pools of recombinant plasmids; lane 17: unlabelled DNA probe; lane 18: DNA markers.

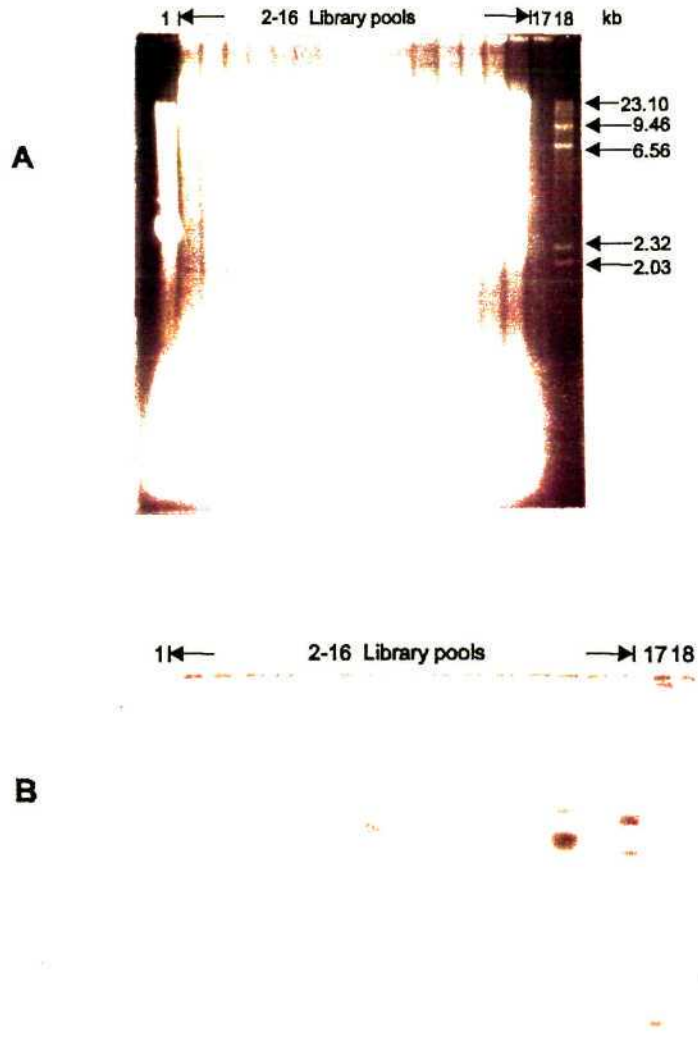


Figure 14 Screening of the *T. congolense* genomic library constructed in pUC119. Agarose gel electrophoresis of the pools of digested plasmid DNA (A) and the corresponding autoradiogram (B) following hybridisation with an [α-^{32}P]-dCTP-labelled DNA probe from *T. congolense*. Lane 1: linearised pUC119; lanes 2-16: pools of recombinant plasmids; lane 17: unlabelled DNA probe; lane 18: DNA markers.

5.6 Screening of partial genomic libraries by colony hybridisation

A partial *T. congolense* library consisting of 12 pools was screened by transformation into *E. coli* MV1184 and subsequent colony hybridisation. A total of 76 putative clones were identified and confirmed by dot blotting using a DIG-labelled DNA probe (A representative result is shown in Figure 15).

A selection of 30 individual recombinants from the dot blot were probed by Southern hybridisation using a DIG-labelled DNA probe. Twenty three putative clones contained DNA inserts of interest in the two to seven kb range (Figure 16). Initial restriction endonuclease mapping of selected clones was then undertaken with the aim of identifying clones that contained the cysteine protease gene and the upstream region.

Figure 15 Colony hybridisation of a partial *T. congolense* library constructed in pUC119. Representative autoradiograms of Magna lift nylon membrane filters showing specific hybridisation of the DNA probe from *T. congolense*, labelled with [α-^{32}P]-dCTP by nick translation, to bacterial colonies lysed *in situ.*

A partial *T. brucei* library consisting of 30 plasmid pools with clones of interest was screened by colony hybridisation. However, colony hybridisation did not reveal any individual clones that hybridised with the probe. Repeated screening of the library pools revealed clones that hybridised to the probe. As the screening of the *T. brucei* pools was reproducible, it is suggested that following transformation into the *E. coli* MV1184 cells a lethal product is produced. This could be overcome by transformation and screening into a *rec*A$^-$ strain, namely DH1 (Hanahan, 1983) or HB101. Due to time constraints it was not possible to continue with screening in another bacterial strain.

5.7 Restriction endonuclease mapping of putative clones.

Five clones, designated pCPC1, pCPC2, pCPC3, pCPC4 and pCPC5 (Figure 17), were selected for restriction mapping. Single and double restriction endonuclease digests were performed on selected clones in order to identify clones that contained the region of interest.

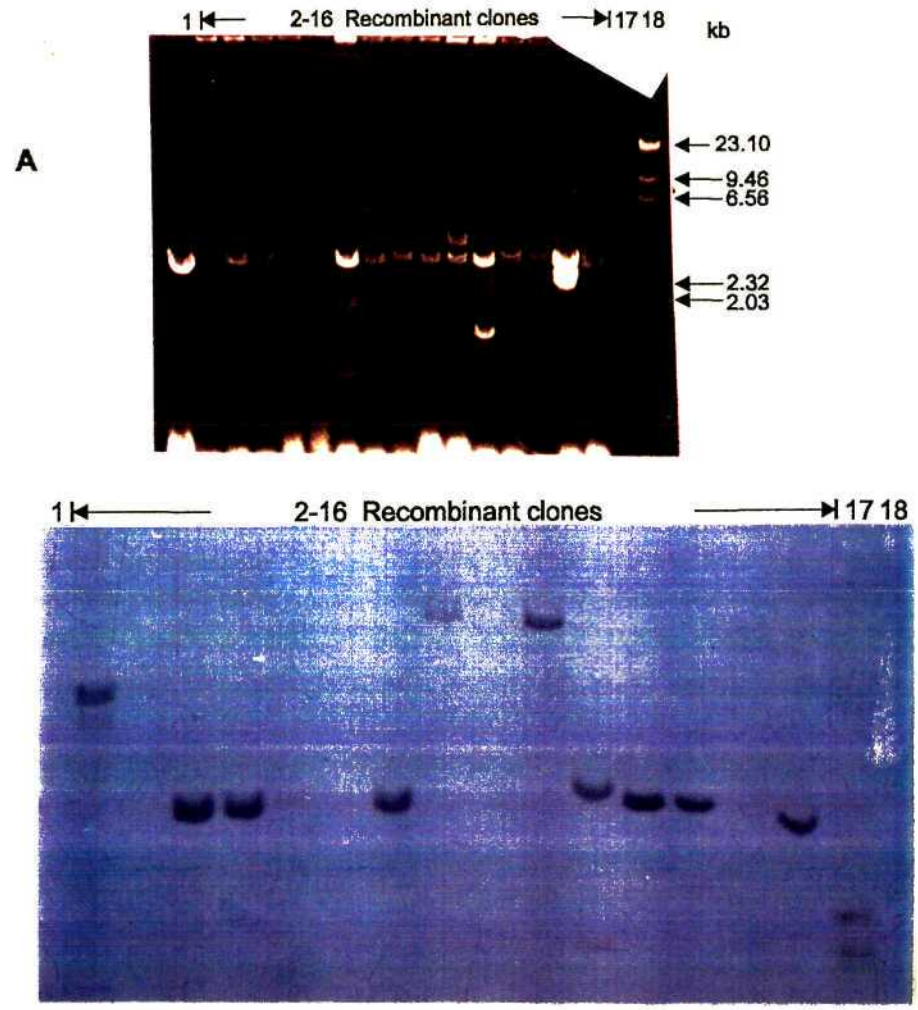


Figure 16 Southern hybridisation of putative *T. congolense* cysteine protease clones. Agarose gel electrophoresis of putative clones (A) and the corresponding Southern blot (B) following hybridisation with a DIG-labelled DNA probe from *T. congolense*. Lane 1: linearised pUC119; lanes 2-16: putative clones; lane 17: unlabelled DNA probe; lane 18: DNA markers.

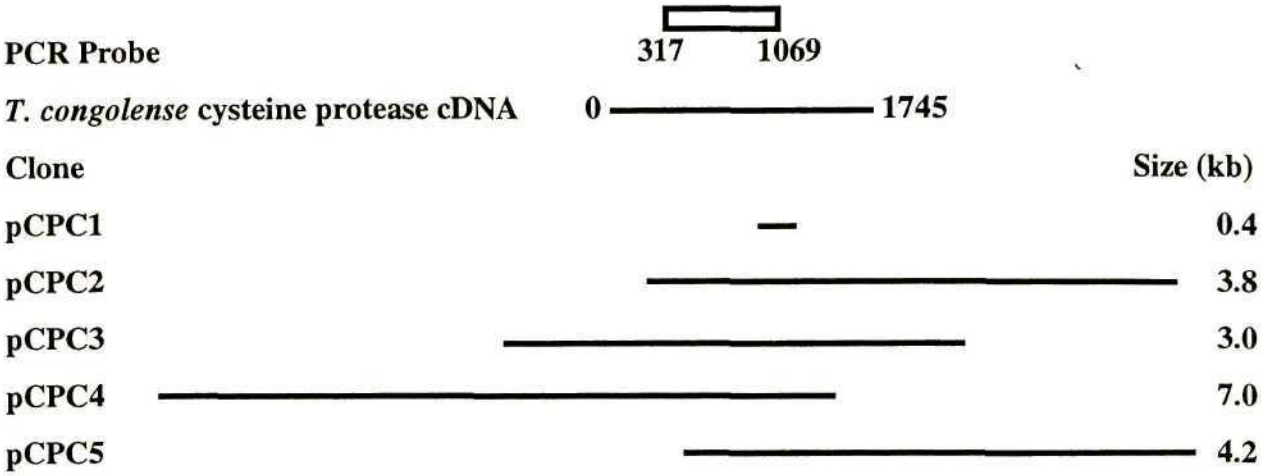


Figure 17 Initial restriction maps of restriction mapped putative clones, pCPC1, pCPC2, pCPC3, pCPC4, and pCPC5.

