Expression of microneme genes in
Eimeria tenella

by

Rachel M. Ryan

ProQuest Number: U642940

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.
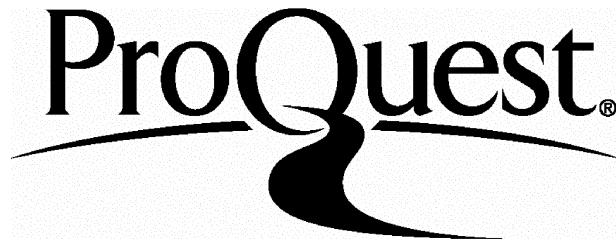
Abstract

Micronemal proteins, which are released from membrane-bound secretory organelles from invasive stages of the protozoan parasite *Eimeria tenella*, are believed to function as specialised adhesins, essential for substrate-dependent parasite motility and invasion of host cells. Little is known about the regulation of microneme protein expression, but ultrastructural studies suggest that micronemes are formed afresh during each successive zoite stage, appearing as the parasites mature.

I set out to examine one broad aspect of the biology of microneme organelle formation, *viz.* the ways in which the appearance of some microneme proteins are regulated. A starting point was the development of an experimental system (oocyst sporulation leading to the formation of invasive sporozoites) so that events occurring during the *de novo* formation of micronemes could be examined. The sporulation of oocysts proved an excellent system to investigate the timings of expression of microneme proteins at both the mRNA and mature protein level. The appearance of microneme proteins, EtMIC1-5, was initially found to be highly co-ordinated as sporozoites matured (~22.5 h). RT-PCR analyses indicated that microneme-specific mRNAs were present after 6-12 h of sporulation and the level of synchronicity observed, suggested that expression was highly regulated. The locations of the microneme-coding genes within the genome was unlikely to be responsible for the co-ordination as the genes were found on different chromosomes.

DNA sequence and 5'RACE analyses of the upstream regions of the genes *EtMIC1-5* were undertaken and initiator regions surrounding the transcriptional start sites of the genes were identified – similar sequences are present in other genes from a variety of apicomplexan parasites. Transient transfection assays of *E. tenella* undertaken to identify the minimum promoter of the genes *EtMIC1* and *2* revealed that 234 and 131 bp, respectively upstream of the ATG start codon were sufficient for expression. The initiator alone of *EtMIC1* was not capable of driving gene expression, whereas the initiator of *EtMIC2* was sufficient for basal level expression. Mutational analysis of the initiators of *EtMIC1* and *2* identified motifs and nucleotides within these motifs that make an important contribution to gene expression.

Table of Contents

Figures

Tables

Abbreviations

A	Adenosine
ADP	Adenosine 5'-diphosphate
AP	Adaptor proteins
APS	Ammonium persulphate
ARF	ADP ribosylation factor
ATP	Adenosine 5'-triphosphate
BAP	*E. coli* alkaline phosphatase
BCIP	Bromo chloro indoyl phosphate
BDM	Butanedione monoxime
BFA	Brefeldin A
BLA	*E. coli* β-lactamase
β-gal	β-galactosidase reporter gene
BSA	Bovine serum albumin
bp	Base pairs
C	Cytosine
CAT	Chloroamphenicol acetyl transferase reporter gene
CBB	Coomassie Brilliant Blue
cbEGF	Calcium binding epidermal growth factor
CD	Cytochalasin D
cDNA	DNA complementary to RNA
CS	Circumsporozoite protein
CTRP	CS and TRAP related protein
dATP	Deoxyadenosine 5'-triphosphate
DBP	Duffy-binding protein
DEPC	Diethyl pyrocarbonate-treated deionised water
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNase	Deoxynuclease
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
EBA-175	Erythrocyte-binding antigen-175 Kda
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
EM	Electron microscopy
EMSA	Electrophoretic mobility shift assay
ER	Endoplasmic reticulum
EtMICX	*E. tenella* microneme (X refers to an number)
FCS	Foetal calf serum
G	Guanine
GAG	Glycosaminoglycans
GCG	Genetics Computer Group
GFP	Green fluorescent protein
GRA	Dense granule protein
GST	Glutathione-S-transferase
HAMs	HAMs nutrient mixture-F12
HSPG	Heparin sulphate proteoglycans
IBP	Initiator binding protein
I domain	Integrin insertion domain
ISG	Immature secretory granule

Inr	Initiator region
IPTG	Isopropyl β-D-thiogalactopyranoside
IVN	Intravacuolar network
Kb	Kilobase(s) or 1000 base pairs
kDa	Kilodalton(s)
LB (borth)	Luria-Bertani (broth)
LDLR	Low-density lipoprotein receptor
Lhs	Left hand side
Mbp	Mega base pair
MDBK	Madin-Derby bovine kidney cells
MICs	Microneme proteins
MMLV-RT	Moloney murine leukemia virus reverse transcriptase
mRNA	Messenger ribonucleic acid
NBL-1	Madin-Derby bovine kidney cell line
NBT	Nitro blue tetrazolium
NcMICX	*N. caninum* microneme protein (X refers to a number)
NcMICX	*N. caninum* microneme gene (X refers to a number)
NSF	N-ethylmaleimide-sensitive fusion protein
NTP	Nucleotide triphosphate
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PM	Plasma membrane
PV	Parasitophorous vacuole
PVM	Parasitophorous vacuole membrane
Py	Pyrimidine
RACE	Rapid amplification of cDNA ends
RAP	Rhoptry associated protein
Rhs	Right hand side
RNA	Ribonucleic acid
RNase	Ribonuclease
ROP	Rhoptry protein
RT-PCR	Reverse transcription-polymerase chain reaction
SDS	Sodium dodecyl sulphate
SO	Sporulated oocysts
Spz	Sporozoite
SML	Small lectin of *S. muris*
SNAP	Soluble NSF attachment protein
SNARE	Integral membrane protein SNAP receptors
T	Thymine
Taq	*Thermus aquaticus*
TBE	Tris-borate-EDTA
TBP	TATA binding protein
TEMED	N,N,N',N'-tetramethylethylenediamine
TFII-X	Transcription factor binding protein of polymerase II (X represents a letter)
TgMICX	*T. gondii* microneme protein (X refers to a number)
TgMICX	*T. gondii* microneme gene (X refers to a number)
TGN	*Trans* Golgi network

T_m	Melting temperature
TRAP	Thrombospondin-related adhesive protein
TRAP-C1	Thrombospondin related adhesive protein-Cryptosporidium 1 of *C. parvum*
TSP	Thrombospondin
UTR	Untranscribed region
v/v	Volume per volume
vWF	von Willibrand Factor type A-domains
w/v	Weight per volume
XGal	5-bromo-4-chloro-3-indolyl β-D-galctopyranoside
5'/3' UTR	5 prime/3 prime untranscribed region
°C	Degrees Celsius

Parasites

C. parvum	*Cryptosporidium parvum*
E. acervulina	*Eimeria acervulina*
E. maxima	*Eimeria maxima*
E. nieschulzi	*Eimeria nieschulzi*
E. tenella	*Eimeria tenella*
E. histolytica	*Entamoeba histolytica*
G. lamblia	*Giardia lamblia*
N. caninum	*Neospora caninum*
P. berghei	*Plasmodium berghei*
P. cynomolgi	*Plasmodium cynomolgi*
P. falciparum	*Plasmodium falciparum*
P. knowlesi	*Plasmodium knowlesi*
P. vivax	*Plasmodium vivax*
P. yoellii	*Plasmodium yoellii*
S. muris	*Sarcocystis muris*
T. gondii	*Toxoplasma gondii*
T. parvum	*Theileria parvum*
T. vaginalis	*Trichomonas vaginalis*
T. brucei	*Trypanosoma brucei*

Bacteria

B. abortus	*Brucella abortus*
C. psittaci	*Chlamydia psittaci*
E. coli	*Escherichia coli*
L. pneumophila	*Legionella pneumophila*
S. typhimurium	*Salmonella typhimurium*

Acknowledgements

I would like to acknowledge my two supervisors Dr. Fiona Tomley and Dr. Martin Shirley for scientific support throughout my PhD. I would also like to thank my external supervisor, Conrad King and the Institute for Animal Health who funded this project.

I would also like to express my gratitude to Karen Billington for her technical expertise and training on various aspects of molecular biology (and of course all the gossip!) and Janene Bumstead for her technical advice on parasitology. I would also like to thank all the members of G1C and G2D for their support and friendship, in particular Julie Clarke, Phil Brown, Leticia Jean, Fionnadh Carroll (for the funny stories), Sarah White and Rahmat Sherrif. Thanks also to Chris Gibbons and Diane Collins in the library, Ian Martins in stores, Elaine Bennett for tissue culture, Claire Powers and Phil Green in the media department and to all the wash-up team.

Finally outside the lab, I am grateful to all the other students of my year, in particular Brian "BK" Dove, Sally "Pants" Rogers, Daniel Gliddon, Terry Gouthrie and Chris Rapier (for listening to me moan on and on!) and especially to my parents John and Catherine, all my siblings (John, Leah, Anthony, Naomi and Rebecca) and Matthew and his family for their strength, patience and encouragement.

Chapter One
Introduction

1.1 Coccidiosis and its control

Protozoa of the phylum Apicomplexa are obligate, intracellular parasites, many of which infect vertebrate hosts. A significant number of genera within this phylum are serious pathogens of man and/or domestic animals, including *Babesia, Cryptosporidium, Cyclospora, Eimeria, Neospora, Plasmodium, Sarcocystis, Theileria* and *Toxoplasma* (Table 1.1). Seven species of *Eimeria* (*E. acervulina, E. brunetti, E. maxima, E. mitis, E. necatrix, E. praecox* and *E. tenella*) parasitise the domestic chicken (*Gallus domesticus*) and cause the intestinal disease coccidiosis. Coccidiosis costs the UK poultry industry approximately £40 million per year (Williams, 1999). Clinical signs include dehydration, anaemia, haemorrhagic enteritis, bloody faeces, reduced weight gain, and heavy infections can lead to high mortality (Joyner and Long, 1974). Subclinical infections cause poor food conversion to meat, and since food comprises 70% of the costs of poultry production, the economic impact is clear.

Current control methods rely on prophylaxis with anti-coccidial drugs, which consist of ionophores and various chemicals. However, control is compromised by problems such as parasite drug resistance (Chapman, 1997), lack of new drug development and most recently the withdrawal of drugs following failure to re-register them. There are also increasing consumer concerns about the possibility of drug residues reaching unacceptable levels in meat products. Recently the breeding and egg-laying sectors of the poultry industry have successfully applied live-attenuated vaccines (Shirley, 1995a). Attenuation was achieved by serial passage of parasites through chickens to select populations that complete their endogenous life-cycles more rapidly than their wild-type parents whilst retaining full immunogenicity (Jeffers, 1975). These "precocious" lines are relatively non-pathogenic and thus suitable for use in live-attenuated vaccines. However cultivation of vaccinal parasites in chickens is time-consuming and expensive, as each batch requires validation by *in vivo* challenge experiments, which significantly increases processing time, and the shelf life is relatively short. A live-attenuated vaccine including all species of *Eimeria* that infect the chicken is probably not economically viable for the vast numbers of birds in the broiler industry (~800 million per annum in the UK) (Anon, 1999). However

Paracox5, which includes single strains of *E. tenella, E. acervulina* and *E. mitis*, and two strains of *E. maxima* has recently been introduced to the broiler market.

It seems likely that the poultry industry will rely increasingly on vaccination, either alone or in combination with chemotherapy. Long-term, sustainable control of coccidiosis is likely to depend on the identification of parasite molecules that can be developed as novel drug targets or vaccine antigens. This in turn will require the elucidation of parasite biochemical pathways and the understanding of complex host-parasite interactions at a molecular level.

1.1.1 The life-cycle of *E. tenella*

The life-cycle of *E. tenella* is homoxenous, with both exogenous and endogenous phases (Fig. 1.1), and is completed in around 12 days (Tyzzer, 1929). The products of fertilisation, the oocysts, are the most accessible stage of the life-cycle since they are shed in large numbers in the faeces. In the presence of oxygen, sporulation occurs over a period of around 2 days during which time the sporont undergoes meiosis and mitosis to produce four sporocysts, each of which contain two infective, haploid sporozoites (Fernando *et al*, 1987). After ingestion by the chicken, the oocyst wall is weakened by changes in moisture and temperature and broken by the grinding action of the gizzard, to release sporocysts into the gut lumen. The stieda body, which blocks the sporocyst opening is degraded by trypsin and, in the presence of bile salts, invasive sporozoites (Fig 1.2A) are released from the sporocysts and rapidly invade enterocytes of the superficial epithelium. From here, sporozoites are transported, apparently within host cells, across the surface epithelium and lamina propria to the glandular crypts, where they invade and develop within crypt enterocytes. The transporting cells were initially thought to be macrophages (Van Doornink and Becker, 1957) but were later identified as intestinal intraepithelial lymphocytes, IEL (Lawn and Rose, 1982; Lee and Alizzi, 1981). It is not clear whether transport within IELs is an obligatory step since histological examination of gut sections has revealed large numbers of free sporozoites within the lamina propria (Vervelde *et al*, 1995). Once inside crypt enterocytes, intracellular sporozoites round up, enlarge and divide by schizogony, a process that involves cytoplasmic expansion, multiple mitotic nuclear divisions and cytokinesis to form around 200-250 1st generation merozoites (Fig. 1.2 B). These are released into the lumen of the crypt, where they invade neighbouring enterocytes and

Phylum	*Class*	*Subclass*	Order	*Suborder*	*Family*	*Subfamily*	*Genus*
Apicomplexa	Sporozoasida	Coccidiasina	Eucoccidiorida	Eimeriorina	Eimeriidae		Caryospora
							Eimeria
							Isospora
					Cryptosporidiidae		*Cryptosporidium*
						Sarcocystinae	*Sarcocystis*
							Frenkelia
						Toxoplastinae	*Besnoitia*
							Hammondia
							Neospora
							Toxoplasma

Table 1.1 Taxonomic arrangement of coccidial parasites.

undergo a second round of schizogony to produce between 200-350 2nd generation merozoites. During early development of the 2nd generation schizont, which is the largest infective stage of the life-cycle, infected enterocytes undergo dramatic morphological changes, which are unique to *E. tenella* infection. The enterocytes become swollen and lose their adhesive connections to neighbouring uninfected cells, allowing them to drop through the basement membrane into the lamina propria (reviewed by Daszak, 1999). The resultant disruption to the epithelial cell layer and rupture of blood vessels in the soft tissue by the developing 2nd generation schizont are the major causes of the disease pathology associated with *E. tenella* infection. 2nd generation merozoites are released from the schizont and invade crypt enterocytes, either *via* the lumen or directly from the tunica propria. A final 3rd round of schizogony gives rise to the final generation of merozoites, which in turn invade enterocytes and develop into the sexual stages (McDonald and Rose, 1987). The majority of these are macrogametocytes (female), which do not proliferate but grow in size, partly due to the formation of two types of wall-forming bodies, which after fertilisation develop into the two-layered, environmentally-resistant oocyst wall. The microgametocytes (male) produce a large number of biflagellate microgametes, which are released into the lumen and invade the macrogametocytes to fertilise the macrogametes. Zygotes are then passed in the faeces as unsporulated oocysts to complete the life-cycle.

1.1.2 Genomic organisation

The nuclear genome of *E. tenella* is approximately 60 Mb in size, arranged within at least 14 chromosomes that range in size from 1 to >6 Mb, as determined by pulsed field gel electrophoresis (Shirley, 1994). The genome is approximately 50% GC rich (http://www.sanger.ac.uk/projects/E_tenella/) and has an abundance of dispersed trinucleotide (GCA/TGC) and telomeric-like heptanucleotide (TTTAGGG) repeats (Shirley, 2000). In addition to the nuclear genome, *Eimeria* species also contain a mitochondrial DNA genome, a plastid-like genome (Dunn *et al.*, 1998) and RNA viral-like molecules (Ellis and Revets, 1990; Lee *et al.*, 1996). Other apicomplexans, including *Neospora*, *Plasmodium* and *Toxoplasma*, contain plastid-like genomes (reviewed by McFadden and Roos, 1999) that comprise 35 Kb circular elements which localise, by *in situ* hybridisation, to a plastid-like organelle called the apicoplast (McFadden *et al.*, 1997; Wilson *et al.*, 1996; Kohler *et al.*, 1997). The apicoplast is hypothesised to have arisen by 'secondary endosymbiosis' whereby an apicomplexan
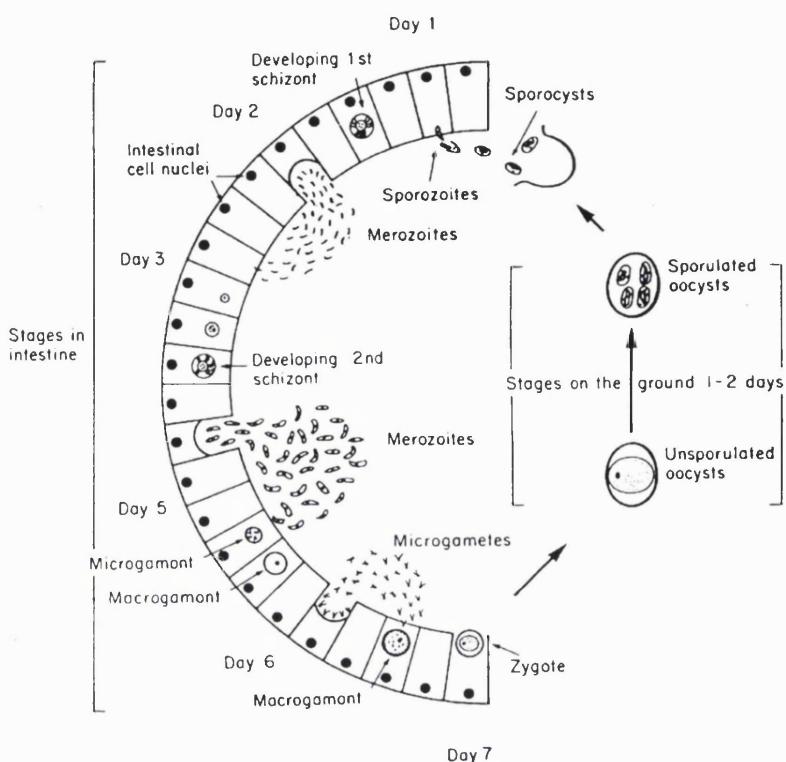
Figure 1.1 The life-cycle of *E. tenella*

In *Eimeria* species the complete life cycle, sporulation, schizogony and gametogony is completed in a single host. Reproduced from original provided by B. J. Millard

ancestor engulfed a eukaryotic alga, which in turn had already engulfed a cyanobacterial-like prokaryote. The morphology of the apicoplast, which is surrounded by four (sometimes three) membranes, supports this hypothesis (Delwiche and Palmer, 1997) and molecular phylogenetic analyses of *tuf*A, and other apicoplast genes suggest a green-algal ancestry (Blanchard, 1999). However, molecular-systematic analyses are more indicative of a red-algal endosymbiont (McFadden and Waller, 1997; McFadden *et al.*, 1997).

The apicoplast is considered a promising drug target, due to the prokaryotic nature of the metabolic pathways housed within it, such as type II fatty acid biosynthesis (Waller *et al.*, 1998). Ciprofloxacin kills *T. gondii* by blocking apicoplast–DNA replication demonstrating that the plastid is an effective target (Fichera and Roos, 1997). The activity of several herbicides and antibacterial antibiotics (e.g. rifampicin and clindamycin) against apicomplexans is thought to be because they inhibit key plastid activities. It has been shown in *Toxoplasma* and *Plasmodium* (Fichera and Roos, 1997; McConkey *et al.*, 1997) that the plastid genome is essential to the parasite, but complete plastid maps yielded little information on essential functions (Wilson *et al.*, 1996). Waller and co-workers (1998) investigated nuclear-encoded genes whose products are targeted to the apicoplast. Such proteins have a bipartite N-terminal signal, which contains a classical signal peptide and an additional hydrophobic region that is sufficient to target recombinant green fluorescent protein (GFP) to the apicoplast of *T. gondii* (Waller *et al.*, 1998). More research is required to validate the role of plastid anabolic pathways and to identify enzymes as targets for parasitidial drugs.

1.2 Invasion by apicomplexan parasites

Conventional phagocytosis was initially thought to be the entry mechanism of *T. gondii*, but it was later shown that the parasite actively penetrated non-phagocytic cells and cultured host cells (Lycke *et al.*, 1965). Jones suggested (1972) that invasion might occur by stimulated or induced phagocytosis following observations that micro-pseudopods were extended by host cells to envelop attached parasites. However, *Toxoplasma* zoites stimulated this response in HeLa cells and fibroblasts, both of which are not normally phagocytic. It was shown, with merozoites of *Plasmodium* and *Toxoplasma*, that invagination of the host cell plasma membrane (PM) occurred during invasion but that the continuity of the host cell PM was not disrupted (Ladda *et*
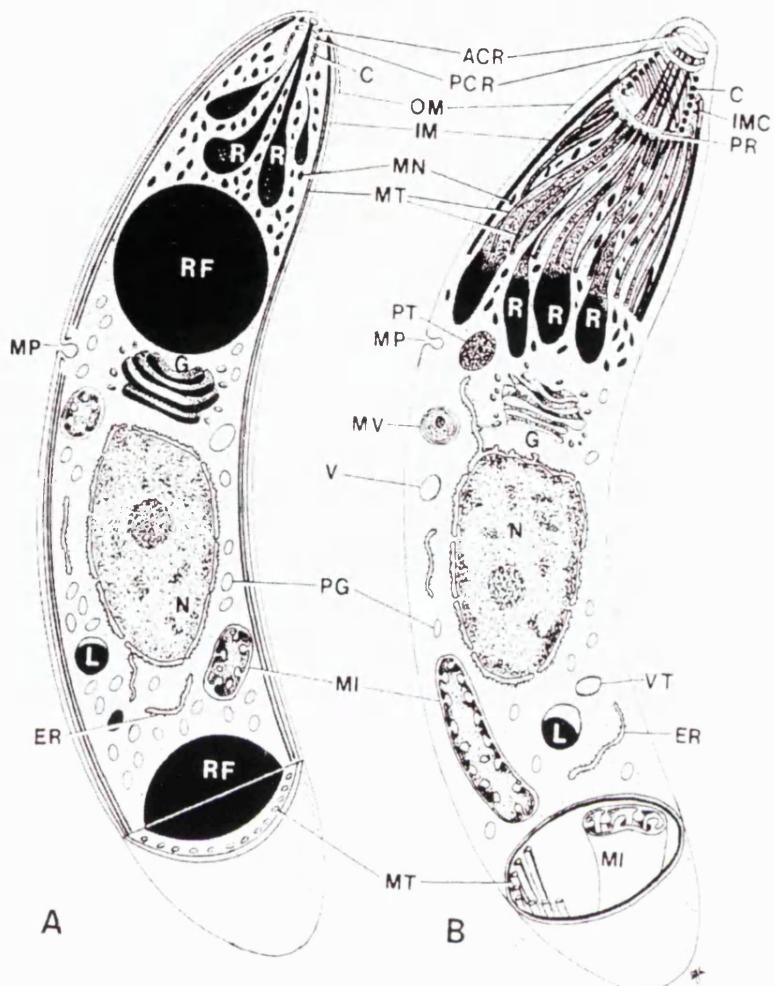
Fig. 1.2 Diagrammatic representation of an eimerian sporozoite (A) and merozoite (B), reproduced from original provided by Ball and Pittilo.

ACR	Anterior preconoidal ring	**PT**	Protein granule
PCR	Posterior precocoidal ring	**MP**	Micropore
C	Conoid	**MV**	Multimembranous vacuole
OM	Outer membrane of pellicle	**V**	Vacuole
IMC	Inner membrane complex	**G**	Golgi body
PR	Polar ring	**N**	Nucleus
IM	Inner membrane of pellicle	**PG**	Polysaccharide granule
MN	Microneme	**MI**	Mitochondrion
MT	Microtubule	**L**	Lipid
R	Rhoptry	**VT**	Vesicle with thick wall
RF	Refractile body	**ER**	Granular endoplasmic reticulum

al., 1969; Aikawa et al., 1977; Aikawa et al., 1978). Observations on the invasion of cultured cells by Eimeria and Toxoplasma zoites supported the hypothesis that parasites invaded by active penetration rather than phagocytosis (Roberts et al., 1970; Nichols and O'Connor, 1981; Danforth et al., 1992). Invasion requires live parasites, yet occurs in glutaraldehyde-fixed or heat-treated macrophages that are incapable of phagocytosis. Invading zoites orientate themselves so that the apical tip enters first, whilst parasites that are phagocytosed are internalised in random orientations. Invasion is rapid, taking 10-15 seconds, whereas phagocytosis takes several minutes. Finally, there is no evidence of membrane ruffling and little rearrangement of host-cell actin-rich microfilaments at the site of entry, unlike phagocytosis, which involves dramatic rearrangements of the host cell cytoskeleton.

The features that lead to successful host cell invasion by most apicomplexan parasites are morphologically conserved and require the involvement of the parasite apical organelles and the parasite actinomyosin system (Aikawa et al., 1978; Morisaki et al., 1995; Dubremetz, 1998a). Invasion (Fig. 1.3) is initiated by contact between the parasite and host cell surface, causing re-orientation of the parasite, so that the apical tip is in direct contact with the host cell PM. Microneme proteins are released at an early stage in the process, when the apical region of the parasite binds to the host cell (Carruthers and Sibley, 1997). As invasion proceeds, microneme proteins move backwards over the zoite surface, essentially "capping" the parasite, and are eventually released from the posterior before internalisation is complete (Tomley et al., 1996; Carruthers and Sibley, 1997). A circumferential zone of attachment is formed, the so-called "moving junction", in which host and zoite membranes remain closely opposed as the parasite moves forward into the host cell (Aikawa et al., 1978). Once the parasite becomes internalised it resides in a tight fitting intracellular or parasitophorous vacuole (PV), which remains distinct from the host cell endosomal system. Rhoptry contents are released during invasion (Saffer et al., 1992; Dubremetz et al., 1993; Carruthers and Sibley, 1997), and some rhoptry proteins probably become integrated into the expanding parasitophorous vacuole membrane (PVM). Once invasion is complete the moving junction is capped off at the posterior of the zoite and becomes part of the PVM (Dubremetz, 1998a). Once the parasite is internalised, dense granule proteins are released and are believed to be involved in modifications of the PV (Entzeroth, 1984; Carruthers and Sibley, 1997).

Although this sequence of events is generally conserved, some interesting variations exist. For example, zoites of *Theileria parva* do not undergo specific reorientation of the apical tip and invaded zoites lie free in the erythrocytic cytoplasm enclosed by an erythrocytic plasma membrane, unlike a PVM (Shaw and Tilney, 1995). In *Babesia* spp., rhoptry discharge occurs during entry into the erythrocyte and the PVM dissolves soon after invasion (Morita *et al.*, 1994). In *Sarcocystis* spp., invading merozoites escape from the primary PV and form a secondary PV (Entzeroth, 1985).

1.2.1 Parasite invasion and motility

Three types of movement have been identified in *Eimeria* sporozoites; bending and pivoting that occur intermittently with decreased frequency through the life of the sporozoite and gliding, which requires contact with the substratum (Russell and Sinden, 1981) and in which no obvious changes in cell shape occur (King, 1988). Time-lapse video microscopy of *T. gondii* zoites has shown that locomotion consists of circular gliding, upright twisting and helical rotation (Hakansson *et al.*, 1999).

It was observed that cationized ferritin binds to the entire surface of *Eimeria* sporozoites and rapidly accumulates at the posterior pole (Dubremetz and Ferreira; 1978). A similar phenomenon was observed in gregarines, using small latex beads (King, 1981). This 'capping' phenomenon, together with parasite motility is inhibited by cytochalasin, which prevents actin polymerisation (Dubremetez and Ferreira, 1978; Russell and Sinden, 1981). It was hypothesised that parasites move by adhering to the substratum *via* surface ligands and then translocating the ligand/substrate complex along the body of the zoite using a microfilament-based contractile system (Russell and Sinden, 1981; King, 1988). Thus, adhesion to a fixed substrate should lead to parasite forward movement and binding to host cell-surface receptors should result in penetration into the host cell. The observation that circumsporozoite protein of *Plasmodium* was secreted from the apical tip onto the zoite surface and moved backwards by a cytochalasin-sensitive process, before being released onto the substrate from the posterior trailing end of the gliding parasite, supported this hypothesis (Stewart and Vanderberg, 1988; Stewart and Vanderberg, 1991). Trails of material, apparently released from the posterior end of gliding zoites have also been observed in *E. nieschulzi* (Entzeroth *et al.*, 1989), *E. tenella* (Tomley *et al.*, 1996; Bumstead and Tomley, 2000), *T. gondii* (Dobrowolski and Sibley, 1996; Hakansson *et*

al., 1999), *P. falciparum,* (Spaccapelo *et al.*, 1997) *P. berghei,* (Stewart and Vanderberg, 1988) and *C. parvum* (Forney *et al.*, 1998).

Since both zoite motility and invasion of host cells is cytochalasin-sensitive, either actin polymerisation or actin filaments must be required for capping. Cytochalasins decrease sporozoite motility (Russell and Sinden, 1981; Dobrowolski and Sibley, 1996; Forney *et al.*, 1998; Hakansson *et al.*, 1999) and inhibit the translocation and release of surface molecules (Stewart and Vanderberg, 1991), as well as preventing invasion of host cells (Ryming and Remington, 1978; Miller *et al.*, 1979). Cytochalasins destabilise microfilaments in a wide range of eukaryotic organisms so it was not immediately clear whether inhibition of invasion implicated host or parasite actin microfilaments. To distinguish between these two possibilities, cytochalasin-resistant *Toxoplasma* and host cell mutants were used. Dobrowolski and Sibley (1996) showed that phagocytosis of *Salmonella typhimurium* by wild-type cells was inhibited by cytochalasin D, whereas uptake proceeds normally in cytochalasin-resistant host cells in the presence of the drug, findings that clearly demonstrated an essential role for host cell microfilaments during the internalisation of *S. typhimurium*. In contrast, invasion of wild-type and cytochalasin D-resistant cells by *T. gondii* was inhibited by cytochalasin D suggesting that an active microfilament in the host was not sufficient for internalisation of the parasite. However tachyzoites of *T. gondii*, resistant to cytochalasin D were capable of invading wild type cells in the presence of the drug, clearly demonstrating that actin-based motility in the tachyzoite, not the host cell, is critical for internalisation. This is likely to be the situation with most apicomplexan zoites. An exception is the spherical sporozoite of *Thieleria*, which has no defined apical complex and invades by circumferential zippering of the parasite and host cell membranes (reviewed by Shaw, 1997).

From a series of studies undertaken to investigate the motor system that powers apicomplexan gliding movements and invasion, the majority of actin found in *Toxoplasma* oocysts was monomeric with very little filamentous (F) actin (Dobrowolski *et al.*, 1997b). Immunoelectron microscopy of osmotically swollen *T. gondii* tachyzoites, revealed that actin was localised to the inner membrane complex, IMC (Dobrowolski *et al.*, 1997b) and was distributed similarly in *Plasmodium* merozoites (Pinder *et al.*, 1998). Shaw and Tilney (1999) were able to artificially induce the formation of F actin in *Toxoplasma* tachyzoites using jasplakinolide,

(which can polymerise and stabilise actin) but only at the apical tip, sometimes forming an apical projection, which suggested that the filamentation was not random, but was induced at specific sites. It was also observed that jasplakinolide could also inhibit gliding and invasion (Poupel and Tardieux, 1999) which suggests that rapid dissociation of actin as well as polymerisation, was required for these processes.

It is highly unlikely that actin filament assembly alone is sufficient to provide the driving force for zoite motility and invasion. Instead actin filaments probably form a scaffold for myosin motors, as proposed by King (1988). More importantly a functional actomyosin was isolated from the oocyst infective stage of *E. tenella* (Preston and King, 1992). Upon addition of ATP to the homogenised oocysts a super-precipitation was observed, typical of actomyosin solution. Furthermore, F-actin was detected by rhodamine-phalloidin staining and beaded filaments were visualised by EM.

Various methods used for identification of apicomplexan myosins include affinity-purified rabbit antiserum generated against a conserved peptide, PCR screens using generic myosin head primers and sequence data from apicomplexan genome projects [Dobrowolski *et al.*, 1997a; Heintzelman and Schwartzman, 1997; Heintzelman and Schwartzman, 1999; Hettmann *et al.*, 2000; Matuschewski *et al.*, 2001). Several apicomplexan myosins have now been cloned and sequenced, including TgM-A, -B, -C, -D and -E from tachyzoites of *T. gondii* (Heintzelman and Schwartzman, 1997; Heintzelman and Schwartzman, 1999; Hettmann *et al.*, 2000; Delbac *et al.*, 2001), PfM-A, from merozoites and ookinetes of *P. falciparum,* (Heintzelman and Schwartzman, 1997) and PbM-A and PyM-A, from sporozoites of *P. berghei* and *P. yoellii,* respectively (Matuschewski *et al.,* 2001).

All apicomplexan myosins examined so far are very similar, suggesting that the diversity of myosins in these parasites is extremely limited (Heintzelman and Schwartzman, 2001). Myosin classification is dependent upon conservation of the various head, neck and tail domains of the molecule. Apicomplexan myosins do not contain the typically conserved glycine residues and often lack the IQ motif which bind calmodulin and calmodulin-related proteins and have been put into a new class, class XIV. They. TgM-A for example, has a typical head and tail domain but no discernible neck domain (Heintzelman and Schwartzman, 1997). These unconventional myosins serve a number of cellular processes, but as yet a filament-

forming myosin (type II) has not been identified. Evidence of a potential role for myosin in gliding motility and invasion came from the use of myosin inhibitors. Butanedione monoxime (BDM), a low affinity inhibitor of myosin ATPase that blocks the action of a variety of myosins but has no effect on actin filaments, inhibited gliding motility and invasion in *T. gondii* (Dobrowolski *et al.*, 1997b); *C. parvum* (Forney *et al.*, 1998) *P. berghei* and *P. yoelii* (Matuschewski *et al.*, 2001). Another myosin light-chain kinase inhibitor, KT5926, also blocked parasite motility (Dobrowolski *et al.*, 1997a).

Endogenous TgM-A localises beneath the PM and its short tail carries the localisation determinant (Hettmann *et al.*, 2000) making myosin-A an appealing candidate for powering gliding motility. More critically there is evidence from co-localisation studies that PyM-A and PbM-A associate with the surface-associated microneme protein TRAP (thrombospondin-related adhesive protein), supporting the hypothesis that myosin plays a role in generating the force needed for gliding motility and host cell invasion, however a direct association between the two has not been identified (Matuschewski *et al.*, 2001). Recently, an inducible knock-out system has been developed for *T. gondii* and has been used to prove that TgM-A is necessary for motility (Meissner *et al.*, 2002).

There is evidence of a cortical actomyosin system in apicomplexan parasites, but its structural organisation remains to be confirmed. The motor can only be localised between the plasma membrane and the inner membrane complex (Dobrowolski and Sibley 1997c). Three-dimensional studies of the cytoskeleton of *E. acervulina* sporozoites showed microtubules extending from the conoid (polar ring) in a straight line to about half way along the zoite and that the IMC contains linear arrays of innermembranous particles, arranged longitudinally and helically following the underlying microtubules (Russell and Sinden, 1982; Dubremetz and Torpier, 1978). This rigid arrangement serves not only as a structural framework, but also provides directionality of the capping process using the linearly arranged IMPs. Also the helical, screw-like path observed during gliding matches the helically coiled microtubules, suggesting a translocation system whereby myosins would run along actin filaments (Russell and Sinden, 1982; Dubremetz and Torpier, 1978).