Plasmid pCPC1 contains an approximately 400 bp fragment internal to the cDNA sequence cDNA. Plasmids pCPC2 and pCPC5 contained approximately 3.8 kb fragment of insert DNA, containing approximately 1.5 kb of the coding sequence and approximately 1.6 kb downstream of the stop codon. These clones lacked the 5' end and the upstream region of the cysteine protease gene. Plasmid pCPC3 contained approximately a 3.0 kb fragment of insert DNA made up of a 0.7 kb fragment upstream of the transcriptional start, the entire cysteine protease cDNA gene fragment (1.7 kb) and 0.6 kb downstream from the stop codon. This clone has a very short region upstream of the transcriptional start and as such was not pursued any further. Plasmid pCPC4 contained approximately 7.0 kb insert DNA with approximately 5.6 kb upstream of the transcriptional start of the cysteine protease gene and most of the coding region (1.4 kb) of the cDNA sequence. This clone contains a long region upstream of the transcriptional start which could enable the identification of the regulatory regions (promoter elements). The restriction endonuclease map of pCPC4 is shown in Figure 18. However, DNA sequence analysis of the insert DNA will confirm the restriction mapping data.

5.8 Project achievements

Trypanosoma brucei and *T. congolense* genomic libraries were constructed in the phagemid pUC 119. A clone containing a seven kb cysteine protease gene fragment was successfully identified and isolated from the *T. congolense* library. A restriction map of this clone was subsequently established. Putative pools were produced repeatedly from the screening of

the *T. brucei* library. However, it was not possible to identify individual putative clones following transformation into the *E. coli* MV1184 strain and colony hybridisation.

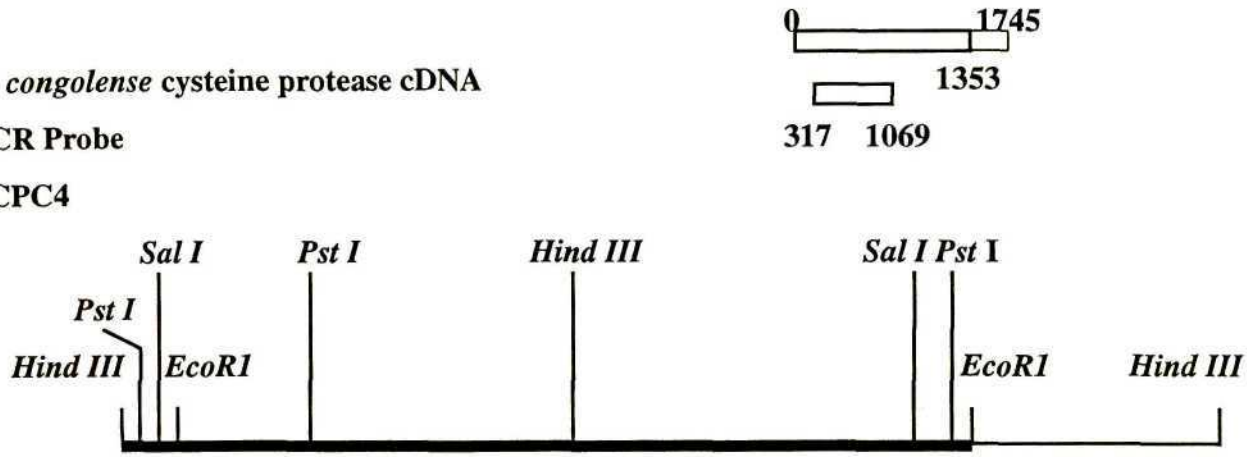


Figure 18 Partial restriction map of pCPC4. Thick line: insert DNA; thin line: vector; dotted box: portion of cDNA sequence that was not cloned into the vector.

5.9 Future Research

The cloned *T. congolense* cysteine protease DNA fragment needs to be sequenced and the resultant data used to confirm the restriction map of the clone. For *T. brucei* two options could be followed. Firstly, if the gene of interest is not amongst the cloned fragments, the cloning strategy could be repeated in order to include the gene of interest. On the other hand, if the gene of interest is unstable when propagated in *E. coli* strain MV1184, the constructed library could be propagated in a different *E. coli* strain (e.g. a *rec*A$^-$ strain, namely DH1 or HB101) in which the recombinant plasmid containing the gene of interest could be stably maintained.

Now that a clone containing the region upstream of the *T. congolense* cysteine protease gene has been isolated, essential regions could be identified by site directed mutagenesis using reporter genes by constructing promoter probe vectors. It will then be possible to investigate the structure and function of the promoter elements of the cysteine protease gene. The sequence data, structure and function of these promoters could also be compared to eukaryotic promoters and also against the promoters that have been identified, to date, in trypanosomes.

REFERENCES

Alecio, M. R., Dann, M. L. and Lowe, G. (1974) The specificity of the S_1' subsite of papain. *Biochemical Journal.* **141,** 495-501

Arnheim, N. and Erlich, H. (1992) Polymerase chain reaction strategy. *Annual Review of Biochemistry.* **61,** 131-156

Asbóth, B, Majer, Z. and Polgár, L. (1988) Cysteine proteases: the S_2P_2 hydrogen bond is more important for catalysis than is the analogous S_1P_1 bond. *Federation of the European Biochemical Societies Letters.* **233,** 339-341

Aslanidis, C. and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Research.* **18,** 6069-6074

Åslund, L., Henriksson, J., Campetella, O., Frasch, A. C. C., Pettersson, V. and Cazzulo, J. J. (1991) The C-terminal extension of the major cysteine proteinase (Cruzipain) from *Trypanosoma cruzi. Molecular and Biochemical Parasitology.* **45,** 345-348

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

Barrett, A. J. and McDonald, J. K. (1980) Mammalian proteases: A Glossary and Bibliography. Vol 1. Academic Press. London, New York, Toronto, Sydney, San Francisco

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

Barrett, A. J. (1986a) An introduction to the proteinases: In "Proteinase inhibitors". (Barrett, A. J. and Salvesen, G., Eds.). pp 3-22. Elsevier Science Publishers. Amsterdam.

Barrett, A. J. (1986b) The classes of proteolytic enzymes: In "Plant proteolytic enzymes". (Dalling, M. J. Ed.). Vol. 1. pp 1-16. CRC Inc. Boca Raton, Florida.

Barrett, A. J. and Rawlings, N. D. (1993) The many evolutionary lines of peptidases. In "Innovations in proteases and their inhibitors". pp 13-30. Walter de Gruyter. Berlin, New York.

Barrett, A. J. and Rawlings, N. D. (1995) Perspectives in biochemistry and biophysics. *Archives of Biochemistry and Biophysics.* **318,** 247-250

Beals, T. P. and Boothroyd, J. C. (1992) Genome organisation and context of a trypanosome variant surface glycoprotein gene family. *Journal of Molecular Biology.* **225,** 961-971

Berkner, K. L. and Folk, W. R. (1977) Polynucleotide kinase exchange reaction. Quantitative assay for restriction endonuclease-generated 5'-phosphoryl termini in DNAs. *Journal of Biological Chemistry.* **252,** 3176

Berger, A. and Schechter, I. (1970) Mapping the active site of papain with the aid of peptide substrates and inhibitors. *Phylosophical Transactions. Royal Society of London Britain.* **257,** 249-264

Bevan, M. (1984) Binary *Agrobacterium* vectors for plant transformation. *Nucleic Acid Research.* **12,** 8711-8721

Beveridge, A. J. (1996) A theoretical study of the active site of papain and S195C rat trypsin: Implications for the low reactivity of mutant serine proteinases. *Protein Science.* **5,** 1355-1365

Blin, N. and Stafford, D. W. (1976) A general method for isolation of high molecular weight DNA from eukaryotes. *Nucleic Acids Research.* **3,** 2303

Bolivar, F. and Backman, K. (1979) Plasmids of *Escherichia coli* as cloning vectors. *Methods in Enzymology.* **68,** 245

Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. and Boyer, H. W. (1977) Construction and characterisation of new cloning vehicles II. A multipurpose cloning system. *Gene.* **2,** 95-113

Boothryod, J. C. (1985) Antigenic variation in African trypanosomes. *Annual Review of Microbiology.* **39,** 475-502

Brisson, N. (1984) Expression of a bacterial gene in plants by using a viral vector: a cloning experiment with caulimovirus. *Nature.* **310,** 511-514

Broach, J. R. (1982) The yeast 2 μm circle. *Cell.* **28,** 203-204

Brown, T. A. (1995) Gene cloning: an introduction. (3rd Edn.). Chapman & Hall. London. Glasgow, Weinheim, New York, Tokyo, Melbourne and Madras

Burke, D. T., Carle, F. G. and Olson, M. V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science.* **236,** 806-812

Burnett, G. F. (1970) Control by insecticides: General considerations: In "The African trypanosomiases", (Mulligan, H. W., Ed.). pp 464-489. Allen and Unwin, London.