1.2.2 The parasitophorous vacuole

Light and electron microscopy studies showed that the PV is formed by invagination of the host cell PM as the zoite invades (Jones *et al.*, 1972; Nichlos *et al.*, 1987). The moving junction may be responsible for selectively sorting host cell PM proteins from the vacuole. Indeed, freeze fracture of *Plasmodium* merozoites during erythrocyte invasion revealed an abundance of IMPs throughout the erythrocyte membrane and at the moving junction. However the newly formed vacuole is depleted of these proteins relative to the host cell PM (Aikawa *et al.*, 1981; Porchet-Hennere and Nicolas, 1983). Despite this, many abundant host cell plasma membrane proteins are found in a newly formed PV, but are rapidly removed and remain absent from mature vacuoles (De Carvalho and De Souza, 1989). Finding that both erythrocyte proteins (McLaren *et al.*, 1977; Aikawa *et al.*, 1981; Dluzewski *et al.*, 1989) and lipids (Dluzewski *et al.*, 1992) were excluded from the PVM, led to the controversial suggestion that the PVM may be formed from compounds stored in the microneme and rhoptry organelles and secreted into the erythrocyte membrane during invasion (Bannister and Dluzewski, 1990; Joiner, 1991; Dluzewski *et al.*, 1992). Ward and co-workers (1993) investigated these findings further by monitoring the incorporation of a fluorescent lipophilic probe and phospholipid analogs into the erythrocyte membrane and followed their fate during PVM formation using low-light-level video fluorescence microscopy after invasion with *P. knowlesi* merozoites. A significant difference was observed in protein labelling but not in lipid labelling (Ward *et al.*, 1993), suggesting that proteins of the erythrocyte surface were excluded from the forming PVM but that lipids were included, demonstrating that the lipid component of the membrane was derived from the host cell PM. Other more recent findings on protein incorporation into the PVM of *T. gondii* have shown that a sorting protocol exists, whereby proteins with a transmembrane sequence were excluded but those anchored to the membrane by a glycosyl phosphatidylinositol (GPI) anchor were incorporated (Mordue *et al.*, 1999b). To further investigate the influence that protein extracellular domains have on this sorting mechanism, mutants of ICAM-1, which has five Ig-like extracellular domains, a transmembrane domain and a cytoplasmic domain, were used. Cells were transfected with wild type ICAM-1 or with a mutant lacking the cytoplasmic tail, or with a GPI-anchored version and transfectants infected with *T. gondii* tachyzoites. Only wild type ICAM-1 was excluded from incorporation into the *T. gondii* PVM demonstrating that the parasite can selectively exclude proteins at the moving junction (Mordue *et al.*,

1999a). Similar processing events are likely to occur in the formation of the PVM during invasion by *Eimeria* spp. and *Plasmodium* spp. Other data to support the PM origin of the PVM were obtained using time-resolved capacitance measurements and video microscopy. Parasites appear to enter the host cell by forcing the formation of a new vacuole, which is continuous with, but distinct from, the host cell PM. If this process involved the synthesis of novel membranes, then the total surface area of the host cell plus the vacuole would be expected to increase. However, measurement of cell capacitance, which is directly correlated to cell surface area, showed no significant increase or decrease occurred in the surface area of the host cell, during internalisation of *T. gondii* tachyzoites. This finding demonstrates that the PVM consists primarily of invaginated host cell membrane (Suss-Toby *et al.*, 1996) but the data are confusing, as the measurements did not return to the initial surface area of the host cell, suggesting that the cell does not regulate the increase in total volume by restoring the corresponding surface area. This in turn implies that the bilayer can withstand some stretching or that the parasite can modify the membrane in such a way so as to increase the surface without increasing the capacitance.

The PV formed during invasion rapidly becomes surrounded by, and associated with, host cell mitochondria and Endoplasmic reticulum (ER) as demonstrated in *T. gondii* (Sinai *et al.*, 1997), the recruitment of which is possibly a function of rhoptry proteins. This phenomenon, termed PVM-organelle association (Sinai *et al.*, 1997) has been observed in other intracellular pathogens, *Chlamydia psittaci* (Matsumoto *et al.*, 1991), *Legionella pneumophila* (Swanson and Isberg, 1995) and *Brucella abortus* (Pizarro-Cerda *et al.*, 1998) which also replicate in intracellular compartments and are ER or mitochondria-associated. Here the function of organelle association is known to be the establishment of replication-permissive compartments and is also likely to involve nutrient acquisition (Swanson and Isberg, 1995; Pizarro-Creda *et al.*, 1998; Matsumoto *et al.*, 1991). The close association between the PV, mitochondria and ER is believed to be an attempt by the parasite to obtain nutrients from the host cell, in order to maintain the high growth rate and replication. Both the mitochondria and ER have been implicated in lipid translocation (Sinai *et al.*, 1997; reviewed in Saliba and Kirk, 2001). Apicomplexa are auxotropic for purines and therefore must obtain purine-containing compounds from the host (Schwartzman and Pfefferkorn, 1982; Wang and Simashkevich, 1981). In *T. gondii* a dense granule protein NTPase, may be involved in
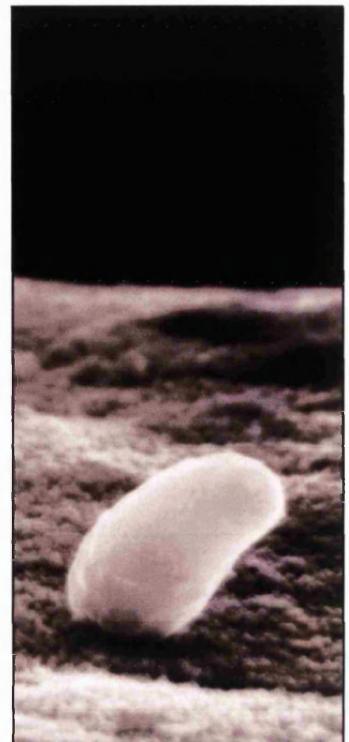
1 **2** **3**

4

Fig. 1.3 Invasion by *E. tenella* sporozoites.

Panels 1, 2 and 3: Scanning electron micrographs of sporozoites of *E. tenella* as they invade caeca. Obtained from Patricia Bland and Elaine Rose, IAH Compton.

Panel 4: Electron micrograph of transverse sections of the apical end of a sporozoite of *E. tenella* as it invades the caeca wall. The moving junction is visible as the parasite squeezes itself into the enterocyte. Obtained from Patricia Bland, Janene Bumstead, Fiona Tomley IAH Compton.

Sporozoite invasion of host cells is initiated by contact between the parasite and host cell surface, causing re-orientation of the parasite, so that its apical tip is in direct contact with the PM (1). Microneme proteins are exocytosed and a circumferential zone of attachment is formed, whereby the host and zoite membranes remain closely opposed as the parasite moves forward, creating the "moving junction" (2 and 3). Invasion continues as the moving junction moves anterior to posterior, with the parasite squeezing itself inside. Once the parasite becomes internalised it resides in a PV. Rhoptry contents are exocytosed and their material probably becomes integrated into the expanding PVM. Once invasion is complete the moving junction is capped off at the posterior of the zoite and becomes part of the PVM. Internalisation of the parasite occurs from the site of contact and proceeds rapidly, the whole process taking approximately 5-10 seconds in *Eimeria* (Dubremetz, 1998). Invasion is an active process and is dependent upon the parasites actomyosin system.

27
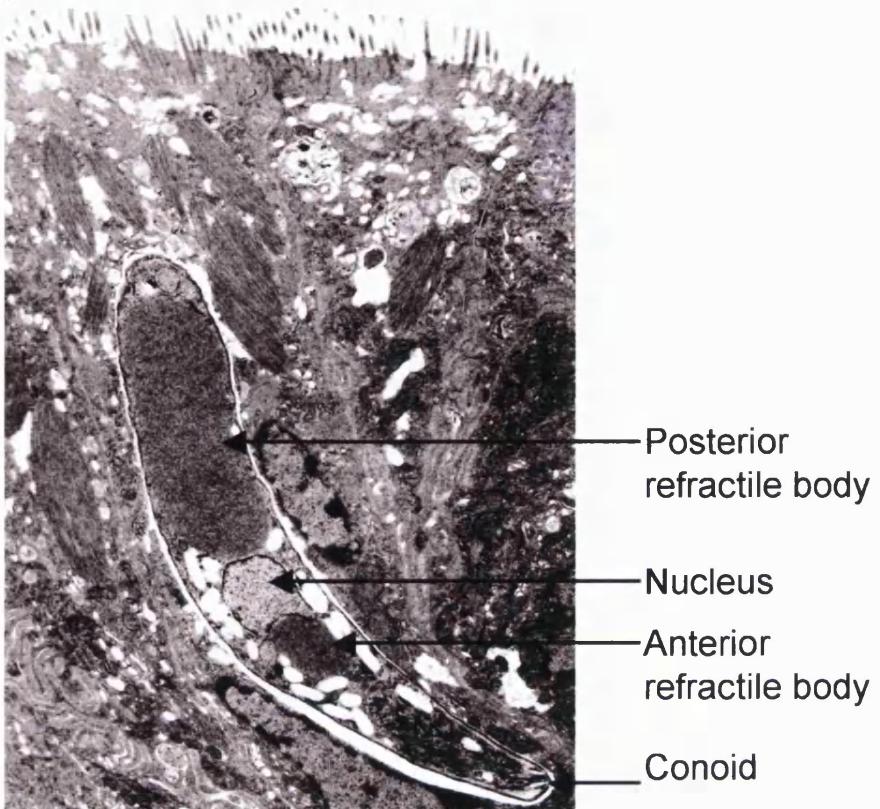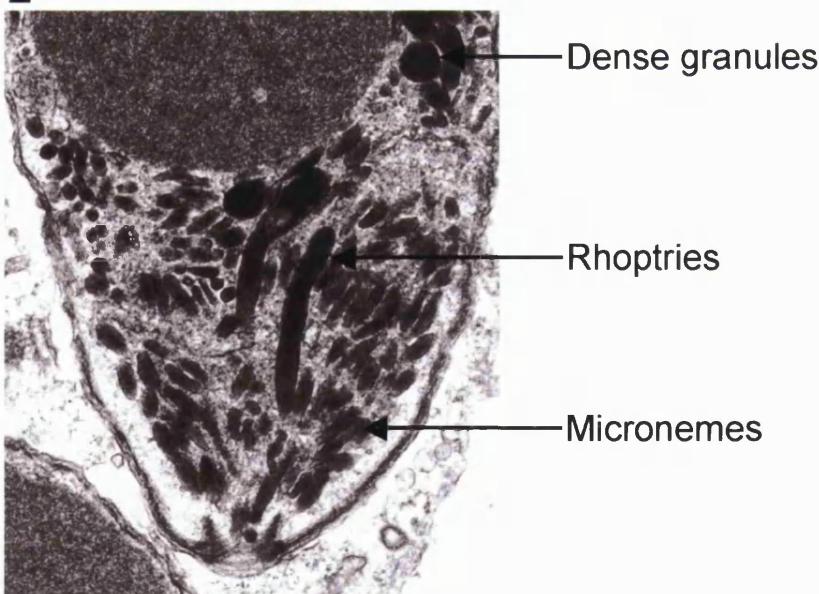
Fig. 1.4 The organelles of *E. tenella* sporozoites.
Panel 1: Scanning transmission electron micrograph of a sporozoite of *E. tenella* in the caeca of a chicken. Obtained from Patricia Bland, Janene Bumstead, Fiona Tomley IAH Compton.
Panel 2: Scanning transmission electron micrograph at higher power of the apical complex of a sporozoite of *E. tenella* in the process of invasion. Obtained from Patricia Bland, Janene Bumstead, Fiona Tomley IAH Compton

this process (Asai *et al.*, 1995). Both *Eimeria* and *Toxoplasma* can incorporate radiolabelled amino acids (Gurnett *et al.*, 1995) and nucleobases (Harris *et al.*, 1988; Schmatz *et al.*, 1986) from extracellular sources, demonstrating the presence of some transport pathways, however these have not been characterised at a functional or molecular level. Schwab (1995) demonstrated that the PVM is not just a physical barrier, but that it also functions as a molecular sieve allowing the exchange of metabolites between the parasite and the host e.g. permeable to charged and uncharged molecules of less than 1400 Da (Schwab *et al.*, 1994). Finding, when a fluorescent compound of <850 Da, was injected into an *Eimeria*-containing vacuole, the compound entered and remained in the PV, but the vacuole was not permeable to molecules above 10,000 Da (Werner-Meier and Entzeroth, 1997). The PV is also entirely segregated from the extracellular fluid as demonstrated by the exclusion of Lucifer yellow from the vacuole, whether it was added to the cells before or after invasion (Joiner *et al.*, 1990). The PV does not fuse with organelles of the host cell endocytic cascade, such as endosomes and lysosomes, again highlighting the difference between parasite invasion and phagocytosis (Mordue *et al.*, 1999b).

1.3 Apical organelles of apicomplexan parasites

1.3.1 Dense granules

The dense granules (Fig. 1.4.2) of apicomplexan parasites are secretory vesicles ranging in size from 100-300 nm (Bannister *et al.*, 1975; Tetley *et al.*, 1998) and are believed to play a role in structural modifications and maintenance of the PV (reviewed in Cesbron-Delauw, 1994; Dubremetz *et al.*, 1998b). Exocytosis from dense granules was first suggested when they were seen to rupture in *P. knowlesi* following invasion (Bannister *et al.*, 1975) and in *S. muris* (Entzeroth, 1984), *P. knowlesi* (Torii *et al.*, 1989) and *T. gondii* (Carruthers and Sibley, 1997). A second dense granule secretory event also occurs after invasion by *T. gondii*, in which membranous tubules, released from a posterior invagination of the zoite, give rise to a structure called the intravacuolar network (IVN) (Sibley *et al.*, 1995), which extends into the PVM and the host cell cytoplasm, connecting with other PVs (Dubremetz *et al.*, 1993) and incorporating dense granule proteins (Torii *et al.*, 1989; Sibley *et al.*, 1995; Lauer *et al.*, 1997).

Several dense granule proteins, GRA, have been identified in apicomplexan parasites, some of which are found as both membrane-associated and soluble forms. In *T.*

gondii, several GRAs are localised to dense granules, including GRA1-7 (Cesbron-Delauw *et al.,* 1989; Jacobs *et al.,* 1998; Lecordier *et al.,* 1993; Lecordier, 1995; Mercier *et al.,* 1993; Mevelec *et al.,* 1992; Ossario *et al.,* 1992) and two isoforms of a nucleotide triphosphate hydrolase, NTPase (Asai *et al.,* 1995; Sibley *et al.,* 1994a).

Little information is available on the composition of dense granules of *Eimeria* (Daszak *et al.,* 1993; Entzeroth *et al.,* 1993). Some caution is required when making comparisons between *Eimeria, T. gondii* and other apicomplexans as secretion of proteins from *T. gondii* dense granules occurs throughout intracellular development (Cesbron-Delauw, 1994), whereas in *Eimeria* exocytosis of dense granules appears to occur only after invasion (Dubremetz *et al.,* 1998b).

1.3.2 Rhoptries

Rhoptry organelles (Fig. 1.4.2) are membrane-bound, club-shaped secretory bodies located in the apical tip of apicomplexan parasites and are believed to play a major role in the formation of the PV. Invasive zoites contain between one and twenty rhoptries (Chobotar and Scholtyseck, 1982) with, for example, *Eimeria* and *Toxoplasma* having ~6-8, *Plasmodium* spp. having 2 (Aikawa, 1971) and *C. parvum* having just 1 rhoptry (Tetley *et al.,* 1998). Secretion of membranous material from the rhoptries *via* the apical tip of zoites has been observed in *Toxoplasma* (Nichols *et al.,* 1981; Porchet-Hennere and Nicolas, 1983; Carruthers and Sibley, 1997), *Plasmodium* (Bannister *et al.,* 1986) and *Eimeria* (Jensen and Edgar, 1976; Perkins, 1991; reviewed in Dubremetz *et al.,* 1998b).

Rhoptries have been biochemically purified from several genera and shown to contain a complex mixture of polypeptides. For example, *E. tenella* sporozoite rhoptries contain ~60 or more polypeptides (Tomley, 1994b) and many have been localised to the PV of cells infected with *E. tenella* (Greif and Entzeroth, 1996). In *P. falciparum* and *P. yoelii,* 12 rhoptry proteins have been identified, including protein complexes of 140/130/110 and 80/40 Kda (Etzion *et al.,* 1991; Khan *et al.,* 2001; Sam-Yellowe and Perkins, 1991). The larger of these two complexes is capable of interacting with the cytoplasmic face of disrupted erythrocyte membranes, which suggests that these ROPs may be translocated across the lipid bilayer during merozoite invasion (Sam-Yellowe and Perkins, 1991). Other proteins identified within the rhoptry organelle of merozoites include the rhoptry-associated proteins RAP 1, 2 and 3. These form a low molecular weight complex within the rhoptries that is thought to play a role in

invasion of erythrocytes *in vitro* (Howard *et al.*, 1998; Schofield *et al.*, 1986). Interestingly, monkeys immunised with RAP1 and 2 are partially protected against malarial challenge (Perrin *et al.*, 1985; Ridley *et al.*, 1990).

A similar number of rhoptry proteins, ROPs, has been identified in *T. gondii* (Leriche and Dubremetz, 1991). Initial evidence of rhoptry involvement in PV formation came from the identification of a ROP that enhanced the penetration of host cells by the parasite (Lycke *et al.*, 1975). Antibodies to this "penetration enhancing factor" had inhibitory effects on invasion *in vitro* (Schwartzman, 1986) and reacted exclusively with a protein, ROP1, which localises to the PVM following invasion (Saffer *et al.*, 1992; Carruthers and Sibley, 1997). However, tachyzoites of *T. gondii* Δrop1 deletion mutants were found to be as invasive as wild-type parasites (Soldati *et al.*, 1995), suggesting functional redundancy amongst rhoptry proteins. Becker and co-workers demonstrated that ROP2 and 4 associate with the cytoplasmic face of the PVM and provided the first direct evidence that rhoptry contents are involved in the formation of the PVM in *T. gondii*, when they showed that antibodies specific for ROP2, reacted with an intact PVM (Beckers *et al.*, 1994).

1.3.3 Micronemes

Micronemes (Fig. 1.4.2) were first described in an EM study of *T. gondii* by Gustafson (1954). Microneme means "thread-like" and refers to the elongate, osmiophilic appearance of the organelles in fixed parasite preparations (Jacobs, 1967). Energy-filtering transmission electron microscopy of thick serial sections of cryopreserved *C. parvum* sporozoites, followed by 3D imaging, showed that micronemes are spherical, uniform in size and situated mainly at the tip of the apical complex alongside the rhoptry neck (Tetley *et al.*, 1998). The number of micronemes varies in zoites of different apicomplexans with *Cryptosporidium*, *Eimeria* and *Sarcocystis* having large numbers and *Toxoplasma* and *Plasmodium* having few. The number of micronemes loosely correlates with how far and fast parasites glide over cultured cells and coated glass slides; an observation that led to early speculation that micronemes may play a role in motility (Dubremetz *et al.*, 1998b). A more detailed description of the contents and function of micronemes is given in section 1.5

1.4 The apicomplexan secretory system

Eukaryotic cell proteins that are destined to function anywhere except the cytosol have to be targeted to their final destination. The secretory pathway ensures the orderly, co-

translational targeting and progression of newly synthesised proteins to intracellular compartments such as the ER, Golgi apparatus, endosomes and secretory vesicles, through to the cell surface and beyond (Fig 1.5). Translation of mRNA begins on ribosomes in the cytosol, but once a hydrophobic N-terminal signal peptide emerges this is recognised by the signal recognition particle (SRP) and peptide elongation is halted momentarily. Once the SRP complex (SRP, signal peptide, mRNA and ribosome) has interacted with its docking protein at the ER membrane, the SRP is released from the ribosome and translation resumes. The docked complex forms an aqueous pore, through which the nascent peptide is co-translationally translocated into the ER lumen where the signal peptide is cleaved. Within the ER, proteins undergo a range of post-translational modifications such as glycosylation, folding and disulphide shuffling. Depending on its final location, a protein may be retained in the ER, transported out of the ER to the Golgi, from where it may be recycled back to the ER or transported through the Golgi apparatus towards the *trans*-Golgi network (TGN). During this journey further modifications, in particular the removal and addition of specific glycans, occurs. At the TGN, further sorting of proteins to their final destination occurs.

The early part of the apicomplexan secretory pathway appears to be very similar to that of higher eukaryotes. Proteins possessing classical N-terminal signal peptides are translocated into the ER lumen and transported, *via* coated secretory vesicles, through the Golgi apparatus to the TGN. However the late part of the apicomplexan secretory pathway is somewhat different to that of higher eukaryotes, since proteins have to be targeted to the specialised secretory organelles of the apical complex.

1.4.1 Trafficking to the ER and Golgi apparatus

There is good evidence that apicomplexan parasites process their proteins through the early secretory pathway in a manner analogous to their higher eukaryotic hosts. Homologues to mammalian components of the ER translocation machinery and to ER chaperones have been identified. For example, BiP, an ER lumenal binding protein, which functions as a soluble molecular chaperone, has been cloned and sequenced from *E. tenella* and *T. gondii* (Dunn *et al.*, 1996; Hager *et al.*, 1999). BiP has a C-terminal HDEL (His-Asp-Glu-Leu) sequence, homologous to HDEL or KDEL sequences in yeast and mammalian cells, which is responsible for its retention in the ER. Fusion of an HDEL motif to green fluorescent protein, GFP, caused it to be

retained in the ER, demonstrating that BiP is functional in *T. gondii* (Hager *et al.*, 1999). A homologue of Erd2, which is the ER receptor for BiP, is also present in the EST dataset for *T. gondii*.

In higher eukaryotes, proteins can be retained in the ER by virtue of ER-retention motifs, but can also be retrieved from the Golgi apparatus by recycling within retrograde transport vesicles (Rothman and Orci, 1992; Hobman *et al.*, 1998; Hong, 1998). Several amino acid motifs involved in this retrieval have been described in higher eukaryotes, including a di-lysine signal (Jackson *et al.*, 1990; Letourneur *et al.*, 1994) a di-arginine signal (Schutze *et al.*, 1994), a tyrosine motif (Mallabiabarrena *et al.*, 1995) and a tryptophan-based sequence (Cosson *et al.*, 1998). Homologues of these retrieval motifs have been found in apicomplexans and in *T. gondii* it has been shown that a C-terminal QKTT sequence can mediate the ER localisation of reporter chimeras (Hoppe and Joiner, 2000). As this sequence resembles the di-lysine motif KKXX of higher eukaryotes it is thought that it could be recognised by *T. gondii* proteins homologous to higher eukaryotic COPI components, and suggests that *T. gondii* and probably other apicomplexans have a vesicle-mediated retrograde Golgi-to-ER transport system.

In higher eukaryotes, the C-terminal acidic patch (DXE) motif has been found to increase export from the ER by 5-10 fold (Nishimura *et al.*, 1999). In *T. gondii*, ER localisation of a reporter molecule B-TGN (*E. coli* alkaline phosphatase fused to the C-terminus of TGN46) is abolished by the addition of a C-terminal acidic motif (DDEVH), suggesting that the parasite has a similar ER export system (Hoppe and Joiner, 2000).

The Golgi apparatus of higher eukaryotes is organized into three functionally distinct regions. The *cis* side, which is closest to the ER, receives newly synthesized proteins, the medial stack is where post-translational modification of proteins occur and the *trans* side has high vesicular activity and is where proteins with different final destinations are sorted (reviewed in Jamieson, 1998). Specialised secretory vesicles mediate the transport of proteins between ER and Golgi, across the Golgi and from the TGN onwards. Three different vesicle coats have been identified to date. COPI-coated vesicles mediate movement of proteins bi-directionally between the ER and Golgi, COPII-coated vesicles are responsible for anterograde transport from the ER to Golgi and clathrin-coated vesicles are involved in protein transport from the TGN to the PM

and in protein transport through the endocytic pathway (Rothman and Orci, 1992; Hobman *et al.*, 1998; Hong, 1998). COP-coated vesicles contain a series of coatomer proteins (α- β- β'- γ- δ- ϵ- ξ-COP) together with ADP ribosylation factor (ARF) and small GTP-binding proteins. Clathrin-coated vesicles contain clathrin, which consists of a heavy-chain protein (180 Kda) and two light chains (33 and 36 Kda) together with adaptor proteins (AP).

Coated vesicles between the apical side of the nucleus and the *cis*-Golgi have been observed by EM and stained with an anti-Tgβ-COP antibody in *T. gondii* (Hager *et al.*, 1999). Following treatment of *T. gondii* with Brefeldin A (BFA), which causes disassemble of the Golgi and redistribution of its membranes to the ER, Golgi localisation of β-COP was dispersed (Hager *et al.*, 1999). Genes homologous to components of COPI and COPII coated vesicles were identified in the *T. gondii* EST database, such as β-COP and ARF-1 (Ajioka, 1998) and ARF-1 was localised to the Golgi by immunofluorescence (Liendo and Joiner, 2000). Transient over-expression of mutant ARF-1 disperses a reporter protein that normally resides in the Golgi and TGN (Liendo and Joiner, 2000). Wild type and mutant ARF-1 were introduced into a stable line of *T. gondii* expressing *E. coli* alkaline phosphatase (BAP) fused with LDLR. The wild-type ARF-1 co-localised with BAP-LDLR at the TGN, while mutant ARF-1 disrupted the localisation with BAP-LDLR consistent with disruption of the TGN, suggesting a role for ARF-1 in maintenance of the TGN in *T. gondii* (Liendo and Joiner, 2000).

In higher eukaryotes, during transport through the Golgi, proteins undergo various post-translational modifications such as proteolytic processing, remodelling of N-linked oligosaccharide side chains and synthesis of O-linked glycans. Glycosidases and glycosyl transferases are capable of synthesising oligosaccharides (Harter and Wieland, 1996). There is good evidence of enzymatic machinery for post-translational modification in *T. gondii*. Analysis of a tachyzoite antigen GRA2, demonstrated that it is a glycoprotein containing exclusively O-linked carbohydrate side chains consisting of a monosaccharide and a disaccharide. The two O-linked carbohydrate species attached to GRA2 were identified as N-acetlygalactosamine and an N-acetyl-galactosamine-containing disaccharide (Schwarz and Tomavo, 1993). Analysis of a *T. gondii* lysate demonstrated that the transferases present are functionally active

(Schwarz and Tomavo, 1993). Also, the surface protein SAG5 contains N-linked oligosaccharides (Odenthal-Schinttler *et al.*, 1993).

1.4.2 Membrane docking and fusion proteins

In higher eukaryotes, proteins required to mediate the targeting and fusion of transport vesicles to appropriate target membranes, include an ATPase, N-ethylmaleimide-sensitive fusion protein (NSF), soluble NSF attachment protein (SNAP), integral membrane protein SNAP receptors (SNARE) and small GTP-binding proteins belonging to the Rab family. Vesicle fusion requires the soluble tetrameric protein, NSF, which works in conjunction with attachment proteins known as SNAPs and which together form a cytoplasmic complex that is recognised by an acceptor membrane. This complex interacts with SNAREs on the acceptor membrane. It has been proposed that transport vesicles contain a vesicle-SNARE, which can bind to a complementary target-SNARE on the destined acceptor membrane (Nichols and Pelham, 1998). The precise function of Rabs, is not fully understood, however they are thought to ensure that vesicle transport is unidirectional, providing specificity in vesicle release, docking and fusion.

The first indication of the presence of a NSF/SNAP/SNARE system in *T. gondii* came from the demonstration that hamster NSF and bovine α-SNAP augment dense granule secretion (Chaturvedi *et al.*, 1999). Rabs identified in *T. gondii* include Rab6 and 11 which are involved in dense granule secretion (Cesbron-Delauw, 1994). Some of these, such as cytosolic regulatory proteins Rab1, 5, 6, 7, 11, NSF and part of the Rab-GDI, have been cloned and sequenced in *T. gondii* (Stedmann and Joiner, 2000; Liendo and Joiner, 2000). Transient over expression of epitope-tagged Rab6 and NSF indicates that these regulatory proteins are localised to the parasite Golgi and are similarly dispersed by BFA treatment (Stedmann and Joiner, 2000; Liendo and Joiner, 2000). Together these results suggest that NSF/SNAP/SNARE/Rab machinery participates in dense granule release in *T. gondii* (Chaturvedi *et al.*, 1999).

1.4.3 Targeting to the dense granules

Karsten and co-workers used stable transfection of *T. gondii* to investigate the secretory pathway. Tachyzoites were transfected with two soluble foreign secretory reporters, *E. coli* β-lactamase (BLA) and BAP, both of which should not contain any targeting information. Both reporter proteins were quantitatively delivered to the dense

granules and secreted into the PV, which suggests that dense granules may form part of the default secretory pathway in *T. gondii* (Karsten *et al.*, 1998). Fusion of alkaline phosphatase to the putative transmembrane domain and cytoplasmic tail of GRA4 resulted in dense granule localisation, followed by secretion into the vacuolar space. On the other hand addition of a GPI anchor to BAP re-routed the protein to the parasite surface.

Dense granule secretion in *T. gondii* is calcium-independent, as increased levels had no effect on the secretion of GRA3 or BLA (Chaturvedi *et al.*, 1999). As no structures similar to immature secretory granules have been observed it seems unlikely that budding of small constitutive secretory organelles is involved.

1.4.4 Post secretory targeting of GRA proteins

The biosynthesis and fate of GRA proteins after secretion has been extensively studied in *T. gondii* (Fig 1.6). Several proteins secreted from the dense granules during the formation of the PV subsequently become tightly associated with membranes of the parasite, the PVM or the IVN (Fig. 1.6). For example, GRA 1, 2, 4 and 6 are secreted into the PV and associate with the IVN which connects the parasite PM with the PVM (Sibley *et al.*, 1986; Achbarou *et al.*, 1991a; Lecordier *et al.*, 1995; Mercier *et al.*, 1993; Labruyere *et al.*, 1999). In contrast GRA5 is closely associated with the PVM (Lecordier et al., 1993) and GRA3 and 7 combine with both the IVN and the PVM (Achbarou *et al.*, 1991a; Jacobs *et al.*, 1998). The basis of the differential targeting of GRAs after secretion is not understood. Amino acid sequences of GRAs reveal little homology to other proteins in the database and moreover they show little homology with one another. Targeting of GRAs is complicated by the fact that despite containing predicted transmembrane domains, many of them are found to be both partially soluble and membrane-associated. For example, GRA5, which contains a transmembrane domain does not associate with membranes of the dense granule or the ER but remains soluble until it is secreted into the PV where it becomes associated with the PVM through its central hydrophobic region (Lecordier *et al.*, 1995). Similarly, GRA4 is found in both partially soluble and membrane-associated forms but, following *in vitro* translation in a rabbit reticulocyte lysate, it clearly adopts a type I integral membrane conformation in microsomal membranes (Karsten *et al.*, 1998).

Analysis of a ∆gra5 *Toxoplasma* knock out resulted in a perfectly formed PVM and IVN and all other GRA proteins investigated (GRA1-7) were expressed and sorted

correctly (Mercier *et al.*, 2001b). Thus GRA5 does not appear to be have a role in targeting other GRAs to their final destinations, nor is it essential for formation of the PVM and IVN. In contrast, a Δgra2 *Toxoplasma* knock-out mutant was partially attenuated for virulence during acute infection (Mercier *et al.*, 1998) and an accumulation of granular material in the vacuolar space, instead of the IVN suggested that GRA2 functions in the organisation of vacuolar components into the network tubular membrane. Deletions of GRA2, which lacked either of its two amphipathic α-helices, were synthesised, secreted and correctly sorted to the IVN but did not stably associate with it (Mercier *et al.*, 1998). Lack of expression of GRAs 2, 5 and 6 did not modify the *in vitro* growth rate of tachyzoites (Mercier *et al.*, 1998; Mercier *et al.*, 2001a; Mercier *et al.*, 2001b). The secondary structure of the soluble protein GRA1, which does not have typical membrane-associated characteristics, reveals a Ca^{2+} binding domain known as an EF-hand and it has been postulated that this protein may function as a Ca^{2+} buffer, modulating the Ca^{2+} concentration, in order to stabilise the PV network (Cesbron-Delauw *et al.*, 1989).

GRA4, 5 and 6 show weak sequence homology to extracellular matrix proteins, suggesting they may have a role in structural modifications of the PV (Lecordier *et al.*, 1993; Mevelec *et al.*, 1992). Labruyere (1999) used cell fractionation and treatment with denaturing agents to show that the association between GRA4 and the IVN membrane was predominately influenced by strong protein-protein interactions, whereas the interaction between GRA6 and the IVN was mediated mostly by hydrophobic interactions. In addition, cross-linking studies demonstrated the formation of a multimeric protein complex between GRA2, 4 and 6 was based upon protein-protein and hydrophobic interactions (Labruyere *et al.*, 1999).

1.4.5 Targeting of proteins to the rhoptries and beyond

In *T. gondii* ROPs are sorted (Fig. 1.6) to an immature rhoptry compartment that might be functionally analogous to an immature secretory granule (ISG) (Soldati *et al.*, 1998). Similarly, *Plasmodium* ROPs are first targeted to a pre-rhoptry organelle (Shaw *et al.*, 1998) and such structures have been observed in *Plasmodium* (Lingelbach and Joiner, 1998).

ROP2 of *T. gondii* contains four tyrosine residues, one of which forms part of the sequence, SEYEQL which is reminiscent of the consensus sequence YXXϕ (X, any amino acid, ϕ an amino acid with a bulky hydrophobic side chain) known to mediate

mammalian protein sorting. Transfection of a construct containing a ROP2 deletion mutant, lacking the YEDL sequence motif, into *T. gondii* localised the transgene to compartments anterior to the parasite nucleus. Similarly, a ROP2 construct, in which the tyrosine was mutated to an alanine, accumulated in a tubular structure anterior to the nucleus distinct from mature rhoptries. BFA treatment showed this compartment to be separate from the Golgi or TGN (Hoppe *et al.*, 2000) and the ROP2 cytoplasmic tail bound in a tyrosine-dependent fashion to murine and *T. gondii* µl chain, a component of the clathrin coat-associated adaptor complex, AP-1 (Hoppe *et al.*, 2000). Clathrin-coated vesicles may therefore be involved in the maturation of rhoptry organelles (Hoppe *et al.*, 2000).

ROP2 of *T. gondii* behaves as an integral membrane protein in the PVM with its amino terminus exposed to the host cytoplasm. Analysis of the amino-terminal domain of the protein that is inserted in the PVM revealed features reminiscent of a mitochondrial matrix-targeting signal. It was shown *in vitro*, that the amino-terminal domain of ROP2 can bind host cell mitochondria and can be partially imported into the mitochondrion. The signal sequence is also capable of directing GFP to the host cell mitochondria. All of these findings suggest a role for ROP2 in the association between the host cell mitochondria and the PVM (Hoppe *et al.*, 2000).

1.4.6 Targeting to the micronemes

Transmembrane MICs carry sorting signals in their cytoplasmic tails. Di Cristina and co workers (2000) analysed the subcellular localisation of epitope-tagged constructs carrying amino acid substitutions or deletions at conserved residues in TgMIC2. Two amino acid motifs within the cytoplasmic domain were identified, both of which are necessary and sufficient for targeting proteins to *T. gondii* micronemes. One motif is based on the amino acid sequence SYHYY and has been found in the cytoplasmic tails of all transmembrane microneme proteins. The second is a stretch of acidic residues, EIEYE, which is also present in TgMIC6 (Reiss *et al.*, 2001). An artificial tail which contained both motifs was sufficient to target SAG1, a surface protein, to tachyzoite micronemes (Di Cristina *et al.*, 2000). Similarly, exchanging the GPI-anchoring signal of SAG1 with the membrane-spanning domain and cytoplasmic tail of TgMIC6 is sufficient to direct the chimeric protein to the micronemes (Reiss *et al.*, 2001). Hoppe and co-workers have shown that a reporter which has BAP fused to the transmembrane domain and cytoplasmic tail of murine Lamp1 (which contains a

YXXϕ motif) partially co-localises with MIC2 in micronemes (Hoppe *et al.*, 2000). Disruption of the GYQTI motif in the Lamp-1 cytoplasmic tail by a tyrosine/alanine substitution abolished microneme staining and resulted in localisation of the reporter to the Golgi.

Deletion analysis of a TgMIC3-GFP fusion showed that in the absence of a sorting signal the protein was delivered to the PV *via* the dense granules (Striepen *et al.*, 2001), in agreement with this being the default secretory pathway for soluble proteins (Karsten *et al.*, 1998). Two distinct domains within TgMIC3, the pro-domain and the C-terminus, possessed sufficient information to target GFP to the micronemes, however deletions in the central cysteine-rich region of TgMIC3 caused the protein to be arrested in various locations (cytoplasm, perinuclear region) within the early secretory pathway (Striepen *et al.*, 2001).