Clark, J. M. (1988) Novel non-templated addition reactions catalysed by prokaryotic and eukaryotic DNA polymerases. *Nucleic Acids Research.* **16,** 1677-9686

Clarke, L. and Carbon, J. (1976) A colony bank containing synthetic Col E1 hybrid plasmids representative of the entire *E. coli* genome. *Cell.* **9,** 91-99

Clayton, C. E., Fueri, J. P., Itzhaki, J. E., Bellofatto, V., Sherman, D. R., Wisdom, G. S., Vijayasarathy, S. and Mowatt, M. R. (1990) Transcription of the procyclic acidic

repetitive protein genes of *Trypanosoma brucei. Molecular and Cellular Biology.* **10,** 3036-3047

Cohen, J. (1996) Protease inhibitors: A tale of two companies. *Science.* **272,** 1882-1883

Collins, J. and Hohn, B. (1978) Cosmids: A type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. *Proceedings of the National Academy of Sciences. U. S. A.* **75,** 4242-4246

Cregg, J. M., Vedvick, T. S. and Raschke, W. C. (1993) Recent advances in the expression of foreign genes in *Pichia pastoris. Biotechnology.* **11,** 905-910

Cross, G. A. M. (1990) Cellular and genetic aspects of antigenic variation in trypanosomes. *Annual Review of Immunology.* **8,** 83-110

Darget, M. and Ehrlich, S. D. (1979) Prolonged incubation in calcium chloride improves the competence of *Eschericia coli* cells. *Gene.* **6,** 23-28

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

Dehrmann, F. M., Elliott, E. and Dennison, C. (1996) Reductive activation markedly increases the stability of cathepsins B and L to extracellular ionic conditions. *Biological Chemistry Hoppe-Seyler.* **377,** 391-394

Dell, A. (1996) Exploring the maze of parasite glycoconjugates. *The Biochemist.* February/March, 4-10

Detzen, G., Bellard, M. Sassone-Corsi, P. and Chambon, P. (1981). A reliable method for the recovery of DNA fragments from agarose and acrylamide gels. *Analytical Biochemistry.* **112,** 295-298

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. and Wolthers, B. G. (1968) Structure of papain. *Nature.* **218,** 929-932

Drenth, J., Jansonius, J. N., Koekoek, R., Sluyterman, L. A. A. and Wolthers, B. G. (1970) Cysteine proteinases. The structure of the papain molecule. *Philosophical Transactions. Royal Society of London.* **257,** 231-236

Drenth, J., Jansonius, J. N., Koekoek, R. and Wolthers, B. G. (1971) Papain. X-Ray structure: In "The enzymes. Hydrolysis: peptide bonds". (Boyer, P. D., Ed.). 3rd Edn. Vol. 3. pp 465-499. Academic Press. New York, London.

Duvall, E. J., Williams, D. M., Mongkolsuk, S. and Lovett, P. S. (1984) Regulatory regions that control expression of two chloramphenicol-inducible *cat* genes in *Bacillus subtilis. Journal of Bacteriology.* **158,** 784-790

Eakin, A. E., Bouvier, J., Sakanari, J. A., Craik, C. S. and McKerrow, J. H. (1990) Amplification and sequencing of genomic DNA fragments encoding cysteine

proteases from protozoan parasites. *Molecular and Biochemical Parasitology.* **39,** 1-8

Fairlamb, A. (1996) Pathways to drug discovery. *The Biochemist.* February/March, 11-16

Feinberg, A. P. and Vogelslein, B. (1983) A technique for labelling DNA restriction endonuclease fragments to high specific activity. *Analytical Biochemistry.* **132,** 6-13

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

Ford, J. (1970) The geographical distribution of *Glossina*: In "The African Trypanosomiases" (Mulligan, H. W. Ed.). pp 274-297. Allen and Unwin, London.

Ford, J. (1971) Trypanosomiases: A continental diseases: In "The role of trypanosomiases in African ecology". pp 466-487. Clarendon Press, Oxford.

Ford, J. (1971) Trypanosomes and trypanosomiases in animals and in men: "In the role of Trypanosomiases in African Ecology". pp 59-90. Clarendon Press, Oxford.

Gordon, R. M. and Willett, K. C. (1958) Studies on the deposition, migration, and development of the blood forms of the trypanosomes belonging to the *Trypanosoma brucei* group. III. The development of *Trypanosoma rhodesiense* from the metacyclic forms as observed in mammalian tissue and in culture. *Annals of Tropical Medicine and Parasitology.* **52,** 346

Gottesdiener, K., Garcia-Anoveros, J., Lee, M. G. S. and Van der Ploeg, L. H. T. (1990) Chromosome organisation of the protozoan *Trypanosoma brucei. Molecular and Cellular Biology* **10,** 6079-6083

Gottesdiener, K., Chung, H.-M., Brown, S. D., Lee, M. G.-S. and Van der Ploeg, L. H. T. (1991) Characterization of VSG gene expression site promoters and promoter-associated DNA rearrangement events. *Molecular and Cellular Biology.* **11,** 2467-2480

Grab, D. J., Ito, S., Kara, U. K. A. and Rovis, L. (1984) Glycosyltransferase activities in Golgi complex and endoplasmic reticulum fractions isolated from African trypanosomes. *The Journal of Cell Biology.* **99,** 569-577

Graham, F. L. (1990) Adenoviruses as expression vectors and recombinant vaccines. *Trends in Biotechnology.* **8,** 20-25

Grubler, U. and Hoffman, B. J. (1983) A simple and very efficient method for generating cDNA libraries. *Gene.* **25,** 263-269

Hajduk, S. L., Cameron, C. R., Barry, J. D. and Vickerman, K. (1981) Antigenic variation in cyclically transmitted *Trypanosoma brucei.* Variable antigen type composition of metacyclic trypanosome populations from the salivary glands of *Glossina morsitans. Parasitology.* **83,** 595-607

Hajduk, S. L., Hager, K. M. and Esko J. D. (1994) Human high density lipoprotein killing of African trypanosomes. *Annual Review of Microbiology.* **48,** 139-162

Hanahan, D. (1983) Studies on transformation of *Eschericia coli lacZ* to the cytochrome c gene of *Sacchoromyces cerevisiae. Proceedings of the National Academy of Sciences. U. S. A.* **78,** 2199

Hartley, B. S. (1960) Proteolytic enzymes. *Annual Review of Biochemistry.* **29,** 45-72

Hershfield, V., Boyer, H. W., Yanofsky, C., Lovett, M. A. and Helinski, D. R. (1974) Plasmid ColE1 as a molecular vehicle for cloning and amplification of DNA. *Proceedings of the National Academy of Science. U. S. A.* **71,** 3445-3459

Hoare, C. A. (1970) The mammalian Trypanosomes of Africa: In "The African trypanosomiases". (Mulligan, H. W. Ed.), pp 16-19. Allen and Unwin, London.

Hoare, C. A. (1972) The trypanosomes of mammals: A zoological monograph. Blackwell Scientific Publications. Oxford and Edinburgh.

Ido, E., Han, H. P., Kezdy, F. J. and Tang J. (1991) Kinetic studies of human immunodeficiency virus type 1 protease and its active-site hydrogen bond mutant A28S. *Journal of Biological Chemistry.* **266,** 24359-24366

International Laboratory for Research on Animal Diseases. (1991). ILRAD 1990: *Annual Report of the International Laboratory for Research on Animal Diseases.* Nairobi: International Laboratory for Research on Animal Diseases. pp 1-6

International Laboratory for Research on Animal Diseases. (1994). ILRAD 1993/4: *Annual Report of the International Laboratory for Research on Animal Diseases.* Nairobi: International Laboratory for Research in Animal Diseases. pp 21-30

Ish-Horowicz, D. and Burke, J. F. (1981). Rapid and efficient cosmid cloning. *Nucleic Acids Research.* **9,** 2989-2998

Jackson, D. G. and Voorheis, H. P. (1985) Release of the variable surface coat glycoprotein from *Trypanosoma brucei* requires the cleavage of a phosphate ester. *The Journal of Biological Chemistry.* **260,** 5179-5183

Kamphuis, I. G., Drenth, J. and Baker, E. N. (1985) Thiol proteases. Comparative studies based on the high-resolution structures of papain and actinidin, and on amino acid sequence information for cathepsins B and H, and stem bromelain. *Journal of Molecular Biology.* **182,** 317-329

Koenig-Martin, E., Yamage, M. and Roditi, I. (1992) A procyclin-associated gene in *Trypanosoma brucei* encodes a polypeptide related to ESAG 6 and 7 proteins. *Molecular and Biochemical Parasitology.* **55,** 135-146

Koga, H., Yamada, H., Nishimura, Y., Kato, K. and Imoto, T. (1990) Comparative study on specificities of rat cathepsin L and papain: Amino acid differences at substrate-binding sites are involved in their specificities. *Journal of Biochemistry.* **108,** 976-982

Koga, H., Yamada, H., Nishimura, Y., Kato, K. and Imoto, T. (1991) Multiple proteolytic action of rat liver cathepsin B: Specificities and pH-dependences of the endo- and exopeptidase activities. *Journal of Biochemistry.* **110,** 179-188

Kreier, J. P. and Baker, J. R. (1987) Parasitic Protozoa. pp 1-89. Allen and Unwin. Wellington, London, Sydney.

Lanham, S. M. and Godfrey, D. G. (1970) Isolation of salivarian trypanosomes from man and other mammals using DEAE-Cellulose. *Experimental Parasitology.* **28,** 521-534

Lonsdale-Eccles, J. D. and Mpimbaza, G. W. N. (1986) Thiol-dependent proteases of African trypanosomes: Analysis by electrophoresis in sodium dodecyl sulphate/polyacrylamide gels co-polymerized with fibrinogen. *European Journal of Biochemistry.* **155,** 469-473

Lonsdale-Eccles, J. D. and Grab, D. J. (1987) Lysosomal and non-lysosomal peptidyl hydrolases of the bloodstream forms of *Trypanosoma brucei brucei. European Journal of Biochemistry.* **169,** 467-475

Lonsdale-Eccles, J. D., Grab, D. J., Russo, D. C. W. and Webster, P. (1993) Endocytosis by *Trypanosoma brucei*: proteases and their possible role in the degradation of anti-VSG antibodies. In "Molecular mechanisms of membrane traffic". NATO ASI Series. (Marre, D. J., Howell, K. E., and Bergeron, J. J. M. Eds.). Vol. 74. pp 315-316 Springer-Verlag. Berlin, Heidelberg.

Lowe, G. and Williams, A. (1965) Papain-catalysed hydrolysis of some hippuric esters. *Biochemical Journal.* **96,** 199-204

Luckow, V. A. and Summers, M. D. (1988) Trends in the development of baculovirus expression vectors. *Biological Technology.* **6,** 47

Mason, P. J. and Williams, J. G. (1985) Hybridisation in the analysis of recombinant DNA: In "Nucleic acid hybridisation: A practical approach". (Hames, B. D. And Higgins, S. J. Eds.). pp 113-137. IRL Press, Oxford, Washington DC.