In general, the sorting of soluble MICs is not as well understood as that of transmembrane MICs. Soluble proteins have no access to the cytoplasmic sorting machinery used by transmembrane proteins and they may have to rely on selective aggregation of regulated secretory proteins, under specific circumstances such as an increase in Ca^{2+} concentration (Soldati *et al.*, 2001). Recent studies have established that some soluble MICs aggregate with transmembrane MICs, which function as escorters to ensure their accurate sorting to the micronemes. TgMIC6, which contains microneme-specific sorting signal(s) in its cytoplasmic tail, functions as an escorter for the accurate targeting of the soluble TgMIC1 and 4. TgMIC6 interacts with TgMIC1 *via* its third epidermal growth factor-like domain (EGF) and TgMIC1 binds directly with TgMIC4. Disruption of TgMIC6 causes both TgMIC1 and 4 to be sorted to the default pathway, which routes them to the dense granules from where they are secreted into the PV (Reiss *et al.*, 2001). These data, along with other studies (Chaturvedi *et al.*, 1999) again suggest that dense granules are the default secretory pathway, and that specific signals are required for targeting to both the micronemes and rhoptries. In the absence of TgMIC1, the TgMIC4/TgMIC6 complex is retained in the perinuclear region/ER/Golgi region suggesting that TgMIC1 is involved in 'quality control', most likely by ensuring correct folding of the complex within the early secretory system (Reiss *et al.*, 2001). Interestingly this phenomenon has been observed elsewhere. For example, in *P. falciparum*, in the absence of RAP1, RAP2 is retained
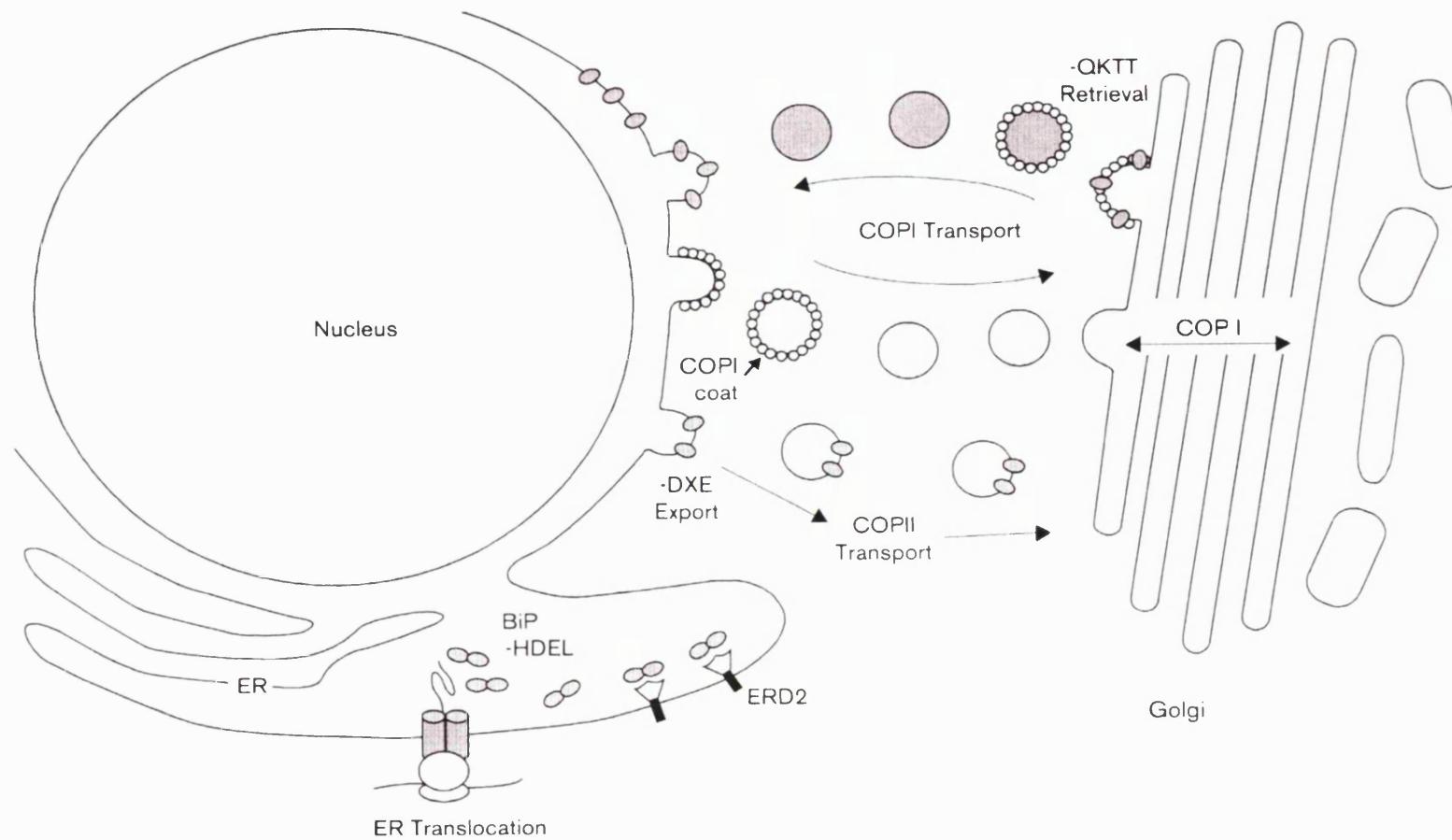
Fig. 1.5 Current model for protein targeting in the early secretory pathway of apicomplexan parasites.
Secreted proteins are translocated into the ER lumen where the chaperone BiP assists in protein folding. BiP resides in the ER due to the interaction of its carboxy-terminal, HDEL sequence with the Erd2 receptor. The export of protein is enhanced by the acidic patch (DXE) and mediated by the COP II coat. Retrieval from the Golgi is achieved by interaction of a modified di-lysine (QKTT) motif with the COP I coat. COP I coated vesicles shuttle proteins between the ER and the Golgi and between the Golgi cisternae.
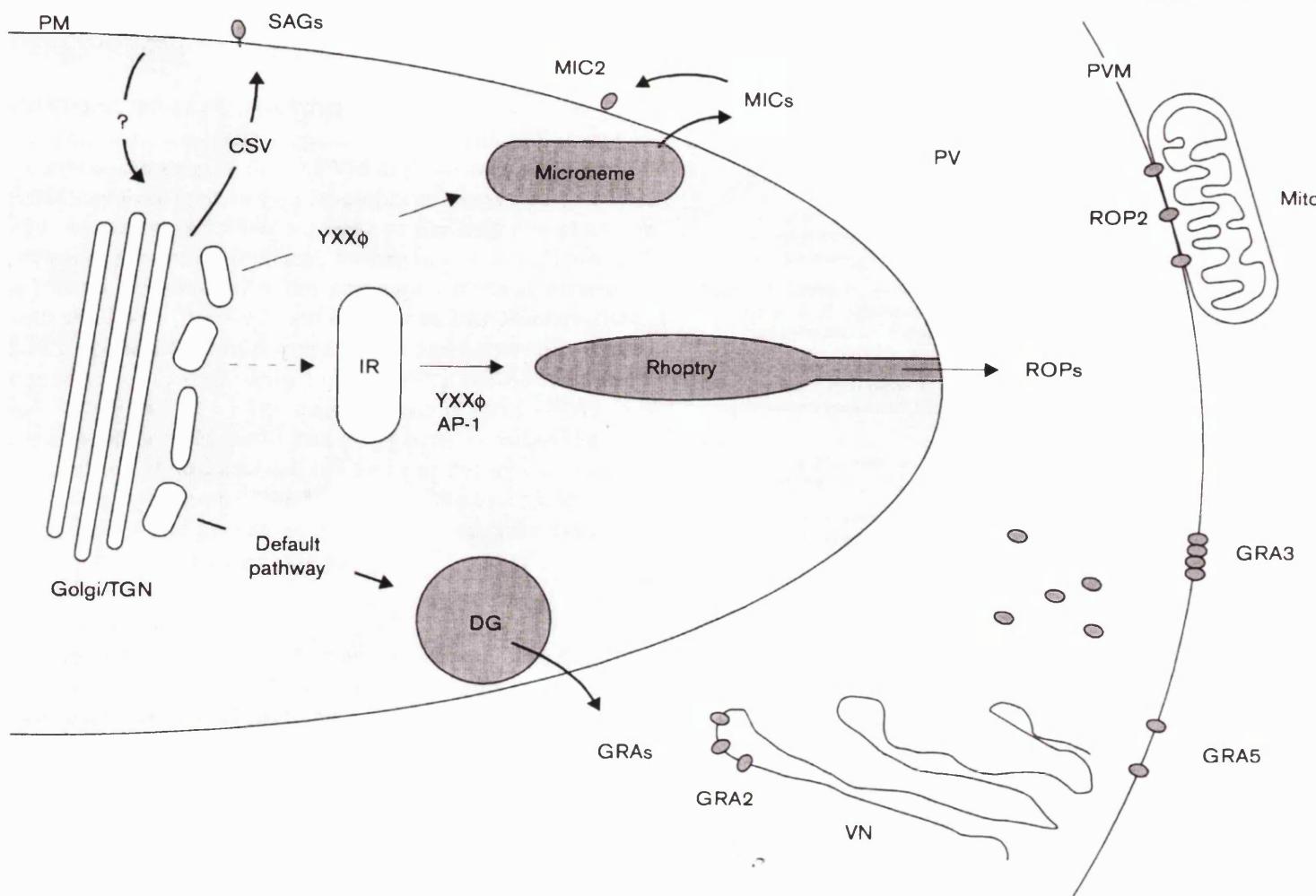
Fig. 1.6 Current model for protein targeting from the Golgi to the secretory organelles, micronemes, rhoptries and dense granules.
Glycosylphosphatidylinositol (GPI) anchored proteins such as SAGs are sorted by constitutive secretory pathway (CSV) to the cell surface. Soluble proteins with no sorting signals are targeted to the dense granules (DG). Their post-secretory localisation can be either to the PVN (GRA2) or the PV (GRA3 and GRA5). ROPs are routed to immature rhoptries (IR) for processing and are targeted to mature rhoptries using tyrosine-motifs (YXXΦ). From here they are released through the rhoptry duct and are inserted into the PVM. Microneme proteins (MIC) are probably sorted from the trans-Golgi network by tyrosine-based motifs (YXXΦ) upon an increase in Ca^{2+} concentration and soluble secreted proteins attach to the parasite surface and are capped backwards over the parasite.

and degraded in the ER (Baldi *et al.*, 2000). Another *Toxoplasma* transmembrane protein TgMIC8, which contains EGF-like domains, functions as an escortor for the soluble adhesin TgMIC3, which also contains multiple EGF-like domains (Garcia-Reguet *et al.*, 2000). It has also been demonstrated that TgMIC2 is physically associated with TgM2AP, a soluble micronemal protein. TgM2AP forms a complex with TgMIC2 within 15 mins of synthesis and this interaction remains while they are in the micronemes, on the parasite surface during invasion and while on the parasite surface membrane (Rabenau *et al.*, 2001). TgM2AP shares primary structural similarities to EtMIC2 of *E. tenella*. Thus assuming it is functionally equivalent to TgM2AP, EtMIC2 should associate with EtMIC1, which is the eimerian equivalent of TgMIC2. This association has been confirmed using co-immune precipitation assays on biotinylated lysates of *E. tenella* sporozoites (Rabeneau *et al.*, 2001). Together all these observations suggest that soluble MICs need to interact with transmembrane MICs in order to be sorted correctly to the microneme.

1.5 Micronemes

Microneme proteins characterised from various genera have been reviewed recently by Tomley and Soldati (2000). The implication of micronemes in the process of binding to and invasion of host cells originated from the discovery of adhesive proteins in *P. falciparum, P. knowlesi* and *P. vivax* (human malarial parasites), which were localised to the micronemes. These proteins were initially identified by antibodies because they partially blocked parasite adhesion to erythrocytes (Camus and Hadley, 1985; Miller *et al.*, 1988; Sim, 1990; Adams *et al.*, 1992) implicating their involvement in binding to host cells. Thereafter highly conserved microneme proteins in apicomplexan parasites were identified, initially in *Plasmodium* spp. (Robson *et al.*, 1988), and thereafter in *Eimeria* (Tomley *et al.*, 1991) and *T. gondii* (Wan *et al.*, 1997). Within the micronemes of sporozoites of *E. tenella*, approximately 11 proteins (ranging in size from <14Kda to >200Kda) have been resolved using both one and two-dimensional gel electrophoresis (Tomley *et al.*, 1996)

1.5.1 Microneme protein adhesive domains

Genes encoding approximately 30 microneme proteins, have been identified in different genera using a variety of methods including, localisation with specific antibodies, screening of cDNA expression libraries and comparisons of DNA sequence homologies. All proteins described so far in the phylum Apicomplexa

contain classic signal peptide sequences and many have a hydrophobic region at the C-terminus, which probably functions as a transmembrane domain. Many of the microneme proteins contain extracellular domains which bear striking conservation to various structural, adhesive domains, such as a thrombospondin type 1 (TSP-1) motif, an integrin insertion (I) domain, an EGF and Apple domains (Fig 1.7). The highly conserved nature of these adhesive proteins suggests that they possibly serve similar invasive and motility functions in the various parasites.

1.5.1.1 TSP domains

The most common repeats found in microneme proteins are the TSP-1 or TSP-1 like motif, and members of the apicomplexan TSP family include TRAP of *P. falciparum*, circumsporozoite protein (CS) of *Plasmodium* spp., the CS protein-TRAP-related protein (CTRP) of *P. falciparum*, EtMIC1 and EtMIC4 of *E. tenella* and Em100 of *E. maxima*, TgMIC2 of *T. gondii* and TRAP-C1 of *C. parvum*. Many members of the TSP family are adhesive molecules, involved in cell-cell and cell-matrix interactions, with TSP possessing the ability to bind to sulphated sugar residues, glycosaminoglycans (GAGs) in particular. The CS protein, an abundant surface protein of *Plasmodium* sporozoites, was the first member of the TRAP family to be isolated (Dame *et al.*, 1984; reviewed in Nussenzweig and Nussenzweig, 1985). This protein is GPI-anchored, contains one copy of TSP-1 and is thought to play a role in sporozoite gliding motility. It is translocated continuously from the anterior to the posterior and trails of the protein, left on glass slides after incubation of motile sporozoites, were revealed with antibodies raised against it (Stewart and Vanderberg, 1988; Stewart and Vanderberg, 1991). CS proteins also bind to highly sulphated, heparin-like oligosaccharides on the basolateral membrane of hepatocytes (Pancake *et al.*, 1992; Cerami *et al.*, 1992; Frevert *et al.*, 1993) and structural studies on parasite ligands demonstrate that TSP-1 binds in a stable fashion to heparan sulphate proteoglycans (HSPGs) (Sinnis and Kim Lee Sim, 1997). Antibodies to CS inhibit sporozoite invasion of HepG2 cells, an hepatocyte-derived cell line that is permissive for invasion and development by *P. berghei* sporozoites *in vitro* (Weiss *et al.*, 1990). The CS protein is the main component of the RTS,S vaccine which has been shown to induce a strong protective immune response in humans (Stoute *et al.*, 1997; Reithinger, 2002).

TRAP was isolated initially from *P. falciparum* by screening a genomic library with a probe corresponding to the TSP-1 motif of CS (Robson *et al.*, 1988). TRAP homologues have subsequently been found in all *Plasmodium* spp., each with a similar structure (Naitza *et al.*, 1988; Robson *et al.*, 1988; Robson *et al.*, 1997; Templeton and Kaslow, 1997). The protein has been localised to both the micronemes and the surface of malarial sporozoites, and has been shown to bind to sulphogalactosyl-cerebrosides (sulphated sugars) and the surface of HepG2, (Muller *et al.*, 1993). A series of nested deletions of TRAP constructs assigned the adhesive properties of the protein to the TSP-1 repeat (Muller *et al.*, 1993), implicating a role for TRAP in adhesion to and invasion of host cells. Heparitinase treatment of human liver sections abolished binding of recombinant TRAP proteins suggesting that in common with CS, HSPGs are receptors for TRAP (Robson *et al.*, 1995). *P. berghei* sporozoites were also shown to shed a continuous trail of TRAP-containing material and antibodies raised against TRAP dramatically blocked parasite motility (Spaccapelo *et al.*, 1997). A *P. berghei* TRAP null mutant line was unable to glide and had reduced invasion ability (Sultan *et al.*, 1997) and transgenic sporozoites of *P. berghei* carrying a PfTRAP gene in place of the endogenous TRAP gene were capable of normal development, motility, invasion and infection. These data provided the first example of functional complementation by a heterologous gene in *Plasmodium* spp. and showed that TRAP does not restrict the infectivity of a range of host cells (Wengelnik *et al.*, 1999). Furthermore, sporozoites expressing a mutant form of PfTRAP with a deletion of TSP-1 showed a reduced ability to invade salivary glands and were unable to glide *in vitro* and *P. berghei* sporozoites expressing PbTRAP with a mutation in the TSP-1 domain also had an impaired ability to invade, although their motility *in vitro* was identical to that of untransfected parasites (Matuschewski *et al.*, 1999). The TRAP gene is thus required for sporozoite motility and infectivity of host target cells and the TSP-1 repeat is of importance in each event.

CTRP of *P. falciparum* has both sequence and structural homology with TRAP and, to a lesser extent, the CS protein (Trottein *et al.*, 1995). CTRP contains seven copies of the TSP-1 repeat, only three of which contain the typical conserved sequence (Trottein *et al.*, 1995) and is expressed in the ookinetes of *Plasmodium* (Yuda *et al.*, 1999b). CTRP knock-out parasites are unable to invade host cells (Dessens *et al.*, 1999; Yuda *et al.*, 1999a) and a similar situation occurs with TRAP knock-out parasites (Sultan *et*

al., 1997), suggesting that TRAP and CTRP may play similar roles in invasion and motility in sporozoites and ookinetes, respectively.

1.5.1.2 I domains

Another motif which has been identified in microneme proteins from more than one genus is the integrin insertion (I) domain. A similar domain is found in soluble plasma proteins which mediate platelet adhesion, and in soluble matrix proteins, such as cartilage matrix protein and collagens.

The I-domain is a region of approximately 200 amino acids that is present in one or more copies in many proteins involved in cell-cell, cell-matrix and matrix-matrix interactions. Analysis of the high-resolution crystal structure of the I-domain of CR3 revealed a classic α/β Rossman fold (Michishita *et al.*, 1993) with Mg^{2+} binding to the surface creating a metal ion-dependent adhesion site (MIDAS). This site may be essential for the adhesive function of the I-domain (Kamata *et al.*, 1999), but binding may be mediated by other regions (Dickeson *et al.*, 1998).

The role of the I-domain in TRAP of *P. falciparum* was investigated using protein expressed as a glutathione S-transferase fusion in combination with a solid phase assay for screening of binding to known I-domain ligands. The PfTRAP I-domain bound to heparin in a metal-ion-dependent fashion *in vitro* and also bound to sulphatide and HepG2 cells (McCormick *et al.*, 1999). Sporozoites expressing a mutant form of PfTRAP, which had a substitution in a conserved residue of the I-domain were motile and capable of invasion (Wengelnik *et al.*, 1999). Thus it appears that TRAP binds heparin-like ligands on hepatocytes using both the TSP-1 and I-domain.

1.5.1.3 EGF-like domains

Another common module identified in micronemal proteins from more than one genus is the EGF-like domain, which is found in a wide range of proteins involved in extracellular events such as adhesion, coagulation and receptor ligand interactions.

A distinct class of these domains has been found to be associated with Ca^{2+} binding (cb). This cbEGF domain can be found in a variety of functionally diverse proteins such as fibrillin-1 (involved in extracellular matrix adhesion), factor IX and X (involved in the control of blood coagulation) and low density lipoprotein receptor, LDLR (involved in cholesterol uptake). Amino acid mutations in the Ca^{2+} binding domains of these proteins can lead to diseases such as Marfan syndrome, Familial

hypercholesterolemia, Haemophilia B and Proteins S deficiency (Downing *et al.*, 1996). Hence Ca^{2+} binding plays a role in maintaining the structural integrity of the EGF-containing proteins and is required for their correct function.

The novel EtMIC4 microneme protein of *E. tenella* has recently been cloned and sequenced (Tomley *et al.*, 2001). This 218 Kda protein contains 31 EGF-like repeats, 22 of which contain residues characteristic of the cbEGF adhesive domains, however it is not known if the cbEGF repeats do bind Ca^{2+}. The presence of such a large number of Ca^{2+} binding and non-Ca^{2+} binding EGF-like repeats is likely to give the EtMIC4 protein a stable conformation needed in its adhesive function (Tomley *et al.*, 2001). A number of micronemal proteins isolated from *T. gondii* contain multiple EGF-like domains, including four transmembrane proteins TgMIC6, 7, 8 and 9 and a single soluble adhesin, TgMIC3. Interestingly, TgMIC6 serves as an escortor for the soluble adhesins TgMIC1 and 4 (Reiss *et al.*, 2001). TgMIC7 and 8 contain five and ten EGF-like motifs respectively and both are expressed in bradyzoites (Meissner *et al.*, 2001). TgMIC8 possesses a lectin-like domain, similar to TgMIC3, (soluble adhesin with five overlapping EGF-like domains) (Garcia-Reguet *et al.*, 2000) and serves as an escortor for TgMIC3 (Meissner *et al.*, 2001). Interestingly its cysteine-rich lectin-like domain has been shown to play a role in binding to host cells (Soldati *et al.*, 2001), however no such lectin-like binding domain has been observed in *Eimeria* micronemal proteins, although NcMIC3 does contain such a domain (Sonda *et al.*, 2000).

1.5.1.4 Apple domains

The Apple domain, which is thought to play a role in adhesion interactions, is now known to be present in a small group of microneme proteins. Apple domains were first identified in the heavy chain of adhesive domains of two soluble plasma glycoproteins, coagulation factor XI (FXI) (Fujikawa *et al.*, 1986) and plasma pre-kallikrein (PK) (Chung *et al.*, 1986). These glycoproteins circulate in the plasma and are involved in the blood coagulation cascade which helps form blood clots. The Apple domain consists of four tandemly arranged repeat regions each stabilised by three intramolecular disulphide bridges (McMullen *et al.*, 1991). Each Apple domain can bind one or more ligand, such as cell surface receptors or co-factors of the coagulation cascade (Baglia *et al.*, 1990; Baglia *et al.*, 1991; Baglia *et al.*, 1993a; Baglia *et al.*, 1993b; Baglia and Walsh, 1996; Ho *et al.*, 1998).

The *E. tenella* microneme protein EtMIC5 was isolated initially from a sporozoite cDNA library, and has a predicted primary sequence containing 8 non-identical tandem repeat regions. The glycoprotein has a hydrophobic leader peptide of eleven cysteine-rich regions which share sequence similarity to the adhesive Apple domains of FXI and PK (Brown *et al.*, 2000b) but no predicted membrane-spanning region. The presence of three intramolecular disulphide bridges has been confirmed by mass spectrometry on a recombinant-expressed putative Apple domain of EtMIC5 (Brown *et al.*, 2001). Another Apple domain containing protein has been found in *E. acervulina*. This soluble protein contains seven full and one partial Apple domains arranged in tandem repeats (van den Boogaart, 1998) but its location has not been confirmed to any apical organelle. Other recently identified microneme proteins that contain Apple domains include a small lectin of *S. muris* (SML) and an adhesive protein, TgMIC4, from *T. gondii*. SML (Eschenbacher *et al.*, 1993) which functions as a dimeric lectin, consists of two Apple domains (Klein *et al.*, 1998) and has high affinity to N-acetyl-galactosamine and galactose (Klein *et al.*, 1998). TgMIC4 has six Apple domains and has been formally localised to the micronemes of all invasive stages (sporozoite, merozoite, tachyzoite and bradyzoite). Deletion mutants of TgMIC4 assigned the adhesive properties of the molecule to the final A6 region of the Apple domains. Removal of the last 12 amino acids at the C-terminus completely abolished TgMIC4 binding to host cells (Brecht *et al.*, 2000). An intact Apple structure, containing three disulphide bridges is also critical for adhesion, as pre-treatment of ESA with 2-mercaptoethanesulphonic acid (a strong reducing agent) resulted in complete failure of TgMIC4 to bind to host cells (Brecht *et al.*, 2000). As members of each Apicomplexa genus contain similar proteins, such as those containing the I-domain and the TSP-1 like domain, it would seem likely that proteins containing Apple domains will be found throughout the phylum Apicomplexa as more sequences becomes available. However, no homologues have yet to be identified in *P. falciparum* (genome sequencing project), *P. vivax*, *P. berghei* (ESTs), *Neospora*, *Cryptosporidium* or *Theileria*. Interestingly, *Eimeria, Sarcocystis* and *Toxoplasma* all invade their hosts *via* the digestive tract, whereas sporozoites of *Plasmodium* spp. are spread by direct injection into the blood and would not come into contact with the same range of host cell types. The presence of these various repeats in such a large number of proteins from distinctly different organisms indicates that these structurally

Cryptosporidium
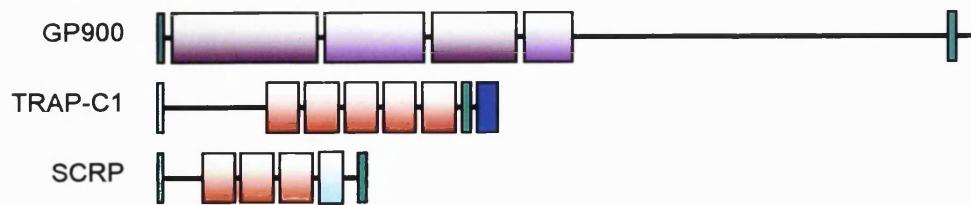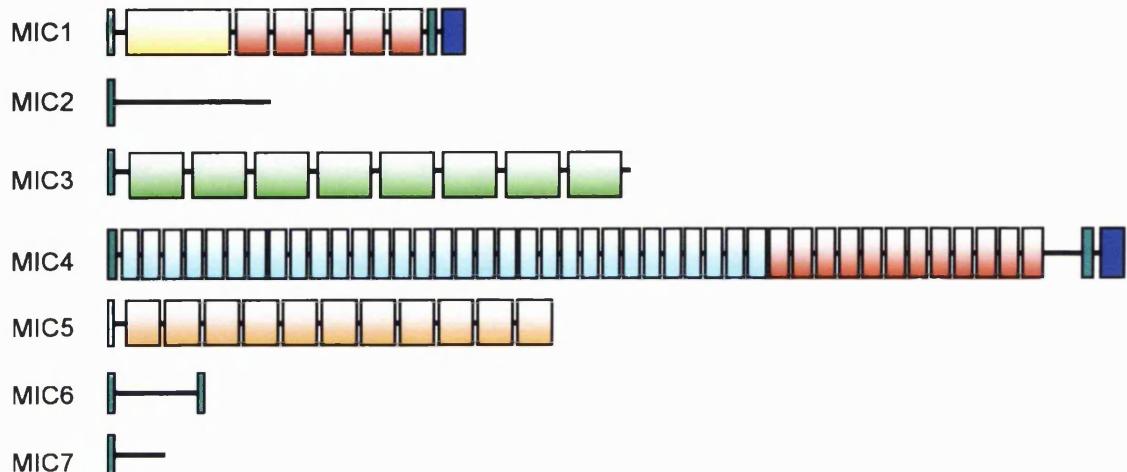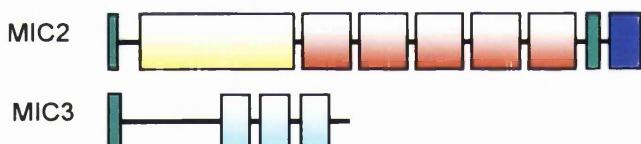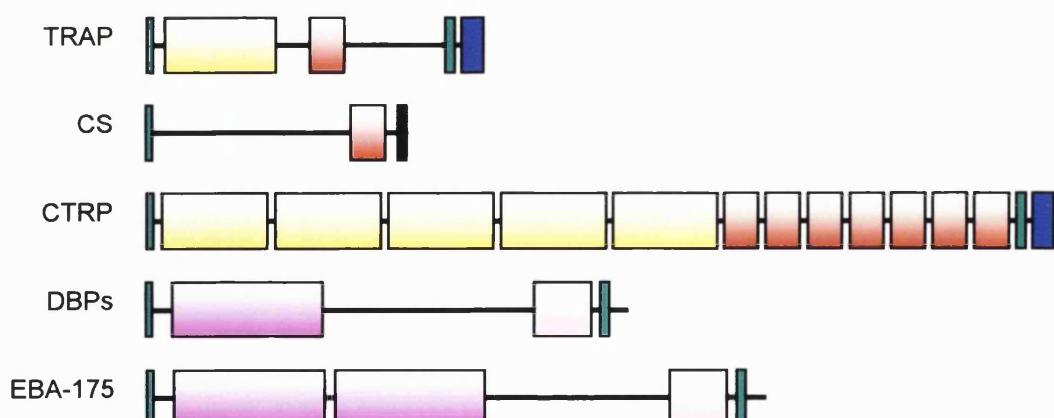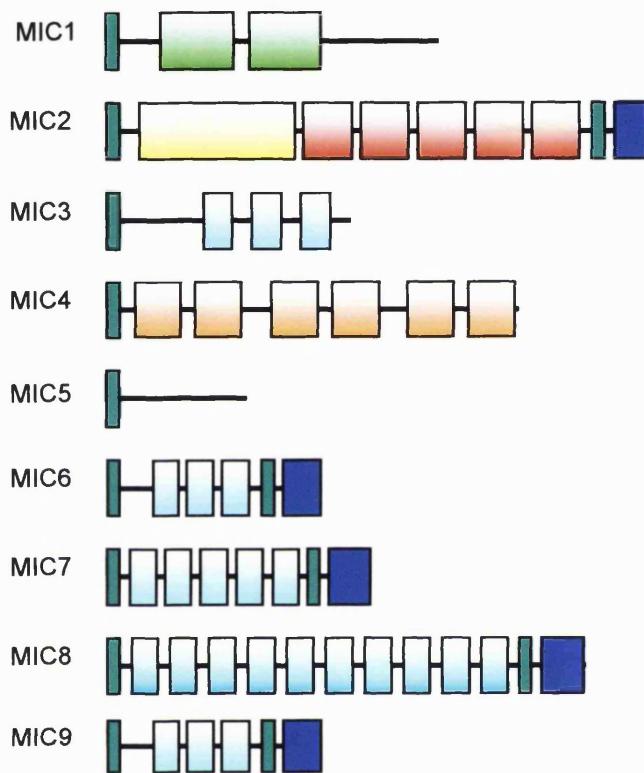
GP900

TRAP-C1

SCRP

Eimeria

MIC1

MIC2

MIC3

MIC4

MIC5

MIC6

MIC7

Neospora

MIC2

MIC3

Plasmodium

TRAP

CS

CTRP

DBPs

EBA-175

Sarcocystis

SML

SM70

Fig. 1.7 Schematic representation of some of the micronemal proteins in apicomplexan parasites.
For each genus, a single representation of each type of protein is shown. Abbreviations for each of the modules are included in the key.

conserved motifs must have some functional significance in motility and invasion.

1.5.2 Function in recognition and attachment

Recognition and attachment to a host cell is a prerequisite for invasion and is usually mediated by surface proteins, called adhesins, that bind to protein or carbohydrate epitopes present on the host cell surface. Many studies have been geared towards identifying sugar-binding proteins that might be involved in the process of recognition and attachment as carbohydrates are found modifying most cell surface proteins. Many of the microneme proteins identified in apicomplexan parasites have sugar binding properties and bind to carbohydrate molecules on the host cell surface membrane. Proteoglycans are present within the extracellular matrix and on the surface of all animal cells and consist of a core protein with one or more covalently attached glycosaminoglycans chains (GAGs). GAGs consist of linear polymers composed of repeating disaccharide units one of which is always an amino sugar (N-acetylglucosamine or N-acetylgalastoamine) and the other is often a uronic acid (glucuronic acid or iduronic acid). During polymerisation the chains undergo various sulphation and epimerisation reactions and in conjunction with the length of the polymer, proteoglycans have tremendous structural heterogeneity, all of which have unique biological characteristics.

Another microneme protein with sugar-binding properties is GP900, a glycoprotein of *C. parvum*. Sequence analysis predicts a mucin-like transmembrane glycoprotein composed of distal cysteine-rich domains separated by polythreonine domains and a large N-glycosylated core region (Barnes *et al.*, 1998). The two mucin-like domains and domain 3 of the cysteine-rich region are believed to play a role in invasion. GP900 has a specific galactose-N-acetlygalactosamine lectin activity and glycoprotein inhibitors directed against this lectin reduce attachment of *C. parvum* sporozoites to host cells *in vitro* (Joe *et al.*, 1994; Joe *et al.*, 1998).

Polypeptides of *T. gondii* also have lectin activity. Whole tachyzoites agglutinate rabbit erythrocytes *in vitro* and the reaction is inhibited by the sulphated polysaccharide-heparin, dextran sulphate and fucoidin (Ortega-Barria and Boothroyd, 1999). Moreover the infectivity of tachyzoites is inhibited by high concentrations of dextran sulphate and fucoidin and inhibition of GAG biosynthesis and sulphation reduces *Toxoplasma* infectivity (Ortega-Barria and Boothroyd, 1999). The microneme protein TgMIC2 may be capable of binding to many ligands, as recombinant forms

can bind to heparin, a reaction which is inhibited by chondroitin and dextran sulphate (Carruthers *et al.*, 2000). This observation is consistent with the suggestion that adhesion by *T. gondii* parasites should involve the recognition of ubiquitous surface-exposed host molecules as this parasite can invade a broad range of cells within a mammalian host.

The *P. falciparum* homologue of TgMIC2, PfTRAP, has similar lectin-binding properties. The use of TRAP constructs demonstrated that this micronemal protein binds to sulphogalactosyl-cerebrosides and to the surface of the immortalised hepatocyte cell line, HepG2 (Muller *et al.*, 1993). It appears that TRAP binds heparin-like ligands on hepatocytes and a series of nested deletions assigned the adhesive properties of the protein to the TSP-1 repeat (Muller *et al.*, 1993).

A number of ligands and complementary erythrocyte receptors have been identified in merozoites of *Plasmodium* spp. These are believed to be involved in attachment of the parasite to the host cell and formation of a tight junction and are dependent upon sugar-binding properties. Members of the erythrocyte-binding protein (EBP) family have been localised to the microneme organelles of merozoites (Sim, 1990; Sim *et al.*, 1992). The first EBP identified (EBP-175) was found to bind to both merozoites and erythrocytes (Camus and Hadley, 1985) and antisera raised against its 5' cysteine-rich regions inhibited invasion of erythrocytes by merozoites (Sim, 1990). EBP-175 recognises sialic acid residues and binds specifically to glycophorin A on the surface of erythrocytes (Sim *et al.*, 1994) using twelve amino acids in the first cysteine-rich region (Jakobsen *et al.*, 1998). Other members of the EBP family from *P. vivax* and *P. knowlesi* having a high degree of structural homology are Duffy binding proteins (DBPs). The 5' cysteine-rich regions of these proteins also mediate binding between themselves and receptors on erythrocytes, but recognise only glycoproteins carrying the Duffy blood group (Chitnis and Miller, 1994).

As already outlined in section 1.1.1, *Eimeria* spp. are not only host specific with regard to infection but are also site specific (Long and Joyner, 1984; Sharma *et al.*, 1996; Strout *et al.*, 1994). Lectins have been identified on sporozoites of three different species of *Eimeria*, (*E. acervulina*, *E. maxima*, and *E. tenella*), all of which infect different intestinal sites within the host. These lectins were only found on the primary infective stage and the lectin for each species had different sugar specificities. Thus it is possible that lectin binding may play a role in determining the site of

infection within the intestine of the host (Baba *et al.*, 1996; John *et al.*, 1999; Strout *et al.*, 1994), unlike the lectins of *T. gondii* which possibly have a much broader sugar-binding specificities.

1.5.3 Secretion

As outlined already (section 1.2), for parasite invasion to proceed an attachment between the apical tip of the zoite and the host cell surface must occur. Many studies suggest that microneme proteins are responsible for mediating this attachment. For example, several red blood cell-binding proteins of *Plasmodium* (Adams *et al.*, 1990; Sim *et al.*, 1992) and members of the TRAP family of adhesive proteins and adhesive *T. gondii* proteins (Wan *et al.*, 1997) were located to micronemes of the apical tip.

During invasion, microneme organelles are stimulated to release their contents through the apical tip of the zoite (Carruthers *et al.*, 1999a). Microneme secretion has been investigated in many apicomplexan parasites but was first demonstrated in cystozoites of *S. muris* with the aid of a monoclonal antibody raised against microneme proteins SM 16/17 and SML. Confocal laser microscopy revealed that microneme secretion occurs *via* the apical tip and the newly released proteins bind to the host cell surface at the site of infection and spread along the cell surface (Entzeroth *et al.*, 1992). A similar finding was observed in sporozoites of *E. tenella* when EtMIC2 was translocated to the sporozoite apical surface and transiently dispersed over the entire surface of the infected cell during the invasion process (Tomley *et al.*, 1996).

This phenomenon was also observed in *T. gondii* when TgMIC2 was secreted from the apical tip at the junction between the parasite and host cell surface and, after initially occupying the apical surface, was rapidly treadmilled to the posterior end of the parasite and released into the supernatant (Carruthers *et al.*, 1999a). EtMIC2 was also translocated backwards over *E. tenella* sporozoites and released from the posterior, however in cytochalasin-treated sporozoites the protein was secreted but not capped over the parasite (Bumstead and Tomley, 2000). Results with both TgMIC2 and EtMIC2 are consistent with the capping model proposed earlier, in which transmembrane proteins are translocated backwards over the parasite and just before entry of the parasite into the PV, are shed from the posterior.

Regulated secretion has been studied extensively in numerous eukaryotic cells. Usually an external stimulus is required to trigger a signal transduction cascade, with the eventual secretion from an organelle. Secretion of microneme proteins TgMIC2

and NcMIC2 were initially shown to be temperature dependent with inhibition below 25°C (Wan *et al.*, 1997; Carruthers *et al.*, 1999b; Lovett *et al.*, 2000). Microneme secretion may also be inhibited by the partial depletion of microneme proteins by temperature cycling (37°C for 10 min, centrifugation, resuspension in fresh media) (Carruthers *et al.*, 1999b), and treatment with a protein kinase inhibitor, staurosporine (Carruthers *et al.*, 1999b; Janene Bumstead, IAH, personal communication).

Secretion has been shown to be dependent upon the concentration of intracellular free Ca^{2+}. For example, secretion of TgMIC2 from tachyzoites can be triggered by treatment with Ca^{2+} ionophores, in the presence or absence of external Ca^{2+} (Carruthers and Sibley, 1999). The Ca^{2+} ionophores, ionomycin and A23187, which release Ca^{2+} from non-acidic compartments, also induce secretion of NcMIC2 from tachyzoites (Lovett *et al.*, 2000). Chelation of intracellular Ca^{2+} using 1,2-bis (*O*-aminophenoxi) ethane-N,N,N',N',-tetraacetic acid (BAPTA) tetraacetoxymethyl ester (AM) can prevent microneme secretion (Carruthers, and Sibley, 1999; Lovett *et al.*, 2000), but this can be partially reversed by the addition of ionophore A23187, which elevates intracellular Ca^{2+} (Carruthers and Sibley, 1999). The release of microneme proteins was stimulated in *Toxoplasma* lysates following treatment with thapsigargin or ammonium chloride, which mobilise Ca^{2+} from intracellular stores such as the ER (Carruthers and Sibley, 1999; Lovett *et al.*, 2000). Thapsigargin inhibits the ER ATP-dependent Ca^{2+} transporter which leads to increased levels of cytoplasmic Ca^{2+} (Thastrup, 1989) and upon dissociation of ammonium chloride, Ca^{2+} levels are increased due to the release of Ca^{2+} from acidocalcisomes (Moreno and Zhong, 1996). Both of these products also stimulated NcMIC2 secretion (Lovett et al., 2000) indicating that intracellular stores of Ca^{2+} alone are sufficient to stimulate the release of some microneme proteins. Further stimulators of microneme secretion include ethanol and acetaldehyde (Carruthers *et al.*, 1999b; Lovett *et al.*, 2000), which induce secretion by mobilising Ca^{2+} from thapsigargin-insensitive compartments and leading to elevated intracellular Ca^{2+}. Microneme discharge can also be induced in a regulated fashion by the addition of foetal calf serum or highly purified bovine or chicken albumins (Bumstead and Tomley, 2000). Overall, these observations suggest that microneme discharge is a stimulus-coupled secretion system and that exocytosis is controlled by a Ca^{2+}-dependent pathway. Also, an increase in intracellular Ca^{2+} occurs in tachyzoites when they attach to host cells and this increase is required for invasion (Moreno *et al.*, 1994; Docampo and Moreno, 1996; Lu *et al.*, 1997; Vieira and

Moreno, 2000). It should be noted that results so far suggest that an increase in intracellular Ca^{2+} is not necessary for recognition of host cells, but is required for attachment and invasion.

1.5.4 Processing

Many microneme proteins are processed either before or after they are released from the parasite. Proteins such as TgMIC3, 5 and 6 are processed proteolytically during their transport through the secretory pathway, whereas others such as TgMIC2 and 4 are processed upon discharge from the micronemes. Proteins such as TgMIC3, SML and EtMIC5 are processed once, whereas others such as TgMIC2, 3 and 6 under go more than one processing event. Information on the proteases involved and the biological significance of the processing events is under investigation.

The first example of intracellular processing within apicomplexan parasites was observed in TgMIC3, which was initially characterised as a 90 Kda, heterodimer of two 38 Kda polypeptides, linked by disulphide bridges. Using both metabolic labelling and immunoprecipitation assays a 40 Kda precursor polypeptide was shown to be processed into a 38 Kda mature polypeptide at the N-terminus (Achbarou *et al.*, 1991b). Processing was found to be sensitive to BFA indicating that this event is likely to occur in the TGN. Deletion of the pre-sequence induced a mistargeting of the proteins, demonstrating that it is required for the correct targeting to the TGN or beyond (Leburn, Soete and Dubremetz, unpublished).

cDNA sequence analysis of EtMIC5, a micronemal glycoprotein, predicted an open reading frame of 932 amino acids (Brown *et al.*, 2000b). Comparisons of this sequence with the N-terminal peptide sequence of the protein derived from purified micronemes showed that is also processed to remove 26 residues such that the N-terminus of mature EtMIC5 begins at residue 52 of the open reading frame. Residues 1-25 show a typical secretory signal peptide, with a predicted cleavage site. It appears that EtMIC5 undergoes intracellular, N-terminal, post-translational processing (Brown *et al.*, 2000b). EtMIC5 and TgMIC3 undergo single processing events, yet many MICs undergo multiple processing events.

Another micronemal protein, which undergoes post-translational modifications by proteolytic processing, is TgMIC5. The protein is expressed initially as a preprotein, which is processed to a proprotein by a signal peptidase before being further processed to a mature 22 Kda protein. BFA treatment blocked the production of the two

intermediate species, 23 and 22.5 Kda as well as the mature protein, suggesting that the processing events occurred in the TGN or beyond (Brydges *et al.*, 2000).

TgMIC6 is processed initially at the N-terminus whilst within the parasite, to remove the pre-sequence, and again at the C-terminus once the protein is released from the microneme. Addition of BFA inhibited TgMIC6 processing, and detection of the pro-protein in both the ER and the Golgi, and absence of staining in the microneme by IF analysis, suggested the processing may occur in the TGN (Reiss *et al.*, 2001).

Other *T. gondii* MICs, which undergo multiple processing events are TgMIC2 and 4 (Carruthers *et al.*, 2000; Reiss *et al.*, 2001). TgMIC2 is a transmembrane protein which is processed at both the N and C-termini on the surface of the parasite (Carruthers, 2002). TgMIC4 is a soluble protein which is processed initially in the TGN to release a single Apple domain and again on the surface of the parasite resulting in the dissociation of the TgMIC1, 4 and 6 complex (Reiss *et al.*, 2001). The significance of the different processing events needs further investigation although a potential role may be activation or deactivation of specific adhesive properties as the full-length, cellular form of TgMIC2 (i.e. the unprocessed form) could bind to the human foreskin fibroblast (HFF) cells and not the processed, secreted form (Carruthers *et al.*, 1999a).

1.5.5 Transmembrane and cytoplasmic tail

An alignment of the transmembrane and cytoplasmic tail (TM-CT) of MICs in *E. tenella, P. falciparum, P. berghei, S. muris* and *T. gondii* demonstrated the amino-acid sequence of the hydrophobic, transmembrane region is well conserved. Just upstream of the transmembrane domain there is a striking conservation of amino acid residues within a glycine-rich region and a tyrosine residue is usually located immediately downstream of the transmembrane region. Since MICs are exocytosed onto the parasite surface, translocated backwards towards the posterior of the parasite and finally released, the conservation of the membrane-spanning region could reflect the presence of recognition sites needed for proteolytic processing, possibly for shedding of the protein from the parasite surface during invasion (Di Cristina *et al.*, 2000). The proteolytic cleavage sites of TgMIC2, 6 and 12 have been mapped to the conserved membrane-spanning domain (Opitz *et al.*, 2002).

The short cytoplasmic tails of MICs are uniform in length, rich in glutamic acid and often have conserved tryptophan or other aromatic residues. The conserved tryptophan

and the stretch of acidic residues within the cytoplasmic tail of PbTRAP are dispensable for expression of the protein and for its translocation to the parasite surface membrane, but are necessary for gliding motility and invasion (Kappe *et al.*, 1999). When the cytoplasmic tail of PbTRAP was replaced with that of TgMIC2, the protein was still targeted to the correct location (Kappe *et al.*, 1999) suggesting that the signals required for distribution and trafficking of MICs to both the organelle and their final destination are held within the cytoplasmic tail. Thus the cytoplasmic tail may contain signals required for the correct sorting of the proteins and may contain a region necessary in generating the force required for invasion and gliding motility (i.e. it may serve as an anchor on the parasite actomyosin to transfer mechanical force across the plasma membrane).

1.6 Project aims

The broad aim of this work was to understand more fully the processes by which the appearance of proteins within the microneme organelles are regulated.

The more specific aims were to:

- Develop and exploit the system of oocyst sporulation (i.e. transition from an undifferentiated oocyst to the formation of eight sporozoites) to investigate the timings of components of the microneme organelle at both the mRNA and protein levels, during *de novo* formation of microneme organelles.

- Define the chromosomal location of genes encoding micronemal proteins.

- Define the upstream genomic regions of microneme-coding genes to identify putative regulatory regions such as transcription start sites, TATA and CCAAT boxes.

- Define the minimum sequence required for the expression of microneme-coding genes using a transient transfection system.

- Define critical nucleotides/regions within putative regulatory motifs through mutation and/or deletion analyses.

Chapter Two
Materials and Methods

General parasitology and molecular biology methods used are outlined in this chapter. Methods specific to individual studies are detailed in the relevant chapters. Details of buffers, solutions, culture media and media supplements not given in this or subsequent chapters are outlined in Appendix A.

2.1 General parasitology methods

2.1.1 Parasites

The Wisconsin (Wis) strain and its precocious derivative, Wisconsin F96 (WisF96), were kindly provided by Dr. T. K. Jeffers and were used for protein and mRNA extraction, pulsed field gel electrophoresis and transient transfection assays. An aprinocid-resistant line (Shirley and Harvey, 1999) was derived from the Weybridge (Wey) strain (kindly provided by Janet Catchpole, VLA, Weybridge) and used in pulsed field gel electrophoresis. All parasites were maintained by passage through six-weeks-old Light Sussex chickens.

2.1.2 Recovery of oocysts from infected caeca

Birds were infected by oral inoculation with 1×10^3 sporulated oocysts of the Wey or Wis strains or 1×10^6 of the WisF96 precocious line, respectively. Seven days post infection (or four days for WisF96) the birds were killed by cervical dislocation, and the caeca removed using sterile scissors. After cutting longitudinally, the caecal contents and tissue were removed by scraping the caecal wall with a glass slide, transferred to a Waring blender and macerated in 100 ml of phosphate-buffered saline, pH 8.0 (PBS 8.0) for 2-3 min. The volumes described are typical for removal of caeca from 15 to 20 birds. The homogenate was transferred to a conical flask and made up to 200 ml with PBS pH8.0. Trypsin (Difco 1:250 powder; Sigma) was added to a final concentration of ~1.5% w/v. The solution was mixed thoroughly and the flask incubated at 41°C for 45-60 min to release oocysts from host cells and caecal contents. The enzyme-digested homogenate was centrifuged in an MSE-GF8 centrifuge at ~1000 g for 10 min, the supernatant removed and the pellet of unsporulated oocysts washed three times in water.

2.1.3 Sporulation

Oocysts were resuspended in 5 L conical flasks, containing no more than 2 L of liquid per flask, at a concentration of ~0.25 x 10^6 per ml in 2% w/v potassium dichromate. They were allowed to sporulate for 72 h at room temperature, with forced aeration through sterile airlines, using a pump (GHOST 3, Waterlife). To further assist sporulation, the culture was stirred continuously on a magnetic stirrer.

2.1.4 Recovery and purification of sporulated oocysts

Potassium dichromate was removed by at least 4 washes in tap water followed by centrifugation (as in section 2.1.2). Oocysts were suspended in 100 ml of 10% sodium hypochlorite (12% w/v available chlorine, BDH) and placed on ice for ~ 5 min. Surface-sterilised oocysts were then washed and centrifuged as before and the pellet suspended in 200 ml saturated sodium chloride solution and dispensed into 50 ml glass tubes. A 1 cm layer of water was carefully pipetted on top and the mixture was again centrifuged. The white layer of sporulated oocysts was recovered from the salt-water interface using a transfer pipette and after three further washes the oocysts were suspended in 100 ml of PBS pH7.6. The numbers of oocysts were determined using a Fuchs Rosenthal counting chamber (Shirley, 1995b) and stored at 4°C until use.

2.1.5 Preparation of sporozoites

Sporulated oocysts were centrifuged in a MSE Centaur 2 at 1000 g for 10 min in a sterile 50 ml Falcon tube. The supernatant was discarded and the pellet of oocysts suspended in a minimal volume of PBS pH7.6, just enough to allow a flowing suspension. An equal volume of # 8 ballotini glass beads (Jencons; 0.5 mm diameter) was added to the mixture, which was then vortexed sufficiently to break oocysts without damage to the majority of the newly released sporocysts. The progress of breakage was monitored microscopically and vortexing was stopped when free sporozoites could be seen. Parasites were removed from the glass beads by repeated washing in PBS pH7.6 and transferred into a 500 ml conical flask where they were suspended at 2 x 10^6 ml^{-1} in a filter-sterilised solution of 0.5% (w/v) trypsin, 1% (w/v) taurocholic acid, 10mM $MgCl_2$ in PBS pH7.6 or Hanks' balanced salts solution. The flask was incubated in a 41°C water bath until most sporozoites had excysted, which took approximately 1.5 h. The sporozoite mixture was centrifuged as above, the pellet washed twice in PBS pH7.6 and then suspended in 10 ml PBS pH8.0, supplemented with 1% glucose (Schmatz *et al.*, 1984).

Sporozoites were purified away from oocysts, sporocysts and debris by passage through columns of nylon wool and DE-52 resin (Diethylaminoethyl cellulose, pre-swollen microgranular anion exchange; Whatman) as described previously (Schmatz *et al.*, 1984; Shirley, 1995b). Briefly the resin was washed repeatedly with ten-fold excess volumes of PBS pH8.0, and then resuspended in 2 volumes of PBS and re-adjusted to pH8.0 with 5% (w/v) orthophosphoric acid. Columns were prepared by pressing "teased" nylon wool into the barrel of 20 or 50 ml syringes to a height of 5cm and, just prior to use, the nylon wool was pre-wetted with 3 ml of PBS pH8.0 with 1% glucose and covered with a 3 cm height of DE-52. The column was equilibrated with 20 ml of PBS pH8.0 containing 1% (w/v) glucose (flow buffer), 10 ml of which was allowed to flow through. The mixture of oocysts, sporocysts and sporozoites was poured on top and the eluate monitored microscopically for the presence of sporozoites (Shirley, 1995b). More flow buffer was added as required and the eluted sporozoites were centrifuged at 1000 g for 10 min, and finally suspended in PBS pH8.0 containing 1% glucose. The numbers of sporozoites were determined using a Fuchs Rosenthal counting chamber and the parasites were stored on ice. All solutions, wool and tubes used were sterilised before use.

2.2 General molecular biology methods

2.2.1 Preparation of plasmid and cosmid DNA

2.2.1.1 Small scale

Plasmid and cosmid DNA (10–15 µg) was extracted from *E. coli* XL1-Blue MRF' cells using commercial QIAprep spin miniprep kits (QIAgen), according to manufacturers instructions. Larger quantities of DNA, typically 100 or 500 µg, were obtained using either midi or maxiprep kits (QIAgen), respectively. All procedures featured alkaline lysis of cells, RNase A treatment, neutralisation, precipitation and pelleting of protein and chromosomal DNA with a final purification of DNA through silica membrane spin columns and elution in a low salt buffer or distilled water.

2.2.1.2 Large scale

Larger quantities (≥ 1 mg) of plasmid or cosmid DNA were obtained through triton lysis of *E. coli* cells and ultra centrifugation through caesium chloride gradients, essentially as described by Sambrook (1989). Briefly, a single colony of bacteria harbouring the desired plasmid or cosmid was used to inoculate 200 ml of LB broth

(supplemented with the appropriate antibiotic) and incubated at 37°C overnight with shaking at 200 rpm. Bacteria were centrifuged at ~5,000g for 10 min at 4°C in a Sorvall 5S plus centrifuge. The pellet was suspended in 2.5 ml of 25% (w/v) sucrose in 50 mM Tris-HCl, pH 8.0 and transferred to sterile 50 ml oakridge tubes. To each tube, 0.5 ml of 0.5% (w/v) lysozyme, in 0.25 M Tris-HCl pH 8.0 and 1.0 ml of 0.25M EDTA pH 8.0 were added and the samples were gently swirled and incubated at room temperature for 5 min. Next, 4 ml of 1% Triton X-100 in 50 mM Tris-HCl pH 8.0; 65.5 mM EDTA was added and the tubes inverted until the samples became glutinous, after which they were centrifuged at 25,000 g for 30 min at 4°C in a Sorvall 5S plus centrifuge. The supernatants were decanted into 20 ml universals and volumes made up to 8 ml with milliQ water. Finally, 8 g of $CsCl_2$ (Sigma) and 1 ml of ethidium bromide (5 mg/ml^{-1}) were added to each sample and mixed to dissolve. The samples were pipetted into 11.5 ml ultra centrifuge tubes (TFT65-13; Sorvall), overlaid with mineral oil, pairs of tubes balanced to within +/- 0.05 g of each other and centrifuged in a T-1270 rotor (Sorvall) in an OTD65B ultra centrifuge (Sorvall) at ~98,500 g for 48-72 h at 10°C.

DNA bands were visualised under long wave UV light and the superhelical plasmid fraction was recovered through the side of the tube after two 18 gauge needles had been inserted, one through the neck of the tube and the other just below the desired DNA band. The upper and lower needles were removed in-sequence and the dripping fraction was collected. The plasmid DNA was then washed repeatedly with equal volumes of water-saturated butan-1-ol and the upper layer (containing ethidium bromide) was removed each time. Harvested DNA was dialysed (Sigma) against milliQ water at 4°C for 24 h.

Plasmid DNA was analysed in 1% (w/v) ethidium bromide stained (0.5 µg/ml^{-1}) agarose (Sigma) gels and quantified by determining absorbance values at 260 and 280 nm using a spectrophotometer (BioRad). The DNA was then dispensed into 100 µg aliquots under 95% ethanol and 0.3 M sodium acetate, pH 4.8 and stored at –20°C.

2.2.2 Polymerase chain reaction (PCR)

Polymerase chain reactions were carried out essentially as described by Mullis & Faloona (1987) and Innis and Gelfand (1990). Reactions were set up in 50 µl volumes with approximately 50 ng of plasmid or cosmid DNA; 12.5 pmol of each primer

(MWG Biotech or Life Technologies); 10 nmol of each dNTP (100 mM dNTP set), 2.5 U Taq DNA polymerase and 1 x reaction buffer (Amersham Pharmacia Biotech).

Template DNA was amplified using 1 cycle at 95°C for 3 min, 30 cycles at 95°C for 15 s, 55°C for 30 s and 72°C for 90 s and one cycle at 72°C as a final extension for 10 mins. The annealing temperature varied depending on the primer combination used but was usually set to approximately 5°C below the melting temperature (Tm). The extension time was modified according to the expected product size and ranged between 1 min for fragments between 100 and 1000 bp and a minimum of 1 min 40 s for products up to 2000 bp.

Recombinant Taq polymerase (from *E. coli* expressing a cloned *Thermus aquaticus* DNA polymerase gene, PE Applied Biosystems) was used for sequencing reactions. Pfu polymerase (cloned from *Pyrococcus furiosus,* Amersham Pharmacia Biotech) was used for amplification of DNA generated by mutational PCRs. If it was necessary to remove primers, polymerases, nucleotides or salts from the PCR-amplified DNA fragments a QIAquick PCR purification kit (QIAgen) was used according to the manufacturer's instructions.

2.2.3 Restriction endonuclease digestion

Plasmid and cosmid DNAs were digested using restriction endonucleases according to the manufacturers' instructions (Gibco BRL, Promega, New England Biolabs). Single digests were carried out in 30-100 μl volumes, depending upon the quantity of DNA to be digested. Digests were carried out using between 2 and 20 U of endonuclease per 1 μg of DNA in the corresponding reaction buffer at the recommended temperature for 1 h. For electrophoresis a fifth volume of loading dye (30% glycerol, 5 x TBE, 0.25% bromophenol blue dye) was added to each sample.

2.2.4 Gel purification of DNA fragments

DNA fragments were electrophoresed through ethidium bromide-stained (0.5μg ml⁻¹) agarose gels (0.7-1.5 % w/v; SeaPlaque low-melting point agarose, Flowgen), in 1 x tris-borate-EDTA buffer (TBE). DNA fragments were visualised on a transilluminator (long wave UV), excised from the gel using a sterile scalpel blade, with a minimum volume of agarose, and the DNA recovered using a QIAquick spin-column gel extraction kit (QIAgen), according to the manufacturer's instructions. Briefly, agarose was melted at 50°C for 10 min, DNA absorbed onto a silica membrane in the presence

of high salt, while contaminants passed through the column. Impurities were removed and DNA was eluted in water.

2.2.5 Ligation of DNA

Ligations were set up in 10 µl volumes in 0.5 ml Eppendorfs. Each ligation contained gel extracted DNA (150 ng), 50 ng of the appropriate vector DNA, (pGEM®-T Easy vector or pSV-β-gal vector; both from Promega) 0.5 µl T4 DNA ligase (3 Weiss U/µl) and the correct volume of 1x or 5x buffer to give a 1 x ligation reaction buffer (Promega). Ligation reactions were incubated overnight at 16°C or at room temperature for 3 h. Controls comprising of a positive (supplied with kit), background (no insert DNA) and a no ligase treatments were included each time ligations were performed. The molar ratios of insert to vector DNA were usually kept at 3:1. If plasmids were digested with only a single restriction endonuclease, alkaline phosphatase treatment was carried out subsequently to prevent self-ligation, before the plasmid was included in a ligation reaction. 2 U of calf intestinal alkaline phosphatase in 50 µl of 1x alkaline phosphatase reaction buffer (both from Roche Molecular Biochemicals) were added to the plasmid and incubated at 37°C for 40 min to catalyse the removal of 5' phosphate residues from the DNA, thus preventing self-ligation. The plasmid was then purified from contaminating salts using a QIAquick PCR purification kit (QIAgen).

2.2.6 Transformation of DNA with *E. coli* cells

The transformation of *E. coli* XL1-Blue MRF' or JM109 cells was performed essentially as described by Hanahan (1983) and Sambrook (1989). Briefly, 10 ml of LB-broth was inoculated with a single colony of *E. coli* XL1-Blue MRF' and incubated overnight at 37°C while shaking. A 400µl aliquot of the overnight culture was used to inoculate 40 ml of LB-broth supplemented with 10 mM MgSO$_4$, 10 mM MgCl$_2$, 2.5 mM KCl, in a 500 ml baffled flask and re-incubated at 37°C, with shaking until the culture reached an A$_{600}$ of 0.3 (or above). 30 ml of cells were centrifuged at 1000 g for 10 min in a Centaur 2 centrifuge. The supernatant was removed and the pellet suspended gently in 2.5 ml of transformation buffer (TFB) and incubated for 15 min on ice. 90 µl of dimethyl sulphoxide (DMSO), 90 µl of dithiothreitol (DTT; 2.5 M in 40 mM potassium acetate, pH 6.0) and 90 µl of DMSO were then added, swirled gently and placed on ice for 5 min, 10 min and 5 min respectively. For each

transformation, 200 μl of competent cells were added to a ligation reaction in an Eppendorf tube and incubated on ice for 30-45 min, after which they were heat pulsed in a 42°C waterbath for 90 seconds and placed on ice. Each mixture of competent cells and DNA was added to 800μl of SOC medium. The transformed cells were then shaken for 30 min at 37°C and spread (in 100 and 900 μl aliquots) on LB-agar plates supplemented with the appropriate antibiotic and 80 μg/ml of X-Gal and 0.5 mM IPTG if required. The plates were then incubated upside down at 37°C overnight.

2.2.7 DNA sequencing

2.2.7.1 Automated sequencing reactions and precipitations

Sequencing reactions were set up using the chain termination method and each reaction contained 500 ng plasmid DNA or up to 2 μg cosmid DNA template, 2.5 pmol specific primer, 2.0 μl Ready Reaction Mix (ABI PRISM™ BigDye™ Termination Cycle Sequencing Ready Reaction Kit with AmpliTaq^R DNA polymerase, FS, PE Applied Biosystems), 2.0 μl of BD HalfTERM (Genpak Ltd.) and milliQ H_2O to a final volume of 10 μl.

The reactions were mixed and placed in a preheated (95°C) thermal cycle block, with a heated lid (Mastercycle gradient, Eppendorf) and the DNA was PCR-amplified through 24 cycles of 95°C for 40 s; 50°C for 15 s and 60°C for 4 min.

The extension products were precipitated by incubation on ice for 15 min with 10 μl sterile H_2O, 2.0 μl 3M sodium acetate (pH 4.8) and 50 μl 95% ethanol. The samples were centrifuged for 30 min at 15,000 g and the supernatants removed. DNA pellets were washed in 200 μl 70% (v/v) ethanol, vortexed briefly and centrifuged at 15,000 g for 10 min. The wash solution was removed and the DNA pellet air-dried.

2.2.7.3 Automated sequencing gels

A 6%, 8M urea, sequencing gel solution was made by mixing 27g urea, 39.4 μl H_2O and 7.1 ml of Long Ranger 50% acrylamide stock solution (Flowgen). 1.5 g of amberlite resin (BioRad, 20-50 mesh) was added and the mixture was stirred gently for 5 min. This mixture and 7.5 ml of 10 x TBE were filtered through a 0.2 μm filter into a collection chamber and degassed under vacuum. Thereafter 375 μl of 10% (w/v) APS and 52.5 μl of TEMED (N, N, N', N'-tetramethylethylenediamine) were added to polymerise the gel and the mixture was poured into a 377A sequencing gel assembly using a 50 ml syringe. Once the gel had polymerised (~3 h) the assembly was

positioned in an ABI model 377A automated sequencer. Running buffer (1 x TBE) was added and used to flush the wells. Prior to loading, the DNA pellets were resuspended in formamide loading dye (Amersham Life Sciences) and heated for 5 min. at 90°C. The samples were loaded onto the 6% polyacrylamide gel and electrophoresed for 10 h at 50 V.

The apparatus was set up according to ABI protocols using filter set A (λ = 531, 560, 580, 610 nm) and the data were collected using the 377A collection package. The contiguous sequence data were assembled and analysed using Staden (GAP 4) software (Staden, 1982), running on a Unix system.

2.2.7.4 Manual sequencing reactions

Sequencing reactions were set up using a Thermo Sequenase Cycle Sequencing Kit from USB, that incorporated Thermo sequenase DNA polymerase (Tabor & Richardson, 1995). Each reaction contained 500 ng plasmid DNA template, 0.5 pmol specific primer, 2.0 μl reaction buffer, 1.0 μl of both 7-deaza-dGTP cycle mix and dCTP cycle mix, 0.5 μl [α-^{35}S]dATP (10 μCi/μl), 2.0 μl Thermo sequenase DNA polymerase and sterile H_2O to a final volume of 17.5 μl.

The reactions were mixed and placed in a preheated (95°C) thermal cycle block, with heated lid (Mastercycle gradient, Eppendorf) and the DNA was PCR-amplified through 60 cycles of 95°C for 15 s and 60°C for 30 s. When the labelling cycle was complete, 3.5 μl of the sequencing reaction was added to 4.0 μl of each ddA/C/G/T termination mix. Each reaction was mixed and placed in a pre-heated (95°C) thermal cycle block and the DNA was amplified through 30 cycles of 95°C for 30 s and 72°C for 90 s. Finally 4.0 μl of Stop solution was added to each of the termination reactions.

2.2.7.5 Manual sequencing gel

A 100 ml, 6%, 6M urea polyacrylamide gel in 1 x TBE at pH8.3 was made by mixing 80 ml SequaGel XR monomer concentrate with 20 ml SequaGel complete buffer (National Diagnostics) and 800 μl of 10% (w/v) ammonium persulphate (Sigma). The gel was poured into a manual sequencing gel assembly using a 50 ml syringe and once the gel had polymerised (~2 h) the assembly was positioned in a hooded vertical sequencing apparatus (Flowgen). Running buffer (1 x TBE) was added and used to flush the wells. DNA sequencing reactions were heated for 2 min at 75°C prior to loading and the gel was electrophoresed for 10 h at 50 V.

After electrophoresis the assembly was dismantled and the gel fixed by soaking in a mixture of 10% (v/v) acetic acid, 10% (v/v) ethanol and 10% (v/v) methanol for 1 h. The gel was then placed onto 3MM Whatman filter paper, covered in food barrier wrap (Saran) and dried at 80°C under vacuum for 2 h. The Saran wrap was then removed and the gel exposed to X-ray film (Biomax™ MR, Kodak) at room temperature for a minimum of 3 h in a photographic cassette (Harmer, London).

2.2.8 DNA detection

2.2.8.1 Southern blotting

Gels were irradiated for 1 min with UV light at 254 nm prior to transfer of DNA to Hybond-N-nylon membrane according to the manufacture's instructions. Blotting was conducted essentially as described in Southern (1975) and Sambrook (1989). Gels were shaken gently in denaturing and neutralisation buffers for 1 h respectively. DNA fragments were capillary blotted onto nylon membrane under weight using neutralisation buffer as the transfer buffer. DNA was immobilised on the filter using a Stratalinker^R (Stratagene), 120,000 μJ of UV energy or by baking at 80°C for 3 h.

2.2.8.2 Hybridisation of radiolabelled DNA probes

Filters were prepared for probing by initial incubation at 65°C in a rotisserie-style hybridisation oven (Hybaid) with gentle rotation in a series of solutions: milliQ H_2O for 15 min, 2 x SSC for 15 min and finally hybridisation buffer comprising 1 x SSC and 5% SDS and pre-hybridisation buffer (Church and Gilbert, 1984) for 15 min.

DNA fragments were radiolabelled with α ^{32}P dCTP (10 mCi ml^{-1}/3000 Ci mmol^{-1}; Amersham Pharmacia Biotech) using a Prime-it^R Random Primer Labelling Kit (Stratagene), which incorporates radio-label into DNA fragments using random 9-mer primers and the exonuclease-deficient Klenow fragment of *E. coli* DNA polymerase I. Filters were incubated with radiolabelled DNA for at least 4 h at 65°C, after which they were washed in 3 x 20 min changes with 0.1% (w/v) SDS and 0.1 x SSC, wrapped in Saran wrap and exposed to X-ray film (XB-200, X-ograph Imaging Systems) at –70°C for a minimum of 3 h, using intensifying screens.

2.2.9 Detection of mRNA

2.2.9.1 Total RNA extraction from a time-course of sporulating oocysts

Total RNA was recovered at each time point from 25×10^6 oocysts using a detergent-based PureScript total RNA isolation kit (Gentra Systems, supplied from Flowgen). All tubes, solutions etc. used throughout this procedure were deemed to be RNase-free by commercial suppliers. Oocysts were centrifuged at 1000 g for 10 min, the supernatant removed and resuspended in 200 µl of H_2O (Sigma) and an equal volume of #8 glass beads (Jencons). The mixture was vortexed vigorously and the oocysts were inspected periodically under a microscope to determine the extent of breakage. Once the oocysts had been broken, the samples were centrifuged for 2 min as above, and the supernatant removed. 1 ml of lysis buffer was added to the supernatant and the solution pipetted up and down 3 times to lyse the sporocysts and sporozoites. 330 µl of protein-DNA precipitation solution was added and the tubes were inverted 10 times followed by a 10 min incubation on ice. The mixture was centrifuged at 13,000 g for 5 min and the supernatant removed and added to 1 ml of 100% isopropanol (BDH). After inverting 50 times, the tube was centrifuged for 15 min at 13,000 g. The supernatant was removed and the RNA pellet was washed in 70% ethanol and allowed to air dry for 15 min. Finally, the RNA was resuspended in 50 µl of diethyl pyrocarbonate-treated deionised water, placed on ice for 30 min and vortexed briefly (5 s). The concentration of RNA was quantified by determining absorbance values at 260 and 280 nm using a SP6-550 UV/VIS spectrophotometer (PYE Unicam) and the quality was examined by electrophoresis in a 1% agarose ethidium bromide (0.5 µg/ml^{-1}) stained gel. Residual genomic DNA in the preparation was removed by adding 4 U of RNase-free DNase I (Invitrogen) to 1 µg total RNA and incubating at 37°C for 10 min. DNase was inactivated by incubation at 65°C for 5 min and the total RNA was stored at −70°C in diethylpyrocarbonate-treated (DEPC) water until use.

2.2.10 Protein analyses

2.2.10.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were analysed by SDS-PAGE through polyacrylamide gels, prepared essentially as described by Laemmli (1970) and Sambrook (1989). Separating gels were prepared with 10% or 15% (w/v) 29:1 acrylamide; 375 mM Tris-HCl, pH 8.8;

0.1% (w/v) each of sodium dodecyl sulphate and ammonium persulphate (APS) and finally 0.1% (v/v) TEMED. The separating mixture was poured using a BioRad minigel apparatus (80% full). The mixture was over-laid with water-saturated butan-1-ol (to give a level surface) and allowed to polymerise for ~30 min. Stacking gels were prepared with 5% (w/v) 29:1 acrylamide; 0.125M Tris-HCl, pH 6.8; 0.1% (w/v) of each SDS and APS and finally 0.1% (v/v) TEMED. Once the gel had polymerised, the butan-1-ol was removed and a 5% stacking gel was placed on top followed by insertion of a 10 or 15 well BioRad minigel comb (15 mm).

Protein samples were resuspended in concentrated SDS loading buffer (x 3) containing 100 mM DTT and denatured at 95°C for ~5 min. Prestained SDS-PAGE standards were used in all gels (5 µl, pre-heated at 37°C for ~5 min). The gel assembly was placed into a BioRad minigel apparatus and covered with Tris-glycine running buffer. Once fully assembled, the protein samples (~500 ng) and markers were loaded using a "Precision Micro Syringe" (Camlab) and were run at 100V (~15 V cm^{-1}) until the samples had passed through the stacking gel. The voltage was then increased to 150 V (~22 V cm^{-1}) until the dye reached the end of the gel. BioRad precast gels (10% or 10-20% (w/v) acrylamide) were used occasionally and according to the manufacturer's instructions.

2.2.10.2 Gel staining and drying

Gels were immersed into Coomassie Brilliant Blue R-250 dye (CBB) for 30 min and destained in several changes of 50% (v/v) methanol and 16.7% (v/v) acetic acid to visualise the protein bands. CBB staining has a detection limit of ~100 ng of protein (in one band) and protein estimations were aided by the addition of SDS-PAGE standards containing various proteins of known concentrations in adjacent wells. For later reference the gels were soaked overnight in 20% (v/v) methanol and 3% (v/v) glycerol, laid on 3MM paper covered, with Saran wrap and dried for ~2 h on a gel-dryer (BioRad) at 70°C under vacuum.

2.2.10.3 Western blots and alkaline phosphatase detection

Proteins from SDS-PAGE were electro-blotted onto nitrocellulose (0.45 µm pore size; Sartorius AG), using a semi-dry blotter (BioRad Laboratories). The protein gel, nitrocellulose and 1MM paper were soaked in Bjerrum-Sahafer-Niehran transfer buffer for ~15 min. Three pieces of 1MM paper were placed onto the transfer cell (anode) followed by the nitrocellulose, the protein gel and three more pieces of 1MM

paper, taking care to remove air bubbles at each point. The lid (cathode) of the apparatus was placed on top. Proteins from a single mini gel were transferred for 30 min at 15 V with a maximum current of 3 mA/cm^2. Protein blots were stained in 1% (v/v) Ponceau S solution to check transfer of proteins to the nitrocellulose membranes and the stain was removed by washing in PBSa. Non-specific binding sites on the membrane were blocked with a one-hour incubation in 5% w/v milk powder in PBSa. The blot was probed with the desired primary antibody (rabbit antisera against *E. tenella* microneme proteins) for one hour (1:200 or 1:500 dilutions in PBSa; 1% (w/v) BSA and 0.05% (v/v) Tween 20). After extensive washing in PBSa and 0.05% (v/v) Tween 20, the blot was incubated in goat anti-rabbit IgG conjugated to alkaline phosphatase secondary antibody (Sigma) (1:30,000 in PBSa; 1% (w/v) BSA and 0.05% (v/v) Tween 20) for one hour. After further extensive washing in PBSa and 0.05% (v/v) Tween 20, the antigen-antibody interaction was visualised by alkaline phosphatase detection using 0.4 mM nitroblue tetrazolium and 0.4 mM 5-bromo-3-chloro-3-indoyl-phosphatase (NBT-BCIP) in alkaline phosphate buffer. The reaction was stopped by rinsing the blot in distilled water.

2.2.11 Transfection of *E. tenella* sporozoites

2.2.11.1 Preparation of cell monolayers

Madin-Derby Bovine Kidney (MDBK) cells (NBL-1; Flow Laboratories) were grown in HAMs F12 nutrient media (Gibco BRL) to a concentration of 6 x 10^5 ml^{-1}, and dispensed into 24 well plates containing sterile coverslips at 0.6 ml per well (giving 3.6 x 10^4 cells per well). Plates were incubated at 41°C with 5% CO$_2$ for 2 h to allow a semi-confluent monolayer to form.

2.2.11.2 Electroporation

Freshly purified sporozoites obtained as detailed in section 2.1.5, were washed once in incomplete cytomix buffer, and resuspended to give a final concentration of 1 x 10^7 parasites in 700 µl of complete cytomix (Van der Hoff, 1992). For each electroporation, 1 x 10^7 sporozoites and 100 µg of caesium-chloride purified, closed circular plasmid DNA (suspended in 100 µl cytomix) were mixed gently in a cuvette (0.4 cm gap, BioRad). Electroporations were carried out in a BTX model 600 electroporator (BTX, San Diego) set at 2.0 kV and 13 ohms resistance, resulting in a pulse time of around 0.26 ms. Electroporated parasites were left undisturbed for at

least 20 min at room temperature. The MDBK cell culture media was replaced with fresh HAMs F12 in readiness for infection.