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

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

McKerrow, J. H. (1989). Parasite Proteases. *Experimental Parasitology.* **68,** 111-115

McKerrow, J. H. (1993a) The proteases and pathogenicity of parasitic protozoa. *Annual Review of Microbiology.* **47,** 821-853

McKerrow, J. H. (1993b) The diversity of protease function in parasitic infections. *Journal of the Brazilian Association for the Advancement of Science.* **45,** 319-321

Ménard, R., Khouri, H. E., Plouffe, C., Laflamme, P., Dupras, R., Vernet, T., Tessier, D. C., Thomas, D. Y. and Storer, A. C. (1991) Importance of hydrogen-bonding interactions involving the side chain of Asp158 in the catalytic mechanism of papain. *Biochemistry.* **30,** 5531-5538

Ménard, R. (1993) Structure-function studies in the papain family of cysteine proteases. *Journal of the Brazilian Association for the Advancement of Science.* **45,** 292-298

Messing, J. (1983) New M13 vectors for cloning. *Methods in Enzymology.* Academic Press **101,** 20-31

Miller, J. H. (1972) Experiments in molecular genetics. Cold Spring Harbor Laboratory. Cold Spring Harbor, New York.

Molyneux, D. H. and Ashford, R. W. (1983) The biology of trypanosoma and leishmania, parasites of man and domestic animals. pp 3-160. Taylor and Francis, London.

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

Muñoz-Jordán, Davies K. P. and Cross, G. A. M. (1996) Stable expression of mosaic coats of variant surface glycoproteins in *Trypanosoma brucei. Science.* **272,** 1795-1797

Murray, M., Clifford, D. J., Gettinby, G., Snow, W. F. and McIntyre, W. I. M. (1981) Susceptibility to African trypanosomiasis of N'Dama and Zebu cattle in an area of *Glossina morsitans submorsitans* challenge. *The Veterinary Record.* **109,** 503-510

Myler, P. J. (1993) Molecular variation in trypanosomes. *Acta tropica.* **53,** 205-225

North, M. J. (1982) Comparative biochemistry of the proteinases of eukaryotic microorganisms. *Microbiological Reviews.* **46,** 308-340

Norrander, J., Kempe, T. and Messing, J. (1983) Construction of improved M13 vectors using oligonucleotide-directed mutagenesis. *Gene.* **26,** 101-106

Ozaki, L. S. and Cseko, Y. M. T. (1984) Genomic DNA cloning and related techniques: In "Genes and antigens of parasites: A laboratory manual". (Morel, C. M. Ed.), 2nd Edn. pp 165-186, Fundação Oswaldo Cruz, Rio de Janeiro.

Pamer, E. G., Davis, C. E., Eakin, A. and So, M. (1990) Cloning and sequencing of the cysteine protease cDNA from *Trypanosoma brucei rhodesiense. Nucleic Acid Research.* **18,** 6141

Pamer, E. G., Davis, C. E. and So, M. (1991) Expression and deletion analysis of the *Trypanosoma brucei rhodesiense* cysteine protease in *Eschericia coli. Infection and Immunity.* **59,** 1074-1078

Pamer, E. G., So, M. and Davis, C. E. (1989) Identification of a developmentally regulated cysteine protease of *Trypanosoma brucei. Molecular and Biochemical Parasitology.* **33,** 27-32

Paszkowski, J. (1984) Direct gene transfer to plants. *EMBO Journal.* **3,** 2717-2722

Pays, E. (1985) Antigenic variability of African trypanosomes. *Research Immunology.* **136,** 163-166

Pays, E. (1988) DNA recombinations and transposition in trypanosomes. In "Transposition". *Society for General Microbiology Symposium.* (Kingsman, A. J., Kingsman, S. M., and Chater, K. F. Eds.), pp 301-342, Cambridge University Press, Cambridge.

Pays, E. (1993) Genome organisation and control of gene expression in trypanosomatids: In The eukaryotic genome organisation and regulation, (Broda, P M A, Oliver, S G and Sims, P F G, Eds.), pp 127-160, Cambridge University Press, Cambridge.

Pays, E., Vanhamme, L. and Berberof, M. (1994) Genetic controls for the expression of surface antigens in African trypanosomes, *Annual Review of Microbiology.* **48,** 25-52

Potts, W. H. (1970) An appraisal of methods for the control of *Glossina*: In "The African trypanomiases", (Mulligan, H W, ed), pp 564-574. Allen and Unwin, London.

Ratner, M. (1989) Protein expression in yeast. *Biotechnology.* **7,** 1129-1133

Rawlings, N. D. and Barrett, A. J. (1994) Families of serine peptidases. *Methods in Enzymology.* **244,** 19-69

Rigby, P. W. J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) Labelling of deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *Journal of Molecular Biology.* **113,** 237-251

Ritz, H. (1916) Uber rezidive bei experimetaeller. *Archiv für schiffs- u. Tropenhygiene.* **20,** 397-420

Roose, J. P. and Van Noorden, C. J. F. (1995) Synthetic protease inhibitors: Promising compounds to arrest pathobiologic processes. *Journal of Laboratory and Clinical Medicine.* **125,** 433-441

Rothe, M. and Dodt, J. (1992) Studies on the aminopeptidase activity of rat cathepsin H. *European Journal of Biochemistry.* **210,** 759-764

Russo, D. C. W., Grab, D. J., Lonsdale-Eccles, J. D., Shaw, M. K. and Williams, D. J. L. (1994) Directional movement of variable surface glycoprotein-antibody complexes in *Trypanosoma brucei. European Journal of Cell Biology.* **62,** 432-441

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anaemia. *Science.* **230,** 1350-1354

Sakanari, J. A., Staunton, C. E., Eakin, A. E., Craik, C. S. and McKerrow, J. H. (1989) Serine proteases from nematode and protozoan parasites: Isolation of sequence

homologs using generic molecular probes. *Proceedings of the National Academy of Science. U. S. A.* **86,** 4863-4867

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory. Cold Spring, New York.

Sambrook, J., Hanahan, D., Rodgers, L. and Gething, M. -J. (1986) Expression of human tissue-type plasminogen activator from lytic virus vectors and in established cell lines. *Molecular and Biological Medicine.* **3,** 459

Scharf, S. J., Horn, G. T. and Erlich, H. A. (1986) Direct cloning and sequence analysis of enzymatically amplified genomic sequences. *Science.* **233,** 76-78

Scharpe, S., De Meester, I., Hendriks, D., Vanhoof, G., van Sande, M. and Vriend, G. (1991) Proteases and their inhibitors: today and tomorrow. *Biochemie.* **73,** 121-126

Schechter, I. and Berger, A. (1967) On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications.* **27,** 157-162

Schechter, I. and Berger, A. (1968) On the active site of proteases. III. Mapping the active site of papain; specific peptide inhibitors of papain. *Biochemical and Biophysical Research Communications.* **32,** 898-902

Shapiro, T. A. and Englund, P. T. (1995) The structure and replication of kinetoplast DNA. *Annual Review of Microbiology.* **49,** 117-143

Silhavy, T. J. and Beckwith, J. R. (1985) Uses of *lac* fusions for the study of biological problems. *Microbiological Reviews.* **49,** 398-418

Simmons, V., Knight, R. H. and Humphries, K. C. (1964) Separation of trypanosomes from rat blood components. 10th Meeting. International Scientific Committee for Trypanosomiasis Research. Kampala. C.C.T.A. London. pp 81-82

Smith, A. B., Esko, J. D. and Hajduk, S. L. (1995) Killing of trypanosomes by the human haptoglobin-related protein. *Science.* **268,** 284-286

Smith, D. (1996) Unusual mechanisms of gene expression. *The Biochemist.* February/March 21-25

Smith, E. L. and Kimmel, J. R. (1960) Papain: In "The enzymes". (Boyer, P. D., Lardy, H., and Myrbäck, K. Eds.) Vol. 4. pp 133-173. Academic Press. New York and London.

Smith, G. E. and Summers, M. D. (1980) The bidirectional transfer of DNA and RNA to nitrocellulose or diazobenzylomethyl-paper. *Analytical Biochemistry.* **109,** 123-129

Smith, G. E., Summers, M. D. and Fraser, M. J. (1983) Production of human beta interferon in insect cells infected with a baculovirus expression vector. *Molecular and Cellular Biology.* **3,** 2156-2165

Sommer, J. M. and Wang, C. C. (1994) Targeting proteins to the glycosomes of African trypanosomes. *Annual Review of Microbiology.* **48,** 105-138

Southern, E. M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology.* **98,** 503-517

Sreekrishna, K., Nelles, L., Potenz, R., Cruze, J., Mazzaferro, P., Fish, W., Fuke, M., Holden, K., Phelps, D., Wood, P. and Parker, K. (1989) High-level expression, purification, and characterization of recombinant human tumor necrosis factor synthesized in the methylotrophic yeast *Pichia pastoris. Biochemistry.* **28,** 4117-4125

Stephen, L. E. (1986) Trypanosomiasis: A veterinary perspective. (1st Edn.) Pergamon Press, New York.

Sugrue, R. J., Almond, N., Kitchin, P., Richardson, S. M. and Wilderspin, A. F. (1994) Purification of crystallizable recombinant SIVmac251-32H proteinase. *Protein Expression and Purification.* **5,** 76-83

Takahashi, T., Dehdorani, A. H. and Tang, J. (1988) Porcine spleen cathepsin H hydrolyses oligopeptides solely by aminopeptidase activity. *Journal of Biological Chemistry.* **263,** 10952-10957

Tang, J. (1979) Evolution in the structure and function of carboxyl proteases. *Molecular and Cellular Biochemistry.* **26,** 93-109

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

Turk, V. and Bode, W. (1993) Lysosomal cysteine proteases and their inhibitors, cystatins: In "Innovations in proteases and their inhibitors". pp 161-178. Walter de Gruyter. Berlin and New York.