Electroporated parasites were added to 4 ml of warm HAMs F12 medium, supplemented with 10% FCS, penicillin (200 U/ml^{-1}) and streptomycin (20 μg/ ml^{-1}), dispensed into wells containing a monolayer of MDBK and incubated at 41°C with 5% CO_2, allowing the parasites to infect the MDBK cells for at least 22 h.

2.2.11.3 Fixing and visualisation for β-galactosidase

The medium was removed from the monolayer, the cells were washed gently twice with PBS, and fixed in 0.25% glutaraldehyde (made up to 1 mM $MgCl_2$, 0.1 M Na_2HPO_2/NaH_2PO_4 pH 7.0). The fixed monolayer was washed three to four times with PBS, 400 μl of β-galactosidase staining solution added and the plate then incubated at 37°C for ~3 hrs. The reaction was stopped by washing three times in 3% DMSO in PBS and the cells were examined by light microscopy, (Zeiss Diaplan Microscope), photographed and the number of blue parasites present in each well was counted.

2.2.11.4 Preparation of parasite/cell lysates for assay of β-galactosidase

Following a 24 h incubation period after electroporation, sporozoites were assayed for beta-galactosidase using a Luminescent β-galactosidase Reporter system 3 kit from Clontech (Palo Alto, California, USA). Briefly, sporozoites were removed from incubation plates using a rubber policeman, transferred to a microcentrifuge tube and centrifuged for 2 min at 13,000 g. The pellet was washed twice in ice-cold PBSa and resuspended in 75 μl lysis buffer (100 mM potassium phosphate, pH 7.8; 1 mM DTT). Where sporozoites had been allowed to invade a monolayer of MDBK cells, cells were washed twice in ice-cold PBS, and resuspended in 500 μl PBS by scraping with a rubber policeman. Parasite/cell suspensions were frozen rapidly by immersion into a dry ice/methanol bath for 1 min followed by thawing at 37°C for 2 min, this freeze/thaw cycle was repeated twice. The suspension was sonicated at an amplitude of 10 μm for 30 s in an ice bath. Samples were centrifuged at 14,000 g for 20 min and the supernatant was transferred to a fresh microcentrifuge tube and kept on ice.

2.2.11.5 Chemiluminescent assay for β-galactosidase using a luminometer

A 50 μl aliquot of parasite lysate was placed into sample tubes and mixed with 196 μl reaction buffer and 4 μl reaction substrate (both heated to room temperature). The

reactions were incubated at room temperature for 60 min and light emissions were recorded at 5 s intervals using a tube luminometer.   70

Chapter Three
Expression of microneme-specific mRNA and protein during sporulation of *E. tenella* oocysts

3.1 Introduction

The life cycles of *Eimeria* spp. are highly complex and infective motile zoites (sporozoites and merozoites) are produced at various time points throughout. Oocysts are the most easily accessible phase of the eimerian life cycle and during their sporulation the single diploid nucleus undergoes meiosis followed by mitosis resulting in the production of eight newly formed haploid sporozoites (Canning and Anwar, 1968). Four separate developmental and morphological changes have been observed within the sporulating oocysts; condensation of the sporoplasm, cleavage of the sporoplasm, sporocyst formation and sporozoite differentiation (Norton and Chard, 1983). Sporulation has many merits for the study of gene expression during the formation of a discrete, invasive life-cycle stage. For example, it occurs exogenously and can be carried out under controlled, reproducible conditions and samples may be withdrawn for analysis at any time. The newly formed sporozoites are produced within tough cysts that can be rendered surface-sterile and free from contaminating host cell tissue and debris. Sporulation is ideal for time-course experiments which require multiple sampling and high yields of protein and nucleic acids and could be used to study the development of microneme organelles within *E. tenella in vitro* without interference from the host, making large scale biochemical and biological studies possible. In other members of the phylum Apicomplexa it can be difficult to assess parasites which are undergoing differentiation and to synchronise their development. The sporulation of *Eimeria* provides an good entrée into this type of examination and this system could be exploited to study fundamental aspects of apicomplexan gene expression. This project began with an investigation of the timing of expression of components of the microneme organelles during sporulation at the mRNA and protein level followed by chromosomal mapping of five genes that encode microneme proteins.

Invasion of host cells by extracellular motile zoites is a fundamental part of the apicomplexan life cycle and involves the sequential release of proteins from three distinct secretory organelles, the micronemes, the rhoptries and finally the dense

granules (Carruthers and Sibley, 1997). Morphological, biochemical and functional data support the hypothesis that micronemes are important for substrate-dependent zoite motility, host cell recognition and attachment (reviewed by Tomley *et al.*, 2001). Microneme proteins, MICs, are secreted from the apical tip of the zoite and capped backwards over the parasite surface as invasion proceeds (Bumstead and Tomley, 2000; Carruthers *et al*, 1999a). Many MICs are structurally conserved between different genera and several are ligands for the binding of parasites to host cells (Adams *et al*, 1990; Muller *et al*, 1993; Robson *et al*, 1995; Sim *et al*, 1992). Antibodies against MICs can reduce or inhibit parasite invasion of host cells *in vitro* (Barnwell and Galinski, 1989; Sim, 1990; Sharma *et al.*, 1996). Furthermore, conditions which inhibit secretion of MICs, such as low temperature, low levels of intracellular free Ca^{2+} and inhibition of serine/threonine protein kinases, also inhibit parasite invasion of host cells *in vitro* (Carruthers & Sibley, 1999; Wan *et al.*, 1997).

Little is known about the formation of microneme organelles within the zoite nor of the regulation of microneme protein expression during the parasite life cycle. Secretion of MIC2 from *T. gondii* is unaffected by treatment with BFA, which blocks constitutive secretion of protein from the endoplastic reticulum (ER) through the Golgi apparatus (Klausner *et al.*, 1992) thus suggesting that microneme organelles contain a store of proteins which is ready to be rapidly released during invasion (Wan *et al.*, 1997). From ultrastructural studies it is also clear that micronemes are formed afresh during each successive stage of the life-cycle. For example, during first generation schizogony of *Eimeria*, the micronemes, together with the pellicle, conoid and subpellicular microtubules of the invading sporozoite, gradually disappear (Chobotar *et al.*, 1975) and new micronemes, probably originating from the Golgi apparatus, appear late in schizogony when daughter merozoites separate from the residuum (Dubremetz, 1975; Dubremetz, 1979). In agreement with this scenario, the *E. tenella* microneme proteins EtMIC2 and EtMIC5 gradually disappear during early schizogony and are detected again later as the invasive merozoites mature, suggesting that microneme protein expression is co-ordinated and occurs only when micronemes are being assembled in readiness for the next round of host cell invasion (Tomley *et al.*, 1996; Brown *et al.*, 2000b).

An effective approach to further investigate this suggestion, would be a study of microneme gene and protein expression during the formation of an invasive life-cycle

stage of the parasite. Such an approach would provide information on the extent of co-ordination required for the production of microneme organelles within *Eimeria*, as well as information of the timing of expression of microneme-specific mRNA and protein during the organelles development. The data might also distinguish between constitutively and differentially expressed microneme genes and may prove useful for studies on others members of the phylum Apicomplexa.

3.1.1 Background of the work

Eimeria spp. when compared to other apicomplexan parasites, have large numbers of apical organelles, (micronemes and rhoptries) in the invasive stages of the life cycle. Subcellular fractionation of sporozoites can be used to obtain pure or enriched fractions of micronemes. Sporozoites can easily be purified using anion exchange chromotography, broken by sonication and subcellular organelles separated with techniques such as sucrose density gradient ultracentrifugation (Kawazoe *et al.*, 1992). Enriched fractions can be monitored by electron microscopy for the presence of purified micronemes and from such fractions micronemal proteins were separated by SDS-PAGE, various polypeptides excised, and specific antisera against each antigen was raised (Tomley *et al.*, 1991; Kawazoe *et al.*, 1991; Tomley *et al.*, 1996; Brown *et al.*, 2000b). A sporozoite cDNA library, EtH11, was constructed (Tomley *et al.*, 1996) and immunoscreened with antisera raised against various micronemal proteins. Clones found to be recognised by the microneme-specific antiserum were isolated and sequenced (Tomley *et al.*, 1991; Tomley *et al.*, 1996; Wan *et al.*, 1997; Brown *et al.*, 2000b). Following the identification of the cDNAs as all or part of the microneme proteins EtMIC1-5 a number of morphological, localisation, biochemical and functional experiments were carried out to evaluate the roles of these specialised proteins and the use of the microneme specific cDNAs provided the basis for this project.

3.2 Methods

3.2.1 Oocysts sampling

Oocysts required in the time course experiments were recovered from infected caeca, sporulated and purified using the procedure outlined in section 2.1.2 – 2.1.4, except that partially sporulated oocysts were removed at various time points from the culture. Purified oocysts were finally resuspended in 100 ml of 1 mM sodium dithionate to

prevent further sporulation (Wang and Stotish, 1975). To determine the stage of sporulation, oocysts from each sampling point were wet-mounted onto glass slides in either 90% glycerol or 100 mM Tris pH 7.6 and photographed at 400x or 1000x magnification, respectively by differential interference microscopy.

3.2.2 Protein Analyses

3.2.2.1 Extraction of protein from sporulating oocysts

Sporulating oocysts (1×10^7) from each time point were centrifuged at 14,000 g for 2 min and resuspended in a 1.5 ml tube with 300 µl PBS 7.6, 300 µl of #8 glass beads and 30 µl protease inhibition solution (Sigma P2714). The contents of the tube were then vortexed vigorously and oocyst breakage was monitored by microscopic examination. Vortexing continued until no intact oocysts, sporocysts or sporozoites could be seen. Samples were then freeze-thawed 3 times by immersion of the tube in a mixture of dry ice and methanol for 1 min followed by thawing at 37°C for 2 min, and centrifuged as before. The supernatant was removed and the beads were washed in 200 µl PBS 7.6. After centrifugation at 14,000 g for 1 min, the second supernatant was added to the first. The combined supernatants were sonicated (Soniprep 150, MSE) for three bursts of 20 s pulses at 10 µm amplitude while on ice. The concentration of solubilised proteins was determined from their A_{280} absorbance values and the lysates were stored at -70°C.

3.2.3 Pulse field gel electrophoresis (PFGE)

3.2.3.1 Preparation of chromosomal DNA blocks

Blocks for chromosomal DNA were prepared from purified sporozoites as described by (Shirley *et al.*, 1990; Shirley, 1994). Briefly, sporozoites of *E. tenella* Wis and Wey strains were purified through a DE-52/nylon wool column as described in section 2.1.5 and suspended at 5×10^8 sporozoites ml^{-1} in a warmed (41°C) solution of 62.5 mM EDTA and 1% (w/v) low melting point agarose (Seakem, Flowgen). The agarose was transferred to a 100 µl block mould (BioRad), containing wells measuring 10 x 5 x 1 mm and placed on ice to set. The blocks were then removed and placed in NDS buffer supplemented with 1% (w/v) proteinase K and incubated at room temperature for 48 h with one change of solution. The chromosomal blocks were washed twice in 50 mM EDTA buffer and stored at 4°C (Shirley, 1994). The haploid nucleus of *E. tenella* contains approximately 75 fg DNA, so each chromosomal block contained ~

37 μg DNA (Fernando and Pasternak, 1991; Shirley, 1994). Chromosomal blocks of *Schizosaccharomyces pombe* and *Saccharomyces cerevisiae* (BioRad) were used as electrophoresis markers.

3.2.3.2 Conditions of pulse field gel electrophoresis (PFGE)

Chromosomal DNA blocks of *E. tenella* were subjected to PFGE through gels made using 1.1 g (0.55%) chromosomal DNA grade agarose (BioRad) plus 0.1 g (0.05%) electrophoresis grade agarose (Flowgen) and 200 ml 1 x TBE. The mixture was heated and allowed to set on 21 cm square glass plates with wells 3 or 5 mm in length, along a line of origin of 1 cm.

Agarose blocks containing *E. tenella* DNA were cut into 1 mm-width pieces, placed into the wells of the gel and sealed in with 1% molten agarose. The gel was placed into a CHEF tank (BioRad) with 2 L of 1 x TBE and electrophoresed at 45V for 240h with a ramped pulse time of 1800-6500 s; followed by 48h with a pulse time of 2500s and finally for 30h with a pulse time of 300-1700 s. Gels were stained with ethidium bromide (0.15μg ml^{-1}) for 30 min, destained in 1 x TBE for 1-4 h, photographed and marked for orientation.

3.2.4 Detection of mRNA

Total RNA was extracted as described in section 2.2.9.1.

3.2.4.1 Reverse transcription-polymerase chain reaction (RT-PCR)

RT-PCR was performed using a ProSTAR first strand RT-PCR kit (Stratagene) according to the manufacturer's instructions. First strand cDNA was synthesised by the addition of 300 ng of random primers to 8 μg total RNA made up to a final volume of 41 μl with DEPC-treated water and mixed gently. The sample was incubated at 65°C for 5 min, left at room temperature for 10 min, after which 5 μl of 10 x PCR buffer, 1 U RNase block, 20 nmol of each dNTP (diluted from a 100 mM dNTP set) and 50 U Moloney murine leukaemia virus reverse transcriptase were added. The mixture was incubated at 37°C for 1 h followed by 90°C for 5 min and immediately placed on ice. Mock first strand syntheses were carried out in which reverse transcriptase was omitted.

For the PCR step, 3 μl of the first strand cDNA mixture was added to 8 nmol of each dNTP (diluted from a 100 mM dNTP set), 12.5 pmol of each primer (MWG Biotech or Life Technologies), 2.5 U Taq DNA polymerase (Amersham Pharmacia Biotech)

and the reaction volume made up to 50 µl in 1 x PCR buffer (Amersham Pharmacia Biotech). All reactions included oligonucleotide primers specific for individual microneme genes and for *EtACTIN*.

The samples were vortexed, centrifuged briefly, overlaid with 40 µl of mineral oil and PCR-amplified through 1 cycle at 95°C for 3 min, 30 cycles at 95°C for 15 s, 55°C for 30 s and 72°C for 90 s and one extension cycle at 72°C for 10 min. Two control RT-PCRs were carried out for each experiment. First, an internal control PCR using mRNA and primers supplied in the RT-PCR kit to give a 1.3 kb product. Second, the cDNA control reaction, which did not include MMLV-RT, in order to detect any residual, contaminating DNA in the RNA samples extracted from oocysts.

3.3 Results

3.3.1 Synchrony of development and morphology of oocysts during sporulation

It has been reported that oocyst sporulation in some species of *Eimeria* is asynchronous (Ferguson *et al.*, 1978a; Ferguson *et al.*, 1978b) and that internal structures show heterogeneity in their morphological appearance during time-course experiments. To investigate the extent of variation during sporulation of *E. tenella*, two time-course experiments were carried out. In the first experiment oocysts were sampled over a 3 day period and in the second oocysts were sampled over a narrower time-span during the period of sporoblast maturation and sporozoite formation within the oocyst.

To determine the extent of variation during sporulation of *E. tenella*, samples of oocysts from each time point were viewed at low-power by differential interface microscopy. In the first experiment (Fig 3.1) at 0 h sporulation, in the majority of oocysts the cytoplasm of the sporont was contracted away from the oocyst wall and the nucleus was central (Fig 3.1A). At 6 h, 60% of the oocysts appeared similar to those viewed at 0h but in ~40% of the oocysts the cytoplasm had constricted further (Fig 3.1B), suggesting that nuclear divisions were complete (Canning and Anwar, 1968) and of these around half had progressed to the two sporont stage. By 12 h, sporoblasts were visible in the great majority of oocysts (Fig 3.1C) but the stage of blast formation varied from the "pyramid stage" in ~20% of oocysts to a more mature elliptical shape in ~70% of the oocysts. Around 10% of oocysts remained unchanged from those seen at 0h. By 24 h, four sporocysts, each containing two mature sporozoites, as judged by the presence of the large refractile bodies within the

sporozoite, were clearly visible within 90% of the oocysts (Fig 3.1D). No further morphological changes were seen at 36, 48, 60 and 72 h (Fig 3.1E-F), indicating that sporulation had been essentially complete by 24 h. In the second experiment, at 13 h sporoblasts were visible in ~90% of the oocysts but the stage of development varied from ~15% in the "pyramid stage" and ~75% in the final elliptical stage. By 15.5 h around 90% of the oocysts were in the final elliptical stage of development and the remaining 10% appeared unsporulated. By 18 h development in the sporoblasts formation had progressed and steida bodies were visible in the majority of sporoblasts. By 22.5 h four sporocysts each with two mature sporozoites were visible in 90% of the oocysts and the remaining 10% appeared unsporulated. No further morphological changes were seen at 25 and 29.5 h, demonstrating that sporulation in *E. tenella* (Wis) was complete by 22.5 h.

Overall the results from the two experiments revealed that whilst there was some variation in the degree of cytoplasmic contraction at 0 h and more profound heterogeneity at 6 h, a high degree of synchrony had been achieved by 12 h. This synchrony was then maintained throughout the time course with around 10% of the oocysts remaining completely unsporulated and 90% proceeding to full sporulation by 24 h. This time-scale is similar, but not identical, to that previously reported for *E. maxima* by Canning and Anwar (1968) and for *E. tenella* by Norton and Chard (1983). Similar incubation temperatures were used in all the studies, however development of *E. tenella* during the first 6 hours of sporulation was generally faster than that of *E. maxima*. Morphologically, *E. tenella* oocysts at 6 hours of sporulation were at a stage comparable to *E. maxima* oocysts at 11 hours. The reason for variation between the species is not known, but oocysts of *E. tenella* are smaller than those of *E. maxima* and the latter are more difficult to break by mechanical shearing. It is possible that gaseous exchange across the oocyst wall of *E. maxima* is less efficient than for *E. tenella*, which might account for slower sporulation.

To view the timing of the main events of sporulation more closely, high power micrographs of oocysts were taken at each sampling time. In the first experiment, all oocysts at 0 h (Fig 3.2.1A) contained a single sporont with the cytoplasm contracted away from the oocyst wall. At 6 h (Fig 3.2.1B) around 20% of the oocysts had progressed to the 2 sporoblast stage, as depicted by the arrowhead in Fig 3.2.1B, whereas 80% had the same appearance as oocysts at 0 h. By 12 h (Fig 3.2.1C), 80%

A (0 h)   B (6 h)

C (12 h)   D (24 h)

E (36 h)   F (60 h)
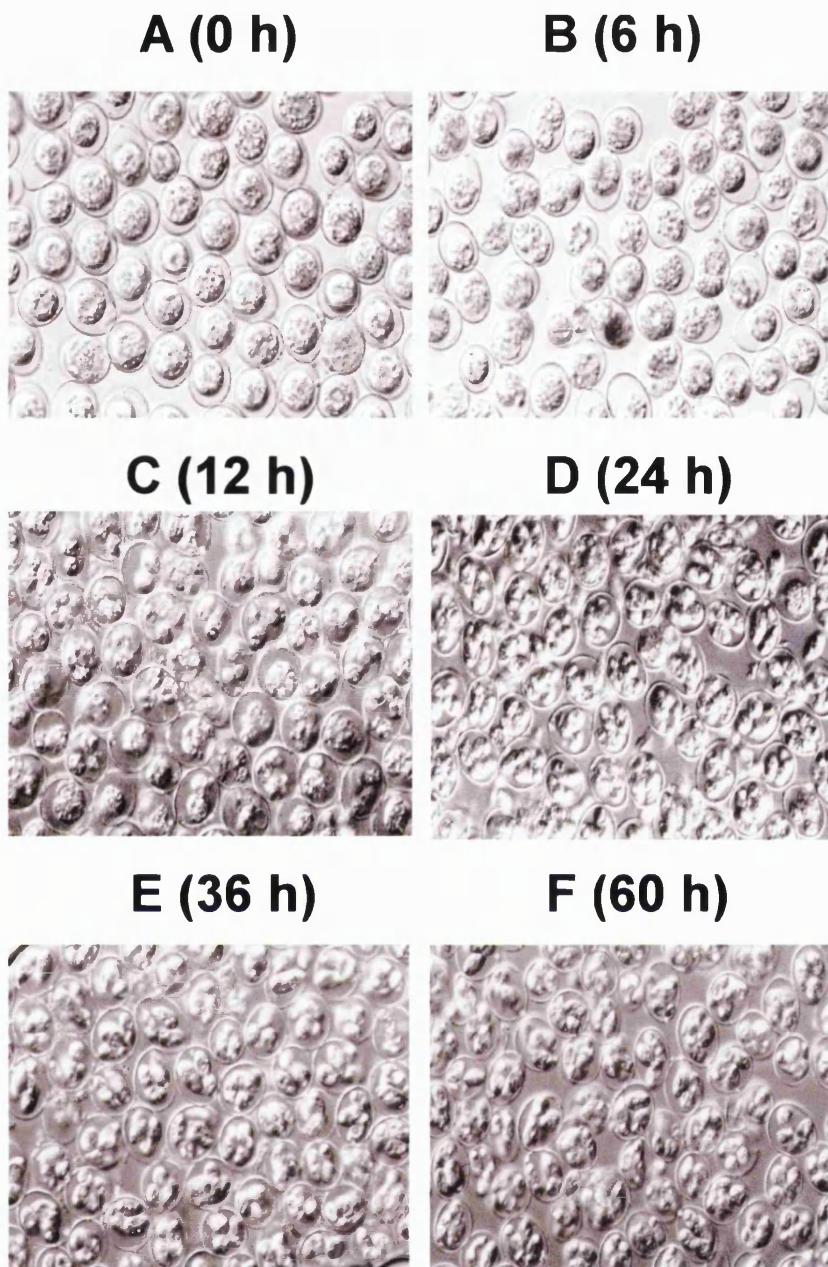
Fig. 3.1 Synchrony of oocysts during sporulation.

To investigate the extent of morphological variation between oocysts of *E. tenella* during sporulation representational fields of oocysts from each sampling point were viewed at low power by differential interface microscopy. It was possible then to determine the level of morphological synchrony. Oocysts were removed at various times during sporulation, pelleted by centrifugation, surface-sterilised with sodium hypochlorite and stored in 1mM sodium dithionite at 4°C. Oocysts from each sampling point were wet-mounted onto glass slides in 90% glycerol and viewed at 400x magnification by differential interference microscopy.

At 0h (A) 100% of oocysts contained a single sporont in which the cytoplasm had contracted away from the oocysts wall and the central position is occupied by the nucleus. At 6h (B) 60% of oocysts are similar to those at 0h but in the remaining 40% the cytoplasm has constricted further (50% of which have progressed to the two sporont stage). At 12h, (C) sporoblasts are visible in the majority of oocysts but the stage of blast formation varies from the "pyramid stage" in ~20% of oocysts to a more mature elliptical shape in ~70% of the oocysts, with the remaining 10% appearing completely unsporulated. By 24h, (D) 90% of the oocysts are mature sporocysts with apparently mature sporozoites, in which a large refractile body is visible. There is no further morphological development beyond this point.

78

1

A (0 h)	B (6 h)	C (12 h)	D (24 h)

E (36 h)	F (48 h)	G (60 h)	H (72 h)

2

I (13 h)	J (15.5 h)	K (18 h)	L (22.5 h)

M (25 h)	N (29.5 h)
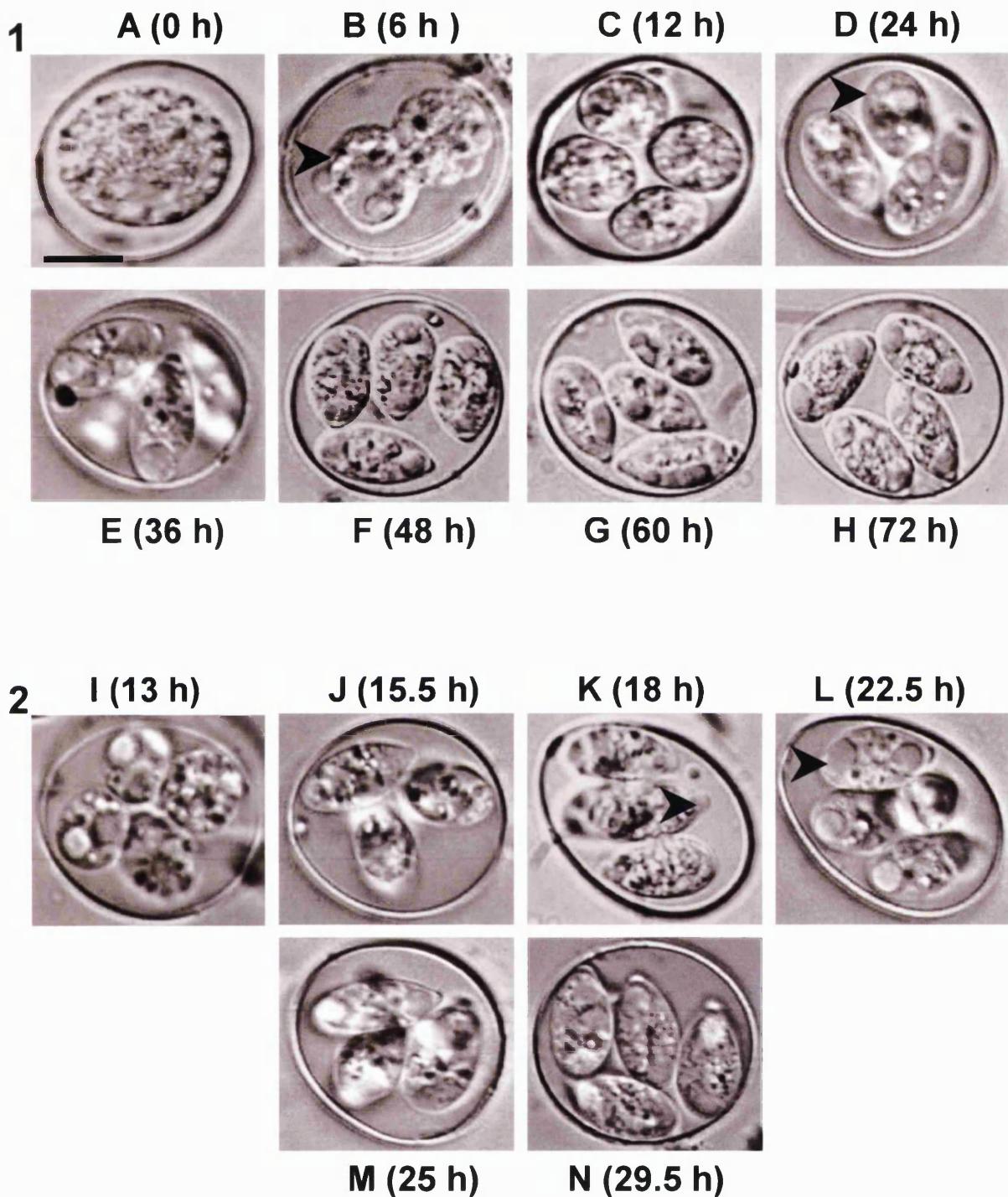
Fig. 3.2 Morphology of *E. tenella* oocysts during sporulation.

Oocysts were removed at various times during sporulation, pelleted by centrifugation, surface-sterilised with 1 mM sodium hypochlorite and stored in 1mM sodium dithionite at 4°C. Oocysts from each sampling point were wet mounted onto glass slides in 100 mM Tris pH 7.6 for photographing at 1000x magnification by differential interference contrast microscopy. Scale bar, 10μ m.

Panel 1: Experiment 1, samples taken from 0-72 h, covering the whole of the first sporulation time course
Panel 2: Experiment 2, samples taken from 13-29.5 h, covering the second time course, during which time sporoblasts and sporozoites mature within the oocyst.

79

of the oocysts had progressed to the 4 sporocyst stage, with no evidence of sporozoite development within the cysts. However, by 24 h (Fig 3.2.1D) 90% of the oocysts contained 4 sporocysts, each of which could be seen to contain 2 mature sporozoites. For all later time-points, the morphological appearance was essentially the same as at 24 h (Fig 3.2.1E, F, G, H). In the second experiment, at 13 h (Fig 3.2.2I) four spherical sporoblasts containing central granular material were visible and by 15 and 18 h (Fig 3.2.2J and K) these blasts had developed into elliptical sporocysts, with had fully formed stieda bodies (arrowhead Fig 3.2.2K) but which did not contain discernible sporozoites. However, by 22.5 h (Fig 3.2.2L) mature sporozoites, each with large refractile bodies (arrowhead), were clearly visible within the sporocysts. No further morphological changes were seen at 25 and 29.5 h (Fig 3.2.2M, N). Thus, under the sporulation conditions used, oocysts of *E. tenella* (Wis) achieved full sporulation, as judged by their morphological appearance, after approximately 22.5 h of incubation.

3.3.2 Appearance of microneme proteins

Cell lysates were prepared from oocysts harvested throughout the two time-course experiments, the proteins were separated by SDS-PAGE (Fig 3.3) and Western blots probed with antibodies specific for the five microneme proteins, EtMIC1-5 (Fig 3.4).

Total protein lysates from all the sampling time points from both time course experiments were loaded onto a single SDS-PAGE gel and stained with Coomassie Brilliant blue to examine gross changes in protein expression over the sampling times (Fig. 3.3). Although equal loadings of all protein samples was intended, as deduced from spectrophotometer readings, it is clear that less protein was loaded from the samples prepared during the second experiment (13, 15.5, 18, 22.5, 25, 29.5 h). Nevertheless, there was good consistency between the two experiments as the following few examples show. Bands A (37 Kda) and B (30 Kda) were detected in all lysates throughout the time course and are likely therefore to correspond to proteins that are constitutively expressed throughout sporulation. Bands C (43 Kda) and D (20 Kda) are representative of proteins that were detected from 1 to 13 h, but not at any of the later time points and are likely to represent proteins important during blastogenesis. In contrast, bands E (50 Kda) and F (15 Kda) were not detected at any of the early time points, but were seen in all lysates from 18 and 22.5 h respectively and thus may be required for late events in the formation of the sporozoite. Overall it
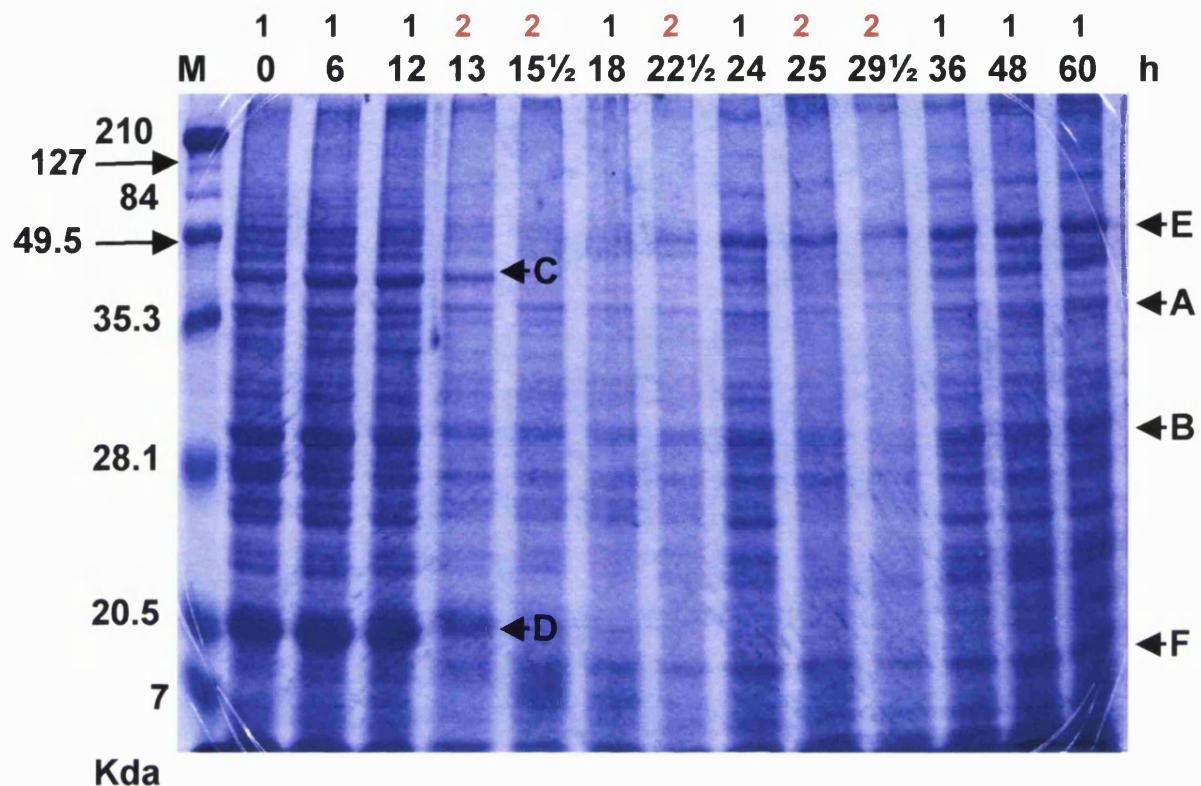
Fig. 3.3 Total lysate analysis of oocysts proteins during sporulation.
Oocysts (10^7) were broken by mechanical shearing and protein samples were examined SDS-PAGE and visualised by staining with Coomassie Blue (section 2.2.10 and 3.2.2). The gel contains molecular weight markers (M; BioRad) in the first lane and approximately 500 ng of protein taken from each time point from both time courses. All oocyst samples taken from both time course experiments are integrated on this gel (in order of sporulation time) to show consistency. Those from experiment one (1) and from experiment two (2) are indicated.
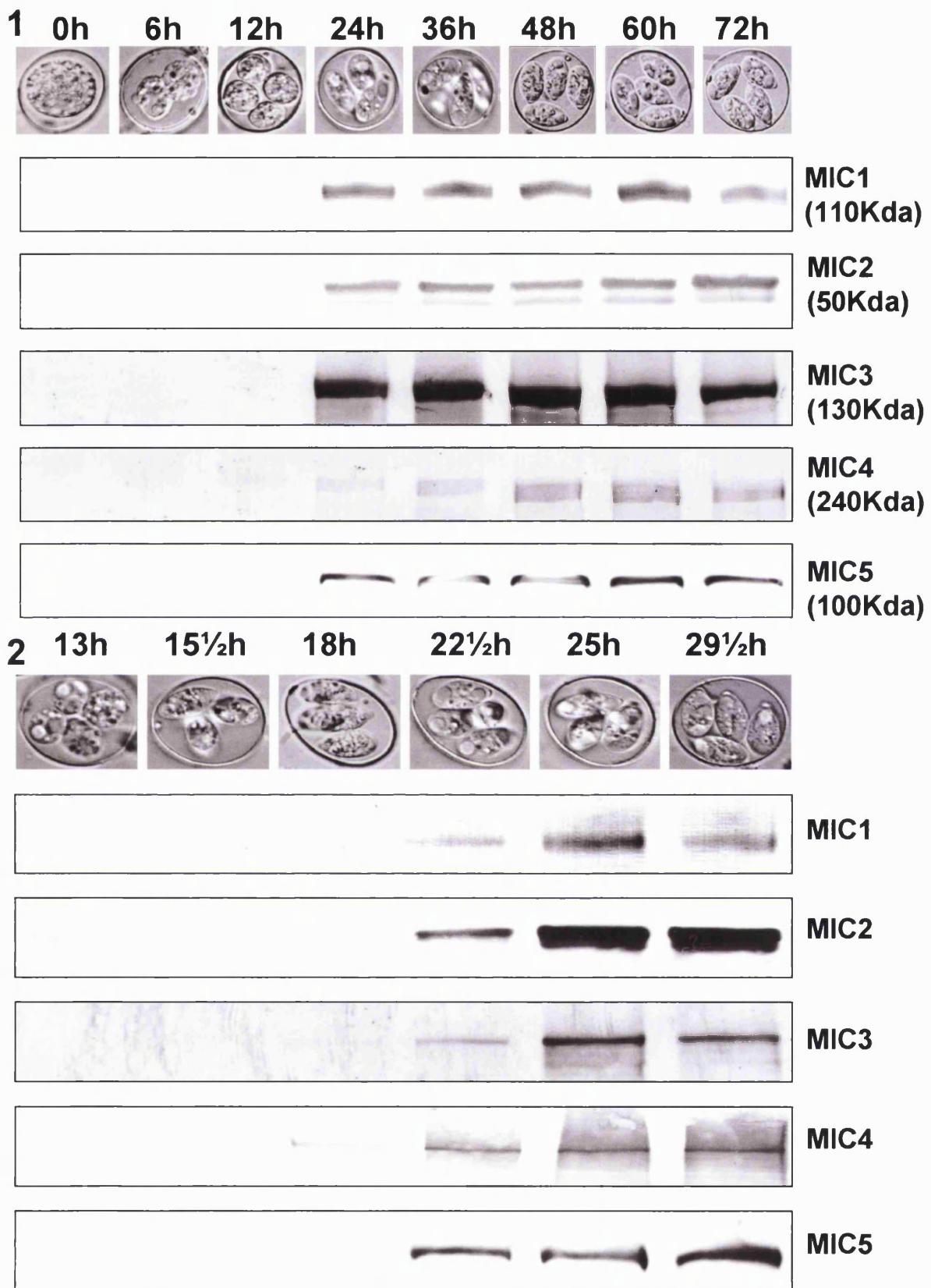
Fig. 3.4 Detection of microneme proteins during oocyst sporulation.