Ullman, A., Jacob, F. and Monod, J. (1967) Characterisation by *in vitro* complementation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Eschericia coli. Journal of Biological Chemistry.* **24,** 339-343

Vallee, B. L. and Auld, D. S. (1990) Active-site zinc ligands and activated H_2O of zinc enzymes. *Proceedings of the National Academy of Sciences. U. S. A.* **87,** 220-224

Vamer, D. H. and Leder, P. (1979) Expression of the chromosomal mouse β-maj-globin gene cloned in SV40. *Nature.* **281,** 35-40

Van der Ploeg, L. H. T., Schwartz, D. C., Cantor, C. R. and Borst, P. (1984) Antigenic variation in *Trypanosoma brucei,* analysed by electrophoretic separation of chromosome sized DNA molecules. *Cell.* **37,** 77-84

Vickerman, K. (1970) Ultrastructure of *Trypanosoma* and relation to function: In "The African trypanosomiases", (Mulligan, H. W., Ed). pp 60-66. Allen and Unwin, London.

Vickerman, K. (1985) Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin.* **41,** 105-114

Vieira, J. and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods in Enzymology.* **153,** 3

Webster, P. (1989) Endocytosis by African trypanosomes. I. Three-dimensional structure of the endocytic organelles in *T. brucei* and *T. congolense. European Journal of Cell Biology.* **49,** 295-302

Webster, P. and Fish, W. R. (1989) Endocytosis by African trypanosomes. II. Occurence in different life-cycle stages and intracellular sorting. *European Journal of Cell Biology.* **49,** 303-310

Webster, P. and Russel, D. G. (1993) The flagellar pocket of trypanosomatids. *Parasitology Today.* **9,** 201-206

Weitz, B. G. F. (1970) Infection and resistance: In "The African trypanosomiases", (Mulligan, H. W., Ed.). pp 97-124. Allen and Unwin, London.

Wellde, B., Lötzsch, R., Deindl, G., Sadun, E., Williams, J. and Warui, G. (1974) *Trypanosoma congolense* I. Clinical observations of experimentally infected cattle. *Experimental parasitology.* **36,** 6-19

Wiederanders, B., Brömme, D., Kirschke, H., von. Figura, K., Schmidt, B. and Peters, C. (1992) Phylogenetic conservation of cysteine proteinases. Cloning and expression of a cDNA coding for human cathepsin S. *The Journal of Biological Chemistry.* **267,** 13708-13713

Wilcox, D. and Man, R. W. (1992) Inhibitors of cysteine proteinases in lysosomes and whole cells. *Biochemistry Journal.* **285,** 495-502

Williams, J. G., North, M. J. and Mahbubani, H. (1985) A developmentally regulated cysteine proteinase in *D. discoideum. EMBO Journal.* **4,** 999-1006

Williamson, J. and Cover, B. (1966) Separation of blood-cell-free trypanosomes and malaria parasites on a sucrose gradient. *Transactions of the Royal Society of Tropical Medicine and Hygiene.* **60,** 425-427

Wondrak, E. M., Louis, J. M., Rocquigny, H., Chermann, J. C. and Roques, B. P. (1993) The gag precursor contains a specific HIV-1 protease cleavage site between the NC (P7) and P1 proteins. *Federation of the European Biochemical Societies Letters.* **333,** 21-24

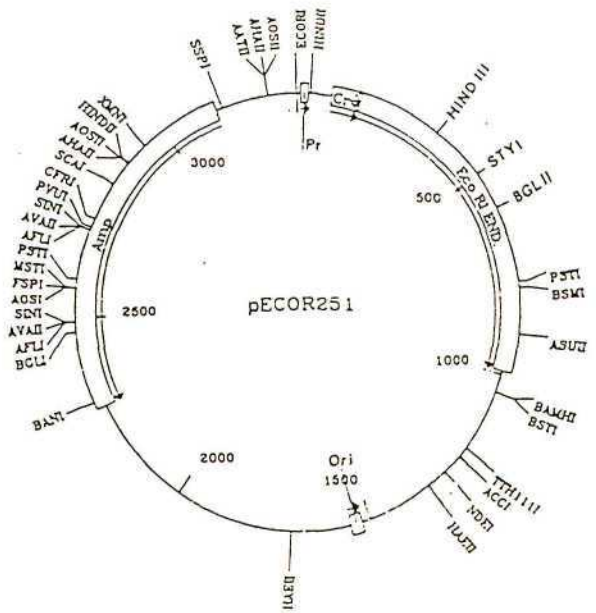
Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene.* **33,** 103-119

Young, R. A. and Davis, R. W. (1983) Efficient isolation of genes using antibody probes. *Proceedings of the National Academy of Sciences. U. S. A.* **80,** 1194-1198

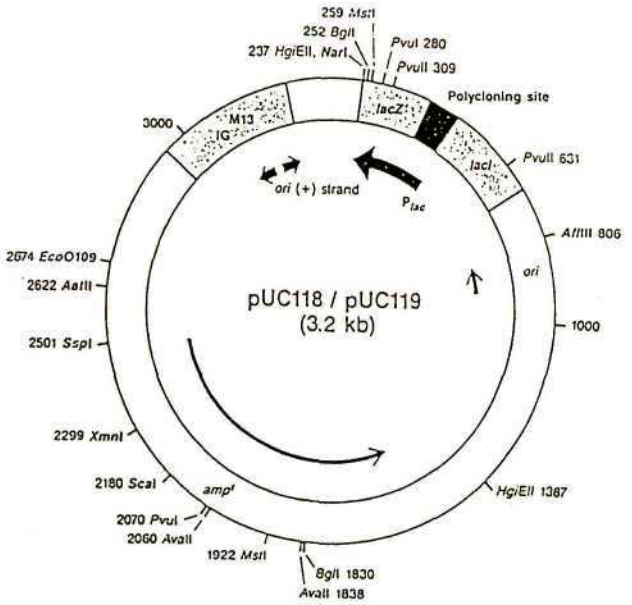
Ziegelbauer, K., Stahl, B., Karas, M., Stierhof, Y-D. and Overath, P. (1993). Proteolytic release of cell surface proteins during differentiation of *Trypanopsoma brucei. Biochemistry.* **32,** 3737-3742

APPENDIX A

RESTRICTION MAPS OF CLONING VECTORS



Restriction map of p*Eco*R251 (Remaut *et al.*, 1983).



Restriction map of pUC119 (Messing *et al.*, 1987).

APPENDIX B

BUFFERS, SOLUTIONS AND MEDIA

All buffers, solutions and media were prepared as outlined in Sambrook *et al.* (1989) unless stated otherwise.

Ampicillin stock solution

Stock solutions of Ap 25 mg/ml were prepared in 20 ml of distilled water (dH_2O), filter sterilised and stored in aliquots at -20°C.

DEAE-cellulose anion exchanger

100 g pre-swollen DEAE-cellulose (Whatman DE-52) was suspended in 500 ml PSG (Appendix B) buffer, pH 8, and adjusted to pH 8.0 with 1 M phosphoric acid. The resin was washed three times with PSG buffer to remove the fines, filtered in a Buchner funnel and resuspended in PSG buffer. The pH of the slurry was adjusted to pH 8 and stored at 4°C until required.

Denatured Herring sperm DNA

Herring sperm DNA (Boehringer Mannheim, Germany) was resuspended in dH_2O at a concentration of 10 mg/ml. The DNA was sheared by passage through an 18-gauge hypodermic needle. The sheared DNA was boiled for 10 min, chilled immediately on ice and stored frozen at -20°C in 1 ml aliquots. The DNA was boiled for 5 min and chilled on ice prior to use.

Denaturing solution (colony hybridisation; Southern transfer of DNA)

NaCl	1.5 M
NaOH	0.5 M

87.66 g of NaCl and 20.00 g of NaOH pellets were made up to 1 l.

Denhardt's solution (50 x)

Ficoll	1 % (w/v)
Polyvinylpyrrolidone-40	1 % (w/v)
BSA	1 % (w/v)

2 g of Ficoll, 2 g of Polyvinylpyrrolidone-40 and 2 g BSA were made up to 200 ml, the solution filter sterilised and stored at -20°C.

DIG alkaline phosphatase-conjugate

Antibody conjugate (750 U/ml) was diluted 1:5000 with DIG detection buffer 2 to give a final concentration of 150 U/ml. Diluted antibody-conjugate solutions were prepared just before use.

DIG colour substrate solution

Colour substrate solution was prepared just before use by combining 45 µl NBT solution (75 mg/ml), 35 µl X-phosphate solution (50 mg/ml) in 10 ml of DIG detection buffer 3.

DIG detection buffer 1

Tris-HCl	100 mM
NaCl	150 mM

100 ml of 1 M Tris-HCl buffer, pH 8.0 and 8.766 g of NaCl were made up in 800 ml of dH_2O, pH adjusted to 7.5 with 1 M NaOH and the solution made up to 1 l and stored at RT.

DIG detection buffer 2

Blocking reagent	1 %

5 g of blocking reagent was made up to 500 ml of DIG detection buffer 1, 1-2 h before use, by heating at 68°C.

DIG detection buffer 3

Tris-HCl	100 mM
NaCl	100 mM
$MgCl_2$	50 mM

20 ml of 1 M Tris-HCl buffer pH 8.0, 1.17 g of NaCl and 2.03 g of $MgCl_2$ were dissolved in 160 ml of dH_2O, adjusted to pH 9.5 with 1 M NaOH and made up to 200 ml. The solution was stored at RT.

DIG hybridisation buffer

SSC	5 X
N-lauroylsarcosine	0.1% (w/v)
SDS	0.02% (w/v)
Blocking reagent	1% (w/v)

125 ml of 20 X SSC, 0.5 g of the N-laurosylsarcosine, 0.1 g of SDS and 5.0 g of the blocking reagent were made in 500 ml of dH_2O, 1-2 h before use by dissolving at 68°C.

DIG wash solution 1

SSC	5 X
SDS	0.1%(w/v)

50 ml of 20 X SSC and 0.5 g of SDS were made up to 500 ml.

DIG wash solution 2

SSC	0.1 X
SDS	0.1% (w/v)

2.5 ml of 20 X SSC solution and 0.5 g of SDS were made up to 500 ml.

DNA sample loading buffer

Bromophenol blue	0.25% (w/v)
Glycerol	50.00% (v/v)
EDTA	0.15% (w/v)

0.025 g bromophenol blue, 5.1 ml of 99% (w/v) glycerol and 0.1 ml of 0.5 M EDTA pH 8.0 were made up to 10 ml and stored at RT.

EDTA (0.5M, pH 8.0)

93.05 g of EDTA was made up in 400 ml of dH_2O, adjusted to pH 8.0 using 1 M NaOH, and made up to 500 ml, sterilised by autoclaving and stored at RT.

Ethidium bromide

10 mg of ethidium bromide was made up in 1 ml of dH_2O and stored in the dark at RT.

Glucose (1M)

19.817 g of glucose were dissolved in 100 ml of dH_2O, sterilised by autoclaving and stored at RT.

High salt elution buffer

Tris-HCl (pH 8.0)	50.00 mM
NaCl	1.00 mM
EDTA (pH 8.0)	10.00 mM

5 ml of 1 M Tris-HCl pH 8.0, 2 ml of 0.5 M EDTA pH 8.0 and 5.844 g of NaCl were made up to 100 ml, sterilised by autoclaving and stored at RT.

Hybridisation solution for radioactively labelled probes

SSC	6 X
Denhardt's reagent	5 X

SDS	0.5% (w/v)

150 ml of 20 X SSC solution, 50 ml of Denhardt's reagent, 25 ml of 10% (w/v) SDS and 5 ml of denatured herring sperm DNA were made up to 500 ml with sterile dH_2O, before use.

IPTG (0.2 g/ml)

2 g of IPTG were dissolved in 10 ml of dH_2O. The solution was sterilised by filtration, split into 1.5 ml aliquots and stored at -20^{0}C.

Low salt elution buffer

Tris-HCl	50.00 mM
NaCl	0.15 mM
EDTA	10.00 mM

5 ml of 1 M Tris-HCl pH 8.0, 2 ml of 0.5 M EDTA pH 8.0 and 0.8766 g of NaCl were made up in 100 ml, sterilised by autoclaving and stored at RT.

Luria agar (LA)

10 g of bacto tryptone, 5 g of yeast extract, 5 g of NaCl and 15 g of bacto agar were made up in 1 l with dH_2O and the solution was sterilised by autoclaving. After cooling the plates were poured.

Luria medium (Luria broth) (LB)

10 g of bacto tryptone, 5 g of yeast extract and 5 g of NaCl were made in 1 l of dH_2O, split into appropriate aliquots, sterilised by autoclaving and cooled at RT.

NaCl (1 M)

5.844 g of NaCl were made up in 100 ml, sterilised by autoclaving and stored at RT.

NaOH (1 M)

5.61 g of NaOH pellets were made up in 100 ml and stored at RT.

Neutralising solution (colony hybridisation; Southern transfer of DNA)

Tris-base	1.0 M
NaCl	1.5 M

121.10 g Tris-base and 87.66g NaCl were dissolved in 800 ml of dH_2O, adjusted to pH 8.0 using cHCl and made up to 1.

Phenol:chloroform (1:1)

Equal vol. of phenol equilibrated with 0.1 M Tris-HCl, pH 8.0, and chloroform were mixed in a dark bottle and left to separate at 4°C. The bottom phenol phase was used.

Phosphate buffered saline glucose (PSG; Lanham and Godfrey, 1970)

Na_2HPO_4	57 mM
$NaH_2PO_4.2H_2O$	3 mM
NaCl	46 mM
Glucose	56 mM

8.088 g Na_2HPO_4, 0.467 g $NaH_2PO_4.2H_2O$ and 2.500 g NaCl were dissolved in 800 ml of dH_2O, adjusted to pH 8.0 using 1 M H_3PO_4 and made up to 1 l. 11 g of glucose was added just before use, and adjusted to pH 8.0.

Potassium acetate

49.1 g of potassium acetate was dissolved in 100 ml and used immediately.

Prewashing solution (colony hybridisation)

SSC	5 X
SDS	0.5% (w/v)
EDTA	1.0 mM

50 ml of 20 X SSC solution, 10 ml of 10% SDS and 400 Tl of 0.5 M EDTA pH 8.0 were made up to 200 ml.

Proteinase K (20 mg/ml)

2 g of proteinase K was made up in 10 ml with sterile dH_2O and was split into 1.5 ml aliquots and stored at -20°C until use.

SDS (10%)

20 g of SDS was made up to 200 ml and stored at RT.

SE buffer

NaCl	0.15 M
EDTA	0.10 M

1.5 ml of 1 M NaCl, 2.0 ml of 0.5 M EDTA pH 8.0 and 50 Tl of proteinase K were made up in 10 ml with sterile dH_2O just before use.

SET buffer (4 X; colony hybridisation)

Tris-HCl pH 8	80.0 M
NaCl	0.6 M
EDTA	4.0 mM

3.504 g of NaCl, 800 Tl of EDTA pH 8.0 and 8 ml 1M Tris-HCl pH 8.0 were made up in 100 ml and sterilised by autoclaving.

Solution I

Glucose	50 mM
Tris-HCl	25 mM
EDTA	10 mM

500 Tl of 1M glucose solution, 250 Tl of 1 M Tris-HCl pH 8.0 solution and 200 µl of 0.5 M EDTA pH 8.0 were made up to 10 ml with sterile dH_2O using a sterile flame technique. The solution was made fresh before use.

Solution II

NaOH	0.2 M
SDS	1.0% (w/v)

2 ml of 1 M NaOH and 1 ml of 10% SDS were made up to 10 ml with sterile dH_2O using a sterile flame technique.

Solution III (3 M with respect to potassium and 5 M with respect to acetate)

5 M Potassium acetate pH 4.8	60 ml
Glacial acetic acid	11.5 ml
Distilled water	28.5 ml

The constituents were made up to 100 ml with dH_2O, sterilised by autoclaving and stored at 4°C.

SSC (20 X)

NaCl	3.0 M
Sodium citrate	0.3 M

175.3 g of NaCl and 88.2 g of sodium citrate.$2H_2O$ were dissolved in 800 ml, adjusted to pH 7.0 with acetic acid, made up to 1 l, sterilised by autoclaving and stored at RT until use.

Sucrose solution (density gradient) (10-40%)

Sucrose	25% (w/v)
NaCl	1000 mM
Tris-HCl (pH 8)	20 mM
EDTA	5 mM

15.00 g of sucrose crystals, 3.51 g of NaCl, 1.20 ml of 1 M Tris-HCl pH 8.0 and 600 µl of 0.5 M EDTA were made up to 60 ml with sterile dH_2O and used immediately.

TE buffer

Tris-HCl (pH 8.0)	10 mM
EDTA (pH 8.0)	1 mM

5 ml of 1 M Tris-HCl pH 8.0 and 1 ml of 0.5 M EDTA pH 8.0 were made up in 400 ml, adjusted to pH 8.0 using NaOH, made up to 500 ml, sterilised by autoclaving and stored at RT.

TE equilibrated phenol

Phenol (Associated Chemicals Enterprise) was melted at 65^0 C. 8-Hydroxyquinoline was added to a final concentration of 0.1% (w/v) and the phenol equilibrated against 0.5 M Tris-HCl buffer, pH 8.0, at RT. The phenol was extracted 3 times with 0.1 M Tris-HCl, pH 8.0., Finally, 0.1 vol. of 0.1 M Tris-HCl pH 8.0 was added and the phenol solution stored in the dark at 4°C.

Tris-acetate buffer (TAE; 50 X)

Tris-base	242.0 g
Glacial acetic acid	57.1 ml
EDTA	37.2 g

242.0 g of Tris-base, 37.2 g of EDTA and 57.1 ml of glacial acetic acid were made up to 1 l, sterilised by autoclaving and stored at RT. The solution was diluted to I X for use.

Tris-HCl buffer (1 M; pH 8.0)

121.1 g of Tris-base was dissolved in 800 ml, adjusted to pH 8.0 using concentrated HCl and cooled to RT. The solution was made up to 1 l and sterilised by autoclaving.

X-gal (20 mg/ml)

200 mg of X-gal was dissolved in 10 ml of dimethylformamide. The solution was split into 1.5 ml aliquots and stored at -20°C in the dark.

APPENDIX C

GENERAL DNA TECHNIQUES

C.1 ISOLATION OF TRYPANOSOMES

Trypanosomes were isolated from rat blood at the peak of parasitaemia. Parasitaemia was monitored by light microscopic examination of an aliquot of blood, obtained from the tail vein of the infected rat. The approximate number of parasites in the blood sample was estimated from the number of observed trypanosomes per microscopic field (X 40 objective) of a wet blood film.

Infected rats were bled by cardiac puncture, under ether anaesthesia using a 10 ml syringe containing 1 ml of 2% (w/v) sodium citrate as an anticoagulant. Trypanosome infected whole blood was mixed with an equal vol. of Percoll. After mixing, the blood-Percoll mixture was centrifuged (25400 x g; 30 min; 4°C) in an angle rotor (30°/40°; Beckman J2-HS). Trypanosomes which form a band near the top of the gradient were collected and mixed with an equal vol. of phosphate buffered saline containing glucose (PSG) pH 8.0 (Appendix B). The cells were pelleted by centrifugation (3020 x g; 20 min; 10°C) in a Beckman J2-HS, JA-20 rotor. The pellet was resuspended in approximately 5 ml PSG, pH 8.0. The resuspended cells were passed through a DEAE-cellulose (Appendix B) column and parasites were eluted using PSG, pH 8.0. The eluted cells were pelleted by centrifugation (594 x g; 10 min; 10°C) using a Sigma 3K20 centrifuge. The motility of the parasites was constantly checked by microscopic examination. Parasite number was determined using a Neubauer haemocytometer. Parasites cells isolated using this method proved to be suitable for genomic DNA isolation.