Oocysts (10^7) were broken by mechanical shearing and protein samples (500 ng) examined by Western blotting using monospecific antibodies against microneme proteins EtMIC1-5

Panel 1: Experiment 1, samples taken from 0-72 h, covering the whole of the first sporulation time course

Panel 2: Experiment 2, samples taken from 13-29.5 h, covering the second time course, during which time sporoblasts and sporozoites mature within the oocyst.

is clear that the expression of many proteins is differentially regulated during the sporulation process and it is perhaps not surprising that the most obvious switches in expression occur at times when major morphological changes are occurring in the oocyst, such as sporoblast formation (~13 h) and sporozoite formation (~18-24 h).

Probing Western blots of proteins from the first time-course experiment with antibodies specific for EtMIC1-5 showed clearly that the expression of these proteins is temporally regulated (Fig 3.4.1). In this experiment, all the MICs were detected from 24 h of sporulation onwards, at which time the sporozoites were already fully formed. To examine the timing of the appearance of MICs more closely, Western blots were carried out using protein samples prepared from the second time-course experiment (Fig 3.4.2). All five MICs were detected from 22.5 h onwards a time which corresponds to the earliest sampling time at which fully formed sporozoites could be seen within the sporocysts. In both time courses, EtMIC4 could be detected, very faintly, from 6 h onwards and EtMIC3 could be detected, also very faintly, at 18h in the second time course. This early, low level expression of EtMIC4 and EtMIC3, was reproducible between several different blots, indicating that these proteins may be switched on earlier than the remaining MICs examined. A more sensitive, chemiluminescent, detection method (ECL plus) was also used, but none of the other MICs was detected at earlier time points (data not shown).

3.3.3 Chromosomal localisation of genes encoding microneme proteins

Since there is a high degree of synchrony in the expression of the five MICs during sporulation, the chromosomal location of genes encoding these proteins was determined to see whether they are clustered together in the genome. In   T. gondii, it has been shown that the genes encoding microneme proteins TgMIC8 and TgMIC9 are located next to each other on the same chromosome (Meissner et al., 2001).

Probes corresponding to cDNAs specific for EtMIC1-5 were hybridised to Southern blots of separated chromosomes of E. tenella (strains Wey and Wis). Most probes hybridised to a single chromosome in each parasite. EtMIC1 was found on chromosome 13, EtMIC2 and EtMIC5 on chromosome 9 and EtMIC4 on chromosome 5. The probe corresponding to EtMIC3 hybridised to chromosome 3 of the Wis strain and to chromosomes 3 and 9 of the Wey strain. This most likely represents cross-hybridisation of the EtMIC3 probe to other, related, sequences in the Wey genome. Thus, there does not appear to be any clustering of the five genes encoding MICs

within the *E. tenella* genome. However, this simple PFGE southern blot experiment does not give any information on the location of the *MIC* genes within the individual chromosomes.

3.3.4 Expression of microneme-specific mRNAs during sporulation

To determine whether the co-ordination of MIC expression during sporulation is likely to be controlled at the level of transcription or translation, total RNA was isolated from oocysts sampled during the time course experiments and from freshly excysted sporozoites, and subjected to specific RT-PCR reactions. A list of the primers used to amplify each gene is given in Table 3.2. Before use, RNA preparations were checked for purity and quality by gel electrophoresis (Fig 3.6) and quantified by spectrophotometry. To serve as positive internal controls, primers specific for *EtACTIN* (expected to be expressed constitutively) were included in each RT-PCR reaction. Controls lacking reverse transcriptase were set up for each reaction to ensure that contaminating residual genomic DNA did not contribute to the PCR signals. No signals were obtained in any of these RT negative controls (data not shown). Messenger RNAs specific for each of *EtMIC1-5* were detected in oocysts from 12 h of sporulation and remained detectable throughout the remainder of the time course (Fig 3.7.1 and 3.7.2). In addition, for *EtMIC3* and *EtMIC4*, distinct signals were also detected at 6 h of sporulation (Fig 3.7.1). Messenger RNAs specific for each of *EtMIC1-5* were also detected in freshly excysted sporozoites (Fig 3.8), which is consistent with the ability to isolate cDNA clones from expression libraries derived from sporozoite mRNA. RT-PCR reactions were carried out several times from batches of RNA prepared on three separate occasions and the same pattern of mRNA expression was detected on each occasion. Thus, it seems that all the microneme-specific mRNAs are expressed at the time of sporoblast formation onwards but those for *EtMIC3* and *EtMIC4* are switched on some hours earlier.

PFGE   EtMIC1   EtMIC2   EtMIC3   EtMIC4   EtMIC5
        (12)   (9)   (3)   (5)   (9)
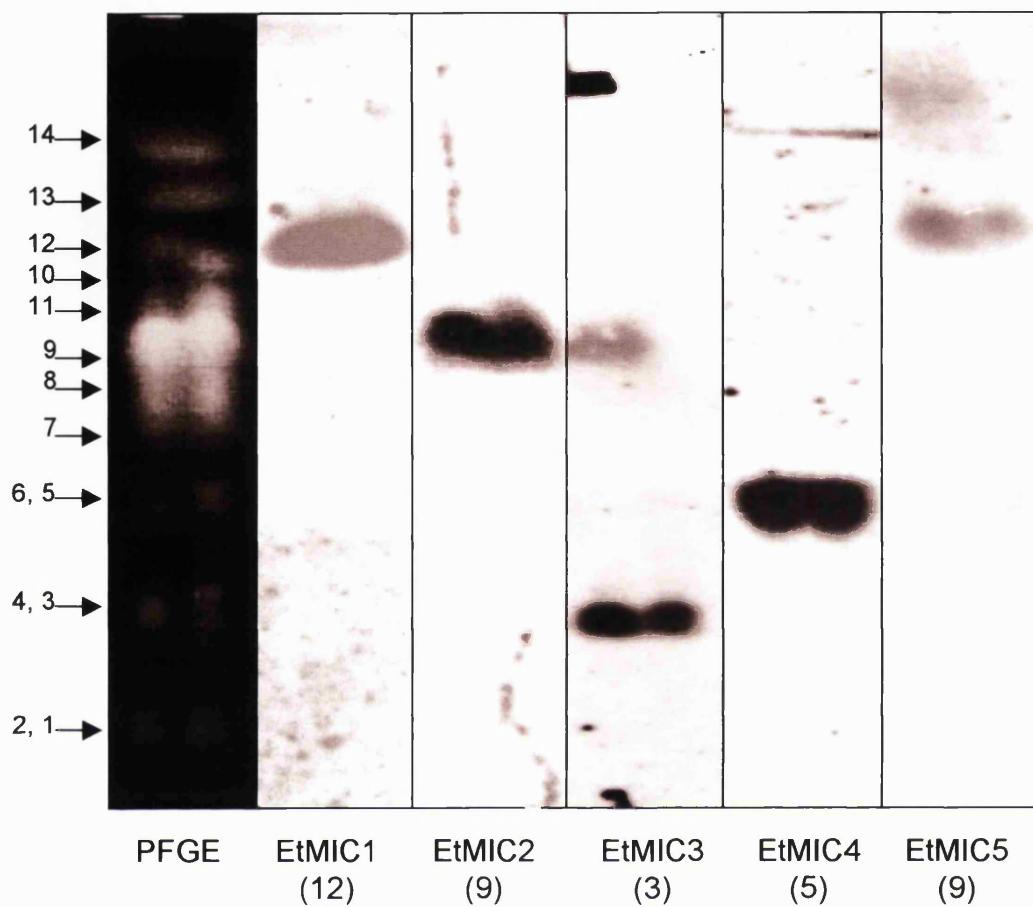
Fig. 3.5 Chromosomal localisations of *EtMIC1-5*.

Chromosomes of two strains of *E. tenella* were separated by pulsed field gel electrophoresis (PFGE, section 3.2.3) and stained with ethidium bromide (left panel, lhs, *E.tenella* Wey; rhs, *E. tenella* Wis). Numbers assigned to chromosomes, which range in size from 1 (chromosomes 1) to 7 Mb (chromosome 14) are indicated by arrows. The line of origin is at the top of the figure and the PFGE conditions gave separation of the major chromosomes without any compression zone just below the origin. Chromosomes of *E. tenella* Wis and Wey from gels subjected to identical PFGE conditions were transferred to filters and probed with sequences from *EtMIC1-5*. Probes and their chromosomal locations are given below the panel and in Table 3.1. The chromosomal locations of *EtMIC3-5*, which hybridised to bands that contain two chromosomes under the conditions shown, were confirmed using gels run under different PFGE conditions (data not shown).

Location of microneme gene	Chromosome number	Size (Mbp)
-	1	1.05
-	2	1.20
EtMIC3	3 (9 in Wey)	2.0
-	4	2.2
EtMIC4	5/6	3/3.5
-	7	3.5
-	8	3.7
EtMIC2 and EtMIC5	9	4.5
-	10	4.5
-	11	5
-	12	5.5
EtMIC1	13	>6
-	14	>6

Table 3.1 Chromosome localisations of *EtMIC1-5*.
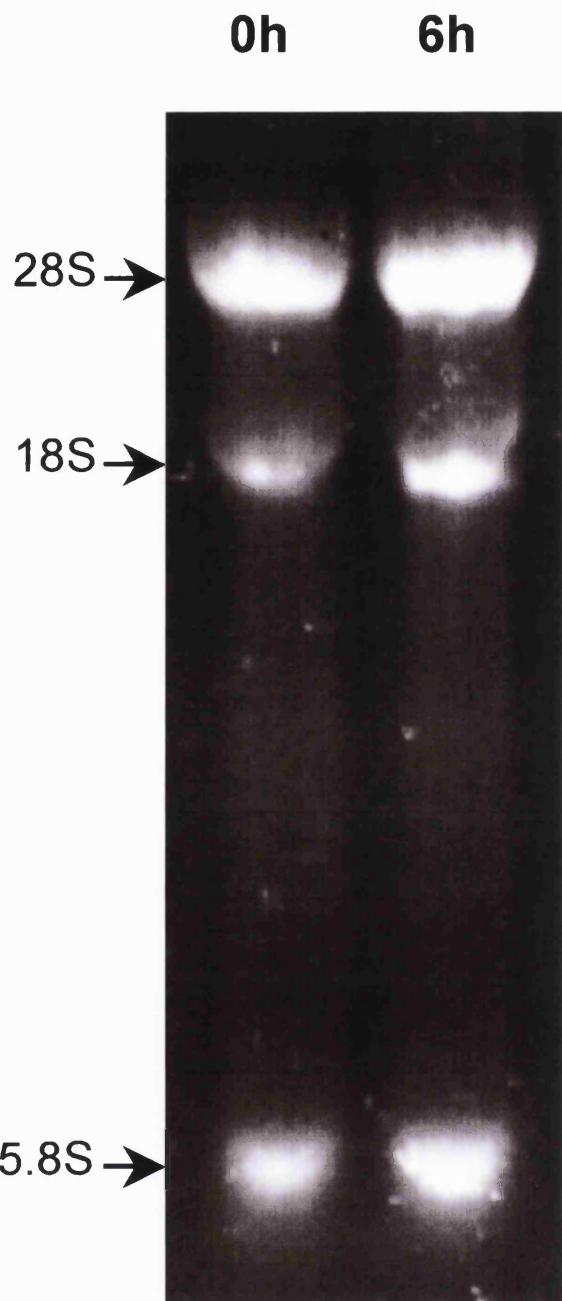
Fig. 3.6 Fidelity of total RNA.
Oocysts (25×10^6) from each sampling point were broken by mechanical shearing and total RNA was extracted using a PureScript total RNA extraction kit, as described in sections 2.2.9 and 3.2.4.1. 50 ng of total RNA from time points 0 and 6 h of sporulation was electrophoresed through a 1% RNase free agarose non-reducing gel and visualised with RNase free ethidium bromide. The large (28S and 18S) and small (5S) ribosomal subunits are indicated.
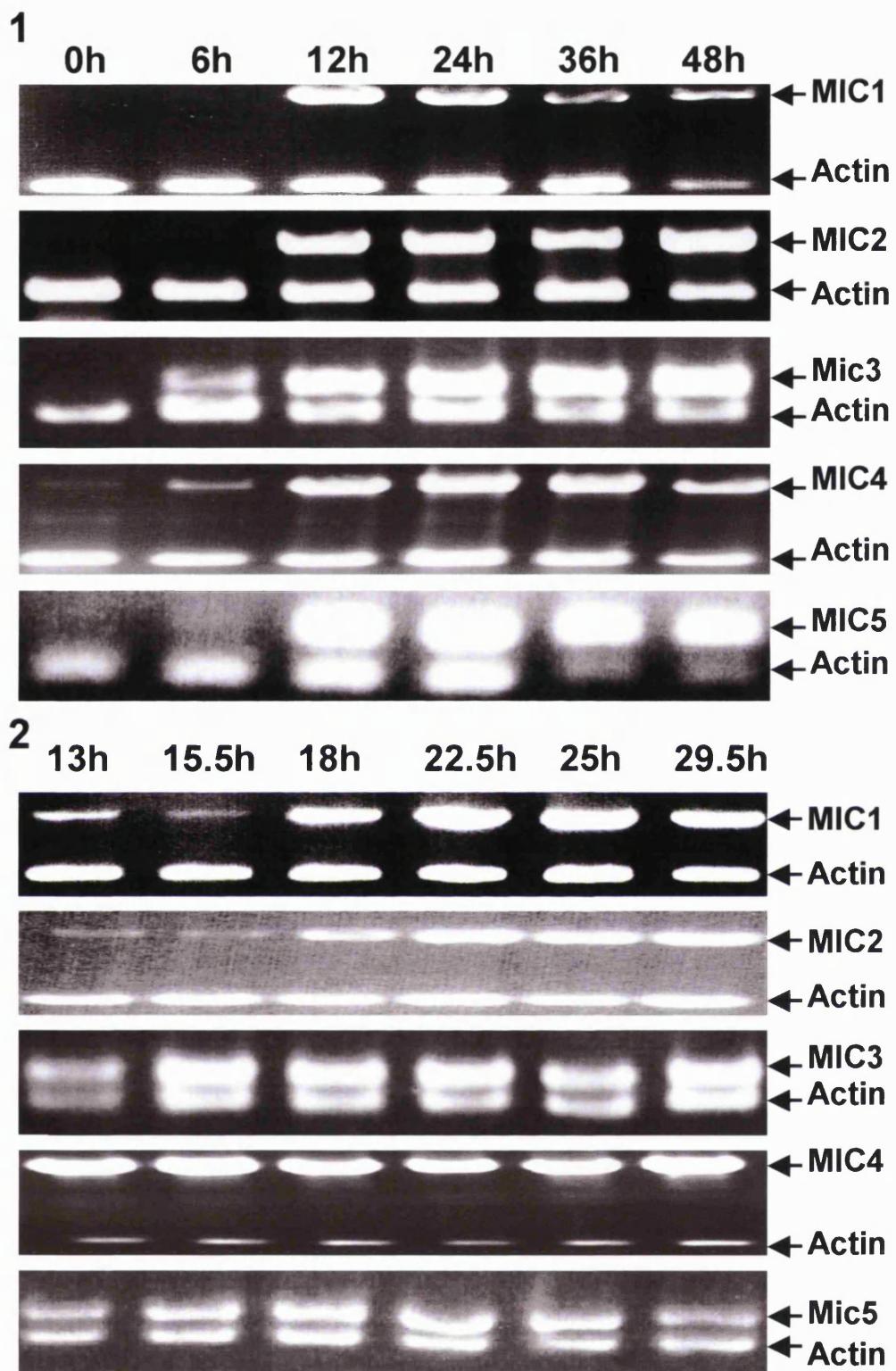
Fig. 3.7 Detection of microneme-specific RNA during oocyst sporulation.
Oocysts were broken by mechanical shearing and total RNA extracted (section 2.2.9). 8μg samples of RNA were subjected to RT-PCR reactions (section 2.2.4.1) using primers specific for *EtACTIN* and for each *EtMIC* gene (Table 3.2). Control reactions, in which RT was omitted, were done on all samples and were negative (data not shown).
Panel 1: RT-PCR results from experiment 1, samples taken from 0-72 h, covering the whole of the first sporulation time course
Panel 2: RT-PCR results from experiment 2, samples taken from 13-29.5 h, covering the second time course, during which time sporoblasts and sporozoites mature within the oocyst.

88
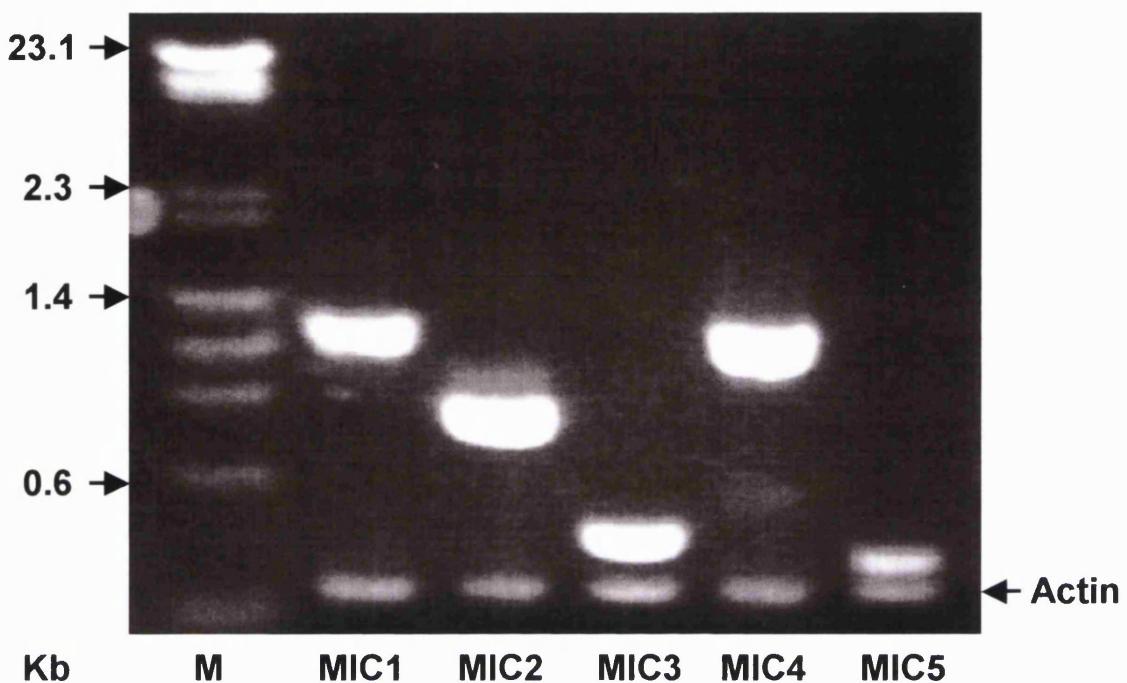
Fig. 3.8 Detection of microneme-specific RNA from sporozoites.
Freshly excysted sporozoites were broken by mechanical shearing and total RNA extracted (section 2.2.9). 8µg samples of RNA were subjected to RT-PCR reactions (2.2.4.1) using primers specific for *EtACTIN* and for each *EtMIC* gene. Control reactions, in which RT was omitted, were done on all samples and were negative (data not shown). The ethidium bromide stained gel shows RT-PCR results from *EtMIC1-5* genes only (the upper product in each lane), with *EtACTIN* acting as the internal control (the smaller product in each lane). Molecular weight markers (M) are indicated in the first lane.

Gene	Primers	Sequence (5' to 3')	Product size
EtMIC1	Mic23	TTGGTCATGACTGACGGC	1209 bp
	Tsp5D	GTGCAAGCTTAGCATGGAACTTCATTGCATC	
EtMIC2	Mic2rr5.1	GAGCGAACGGGACTTCATTG	832 bp
	Mic2rr3.1	ACTCTGCTTGAACCTCTTCC	
EtMIC3	Mic3g	TGTCGCTGTCAATGACCGCTTGAA	451 bp
	Mic3b	GAGGCCGCGGGGCCAGGCTGTGTA	
EtMIC4	Mic4rr5.1	CCACGCCTCTTGTGCCAACA	1127 bp
	Mic4rr3.1	GAAGGTGGTGTTGTCGTCGC	
EtMIC5	Pjb6	TTCCGTCAGGGCGTTGGATAC	400 bp
	Pjb7	ACTTCGTAGGCCGAAGGGCTG	
EtACTIN	Act1	CTGTGAGAAGAACCGGGTGCTCTTC	350 bp
	Act8rr	CGTGCGAAAATGCCGGACGAAGAG	

Table 3.2 Oligonucleotide primers for RT-PCR analysis. Combinations of primers used for the amplification of microneme-specific products from cDNA (amplified by random primers) in RT-PCR analyses, including primer sequence and predicted size of products.

3.4 Discussion

The utility of oocyst sporulation in *E. tenella* for examining the expression of genes encoding proteins that reside in the microneme sub-cellular organelle has been explored. Dramatic morphological changes occur within the oocyst during sporulation and these culminate in the production of invasive sporozoites. During the time of sporocyst and sporozoite formation, sporulation is highly synchronised and around 90% of the culture proceeds to full sporulation by 24 h. This is in agreement with timings that were reported previously, some 24.8 h for 95% of *E. tenella* (Wis) oocysts to complete sporulation (Norton and Chard, 1983). Thus, samples taken from time-course experiments should give valuable insights into patterns of specific gene expression during the differentiation process.

From Western blotting studies it is clear that the timing of expression of microneme proteins is highly co-ordinated. Only one protein, EtMIC4 was detected at low levels before the time sporozoites were reaching maturation within the sporocyst (22.5 h). From 22.5 h onwards, all five of the microneme proteins examined were present. This is similar to the timing of the appearance of EtMIC2 and EtMIC5 during first generation schizogony, when the proteins were detected only from the time at which daughter merozoites were forming (Tomley *et al.*, 1996; Brown, 2000a). Thus it appears that microneme proteins are made only at the time when, presumably, microneme organelle biogenesis occurs. Since all five MICs displayed such a high level of synchronicity, their expression is clearly highly regulated, but an absence of clustering of the genes within in the *E. tenella* genome, suggests that this regulation is not due to positional effects.

Another protozoan parasite, which displays simultaneous transcription of multiple genes is *P. falciparum*. Cytoadherence of infected erythrocytes to the vascular endothelium is mediated by a family of highly polymorphic proteins known as erythrocyte-membrane-protein-1 (PfEMP-1) (Roberts *et al.*, 1992; Smith *et al.*, 1995). These proteins are antigenically unique and the switching of the individual proteins is an important mechanism for the maintenance of chronic interactions (Borst *et al.*, 1995; Bull *et al.*, 1995). Approximately 50 *var* genes encode these proteins (Baruch *et al.*, 1995) which are separated throughout the *P. falciparum* genome on multiple chromosomes (Hernandez *et al.*, 1997; Rubio *et al.*, 1996). Synchronous transcription of several *var* genes from different chromosomes has been detected in infected

erythrocytes (Chen *et al.*, 1998). Similar expression patterns have been observed in other gene families of *P. falciparum* such as *stevor* (subtelomeric, variable open reading frames) and *rif.* RT-PCR analysis of genes from both families has demonstrated simultaneous transcription from multiple genes and the genes have been localised to various different chromosomes (Cheng *et al.*, 1998)

To determine whether oocyst sporulation time courses are useful for examining gene expression at the mRNA level, a series of RT-PCR reactions were carried out. The positive control for the study was *EtACTIN* (a single copy gene encoding the *E. tenella* actin protein), and an RT-PCR product of the predicted size was obtained with RNA samples taken throughout the time-courses, indicating that the gene is expressed constitutively. In contrast, primers for *EtMIC1-5* revealed the presence of mRNAs for *EtMIC1-5* from 6 or 12 h into sporulation, indicating that there is regulation of expression between the unsporulated and the sporulating oocyst stages. Whether this temporal regulation is due to differences in transcription between the stages or is due to post-transcriptional effects, such as differential mRNA turnover, message stability, translation or protein degradation remains to be determined. Interestingly, mRNAs for *EtMIC3* and *4* were detected earlier than those for *EtMIC1, 2* and *5* (Fig 3.6) which correlates with the slightly earlier appearance of these two proteins by Western blotting. None of the MICs were detected until 10-12 h after detection of their specific mRNAs, indicating that post-transcriptional factors are important in the regulation of MIC expression. The apparent co-ordination of mRNA expression for the different microneme genes suggests that there may be transcriptional control of their regulation and this could be governed by DNA sequences that flank the transcriptional start sites and influence the binding of transcriptional complexes.

Overall, this study has shown that the oocyst offers a convenient, synchronous system for analysing the relative abundance of specific products of gene expression during one phase of parasite differentiation. The end products of the differentiating diploid nucleus are eight sporozoites, the first invasive stages of the life-cycle. This sporulation model has now been further exploited in our laboratory for examining the expression for eimepsin, an *E. tenella* aspartyl proteinase, which is also developmentally regulated during sporulation (Jean *et al.*, 2001).

Further analysis of the transcription kinetics of the genes encoding MICs would require metabolic labelling and/or inhibitor studies of mRNA, or the development of

an *in vitro* transcription system for this parasite. A potential disadvantage of the oocyst is its impermeability to exogenous metabolites. Oocysts have a tough, two-layered wall composed of glycoprotein and lipid (67% protein, 19% carbohydrate, 14% lipid), which is resistant to treatment with harsh chemicals. To date the oocyst wall has been found to be permeable to only a small number of uncharged or hydrophobic molecules such as gases (e.g. ammonia) or organic solvents. A range of studies has shown major changes in mRNA and protein abundance during sporulation (Abrahamsen *et al.*, 1994; Ellis and Thurlby, 1991; Sutton et al., 1989; Tomley, 1994a). After treatment of oocysts with sodium hypochlorite and DMSO, Wang and Stotish reported that labelled nucleotides and amino acids could be introduced into the oocyst (1975) but attempts by myself to reproduce this method for the incorporation of uridine into RNA and leucine into protein failed consistently (data not shown).

As metabolic labelling was unsuccessful and an *in vitro* transcription system was not available for the parasite, no further attempts were made in this study to analyse mRNA synthesis. However, since the results from the time-course experiments strongly suggested that the transcription of genes encoding MICs is differentially regulated during sporulation, a detailed examination of the sequences that flank these genes was undertaken in order to identify and characterise regulatory regions.

Chapter Four
Identification and analysis of flanking genomic regions of the genes encoding the microneme proteins EtMIC 1-5

4.1 Introduction

The stages at which a cell can regulate the protein it makes include (1) controlling when and how often a gene is transcribed (transcriptional control), (2) controlling how the primary RNA transcript is spliced or processed (RNA processing control), (3) selecting when completed mRNAs in the cell nucleus are exported to the cytoplasm (RNA transport control), (4) selecting when mRNAs in the cytoplasm are translated by ribosomes (translational control), (5) selectively destabilising certain mRNA molecules in the cytoplasm (mRNA degradation control) and finally (6) selectively activating, deactivating or compartmentalising specific protein molecules after they have been made (protein activity control).

In higher eukaryotes, three RNA polymerases (pol I, II and III) transcribe nuclear genes. Pol II transcribes mRNAs that will be translated into proteins, pol I transcribes genes for ribosomal RNA and pol III synthesises small stable RNAs. In order for pol II to transcribe a gene, over 20 transcription factors (TF) proteins must be assembled at the promoter. An extensive body of work has contributed to the production of a schematic model for the assembly of the transcription initiation complex, based mainly on kinetic assays, native gel electrophoresis and nuclease protection assays (reviewed in Buratowski *et al.*, 1989; Conaway and Conaway, 1993; Zawel and Reinberg, 1993). The overall steps involved include formation of a pre-initiation complex, initiation, promoter clearance, elongation and termination. The first transcription factor protein to bind to the template DNA, is TFIID which is composed of two types of components, a TATA-binding protein (TBP) and TBP-associated factors (TAFs). This complex recognises the promoter region whereby the TBP binds to the minor groove of DNA at a TATA (or Goldberg-Hogness) box, forming the pre-initiation complex. This is the only transcription factor of the initiation complex which makes sequence-specific contact with the template DNA. The TATA box has a consensus sequence "TATa/tAa/t" and is usually located ~25-30 bp upstream from the transcription start site (reviewed in Breathnach and Chambon, 1981; Novina and Roy, 1996). TATA activity can be imparted by a wide variety of AT-rich elements (Singer

et al., 1990), which is why it was thought that various TATA sequences were recognised by multiple distinct TBPs within a cell (Wefald *et al.*, 1990). However some of the TBPs isolated have been shown to functionally interact with numerous AT-rich elements (Zenzie-Gregory *et al.*, 1993). A highly divergent TBP has been characterised in *P. falciparum* (McAndrew *et al.*, 1993), and it seems likely that a TBP exists in *E. tenella* and is involved in promoter function.

Following the binding of TFIID to the template DNA surrounding the TATA box, the complex acts as a binding site for the addition of TFIIA which may activate TBP by relieving a repression that is caused by the TAFs. Next TFIIB binds downstream of the complex (adjacent to the TBP) and is thought to protect the template strand around the transcription start site (between -10 and $+10$). The complex then recruits TFIIF and RNA pol II. TFIIF consist of a large subunit, RAP_{74} and a smaller subunit, RAP_{38}. The large subunit has an ATP-dependent DNA helicase activity required for melting the DNA at the initiation site while the smaller unit binds tightly to the RNA pol II (which contains between 8 and 14 subunits) bringing it into the complex. TFIIE, TFIIH and TFIIJ are next to associate with the complex. TFIIH has multiple activities including an ATPase, helicase and kinase activity, the latter of which is thought to phosphorylate the cytoplasmic tail domain (CTD) of RNA pol II in order to release the TFII factors from it and allow elongation (Bunick *et al.*, 1982; Sawadogo and Roeder, 1984; Parvin and Sharp, 1993; Tyree *et al.*, 1993). The TATA box aligns the RNA pol II on the template DNA strand, *via* the interaction between the TBP and the TATA box, to ensure that it initiates at the correct location and this explains why it is at a fixed location with respect to the transcription start site.

These promoter elements are typical of most eukaryotic genes and are required for their basal expression. The activity of many promoters can be further characterised by additional regulatory elements known as enhancers, which assist initiation. The enhancer must be on the same molecule of DNA, but its position is not fixed and can be several thousand nucleotides upstream or downstream from the transcription start site, even in another gene (Fromm and Berg, 1983).

Not all protein-encoding genes of higher eukaryotes contain TATA boxes; many contain an initiator (Inr) motif, which is functionally analogous to the TATA box in that both are involved in directing accurate transcription initiation (Breathnach and Chambon, 1981; Smale and Baltimore, 1989; Smale, 1994; Hampsey, 1998). Many

genes are quite heterogeneous and can contain both elements (composite), a TATA box and/or an Inr motif (distinct) or neither (null). For transcription initiation to occur in Inr-containing promoters, similar transcription factors to those required for TATA-containing genes are needed. For example, it is known that a TBP of TFIID is required for initiation (Pugh and Tjian, 1990; Pugh and Tjian, 1991) and both RNA pol II (Caracamo *et al.*, 1991) and TFII-I (Roy *et al.*, 1991) have also been implicated. In the initiation process a selector initially binds or "tethers" TFIID to the promoter, controlling the position of the pre-initiation complex, as the components of TFIID do for TATA-containing promoters, then the Inr interacts with an unknown component (thought to be an Inr-binding protein, IBP) of the pre-initiation complex and controls the exact start site. However the process of initiation is not fully understood. To date it is unknown if an IBP exists. Inr motifs localise a transcription start site and mediate their action in the presence or absence of a TATA box or other promoter elements. The activity of an Inr can be carried out directly from its position, which usually overlaps the transcription start site. Alone, the Inr is capable of initiating low levels of transcription and one or more binding sites for activator or repressor proteins are typically present to regulate the efficiency and specificity of transcription. TATA-less promoters are usually associated with "housekeeping" genes and degenerate Inrs serve as promoters in some cases (Smale and Baltimore, 1989; O'Shea-Greenfield and Smale, 1992). The Inr element was initially characterised as a pyrimidine rich sequence (5'-PyPyPyCA^{+1}PyPyPyPyPy-3') where Py represents a pyrimidine and A^{+1} (adenosine) is the start of transcription (Breathnach and Chambon, 1981; Javahery *et al.*, 1994; Smale and Baltimore, 1989; Hariharan and Perry, 1990). Mutational analysis demonstrated that the stretch of pyrimidines in the Inr is not always necessary for efficient transcription, resulting in a loose consensus with the sequence of 5'-PyPyA^{+1}NA/TPyPy-3' (Smale and Baltimore, 1989), where N represents any nucleotide. Also, sequence analyses of well-conserved Inr motifs in >500 eukaryotic genes identified 5'-TCA^{+1}NT-3', as the minimal sequence usually conserved around the transcription start site, however this was not always present (Butcher, 1990).

Very little is known about the regulation of gene expression in protozoan parasites in general, and in apicomplexans in particular. However some studies have begun to define transcription and the motifs critical for promoter function. Trypanosomatids, for example *T. brucei* and *Leishmania* spp., transcribe protein-coding genes in long polycistronic transcription units using either pol II or pol I-like polymerases (Muhich

and Boothroyd, 1988; Tschudi and Ullu, 1988). *Trichomonas vaginalis* does not contain TATA-like elements but does have Inr motifs together with unique upstream elements which have been shown to be sufficient for high transcriptional levels (Quon *et al.*, 1994; Liston and Johnson, 1999). *Entamoeba histolytica* is again different, as its promoters have a tripartite architecture and contain three core promoter elements; a modified TATA box, an Inr and a GAAC motif, all of which play roles in transcription initiation (Singh *et al.*, 1997). Many *Giardia lamblia* promoters contain several degenerate CAAT boxes, simple TTT or AAA elements and modified Inr motifs at the transcription initiation site (Holberton and Marshall *et al.*, 1995; Elmendorf *et al.*, 2001). Promoter analysis of *G. lamblia α-giardin, δ-giardin* and *α2-tubulin* indicates that very short regions of the upstream DNA are sufficient for expression (Singer and Rogers, 1998; Sun and Tai, 1999).

Studies on apicomplexan gene expression have shown that regulation of gene activity occurs at both the transcriptional (GBP130 and KAHRP genes of *P. falciparum;* Lanzer *et al.*, 1992a) and post-transcriptional (MSA-1 gene of *P. falciparum;* Lanzer *et al.*, 1992a) levels. Transcription in *P. falciparum* is sensitive to α-amanitin indicating that RNA pol II is responsible (Lanzer *et al.*, 1992a). Genes encoding RNA pol II and associated transcription factors such as TFIID and TFIIA are in the *P. falciparum* genome. Upstream sequences of protein encoding genes contain motifs similar to the binding sites of common eukaryotic transcription factors (Lanzer *et al.*, 1992b; Lanzer *et al.*, 1993; Su and Wellems, 1994; Alano *et al.*, 1996; Horrocks and Kilbey, 1996) and some have been shown to bind nuclear proteins (Lanzer *et al.*, 1992a; Lanzer *et al.*, 1992b). A region in the promoter of *pcna* has a classical enhancing ability (Crabb and Cowman, 1996; Horrocks and Kilbey, 1996; Su and Wellems, 1994) and enhancers have also been observed in the calmodulin and dhfr-ts promoters (Crabb and Cowman, 1996). In the ookinete (mosquito) stage gene pgs28, sequences from −337 to the ATG start codon, are required for normal gene expression. Furthermore, transcription can proceed independently of a conventional TATA sequence. A similar phenomenon evidently occurs in *T. gondii* as many of the protein-coding genes investigated thus far lack typical eukaryotic promoter elements, such as the TATA box but do contain Inrs. Inr motifs have also been found in a variety of *Toxoplasma* genes, including surface antigen 1 gene (*SAG1*), dense granules protein (*GRA1, 2, 5* and *6*), the tubulin gene (*TUB1*) (Soldati *et al.*, 1995; Mercier *et al.*, 1996) and in *NTPase1* and *3* (Nakaar *et al.*, 1992). Analysis of sequence databases of

Eimeria,   (http://www.sanger.ac.uk/projects/E_tenella/),   Theileria

(http://www.sanger.ac.uk/projects/T_annulata/)   and   Toxoplasma

(http://www.sanger.ac.uk/projects/T_gondii/) revealed the presence of RNA pol II and

transcription factors TFIID and TFIIA.

Other motifs critical for promoter function of some apicomplexan genes include cis-acting polypurine motifs. For example, an A/TGAGAG motif was identified in the promoter region of a variety of protein-coding genes including the histone and tubulin genes of the ciliate protozoan *Tetrahymena pyriformis* (Barahona *et al.*, 1988; Brunk and Sadler, 1990), however in both cases a functional role was not confirmed. Similar polypurine motifs have been observed in *T. gondii* genes including *SAG1*, *GRA1*, *2*, *5* and *6*, *TUB1* and *NTPase1* and *3* and in some cases have been shown to be essential for high expression and for positioning during transcription initiation (Nakaar *et al.*, 1992; Soldati *et al.*, 1995; Mercier *et al.*, 1996).

E. tenella alternates between the chicken and the external environment and displays numerous morphologically distinct development stages during its life-cycle. The assumption that *Eimeria* development entails some degree of co-ordinated gene expression was confirmed using an oocyst sporulation model. It was shown in Chapter Three that microneme proteins EtMIC1-5 are expressed synchronously at the time sporozoites mature, suggesting that their expression is tightly regulated. Since there is no clustering of these genes within the genome, co-ordinated expression is unlikely to be due to positional effects or polycistronic promoters. RT-PCR analyses revealed co-ordination of microneme-specific mRNA expression, suggesting a form of transcriptional control which could be governed by DNA sequences that flank the transcription start sites and influence the binding of transcriptional complexes. The next aim of the project was therefore to map the transcriptional start sites of the microneme-encoding genes using 5'RACE and to analyse the genomic DNA surrounding these sites. As a preliminary step DNA fragments lying upstream of the translation initiation codon and contiguous with the existing characterised cDNAs, were isolated, cloned and sequenced.

4.2 Methods

4.2.1 Rapid amplification of cDNA ends (RACE)

5'RACE reactions were performed using a 5'/3'RACE kit from Boehringer-Mannheim (Germany), according to the manufacturer's instructions. A list of all the

primers used is given in Table 4.2 and a schematic of the 5'RACE protocol is shown in Fig 4.2.1.

First strand cDNA was synthesised from 2 μg total *E. tenella* RNA by the addition of 12.5 pmol gene-specific primer 1 (Sp1), 20 nmol of each dNTP (100 mM dNTP mixture), 12.5 pmol cDNA synthesis primer and 2 U AMV reverse transcriptase in 20 μl 1x cDNA synthesis buffer. The reaction was incubated at 55°C for 60 min followed by 65°C for 10 min to inactivate the AMV reverse transcriptase. First strand cDNA was purified from unincorporated nucleotides and primers using a High Pure PCR Product Purification Kit (Boehringer-Mannheim), according to the manufacturer's instructions. Briefly, cDNA was bound to a glass fiber fleece and impurities such as salt, small oligonucleotides and dimerised primers were washed through with a high salt buffer. Nucleic acids were then eluted from the fleece in water.

A homopolymeric oligonucleotide tail (poly dA) was added to the 3' end of the cDNA by incubating 19 μl of cDNA, 2.5 μl 10x reaction buffer and 5 mM dATP at 94°C for 3 min. Following the addition of 10 U of terminal transferase, the mixture was incubated at 37°C for 30 min, followed by 70°C for 10 min to inactivate the terminal transferase.

PCR reactions were set up with 5.0 μl dA-tailed cDNA, 25 pmol oligo dT-anchor primer, 12.5 pmol gene-specific primer 2 (Sp2 nested), 10 nmol each dNTP (100 mM dNTP mixture) and 2.5 U Taq DNA polymerase (Amersham Pharmacia Biotech) in 50 μl 1x PCR buffer (Amersham Pharmacia Biotech). Products were amplified through 1 cycle at 94°C for 2 min, 10 cycles at 94°C for 15 s; 55°C for 30 s; 72°C for 40 s, 25 cycles at 94°C for 15 s; 55°C for 30 s; 72°C for 40 s, with elongation of 20 s for each cycle, and a final elongation of 7 min at 72°C.

Nested PCR reactions were required to obtain a visible product. These reactions contained 5.0 μl first PCR amplified products, 12.5 pmol anchor primer, 12.5 pmol gene-specific primer 3 (Sp3 nested), 10 nmol each dNTP (100 mM dNTP mixture) and 2.5 U Taq DNA polymerase in 50 μl 1x PCR buffer. Products were amplified as described above. Nested PCRs were also carried out using dilutions (1:20) of the first PCR amplified products to amplify specific messages. Internal controls supplied with the kit to check each step of the protocol were also carried out, giving products of 655 bp, 293 bp and 157 bp.

PCR products were electrophoresed through a 1% low melting temperature agarose gel (Sea Plaque; Promega), visualised by staining with ethidium bromide (0.5 μg/ml), excised from the gel and recovered using a QIAgen gel extraction kit according to the manufacturer's instructions. Purified DNA fragments were ligated into the pGEM T-Easy vector (Promega), transformed into *E. coli* XL1-Blue cells and recombinant plasmids sequenced (see section 2.2.5-2.2.7.3). The annealing temperatures for PCRs were adjusted to approximately 5°C below the melting temperature.

4.2.2 Primer extension

Primer extension reactions were performed using a Primer Extension System – AMV Reverse Transcriptase Kit (Promega) according to the manufacturer's instructions. A primer complementary to a region approximately 100 bp downstream of the predicted transcription start site was end-labelled as follows. 10 pmol Sp1 primer (Table 4.2), 30 μCi of [γ ^{32}P]ATP and 10 U of T4 polynucleotide kinase were made up to 11 μl in 1x T4 polynucleotide kinase buffer and incubated at 37°C for 10 min, followed by 90°C for 2 min to inactivate T4 polynucleotide kinase. The mixture was centrifuged briefly and the final concentration of the primer was brought to 100 fM by adding 90 μl of nuclease-free water.

The end-labelled primer was annealed by incubating 10 μg of total RNA with 100 fmol of ^{32}P labelled primer in 11 μl in 1x AMV primer extension buffer at 58°C for 20 min. The reaction was cooled by placing the tubes at room temperature for 10 min then 2.9 mM sodium pyrophosphatase and 1 U of AMV RT were added. The reaction made up to 20 μl with 1x AMV primer extension buffer and incubated at 42°C for 30 min. 20 μl of loading dye were added to each reaction and heated to 90°C for 10 min before loading 20 μl onto a 6% denaturing polyacrylamide gel. The gel was prepared, electrophoresed and exposed to film as described in section 2.3.

4.3 Results

4.3.1 Isolation of genomic clones

An *E. tenella* (Houghton strain) cosmid library was made in the Supercos vector (Stratagene) from *Sau*3A partially digested genomic DNA (F. Tomley and B. Hoogendoorn, 1997; unpublished). The library, which contains *E. tenella* genomic DNA inserts of between 30 - 42 Kb, was plated onto LB agar plates containing kanomycin, transferred onto nitrocellulose filters and screened by hybridisation to

radiolabelled cDNA clones, encoding all or part of each microneme protein, EtMIC1-5. Between 13 and 37 positive clones were identified for each probe. Each positive clone was grown up and DNA templates were prepared which were then used in a PCR-based screening assay to amplify any genomic sequences lying upstream of the microneme-coding regions which happened to lie close to one end of the cosmid vector (i.e. near the T3 or T7 primers which flank the Supercos *Sau*3A cloning site).

Since genomic DNA had been prepared by partial digestion with *Sau*3A the microneme genes of interest could lie anywhere within the cosmid inserts. If a given gene were to lie close to either end of the vector cloning site it would be possible to amplify the 5' upstream region by PCR using a combination of Supercos and gene-specific primers. However if the gene were to lie some distance from the primers, amplification would be unsuccessful. The screening strategy to amplify the fragment which corresponded to the upstream sequence, used PCRs with combinations of T3 or T7 primers (the sites of which were positioned 5' and 3' respectively to the multicloning site of the cosmid) in conjunction with gene-specific primers which were to the 5' (anti-sense) end, in order.

The position of the gene-specific primers varied depending on the amount of sequence known from cDNA clones isolated and any AT rich regions. In general they were designed to lie no more than 50 bp downstream of the translation initiation site (ATG start codon) to amplify the 5' untranslated region. The amplified products were electrophoresed through 1% ethidium bromide stained agarose gels and the results of a typical set of PCRs can be seen in Fig 4.1. Using this strategy, products of >1500 bp were amplified for all 5' untranscribed regions of *EtMIC1-5*. Approximately 800 bp of the upstream sequence of *EtMIC1* had been previously identified although no upstream sequence for the other microneme encoding genes was available. Amplified fragments were excised, purified, ligated into pGEM T-Easy vector and transformed into competent *E. coli* XL1-Blue MRF' cells. The upstream regions of *EtMIC1, 2, 3,* and *5* (Appendix B) were sequenced using a combination of end-sequencing and primer walking whereas the upstream sequence of EtMIC4 was obtained from the genome sequencing project (http://www.sanger.ac.uk/Projects/E_tenella/). The objective of this was to identify potential promoter elements in the upstream regions of the genes. Between 1.5 and 2.5 Kb of the 5' untranscribed regions were sequenced for *EtMIC1, 2, 3,* and *5,* depending on the size of the amplified product.
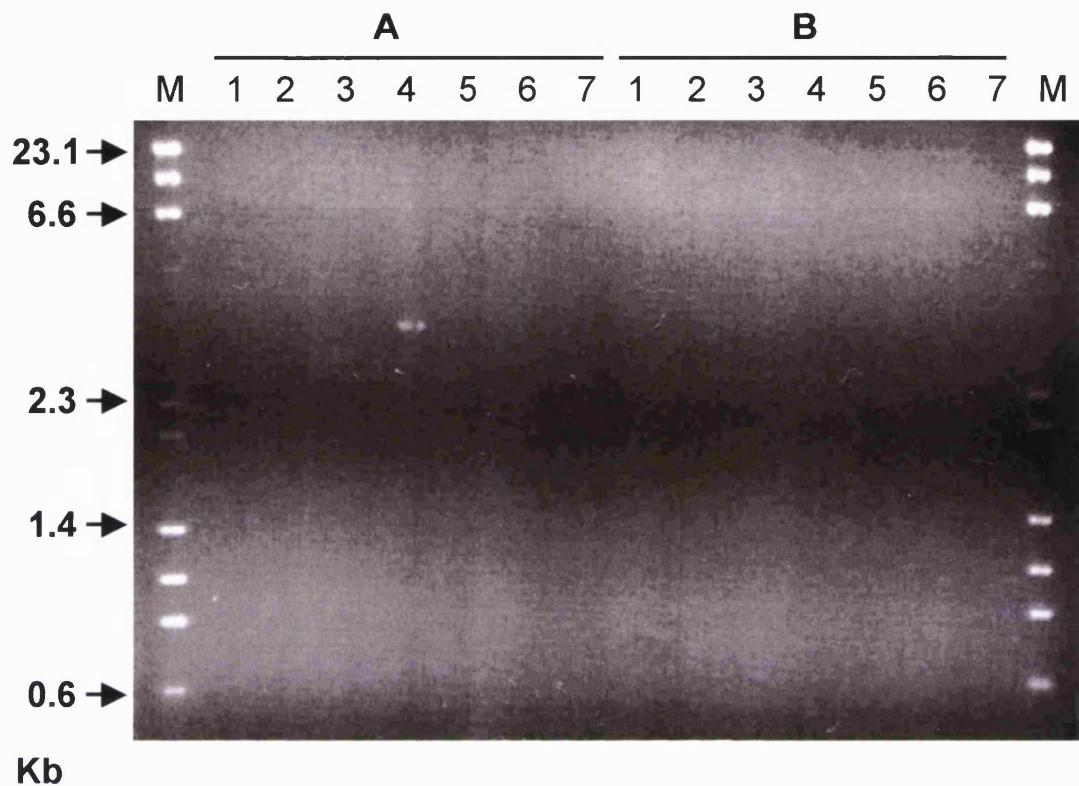
Fig. 4.1 *EtMIC2* genomic DNA amplification to clone the 5' upstream region.
A panel of seven *EtMIC2* positive cosmids (1-7) were PCR amplified using primer mic1rr5 which is complementary to the 5' end of *EtMIC2*, together with either A; the T3 primer or B; the T7 primer, which are complementary to the multicloning site of SuperCos. The products were resolved on a 1% ethidium bromide agarose gel, and a single cosmid (number 4) yielded a specific product of approximately 3.5 kb. This product was excised, cloned into pGEM T-East vector (Promega) and sequenced. The fragment was found to contain the 5' upstream flanking sequence of *EtMIC2*. Molecular weight markers (M) are indicated in the first and last lanes.

4.3.2 Transcription start site mapping of *EtMIC1-5*

In Chapter Three it was shown by RT-PCR that transcripts for *EtMIC1-5* are present in fully sporulated oocysts, freshly excysted sporozoites and in sporulating oocysts from around 12 h onwards. To determine the position of the transcription start sites of *EtMIC1-5*, 5'RACE reactions were carried out using total RNA extracted from fully sporulated oocysts. The RNA was checked for purity and quality by gel electrophoresis and quantified by spectrophotometry. A schematic representation of the 5'RACE procedure can be seen in Fig 4.2.1 and a list of the primers used to amplify each gene is given in Table 4.2. The 5'RACE reactions generated final PCR products of ~130 bp for each of *EtMIC1* and *EtMIC2* and of ~100 bp for *EtMIC3* (Fig 4.2.2). These PCR products were cloned into pGEM-T-Easy vector and aapproximately 15 clones of each were sequenced to identify the transcription start sites of *EtMIC1, 2* and *3*. As can be seen from Table 4.1, sequence analysis localised the transcription start sites of each gene to more than one nucleotide (as indicated by the arrows), however all nucleotides were within 5 bp of each other. For each gene, approximately 75% of products localised the transcription start site to adenosine nucleotides.

The major transcription start site of *EtMIC1* was localised to an adenosine nucleotide, 110 bp upstream of the ATG start codon. The surrounding sequence, between positions 105-112 bp, had a good match to the Inr consensus (**TCA^{+1}CAGT**, the transcription start site is designated **A^{+1}**, with the distance from the ATG start codon indicated in brackets, all shown in bold type). 5'RACE analyses identified two potential transcription start sites in the 5' upstream region of *EtMIC2* (5'-A$^-$ $^{3(+1)}$TTA^{+1}CTCC-3') both of which were localised to adenosine nucleotides. The sequence surrounding, adenosine nucleotide nearest the ATG start codon, had an excellent match to an Inr motif, but the second transcription start site does not lie within a sequence that conforms to an Inr consensus. For ease of discussion the start site nearest the ATG start codon has been numbered +1. The major transcription start site of *EtMIC3* (CTA^{+1}TTTT) was also localised to adenosine nucleotide again within a motif, which has a good match to an Inr consensus, positioned 51-58 bp upstream of the ATG start codon.

Amplified products were detected in the first round of PCR amplification for all five microneme encoding genes but the products for *EtMIC4* and *5* failed to amplify

through the second and third rounds of the procedure. The reactions were repeated with various annealing temperatures (between 50°C and 65°C) using a gradient PCR machine (Eppendorf) and the amount of template total RNA included in the reactions was also increased up to five fold but second round amplification for both *EtMIC4* and *5* was again unsuccessful.

Further attempts were made to obtain the transcription start sites for *EtMIC4* and *5*, and to confirm the results obtained for *EtMIC1, 2* and *3* using primer extension analysis. As with 5'RACE the template:primer ratio and annealing temperature are critical and various conditions were used. The distance between the transcription start site and primer are also critical for the successful detection of the start site and the primer should be no more that 100 bp downstream from it in order to minimise premature termination of reverse transcription due to the formation of RNA secondary structures (Calzone et al., 1988). Primers were chosen ~90 bp downstream from the Inr sites (as demonstrated by 5'RACE), but no signals were detected from any of the reactions. A signal of the correct size (109 bp) was detected from the positive control reaction using the control template included in the kit, but no signals were detected from the *E. tenella* total RNA template using any of the primers.

Primer extension reactions are greatly dependent upon the amount of mRNA present. In both protocols the volume of total RNA included in the reactions was at the maximum recommended by the suppliers (Boehringer-Mannheim and Promega). It is difficult to predict the amount of target, or specific transcript present in the RNA included in each reaction. It has been estimated that 1 μg of poly A^+ RNA would contain 0.392 fmol of each transcript (Chakrabarti *et al.*, 1994; Brown *et al.*, 2000a; Hayward *et al.*, 2000) and that the detection limit of the primer extension kit used was $\geq$ 1.3 fmol (Brown, 2000). In this study attempts were made to map the transcription start site using total RNA which contained less than 1% of poly A^+ RNA. Therefore 10 μg of total RNA would contain approximately 0.0392 fmol of each transcript, such that the amount of total RNA included in the reaction was probably insufficient.

4.3.3 Sequence analysis of genomic DNA (5' untranscribed regions)

Approximately 2400 bp of the upstream genomic sequence of *EtMIC1* was sequenced by end primer sequencing using T3 and T7 of the pGEM-T Easy vector and primer walking, initially using primer mic1rr5, which corresponded to the 5' end of the

1

ATG

5' ——————————————————————————— (A)-3'

← SP1

Synthesis of first strand cDNA
from mRNA using primer SP1

5' ——————————————————————————— (A)-3'
3' ——————————————————————————— 5'

Degradation of mRNA template by
the RNase H activity of the AMV
RT and purification of cDNA

3'AAAA ———————————————————————

Tailing of purified
cDNA using dATP and
terminale transferase

5'TTTTV
3'AAAA ———————————————————————

← SP2

**PCR anchor
primer** →

Amplification of tailed cDNA by
PCR using the oligo dT-anchor
and nested SP2 primers

← SP3

PCR product ready for agarose
gel analysis followed by cloning
and sequence analysis

2

1.4 →
0.6 →
0.3 →

M   MIC2 MIC2   MIC1 MIC1

1.4 →
0.6 →
0.3 →

kb   M   MIC3

Fig. 4.2 Mapping *EtMIC1*, *2* and *3* transcription start sites using 5' RACE PCR analysis of *E. tenella* total RNA.
Panel 1: A schematic representation of the procedure used to determine transcription start sites using 5' RACE
Panel 2: cDNA was synthesised from total RNA extracted from fully sporulated oocysts using a gene specific primer. An oligo dT tail was enzymatically added to the cDNA using terminale transferase and was amplified using an oligo dT-anchor primer and a gene-specific nested primer. A final PCR was performed using a PCR-anchor primer and a third gene-specific nested primer. This was done for each of the genes investigated. The final product was electrophoresed through a 2% low melting point ethidium bromide stained agarose gel, excised, cloned into pGEM-T-Easy vector (Promega) and sequenced to obtain the transcription start sites. The molecular weight marker (M) are indicated in the first lane.

Gene	Sequence (5' – 3')	Bp to ATG start codon
EtMIC1	TCTC$\underline{\textbf{A}^{+1}\text{C}}$AGT	110
EtMIC2	G$\textbf{A}^{-3\,(+1)}$T$\underline{\text{T}\textbf{A}^{+1}\text{C}}$TCCT	94 and 97
EtMIC3	A$\underline{\text{CC}\textbf{A}^{+1}\text{T}}$TTTT	56

Table 4.1 Results of transcription start site mapping by 5'RACE. The transcription start sites were mapped to more than one nucleotide within the above sequences, as indicated by the arrowheads, however in approximately 75% of the clones, sequenced for both *EtMIC1* and *3*, the start sites were mapped to adenosine nucleotides, designated A^{+1} (shown in bold type) and therefore these were regarded as the transcription start sites. Of the clones sequenced to identify the transcription start site of *EtMIC2*, the clones were equally dispersed between two adenosine nucleotides both shown in bold type and are therefore both thought to be transcription start sites. Sequences which conform to the Inr consensus are underlined.

Gene	Primers	Sequence (5' to 3')
EtMIC1	Sp1	GCACCTTCCTGAAGTTCGATGTGCC
	Sp2	CTACCAACATGACATCCAAGAGGC
	Sp3	GCCAGAGCTGGTAGTTGCGCCTGGC
EtMIC2	Sp1	GGTCACTTTGATGGTCTCATCCACGC
	Sp2	GCCTGCCGCAGTTACCTTCATCTAG
	Sp3	GTTCCGGCGCATCTCTCCCACTGAG
EtMIC3	Sp1	GCCGAGTCTACAAATTGTAGAGCCCG
	Sp2	CGAGGCAAATGCCTAAAAGAAGGTC
	Sp3	CACAGGCAGCGCAGTCACGGAAGGC
EtMIC4	Sp1	GGTTGAATGCCTCGCACTGAGTCCC
	Sp2	GCACTGACAAGTGGCTCCGAGGGGG
	Sp3	CACTGGCTACAATCGAAGGCGCCGC
EtMIC5	Sp1	CGCTGCTCAGTTTCCTGCCTTGGG
	Sp2	GAACCACCAAAATGCCACTACCCTC
	Sp3	GATGAGGCTGCAGGTTGCGCCCTCC
	Oligo d(T) primer	GACCACGCGTATCGATGTCGACTTTTTTTTTTTTTTTTV
	PCR anchor primer	GACCACGCGTATCGATGTCGAC

Table 4.2 Oligonucleotide primers used for 5'RACE analysis.

The above combinations of primers were used for the identification of transcription start sites of microneme-encoding genes. Sp1 primers were used for production of microneme-specific cDNAs, which were further amplified using combinations of Sp2 primers with the oligo d(T) primer and again further amplified using combinations of Sp3 and the PCR anchor primer. Primer Sp1 was also used in primer extension reactions.

cDNA sequence. The remainder of the sequences was obtained using mic1rr5.7, mic1rr5.8 and mic1rr5.10 primers. The upstream sequence of all *EtMIC1-5* genes and the primers used for sequence analysis are detailed in Appendix B and C, respectively. The upstream region of *EtMIC1* was approximately ~52% AT-rich and contained two homopolymeric (dA:dT) tracts of 9 bp. In addition to the Inr motif surrounding the transcription start site, the upstream sequence of *EtMIC1* contained a single potential, degenerate TATA box (TTAATAA) at -144 (144 bp upstream from the predicted transcription start site) and multiple potential CCAAT sites between -6 and -310. However the potential TATA box in the upstream region of *EtMIC1* is degenerate and was located 144 bp upstream of the transcription start site, 119 bp more that the distance usually observed (~25-30 bp) between the two elements. The nearest degenerate CCAAT motif (CAAT) was located 63 bp upstream.

The upstream genomic region of *EtMIC1* also contained a AGAGACC motif (–394, direct orientation) and a TGAGACG motif (-6, indirect orientation) which are similar to the T/AGAGACG repeats observed in *T. gondii SAG1, TUB1, GRA1, 2, 5* and *6* and *NTPase 1* and *3*. Much further upstream at -1244 and –662 were two GGAGAGG motifs, similar to those observed in the promoter regions of *NTPase 1* and *3* (Mercier *et al.*, 1996; Nakaar *et al.*, 1992; Soldati and Boothroyd, 1995). At positions -1140 and -410 were GTGTGT and CACACA sequences respectively, which were defined in the promoter of *TamS1* in *T. annulata* (Shiels *et al.*, 2000). A GAAC motif was also identified at -81 with a TATA box upstream of it and an Inr element downstream. Thus the upstream region of *EtMIC1* conformed to a tripartite promoter structure, observed in a variety of protein-encoding genes of *E. histolytica* (Singh *et al.*, 1997).

For *EtMIC2*, approximately 1200 bp of 5' upstream genomic DNA was sequenced initially by end primer sequencing using T3 and T7 of the pGEM T-Easy vector and primer mic2rr5, which corresponded to the 5' end of the cDNA sequence. The remainder of the sequence was obtained by primer-walking using mic2rr5.6 and mic2rr5.5 primers. The upstream region of *EtMIC2* was approximately ~54% AT-rich and contained three homopolymeric (dA:dT) tracts of 9 bp. As well as the Inr motif surrounding one of the predicted transcription start sites, the upstream region of *EtMIC2* contained a single potential TATA box (TATAAT) and a single, degenerate CCAAT site (CAATT) positioned -788 and -98 respectively (see Appendix B). It is highly unlikely that these are functional transcriptional elements due to their distance

from the predicted transcription start site and the large distance between them. Also found in the upstream genomic region of *EtMIC2* at positions -905 and -430 were GTGTGT elements and a AGAGAGC motif (-968, indirect orientation), which were defined in other apicomplexan promoters (Mercier *et al.*, 1996; Nakaar *et al.*, 1992; Soldati and Boothroyd, 1995; Shiels *et al.*, 2000) and in *EtMIC1*. The upstream region of *EtMIC2* also has a potential tripartite promoter architecture as it contained a GAAC motif at -310 with a degenerate TATA box upstream of it and an Inr motif downstream. However the elements were dispersed over a much larger area (approximately five times greater) than that observed for *EtMIC1*.

Approximately 2100 bp of the 5' upstream region of *EtMIC3* was sequenced by end primer sequencing using T3 and T7 of the pGEM T-Easy vector and primer mic3rr5, which corresponded to the 5' end of the cDNA. The remainder of the sequence was obtained by primer walking using mic3rr5aa, mic3rr5ab, mic3rr5ae, mic3rr5ah and mic3rr5af primers. The upstream region of *EtMIC3* was approximately ~53% AT-rich with four homopolymeric (dA:dT) tracts of 9 bp or more. In addition to the Inr motif surrounding the transcription start site, a potential degenerate TATA box (TTAATA) and CCAAT box (CAAAT) were located at positions -195 and -211, respectively. These potential promoter elements were only 11 nucleotides apart, indicating that they may be functional, but are ~200 bp upstream of the predicted transcription start site, a distance greater than what is usually observed (see Appendix B). It is possible that another transcription start site located ~25 bp downstream of the TATA box was not identified by 5'RACE, possibly due to the position of the primer. The upstream genomic region of *EtMIC3* also identified a TGAGAGG motif (-1618) and a GTGTGT motif (-2016) which were initially defined by other apicomplexan promoters (Mercier *et al.*, 1996; Nakaar *et al.*, 1992; Soldati and Boothroyd, 1995; Shiels *et al.*, 2000). The GAAC motif observed in tripartite promoter structure of protein-coding genes of *E. histolytica* (Singh *et al.*, 1997) and in the upstream regions of both *EtMIC1* and *2* was also identified in the upstream region of *EtMIC3* at positions -276 and -235, however in this case the upstream region of *EtMIC3* did not conform to a tripartite architecture, as the GAAC motifs were most further upstream of the predicted Inr motif, rather than the TATA box.

A total of 3090 bp of the upstream genomic sequence of *EtMIC4* was obtained from analysis of the *E. tenella* genome sequencing project. The upstream region of *EtMIC4*

was approximately ~50% AT-rich. Attempts to map the transcription start site(s) of *EtMIC4* by 5'RACE and primer extension failed, and therefore the positions described here are relative to the ATG start codon, which has been designated +1. *EtMIC4* also contains degenerate TATA (TTAATAT) and non-degenerate, consensus CCAAT (CCAAT) boxes located at positions -1452 and -1467, respectively. The two eukaryotic promoter elements are 16 nucleotides apart, which are just below the usual 25-30 bp. Interestingly in this region there is a potential Inr motif between nucleotides -1440 and -1434, $TCA^{+1}GAGA$. If this were a functional Inr element it would locate the transcription start site to exactly 14 nucleotides downstream from the TATA box. As with the first three microneme-coding genes investigated, *EtMIC4* also contains several GAAC motifs (Singh *et al.*, 1997) positioned at −1549, -1555, -1715, -1735, -2690 and -2965. At positions -1884 and -1376 are CACACA motifs (Shiels *et al.*, 2000) and a GCGAGACG motif (-516) and several inverted repeats (GGTCTCA, -401, AGAGAGC, -536, CGTCTCT, -748, AGAGAGG, -1438, TGAGAGAC, -1586, and GGTCTCA, -2426, respectively) were identified which are similar to those observed in *SAG1, TUB1, GRA1, 2, 5* and *6* and *NTPase 1* and *3* (Mercier *et al.*, 1996; Nakaar *et al.*, 1992; Soldati and Boothroyd, 1995).

Approximately 600 bp of upstream genomic sequence of *EtMIC5* (from the ATG start codon) had been sequenced previously (Brown, 2000). An additional 2500 bp of this upstream sequence was obtained initially using primer mic5rr5, which corresponded to the genomic sequence 600 bp upstream and by primer walking using primers mic5rr5.6, mic5rr5.7, mic5rr5.8, mic5rr5.9 and mic5rr5.10. The upstream region of *EtMIC5* was approximately ~53% AT-rich and contained five homopolymeric (dA:dT) tracts of 9 bp or more. A potential consensus TATA box (TATAAA) at position -1093 and a potential degenerate CAAT box (CAAAT) at -1104 were identified as well as several potential transcription Inr motifs (see Appendix B). However in this case the closest potential Inr ($CTA^{+1}GTCT$) to the consensus TATA box was approximately 130 nucleotides downstream of it. Also found in the upstream genomic region of *EtMIC5* was a GGAGAGG motif (-1479) which was similar to the repeats observed in the promoter regions of *NTPase 1* and *3* (Nakaar *et al.*, 1992). At position -740 was CACACA sequence which was defined the promoter of *TamS1* in *T. annulata* (Shiels *et al.*, 2000). No GAAC motifs were observed in the upstream region of *EtMIC5*. Figs 4.3.1 and 4.3.2 show all the information available on microneme-coding genes, *EtMIC1-5*, following sequence identification and 5'RACE analyses.

EtMIC1 Genomic DNA (2374 bp)

-383 to -350

GAAC -81

TCA⁺¹CAGT

ATG cDNA (3327 bp) 3'

5'

GGAGAGG -1244 GTGTGT -1141 GGAGAGG -662 CACACA -410 AGAGACC -394 CAAT -310 CAAT -211 TATA -144 CAAT -124 CAAT -118 CAAT -10 CGTCTCA -6

110 bp

EtMIC2 Genomic DNA (1191 bp)

-197 to -155

GA⁽⁺¹⁾TTA⁺¹CTCC

ATG cDNA (1842 bp) 3'

5'

94 bp

GCTCTCT –968 GTGTGT –905 TATA -788 GTGTGT -430 GAAC -310 Consensus CAAT -98 Consensus CAAT +20

EtMIC3

Genomic DNA (2241 bp)

CTA⁺¹TTTT

ATG cDNA (4787 bp) 3'

5'

55 bp

GTGTGT -2016 TGAGAGG -1618 GAAC -276 GAAC -235 CAAT -211 TATA -195

Fig. 4.3.1 Organisation of the microneme encoding genes *EtMIC1* to *3* determined from DNA sequence and 5'RACE analyses. A thick black line represents each microneme encoding gene, 5' to 3' (*EtMIC1-3*). The ATG translation initiation codon is shown for each gene and the length of the cDNA sequence identified is shown immediately on the right hand side of this. The amount of derived genomic sequence at the 5' end of each gene is given on the left, shown in blue. The transcription start sites (bent arrow) for *EtMIC1, 2* and *3* was confirmed by 5'RACE, some have been located to Inr regions (coloured in red) are regarded as+1. Two transcription start sites are shown for *EtMIC2*. The start site nearest the ATG start codon, which conforms to an Inr region, is regarded as +1 and therefore all the distances discussed are relative to this point. The distance between the ATG start codon and confirmed Inrs are shown with the double headed arrow. Potential transcriptional elements, such as TATA boxes, are also indicated for each gene. Genetic elements are not drawn to scale.

111

EtMIC4 Genomic DNA (2074 bp)

GGTCTCA CACACA   GAAC GAAC CAAT TATA   GCGAGACG
-2426   -1884   -1555 -1549 -1467 -1452   TCA⁺¹GAGA (-1440)   -516   ATG cDNA (6569 bp)

GAAC GAAC CCAAT   GAAC GAAC   CAAT TGAGAGC AGAGAGG CGTCTCT GGTCTCA CAAAT GTGTGT
-2965 -2690 -2196   -1735 -1715   -1456 -1586   -1438   -748   -401   -120   -22

EtMIC5 Genomic DNA (2530 bp)   ATG cDNA (2620 bp)

                                          CTA⁺¹GTCT
        CGAGAGG   CAAT   Consensus   (-956)   CACACA
        -1479   -1104   TATA -1093   -740
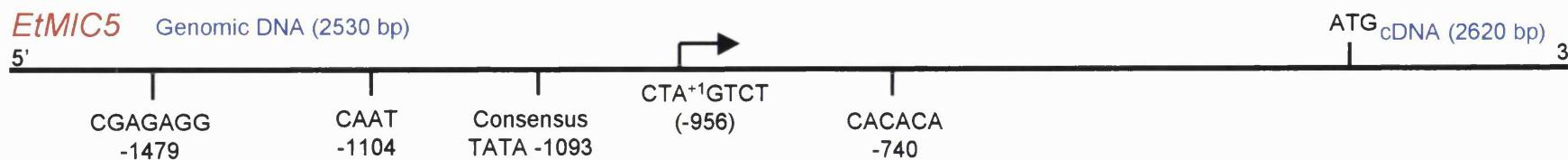
Fig. 4.3.2 Organisation of the microneme encoding genes *EtMIC4* and 5 determined from DNA sequence and 5'RACE analyses. A thick black line represents each microneme encoding gene, 5' to 3' (*EtMIC4-5*). The ATG translation initiation codon is shown for each gene and the length of the cDNA sequence identified is shown immediately on the right hand side of this. The amount of derived genomic sequence at the 5' end of each gene is given on the left, shown in blue. The theoretical transcription start sites (bent arrow) for *EtMIC4* and 5 which have not been confirmed by 5'RACE or primer extension are shown (bent arrow and Inr, coloured in black). As the transcription start sites are not known the ATG start codon has been regarded as +1 and therefore all the distances discussed are relative to this point. Potential transcriptional elements, such as TATA boxes, are also indicated for each gene. Genetic elements are not drawn to scale.

4.4 Discussion

Very little is known about transcriptional regulation in parasitic protozoa. Analyses of eukaryotic promoters and their interacting proteins have been mostly confined to animals, plants and fungi. These organisms use a diverse range of mechanisms to regulate gene transcription. To further investigate the regulation of gene transcription in *E. tenella*, analyses was undertaken of the 5' untranscribed regions of microneme-coding genes, to identify promoter elements and gain an insight into potential mechanisms of gene expression.

Following sequence analyses of the 5' untranscribed genomic region of each gene (*EtMIC1-5*) attempts were made to obtain their respective transcription start sites. 5'RACE analyses localised the major transcription start sites of *EtMIC1* and *3* to adenosine nucleotides 110 and 56 bp upstream from the translation start codon of each gene, respectively, within sequences that conform to Inr motifs. Two transcription start sites were identified for *EtMIC2*, located 94 and 97 bp upstream of the ATG start codon, one of which was positioned within a sequence that conformed to an Inr. Such elements are capable of directing start site placement and initiating transcription in TATA-less promoters.

Transcription start sites which are surrounded by sequences which conform to the Inr consensus have been found in an increasingly large number of protozoan genes (see Table 4.3), including *T. gondii SAG1* and *NTPase1* and *NTPase3*, *T. annulata*, *TamS1*, *G. lamblia α2 tubulin*, *α-giardin* and *δ-giardin* and in various genes of the human protozoan pathogen *E. histolytica* (Nakaar *et al.*, 1992; Soldati and Boothroyd, 1995; Singh *et al.*, 1997; Shiels *et al.*, 2000; Elmendorf *et al.*, 2001).

Inrs have been shown to constitute a simple functional promoter and to be required for expression of the *NTP3* gene in *T. gondii* (Nakaar *et al.*, 1992). The presence of a highly conserved, ubiquitous Inr element used in transcription initiation in a wide variety of protein-coding genes from various parasitic protozoa suggests that such an element appeared early during eukaryotic evolution. Although this promoter element is not conserved in all eukaryotes, little divergence has occurred amongst protozoan parasites. There is evidence that specific Inr-binding proteins interact with Inr-specific sequences (Purnell *et al.*, 1994; Smale, 1997). Purification of nuclear proteins from various protozoan parasites that specifically recognise a functional Inr will advance the understanding of transcription of apicomplexan parasites.

Analyses of the upstream genomic regions of *EtMIC1-5* revealed that all contain single or multiple copies of a purine rich sequence similar to those observed in *SAG1*, *TUB1*, *GRA1*, *2*, *5* and *6* and *NTPase 1* and *3* (see above and Table 4.4). SAG1, a major surface antigen of *Toxoplasma* tachyzoites, is believed to be involved in host cell adhesion and invasion (Mineo *et al.*, 1993) and is distributed homogeneously on the surface of both intracellular and extracellular tachyzoites (Dubremetz *et al.*, 1985; Kasper *et al.*, 1983). *SAG1* contains no higher eukaryotic TATA promoter elements (Burg *et al.*, 1988), but two transcription initiation sites, a major and a minor (the latter conforms to the Inr consensus) have been identified and are separated by ~35 bp (Burg *et al.*, 1988). Using transfection technology, Soldati (1995) identified a stretch of six repeated sequences (of 27 bp) between -35 and -190 bp, of which a purine rich sequence (A/TGAGAGC) forms the core. These repeat elements have been shown to act in an orientation-independent fashion, to be essential for efficient and accurate expression of *SAG1* (at least two repeats are required for basal level expression of *SAG1*) and can specify transcription initiation (Soldati and Boothroyd, 1995). Similar repeats are also critical for the expression of dense granule proteins (Sibley *et al.*, 1994b).

GRA proteins (*GRA1*, *2*, *5*, and *6*), of *T. gondii* also contain a purine rich, repeated element similar to the core motif discussed above (A/TGAGACG). Deletion and mutation analyses demonstrated that a single motif (nearest the transcription start site) is sufficient for basal level expression and the addition of at least one other motif significantly aids expression (Mercier *et al.*, 1996). Deletion of all repeats results in complete loss of expression, indicating that these motifs are critical for transcription (Mercier *et al.*, 1996).

A single copy of this purine rich sequence has been found in the upstream region of *TUB1* in *T. gondii*, 30 bp upstream of the transcription start site (Soldati and Boothroyd, 1995). As a construct containing 70 bp upstream of the transcription start site, including the motif, drove efficient expression of a chloroamphenicol acetyl transferase (CAT) reporter gene (Soldati and Boothroyd, 1995) it seems likely that this motif will play a similar role in *TUB1* as in *SAG1*, however as the promoter region of *TUB1* has not been precisely mapped its real implication cannot be ascertained.