C.2 EXTRACTION OF *E. coli* PLASMID DNA

C.2.1 Small scale ("miniprep") extraction: *E. coli* strains containing plasmids were grown in LB (Appendix B) containing Ap (50 µg/ml; Appendix B) to maintain selective pressure. Plasmid DNA was isolated from 1.5 ml of an overnight culture using the method of Ish-Horowicz and Burke (1981) as modified by Sambrook *et al.* (1989). Bacterial cells, in an eppendorf tube, were pelleted by centrifugation (10 000 x g; 1 min; RT) in a microfuge, resuspended in 100 µl of solution I (Appendix B) and kept at RT for 5 min. 200 µl of solution II (Appendix B) was added, mixed gently by turning the tube end to end and chilled on ice for 5 min. 150 µl of precooled solution III (Appendix B) was added, the

solution mixed thoroughly and chilled on ice for 5 min. The precipitated protein, SDS, and chromosomal DNA was removed by centrifugation (10 000 x g; 5 min; RT) in a microfuge. The supernatant was transferred to a clean microfuge tube and extracted twice with an equal vol. of phenol:chloroform (Appendix B). 2 vol. of cold isopropanol was added to the aqueous phase, mixed thoroughly and chilled on ice for 5 min. The nucleic acids were pelleted by centrifugation (10 000 x g; 15 min; RT) in a microfuge. The resultant pellet was air dried and resuspended in 80 µl of TE buffer (Appendix B). The plasmid DNA obtained (approximately 60 ng/µl) was pure enough, as judged by agarose gel electrophoresis, for digestion by restriction endonucleases.

C.2.2 Medium scale ("midiprep") extraction: For medium scale plasmid preparation, 50 ml of an overnight culture in LB (Appendix B) was harvested. The method of Ish-Horowicz and Burke (1981) as modified by Sambrook *et al.* (1989) was used, but scaled up (5-fold). The DNA pellet obtained after the isopropanol precipitation was resuspended in 500 µl TE (Appendix B). The DNA obtained (100 ng/µl) was sufficiently pure, as determined from agarose gel electrophoresis, for digestion with restriction endonucleases.

C.2.3 Maximum scale (maxiprep) extraction: For large scale plasmid preparation, 200 ml of an overnight culture in LB (Appendix B) was harvested. The method of Ish-Horowicz and Burke (1981) as modified by Sambrook *et al.* (1989) was used, but scaled up (20-fold). The DNA obtained (150 ng/µl) after isopropanol precipitation was resuspended in 2 ml TE. The DNA obtained was sufficiently pure, as determined from agarose gel electrophoresis, for digestion with restriction endonucleases.

C.3 PCR AMPLIFICATION OF DNA

Amplification of the DNA fragments was performed by PCR as outlined in the Perkin-Elmer Cetus GeneAmp PCR Core Reagents instruction manual. Amplification reactions (100 µl) were set up in capped 1.5 ml or 0.5 ml eppendorf microcentrifuge tubes as shown in Table 4. The reaction mixture was made up to 100 µl using sterile distilled water.

The following controls were prepared: 1. No template DNA control; 2. No enzyme control; 3. No primer control. 250 ng of trypanosome genomic DNA was used as a template. The reaction mixture was overlaid with mineral oil to prevent evaporation during PCR synthesis in the Intelligent Heating Block HB01 thermocycler (Biometra). Overlaying with mineral

oil was not required when using the Perkin-Elmer GeneAmp PCR system 2400 thermal cycler.

Table 4 Preparation of reactions for the amplification of the cysteine protease genes from *T. brucei* and *T. congolense*.

Reagent	Concentration	Volume added (µl)	Final concentration
PCR buffer 10 X		10 0	1 X
dNTPs	10.0 µM	2.0	0.200 µM
$MgCl_2$	25.0 mM	10.0	2.500 mM
Primer 1	10.0 µM	2.5	0.250 µM
Primer 2	10.0 µM	2.5	0.250 µM
Template DNA	250.0 µM	1.0	2.500 ng/µl
Taq Enzyme	2.5 U/µl	1.0	0.025 U/µl

The PCR programmed cycles for DNA synthesis were as follows:

Cycle number	Temperature ° C	Time (min)	Process
1 (1X)	97	5	Melting
	50	1	Annealing
	72	1	Extension
2 (3X)	97	1	Melting
	50	1	Annealing
	72	1	Extension
3 (29X)	94	1	Melting
	50	1	Annealing
	72	1	Extension
4 (1X)	94	1	Melting
	50	1	Annealing
	72	6	Extension

C.4 PURIFICATION OF PCR DNA PRODUCTS

Mineral oil was removed from the PCR mixture by addition of an equal vol. of chloroform. The upper aqueous phase containing DNA was collected and the DNA was purified by

extraction with phenol-chloroform, precipitated with isopropanol and dissolved in 100 µl of TE buffer. The DNA was sufficiently pure for hybridisation, as judged by restriction endonuclease digestion and agarose gel electrophoresis.

C.5 RESTRICTION ENDONUCLEASE DIGESTION OF DNA

DNA was analysed employing the established techniques outlined by Sambrook *et al.* (1989). Restriction enzyme digestion of DNA was carried out using appropriate buffers as specified by the commercial suppliers (Boehringer Mannheim; Germany). Restriction digests consisted of 2 µl of 10 x restriction buffer, variable amounts of DNA and 5-10 units of restriction endonuclease in a final volume of 20 µl. For bulk digests all components were scaled up accordingly. The incubation temperatures for restriction digests were as specified by commercial suppliers. Double digests were carried out simultaneously if the salt and temperature requirements were the same. Routinely, plasmid DNA digests were incubated for a minimum of 2 h and chromosomal DNA digests were incubated overnight to ensure complete digestion. Optimum conditions for partial *Sau*3A restriction digests were determined by varying the incubation time and amount of enzyme. Enzyme activity was stopped by the addition of DNA sample loading buffer, containing 0.15% (w/v) EDTA (Appendix B), and the DNA was analysed by electrophoresis. For further analysis, digested DNA was purified by phenol:chloroform (Appendix B) extraction, precipitated by isopropanol, and resuspended in TE buffer before use.

C.6 DNA GEL ELECTROPHORESIS

Electrophoresis was carried out using a horizontal gel system with 1 X Tris-acetate (TAE) buffer (Appendix B). The concentration of agarose (Sigma type II) varied from 0.8% (w/v) for the analysis of large DNA fragments to 2% (w/v) for that of smaller DNA fragments (Sambrook *et al.*, 1989). An agarose concentration of 1% (w/v) was used for routine analysis of DNA. Electrophoresis was carried out towards the anode at 10 V/cm for rapid analysis in minigels or 2.7 V/cm overnight through large gels. DNA markers used were lambda DNA digested with *Hind*III and a 1 kilobase ladder (GIBCO BRL). Gels were stained with ethidium bromide (0.5 µg/ml) and the DNA visualised using a Fotodyne 300 Series Foto UV DNA transilluminator at a wavelength of 300 nm. The gels were photographed using a Fotodyne Polaroid Camera with Polaroid type 667 film. Exposures

were of 1-5 sec duration at an aperture of f-4.5. Molecular masses were determined by measuring the mobilities of individual bands relative to the size markers.

C.7 DNA FRACTIONATION BY SUCROSE GRADIENT CENTRIFUGATION

The standard method described by Sambrook *et al.* (1989) was followed. Gradients were prepared in Beckman L8-M ultracentrifuge tubes (13.5 ml). Sucrose gradients (10-40%) were established after two cycles of freezing (-20°C) and thawing (4°C) of a 25% (w/v) sucrose solution (Appendix B). Samples (20 µg in 250 µl) were layered to the gradients and centrifuged at (115915 x g; 20 h; 15°C) in a Beckman L8-M ultracentrifuge. Fractions (250 µl) were eluted from the bottom of the tube and the DNA was then precipitated with isopropanol. The DNA pellet was resuspended in TE buffer (80 µl) and fractions were analysed by agarose gel electrophoresis.

C.8 ELUTION OF DNA FROM AGAROSE GELS

DNA fragments were purified from agarose gels by elution onto a membrane employing the method of Dretzen *et al.* (1981), as modified by Sambrook *et al.* (1989), for use of a DEAE membrane (NA-45; 0.45 µm membrane; Schleicher and Schuell; Germany). DNA fragments to be purified were resolved by electrophoresis and visualised by ethidium bromide fluorescence using long wave (300 nm) UV light. A horizontal incision was made directly ahead of the band to be purified and a piece of DEAE membrane of an appropriate size was inserted into the slit. The DEAE membrane was first activated by soaking the membrane for 5 min in 10 mM EDTA pH 8.0, followed by soaking for 5 min in 0.5 M NaOH and finally washing the membrane six times in sterile distilled water. Contamination by DNA fragments directly behind the DNA fragment to be purified was prevented by insertion of a piece of membrane directly behind the DNA band of interest. Electrophoresis was continued at 10 V/cm until the DNA was transferred onto the membrane. Transfer of the DNA was confirmed by observation on a UV transilluminator, at a wavelength of 300 nm. The membrane was washed free of contaminants in a buffer of low ionic strength (Appendix B). Finally, the DNA was eluted from the membrane in a buffer of high ionic strength (Appendix B) for a period of 15-30 min at 65°C. Total elution of DNA was verified by observation of the DEAE membrane under UV illumination. DNA in solution was purified by phenol:chloroform extraction followed by isopropanol precipitation and resuspended in TE buffer.

C.9 LIGATION OF DNA

During the construction of the genomic libraries in both parasites, a variable vector to insert molar ratio and a variety of concentrations of insert DNA molecules were used. Generally, ligation reactions contained 100-300 ng DNA, 2 µl ATP (10 mM), 2 µl 10 x ligation buffer and 1 unit of T_4 DNA ligase (Boerhinger Mannheim) in a final volume of 20 µl. Reaction mixtures were incubated at room temperature for 2-15 h.

C.10 PREPARATION OF COMPETENT *E. COLI* CELLS

Competent bacterial cells were prepared by the method of Cohen *et al.* (1972) with modifications as described by Dagert and Ehrlich (1979). An overnight preculture (10 ml) of *E. coli* cells was used to inoculate 100 ml of LB. The culture was incubated at 37°C, with aeration until early exponential phase of growth was reached (A_{600} = 0.4-0.5). The bacterial culture was chilled rapidly and all subsequent preparative procedures were carried out at 4°C. Cells were harvested by centrifugation (1649 x g; 10 min; 4°C), resuspended in 1/2 the growth vol. of cold 100 mM $CaCl_2$ solution and chilled on ice for 15 min. The cells were harvested (1649 x g; 10 min; 4°C) and resuspended in 1/100 the growth vol. of cold 100 mM $CaCl_2$ and aged for 2-4 h depending on the specific bacterial strain.

C.11 TRANSFORMATION OF COMPETENT CELLS

Competent *E. coli* cells (100 µl) were transformed with ligation reaction mixtures containing 120-140 ng of DNA. The transformation mixture was maintained at 4°C for 10 min to allow DNA to bind to the cell surface, then heated at 42°C for 3 min to induce uptake of DNA by competent cells and once again chilled at 4°C. To each sample 1 ml of LB was added and the transformed cells were incubated at 37°C with shaking for 1 h. Generally, 40-100 µl of the expression mixture was plated onto LA agar plates containing Ap as a selection pressure. The plates were incubated aerobically at 37°C overnight. The expression mixture was stored at 4°C and plated at a higher or lower density where necessary. Putative recombinants were analysed by small scale plasmid extractions (Ish-Horowic and Burke, 1981) and restriction endonuclease analysis.

C.12 TRANSFER OF DNA ONTO NYLON MEMBRANE

Transfer of DNA from agarose gels to Magna charge nylon membrane (Micron Separations Inc.) was achieved employing the method of Smith and Summers (1980) as modified by Boehringer Mannheim (Germany). After electrophoresis, DNA in the agarose gel was

denatured by soaking the gel in several vol. of 0.5 M NaOH and 1.5 M NaCl for 1 h at RT with constant shaking. The gel was neutralised by soaking in several vol. of 1 M Tris-Cl pH 8.0 and 1.5 M NaCl for 1 h at RT with constant shaking. The gel was placed inverted on a clean glass surface, wrapped with Whatman 3MM paper, in a plastic container with 10 X SSC solution. Magna charge nylon membrane was cut to the gel dimensions, soaked in 2 X SSC solution, and placed on the gel. Two pieces of Whatman paper, also cut to the gel dimensions and soaked in 2 x SSC were placed on top of the membrane. A 5 cm layer of dry paper towel was placed on top of the filter paper and a light weight (approximately 2 kg) was placed on top of the paper towel layer and transfer allowed to proceed for 12 to 24 h. The filter was soaked in 6 x SSC at room temperature for 5 min, air dried and DNA fixed by exposure to UV light at a wavelength of 300 nm for 5 min.

C.13 COLONY HYBRIDISATION

Selected plasmid pools, which contained putative clones, as established from Southern blotting, were transformed into competent *E. coli* MV1184. The expression mixture was plated out onto LA containing the Ap and plates were incubated at 37°C until the colonies were just visible to the naked eye. Colonies were transferred from the LA medium to the Magna transfer membrane disc filters by laying each dry nitro-cellulose filter carefully onto the corresponding agar surface. Filters were pricked asymmetrically in order to "key" the replica to the original plate. The filters were peeled off and laid, colony side up, on a fresh agar plate containing Ap. The replicas and original agar plates (master plates) were incubated at 37°C until the colonies were 1-2 mm in diameters. The master plates were kept at 4°C until screening was complete. Forceps were used to peel each filter off its plate and the filters were held colony side up on a stack of three sheets of 3MM paper (Whatman) saturated with denaturation solution (Appendix B). After 5 min the filters were transferred to a stack of 3MM paper saturated with neutralisation buffer (Appendix B) for 5 min. The filters were transferred to a stack of 3MM paper saturated with 4 x SET buffer (Appendix B), placed on dry 3MM paper and left at room temperature until they were dry (15-30 min). DNA was fixed by exposure to UV light for 5 min and the filters were kept at -20°C for hybridisation.

C.14 RADIOACTIVE LABELLING OF DNA PROBES

Reagents for labelling by nick translation (Rigby *et al.*, 1977) were purchased from Amersham (United Kingdom) in kit form. The probe was labelled with [α-^{32}P]dCTP (ICN Pharmaceuticals Inc.). The unlabelled DNA probe (100 ng) was denatured by heating in a boiling water bath for 10 min followed by chilling on ice. After mixing, the reaction mixture was maintained at RT for 3 h. Labelled DNA was separated from unincorporated [α-^{32}P]dCTP by selective precipitation. To achieve this, 50 µl of 4 M ammonium acetate, pH 4.5, was added to the reaction mixture. The mixture was vortexed and 200 µl of ice cold isopropanol added, mixed, and chilled for 15 min at -70°C. After warming the mixture in a water bath at 37°C for 2 min, the labelled DNA was pelleted by centrifugation for 15 min in a microfuge. The pellet was washed once in 0.5 ml of 0.67 M ammonium acetate, pH 4.5, 67% ethanol, and finally 90% ethanol by gentle shaking, centrifugation in the microfuge and aspiration. After drying the pellet, DNA was redissolved in TE buffer and used immediately. The activity of the labelled DNA was measured in a Packard Tri-Carb^{R} 1500 Liquid Scintillation Analyser.

C.15 PREPARATION OF DIGOXIGENIN-LABELLED DNA PROBES

The DIG DNA labelling and detection system was supplied by Boeringer Mannheim (Germany) in kit form. The nucleotide analogue digoxigenin-11-dUTP was incorporated into the PCR synthesised DNA fragment by a random primed labelling technique and the hybrid homologues were detected by an enzyme linked immunosorbent assay. The DNA fragment (500 ng) was denatured by heating in a boiling water bath for 10 min and rapidly cooling on ice. The DNA was maintained on ice and the following reagents were added namely, 2 µl hexanucleotide mixture, 2 µl dNTP labelling mixture, 1 µl Klenow enzyme (2 units), and the mixture was made up to 20 µl with distilled water. The reaction mixture was centrifuged briefly (20 sec) and incubated at 37°C for 12 to 16 h. The reaction was stopped by the addition of 2 µl of 0.2 M EDTA, pH 8.0. For Precipitation of DIG labelled DNA, 2.5 µl of 4 M LiCl and 75 µl of prechilled ethanol was added to the mixture. After thorough mixing the solution was stored at -20°C for 1 h and the labelled DNA was pelleted by centrifugation (10 000 x g; 15 min, RT). The pellet was washed once with 100 µl of 70% ethanol, air dried and resuspended in 50 µl TE buffer pH 7.6. The labelled probe was quantified as outlined by the suppliers and stored frozen (-20°C) until required.

C.16 HYBRIDISATION OF DNA

Standard procedures for hybridisation of Southern membranes (Sambrook *et al,* 1989) and disc filters (Mason and Williams, 1985) were followed. The prepared membranes were soaked in 6 x SSC and transferred to a sealable plastic container. Prehybridisation was carried out at 68°C for 3-6 h in an excess vol. of prehybridisation solution (Appendix B). Excess fluid was decanted, the denatured labelled probe approximately 1.6 x 10^7 cpm, was added and hybridisation allowed to proceed at 68^0C for 12-16 h. Following hybridisation with a radioactively labelled probe, the membrane was washed once with 2 x SSC containing 0.5% (w/v) SDS for 5 min at RT followed by 1 wash at RT for 15 min in 2 x SSC containing 0.1% (w/v) SDS for 15 min with occasional shaking. The wet membrane, in a sealable plastic bag, was exposed under Fuji Medical X-ray film in an Amersham X-ray Hypercassette with hyperscreen X-ray intensifying screens. Exposure was for 1-2 days at -70°C. The film was developed using Polycon A developer and fixer according to the manufacturers' (Polycon A) specifications.

Magna lift nylon membrane disc filters containing replicas of bacterial colonies were floated on the surface of a solution of 2 x SSC until they became thoroughly wetted from beneath and then submerged for 5 min. The disc filters were incubated in 200 ml of prewashing solution (Appendix B) at 50°C for 30 min. The bacterial debris was gently scraped from the surfaces of the filters using Kimwipes soaked in prewashing solution. Prehybridisation and hybridisation of the filters were performed as for the Magna charge membranes. Following hybridisation, the filters were washed two to four times (depending on the amount of activity on the membrane) in a large vol. (300-500 ml) of 2 x SSC and 0.1% SDS at room temperature The filters were then washed once or twice in an excess vol. (300-500 ml) of 1 x SSC containing 0.1% SDS at 68°C. The filters were exposed to X-ray film and developed as for Magna charge membranes.

When employing the DIG labelling and detection system, the genomic Southern blots were washed twice in wash solution I (Appendix B) for 5 min at room temperature, followed by two 15 min washes in wash solution II (Appendix B) at 68°C. Finally, the Southern blots were washed in DIG detection buffer 1 (Appendix B) for 1 min. The genomic blots were then soaked in DIG detection buffer 2 (Appendix B) for 30 min, followed by a 30 min incubation in diluted antibody-conjugate (150 U/ml; Appendix B) solution. The unbound

antibody-conjugate was removed by washing the membranes twice in detection buffer 1 for 15 min at room temperature. The membrane filters were equilibrated in detection buffer 3 (Appendix B) for 2 min, and incubated in the dark with freshly prepared colour-substrate (Appendix B) solution. The enzyme reaction was stopped by washing the membrane in detection buffer 4 (TE) for 5 min and the banding pattern was recorded by photography.

APPENDIX D AMINO ACID SYMBOLS

Amino acid	3-letter code	1-letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V