The upstream genomic regions of the genes encoding *NTPase 1* and *3* contain potential regulatory elements similar to the conserved purine rich repeats discussed above (Nakaar *et al.*, 1992). NTPase is used by *T. gondii* to salvage purines from the

host cell (Bermudes *et al.*, 1994; Schwab *et al.*, 1994) and comprises ~2% of total parasite protein. The enzyme is contained in the dense granules, where it is secreted into the PV space upon invasion (Sibley *et al.*, 1994a; Vasconcelos *et al.*, 1996). Purine rich elements, AGAGACGC and GGAGAGG, have been observed in the upstream region of both *NTPase 1* and *3* ~1500 bp upstream of the translational start codon (Nakaar *et al.*, 1992). To investigate their involvement in gene expression deletion constructs linked to CAT reporter gene were used. However the presence or absence of these elements did not alter CAT activity and therefore they are probably not required for high gene expression (Nakaar *et al.*, 1992).

It has been postulated that the multiple repeats observed in the upstream region of *SAG1* play a role in the transcription initiation process (Soldati and Boothroyd, 1995). It has also been suggested that they act like SP1 elements and determine the transcription start site by protein-protein interaction in a TATA-less promoter (Kadonaga *et al.*, 1986). GC-rich sequences are recognised by SP1 and are usually found in tandem repeats and in either orientation. The purine rich repeats of *SAG1* stimulated expression when introduced into a heterologous promoter (*TUB1*) in either orientation, 70 bp upstream of the transcription start site. However they were not able to produce the same level of expression when positioned 230 bp upstream of the transcription start site (Soldati and Boothroyd 1995), indicating that they were position sensitive, thus supporting the hypothesis of their involvement in transcription initiation. The most important SP1-responsive element in gene expression is the one nearest the transcription start site (Kadonaga *et al.*, 1986) as with the repeats of the GRA proteins.

As shown in Table 4.3, the transcription start site of *TamS1* in *T. annulata* also has homology with the consensus eukaryotic Inr motif. Further analysis of the upstream region by electrophoretic mobility-shift assay identified a 23 bp sequence which specifically bound factors from parasite-enriched nuclear extracts. Three complexes were shown to bind to a 9 bp motif which had a core binding site of 5'-TTTGTAGGG-3' (Shiels *et al.*, 2000). Analysis of the upstream regions of all five microneme-encoding genes did not identify this or a similar sequence. This core motif has not been found in promoter regions of any other apicomplexan genes or those of higher eukaryotics and may therefore be specific to *Theileria*. Also observed in the upstream region of *TamS1* were sequences CACACA and GTGTGT, the former of which is involved in stabilising complex-a formation. The latter has been observed in

the SV40 core enhancer region (Weiher *et al.*, 1983) and related motifs have been found in the upstream regions of *Plasmodium* (Lanzer *et al.*, 1993) and *T. gondii* (Gross *et al.*, 1996). The CACACA motif was observed in the upstream genomic region of *EtMIC1, 4* and *5* and the GTGTGT motif was identified in *EtMIC1, 2* (twice), *3* and *4*. Whether this motif plays a significant role in the transcription of *Eimeria* or other (apicomplexans) is not known.

E. histolytica, one of the earlier diverging eukaryotes, is a protozoan parasite, that causes amoebic colitis and liver abscesses. Analyses of the 5' upstream region of 37 protein-coding genes identified three core promoter motifs including a degenerate TATA box, an Inr motif and a GAAC element, which were all within 30 bp of each other. This unusual GAAC motif controls the rate of gene expression, the endogenous transcription start site and is able to direct a new initiation site, independent of both the TATA box and the Inr (Singh and Rogers, 1998; Singh *et al.*, 1997). It is also capable of mediating transcriptional activation of some upstream regulatory elements and is involved in protein complex assembly at the core promoter, although it is unable to recruit a higher-order protein complex alone (Singh *et al.*, 2002). Mutational analyses of the GAAC element in the promoter region of various genes in *E. histolytica* indicated it plays an important role in gene expression (Prurdy *et al.*, 1996; Singh *et al.*, 1997). *Hgl5* of *E. histolytica*, which has the tripartite core promoter, codes for galactose- and *N*-acetyl-D-galactosamine-inhibitable lectin. It has a major role in pathogenesis including host cell attachment, invasion and killing of the target cell (Gilchirst and Petri Jr, 1999; Huston *et al.*, 2000; Petri Jr *et al.*, 1987; Vines *et al.*, 1998). Singh and co-workers demonstrated that nuclear protein complex assembly on the core promoter of the *hgl5* gene requires both the TATA box and GAAC regions (Singh *et al.*, 2002) and proposed these motifs function in aiding the recruitment of the TFIID to the TATA box facilitating organised and regulated gene expression. In many of the protein-coding genes of *E. histolytica* the GAAC motif was positioned approximately 15 bp upstream of the Inr element, with the degenerate TATA box a further 15 bp upstream. Motifs with this consensus were found within the microneme encoding genes of *E. tenella* but were positioned several hundred nucleotides upstream of the predicted transcription start site so it seems unlikely that these elements in microneme encoding genes function in a similar way, if at all.

Sequence comparisons in this study has revealed some common conserved elements that might potentially interact with transcription machinery such as TFIID. However

in many cases there seems to be a lack of repeat elements or the distances between the motifs (such as the GAAC) are far greater than those observed in other promoters. With the recent development of a transient transfection system for *E. tenella* it is possible to characterise DNA elements within the 5' untranscribed regions that are used by the parasite for gene expression. Therefore it was decided to use this system to investigate if any *cis*-acting elements are involved in gene expression.

Pol II promoter	Inr region	Reference
T-cell receptor, variable chain β (Vβ)	ACTCTCACTTTCT	Anderson *et al.*, 1988
Lymphocyte function associated antigen 1 (LFA 1α)	ATGATCATTTTCC	Nueda *et al.*, 1993
Mouse acetylcholinesterase (ACHE)	GCTGTCACTGTCG	Li *et al.*, 1993
Human and rabbit acetylcholinesterase (BCHE)	TTTGTCAGTAACA	Jbilo *et al.*, 1994
Human amyloid β-protein precursor (APP)	TCCGTCAGTTTCC	Quitschke *et al.*, 1996
G. lamblia, α-giardin (α-gar)	CATTTAGAAAATG	Elmendorf *et al.*, 2001
G. lamblia, δ-giardin (δ-gar)	CCAGAAAAAAATG	Elmendorf *et al.*, 2001
G. lamblia, α2-tubulin (α2-tub)	GGAAATAAAAATG	Elmendorf *et al.*, 2001
T. gondii, surface antigen 1 (SAG1)	GTAGTCACTGCTG	Soldati & Boothroyd, 1995
T. gondii, nucleotide triphosphate hydrolase 1 (NTP1)	CTGTTCAGTTTTT	Nakaar *et al.*, 1992
T. gondii, nucleotide triphosphate hydrolase 3 (NTP3)	CTGTTCAGTTTTT	Nakaar *et al.*, 1992
T. annulata, merozoite surface antigen 1 (TamS1)	CACTTCACTTATA	Sheils *et al.*, 2000
E. histolytica, Gal-GalNAc-specific lectin (hgl5)	AGAAAGACAAAAT	Singh *et al.*, 1997
***E. tenella*, microneme protein 1 (EtMIC1)**	**CGTCTCACAGTAG**	
***E. tenella*, microneme protein 2 (EtMIC2)**	**AAGATTACTCCTT**	
***E. tenella*, microneme protein 3 (EtMIC3)**	**CACCTTATTTTTC**	
Consensus	PyPyANT/APyPy	

Table 4.2 Prevalence of Inr consensus in natural Pol II promoters. The predicted transcription start sites are coloured in blue. In the case of *T. gondii SAG1*, there are two transcription start sites, a major and a minor, however only the minor site is shown here. The initiation sites of *G. lamblia* genes are exceptionally close to the translation initiation codon (ATG shown in red), 3 nucleotides upstream. The sequence which conforms to the Inr motif is underlined.

Gene	Sequence	Direct/Indirect	Position	Reference
GRA1	CGACACG	D	-349	Mercier et al, 1996
	TGAGACG	I	-251	
	CGTGACG	I	-210	
	TGAGACG	D	-125	
	TGAGACG	I	-90	
	CGAGACG	I	-56	
GRA2	TGCGACG	I	-208	
	AGAGACG	I	-172	
	AGAGATG	I	-95	
	AGAGACG	D	-84	
	AGAGACG	D	-72	
	AGCGACG	I	-44	
GRA5	AGACACG	D	-123	
	AGAGACG	D	-47	
GRA6	AGACACG	D	-245	
	TGCGACG	D	-158	
	AGCGACG	D	-121	
	AGCGACG	I	-85	
	TGCGACG	D	-35	
TUB1	GGAGATG	D	-30	(Nagel and Boothroyd, 1988)
SAG1	TGAGACG	D	-193	(Soldati and Boothroyd, 1995)
	TGACACG	D	-166	
	TGAGACG	D	-139	
	TGAGACG	D	-112	
	TGAGACG	D	-85	
EtMIC1	**AGAGACC**	**D**	**-394**	
	TGAGACG	**I**	**-6**	
EtMIC2	**AGAGAGC**	**I**	**-968**	
EtMIC3	**TGAGAGG**	**D**	**-1618**	
EtMIC4	**GGTCTCA**	**I**	**-401**	
	CGTCTCT	**D**	**-748**	
	AGAGAGG	**D**	**-1438**	
	TGAGAGC	**D**	**-1586**	
	GGTCTCA	**I**	**-2426**	
EtMIC5	**CGAGAGG**	**D**	**-1479**	
Consensus	A/TGAGACG			

Table 4.4 Repeat sequences in the upstream regions of Apicomplexan genes. Such motifs appear in the direct (D), A/TGAGACG, or the indirect (I), CGTCTCT/A, orientation in the upstream region of many apicomplexan genes. The consensus sequence is shown at the end of the table. All microneme-encoding genes are shown in bold type and residues matching the consensus sequence are coloured in blue.

Chapter Five
Analysis of untranscribed regions of five microneme genes using transient transfection in *Eimeria tenella*

5.1 Introduction

In Chapter Three, it was shown that microneme-specific mRNAs are co-ordinately expressed during oocyst sporulation and in Chapter Four possible *cis*-acting elements were identified, including potential Inr motifs at the transcriptional initiation sites, TATA boxes, purine rich elements and GAAC motifs.

Recently, one of the most exciting breakthroughs in parasitology has been the development of transient and stable transfection systems, which are extremely useful for studying biochemistry and cell biology. Stable transfection and gene replacement were first established in *Leishmania* species and *T. brucei* (Lee and Van der Ploeg, 1990; Laban *et al.*, 1990; Cruz and Beverly, 1990; Ashbroek, 1990). The first member of the phylum Apicomplexa to be transfected was *T. gondii* (Donald and Roos, 1993; Kim *et al.*, 1993; Sibley *et al.*, 1994a; Soldati and Boothroyd, 1993) closely followed by *P. gallinaceum* (Goonewardene *et al.*, 1993), *P. falciparum* (Wu *et al.*, 1995) and *P. berghei* (van Dijk *et al.*, 1996; van Dijk *et al.*, 1995).

Transient transfection has been utilised to characterise DNA elements, within the 5' and 3' untranscribed regions, that are used by parasites for gene expression. A transient transfection system has been developed for *E. tenella* (Kelleher and Tomley, 1998). This involved electroporating *E. tenella* sporozoites with a plasmid containing approximately 700 bp of the 5' untranslated sequence of the microneme protein EtMIC1, fused to the *E. coli lacZ* gene (which codes for β-galactosidase). Transfected sporozoites expressing β-galactosidase (as visualised by staining with X-gal) were able to invade host cells (MBDK), differentiate into schizonts and form merozoites (Kelleher and Tomley, 1998).

Transfection technology could be used to investigate the ability of *Eimeria* sequences to drive the expression of a reporter gene product, to identify minimal promoters required for expression and to identify additional sequences that enhance expression. It could also be used to determine whether the predicted Inr regions found to surround the transcriptional start sites of three of the microneme genes (*EtMIC1, 2* and *3*) are sufficient to drive gene expression. Ideally a comprehensive dissection of the elements

involved in gene transcription is required to gain an insight into understanding the regulation of microneme-coding genes in *E. tenella*. As an initial step we chose to dissect the upstream region of *EtMIC1*, as this region had been previously shown to drive expression of β-galactosidase in transient transfection assays. This chapter presents the results of the deletion analyses of the 5' untranscribed regions of *EtMIC1, 2* and *3*.

5.2 Methods

Sporozoites were purified as detailed in section 2.1.2 - 2.1.4 and counted in a Fuchs Rosenthal counting chamber. The sporozoites were washed once in incomplete cytomix and then resuspended in complete cytomix at a concentration of 1.4×10^7 per ml and stored on ice prior to electroporation.

5.2.1 Molecular constructs

Genomic cosmid clones of *E. tenella* (H) containing the microneme genes *EtMIC1*, *EtMIC2* and *EtMIC3* including their upstream untranslated sequences were used as template DNAs for PCR amplification. Each PCR contained 50 ng of cosmid template DNA, 12.5 pmol of each primer (see Appendix B), 10 nmol dNTP's (100 mM dNTP mixture of each), 2.5 U of *Taq* DNA polymerase and made up to 50 µl 1x PCR buffer (Amersham Pharmacia Biotech). All DNA templates were denatured for 3 min at 94°C and amplified through 30 cycles of 30 s at 94°C; 50 s at 62°C; 1 min 40 s at 72°C and finally extended at 72°C for 10 min.

Amplified products were digested with 5 U of *Kpn*I in 50 µl of 1x React 4 buffer (Gibco BRL) and incubated for 1 h at 37°C. The digested products were electrophoresed through a 1% agarose gel and visualised with ethidium bromide (see Fig 5.1 for an example). Bands of the correct size were cut out and gel extracted using a QIAgen gel extraction kit, according to the manufacturer's instructions (see section 2.2.4).

5 µg of plasmid pSV-β-gal (Promega) was digested at 37°C for 1 h with 10 U of *Kpn*I in 100 µl of 1x React 4 buffer (Gibco BRL). The digested products were purified using a QIAgen PCR purification kit, according to the manufacturer's instructions (see section 2.2.2). 2 U of calf intestinal alkaline phosphatase in 50 µl of 1x alkaline phosphatase reaction buffer (Roche Molecular Biochemicals) were added to the purified DNA fragments and incubated at 37°C for 40 min. The fragments were then electrophoresed through a 1% low melting agarose (SeaPlaque; Flowgen) gel and

Double stranded cosmid DNA → **PCR (primers designed with synthetic _Kpn_ 1 restrictions sites)** → **Restriction endonuclease digestion** → **Gel extraction of PCR products**

pSV- βgal Vector (Promega) → **Restriction endonuclease digestion** → **Purification of DNA** → **Addition of calf intestinal phosphatase** → **Gel extraction of upper band**

Gel 1 markers: ←613, ←410, ←234, ←147, ←99 bp

Gel 2 markers: ←6590, ←500 bp

Ligation & transformation → **Check orientation (PCR or sequencing)**

Fig. 5.1 Schematic diagram of the protocol used to produce truncated molecular constructs using the 5' flanking region of _EtMIC1_.

The ethidium bromide-stained gel (1) shows PCR-amplified products of the 5' upstream region of _EtMIC1_, ranging in size from 99 to 613 base pairs. The ethidium bromide-stained gel (2) shows _Kpn_I digested pSV-βgal vector. In both gels the molecular weight markers (M) are in the first lane. The size of the amplified or digested products are shown in base pairs. A similar process was used to produce truncated constructs for _EtMIC2_, and _3_.

122

visualised with ethidium bromide (see Fig 5.1 for an example). The upper 6254 bp band was cut out and purified with a QIAgen gel extraction kit according to the manufacturer's instructions (the lower 567 bp band was discarded).

KpnI-digested PCR products were each ligated into KpnI-digested pSV-β-gal and transformed into E. coli XL1-Blue cells (see section 2.2.3 - 2.2.6) All constructs were sequenced to check that the flanking sequences were in the correct orientation and fused in frame to lacZ.

Caesium chloride purified DNA of each construct was resuspended at a final concentration of 100 μg under ethanol (100 μl DNA, 250 μl 95% ethanol and 10 μl 3 M sodium acetate, pH 4.8) and stored at -20°C. When required for transfection, DNA was centrifuged at 13,000g for 5 min, the supernatant removed and the pellet washed in 70% ethanol. Finally the dried DNA pellet was resuspended in 100 μl of complete cytomix (see section 2.2.11).

5.3 Results

5.3.1 Promoter mapping by 5' deletion analysis of the untranscribed region of EtMIC1

Sporozoites of E. tenella were electroporated with plasmids containing 5' upstream sequences from EtMIC1 fused to lacZ, and allowed to infect Madin-Derby Bovine Kidney (MBDK) cells. After ~24 h, cells were fixed and the expression of β-galactosidase was visualised by X-gal staining, for ~3 h (see Fig 5.2 for an example of X-gal staining). The minimal sequence required for expression was first determined by transfecting sporozoites with plasmids bearing progressively larger deletions in the EtMIC1 5' upstream region (Fig. 5.3). As differences between the constructs all lie outside the transcribed region, changes in β-galactosidase activity levels are presumed to reflect relative promoter strength.

Digestion of pSV-βgal with KpnI removed the lac Z translational start site and 567 bp of upstream sequence, including the SV40 early promoter/enhancer sequences. Sequences downstream of the lacZ coding region were left intact including the first SV40 poly (A) addition signal which is 99 bp downstream of the lacZ translational stop codon. It had been shown previously that introduction of the 3' untranscribed region of EtMIC1 reduced the number of blue parasites that were obtained by transient transfection (Kelleher and Tomley, 1998) so no downstream sequences were included in any constructs in the current study.

Pmic-1-βgal, which contains 761 bp of 5' upstream sequence (651 bp upstream of the transcription start site plus 110 bp of transcribed sequence upstream of the translational start site) and the translational initiation codon of *EtMIC1* fused to *lacZ*, readily directs expression of β-galactosidase as shown previously (Kelleher and Tomley, 1998). For each experiment at least two separate transfections experiments were performed with a minimum of eight replicas for each plasmid investigated. The results described here and shown in the figures as bar charts represent the mean number of transfected sporozoites in each experiment and the error bars on the figures indicate the standard deviation from the mean. Pmic1-βgal-5'613, which contains 613 bp of the upstream sequence of *EtMIC1* (503 bp of untranscribed and 110 bp of transcribed sequence) gave comparable numbers of blue parasites to pmic-1-βgal, suggesting that sequences between -651 and -503 upstream of the transcriptional start site have no effect on expression (Fig 5.4). The next deletion construct, pmic1-βgal-5'410, (300 bp of untranscribed and 110 bp of transcribed sequence) resulted in a 4-fold reduction in the number of blue parasites (average reduction from three experiments, each with eight replicas), suggesting that sequences between -300 and -503 contribute to expression of *EtMIC1*. This region contains a AGAGACC motif, similar to the A/TGAGAGC found in the upstream regions of *SAG1, GRA1, 2, 5* and *6* and *TUB1* (Soldati *et al.*, 1995; Mercier *et al.*, 1996) and the AGAGACGC found upstream of *NTPase1* and *3* (Nakaar *et al.*, 1992) of *T. gondii*. The next smallest construct, pmic1-βgal-5'234, (124 bp of untranscribed and 110 bp of transcribed sequence) resulted in a further 2-fold reduction in the number of blue parasites (average reduction from three experiments), suggesting that sequences between -300 and -124 also have a role in expression. After transfection of sporozoites with the two remaining constructs, pmic1-βgal-5'147, which included the predicted transcription start site (37 bp of untranscribed and 110 bp of transcribed sequence) and pmic1-βgal-5'99, which lacked the predicted transcription start site (99 bp of transcribed sequence only) respectively, no blue parasites were detected in any of the three separate transient transfection experiments. Thus these transient transfection assays defined the minimal promoter sequence of *EtMIC1* as lying between -124 and -37 bp upstream of the transcriptional start site. Interestingly the sequence covering the putative Inr motif, which is included in plasmid pmic1-βgal-5'147, is insufficient to produce detectable β-galactosidase activity in this assay.
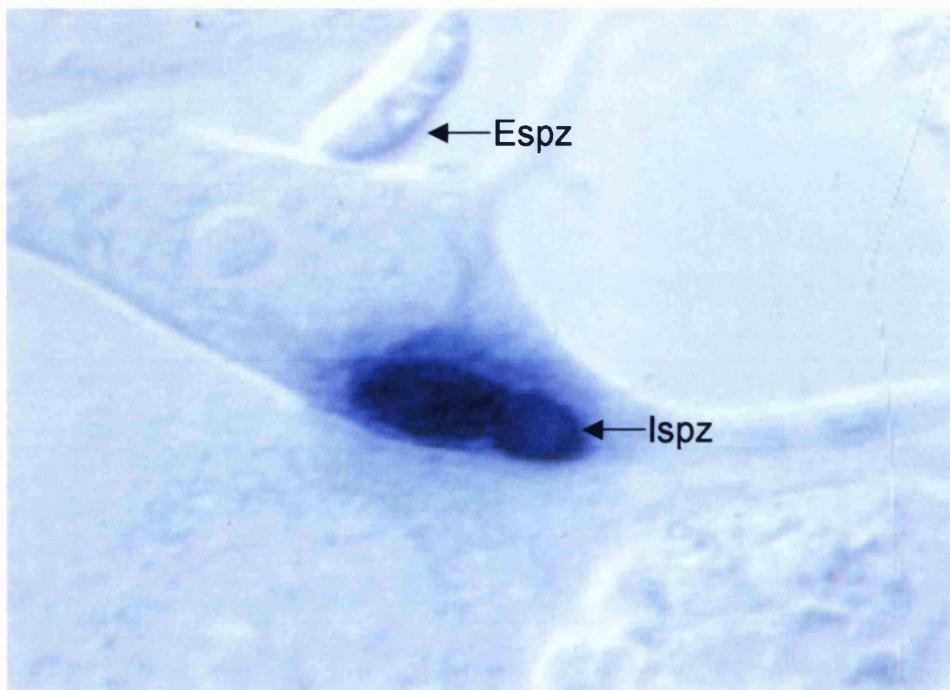
Fig. 5.2 Visualisation of β-galactosidase activity in developing sporozoites of *E. tenella* (Wis F96 strain).
Freshly excysted sporozoites were electroporated with 100 μg of plasmid and allowed to infect monolayers of MBDK cells at 41°C for 24 h, before fixing in 5% glutaraldehyde and staining with Xgal. A transfected intracellular sporozoite (Ispz) and an extracellular untransfected sporozoite (Espz) are shown.
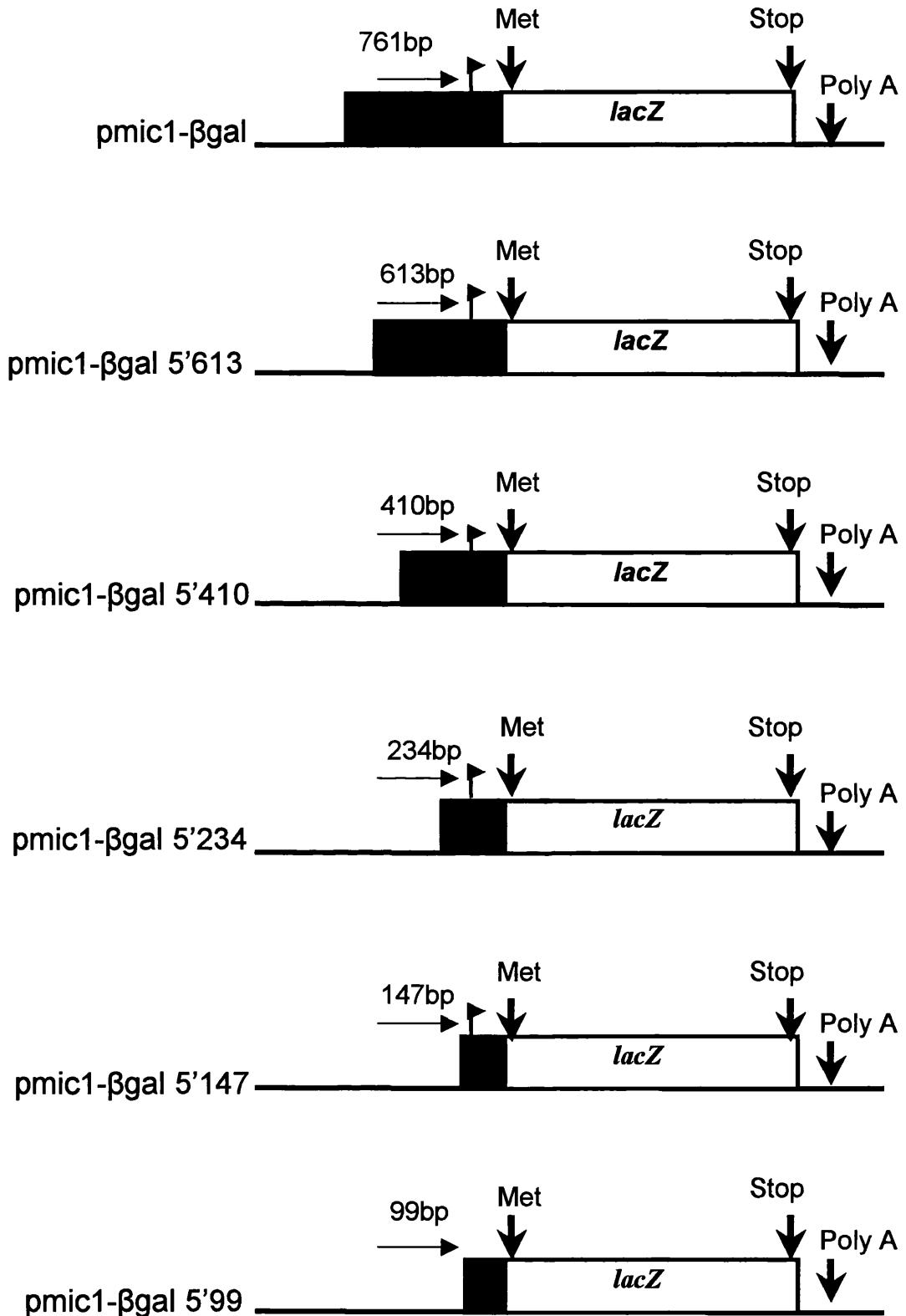
Fig. 5.3 Schematic diagram of initial set of *EtMIC1* plasmid constructs used in transient transfection experiments.

Filled boxes represent 5' upstream sequences of the microneme gene, *EtMIC1*. The open box illustrates coding regions of the *lacZ* gene and the thin line designates pSV-βgal sequence (Promega), which includes the first SV40 poly(A) addition signal which is 99 bp beyond the authentic *lacZ* translational stop codon. Arrows show the orientation of upstream UTRs (3' at arrow head). Bold arrow heads show the positions of initiation codons (Met, from *EtMIC1*), termination codons (Stop, from pSV-βgal vector) and proposed polyadenylation signals (Poly A, from pSV-βgal vector). The flag represents the transcription start site. Genetic elements are not drawn to scale.

126
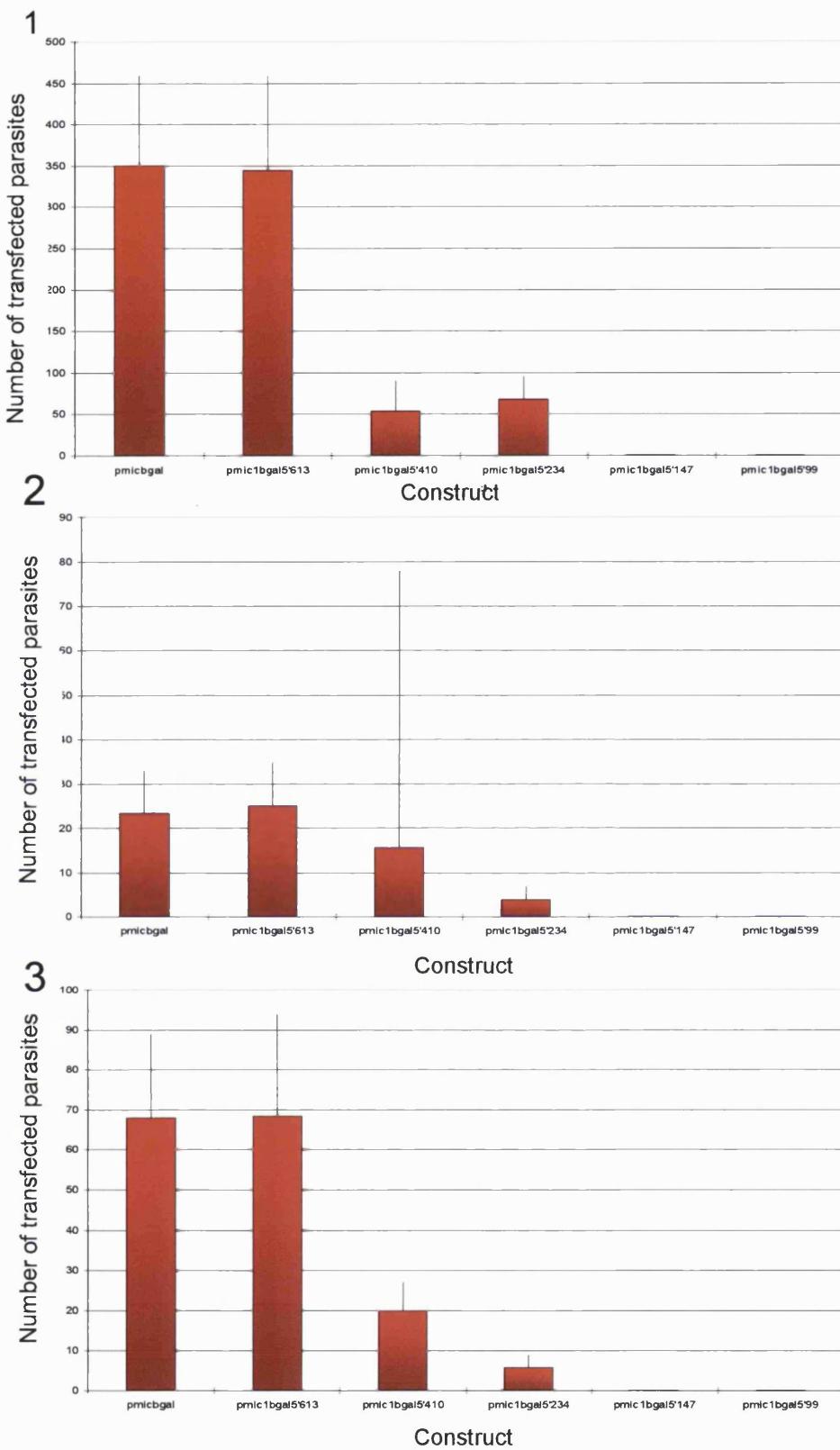
Fig. 5.4 Transient transfection analyses of *E. tenella EtMIC1* promoter.
Functional analysis of *EtMIC1* 5' upstream region by transient transfection of sporozoites of *E. tenella*. Various fragments from the *EtMIC1* upstream region were cloned upstream of a promoterless *lacZ* gene in pSV-βgal. Sporozoites were electroporated with 100 μg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. The number of blue (transfected) sporozoites was counted and as the differences between the constructs all lie outside the transcribed region, changes in the number of transfected sporozoites are presumed to reflect relative promoter strength. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

127
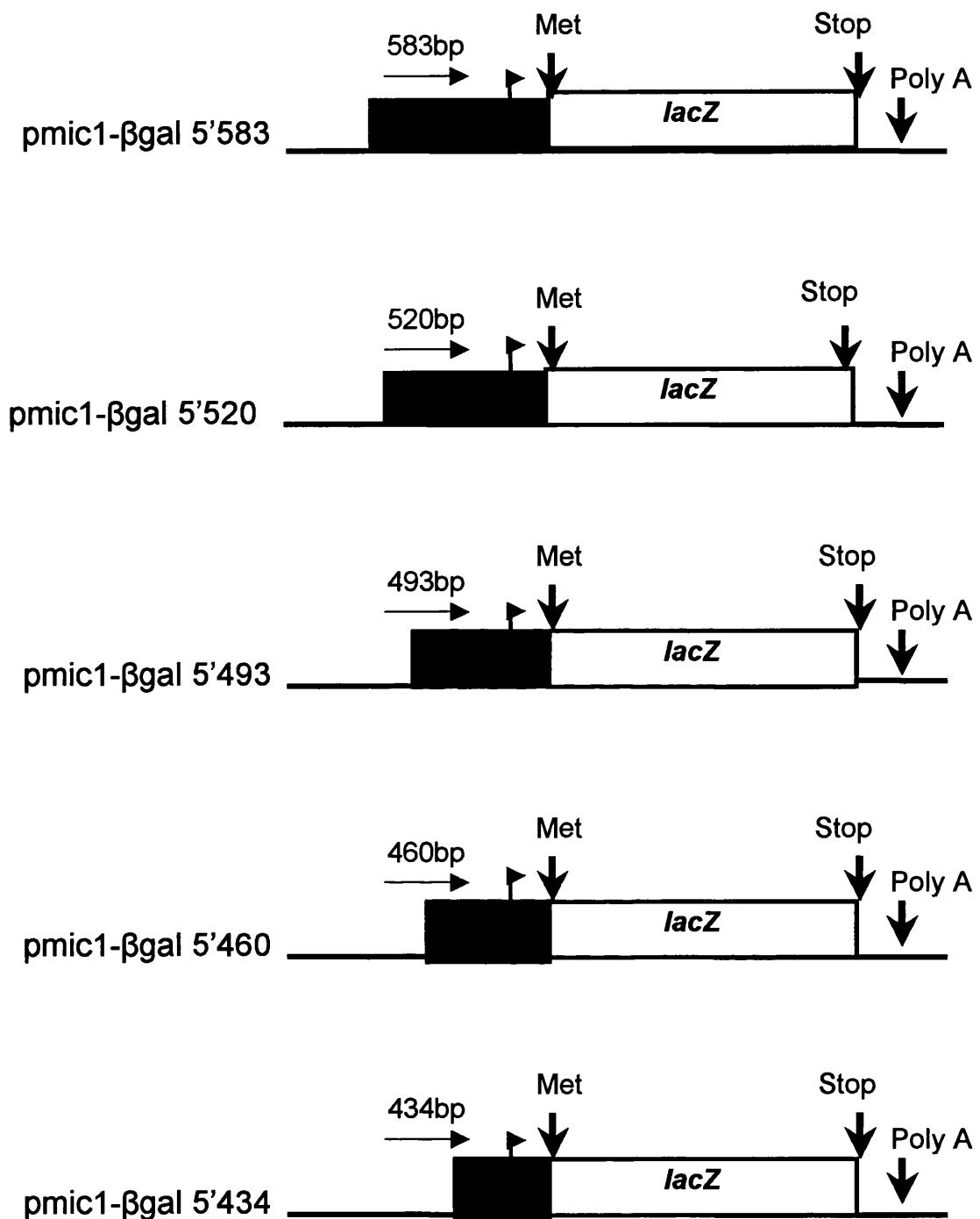
Fig. 5.5 Schematic diagram of EtMIC1 plasmid constructs used in transient transfection experiments.

Filled boxes represent 5' upstream sequences of the microneme gene, *EtMIC1*. The open box illustrates the coding region of the *lacZ* gene and the thin line designates pSV-βgal sequence (Promega), which includes the first SV40 poly(A) addition signal which is 99 bp beyond the authentic *lacZ* translational stop codon. Arrows show the orientation of upstream UTRs (3' at arrow head). Bold arrow heads show the positions of initiation codons (Met, from *EtMIC1*), termination codons (Stop, from pSV-βgal vector) and proposed polyadenylation signals (Poly A, from pSV-βgal vector). The flag represents the transcription start site in the upstream region of *EtMIC1*. Genetic elements are not drawn to scale.

1

2

Fig. 5.6 Transient transfection analyses of *E. tenella EtMIC1* promoter.
Functional analysis of *EtMIC1* 5' flanking region by transient transfection of sporozoites of *E. tenella*. Various fragments from the *EtMIC1* upstream region was cloned upstream of a promoterless *lacZ* gene in pSV-βgal. Sporozoites were electroporated with 100 μg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. The number of blue (transfected) sporozoites was counted. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

129

Because the region between -503 and -300 bp upstream of the transcriptional start site of the *EtMIC1* promoter, clearly influenced the number of blue parasites obtained, a further series of deletion constructs was generated to analyse this region (Fig 5.5). In two separate transfection experiments constructs pmic1-βgal-5'583 (473 bp of untranscribed and 110 bp of transcribed sequence) and pmic1-βgal-5'520 (410 bp of untranscribed and 110 bp of transcribed sequence) gave similar numbers of blue parasites to pmic1-βgal-5'613, suggesting that upstream sequences between -410 and -503 were not important for expression (Fig. 5.6). The next smallest deletion construct, pmic1-βgal-5'493, (383 bp of untranscribed and 110 bp of transcribed sequence) which had the AGAGACC motif deleted again gave similar numbers of blue parasites when compared to pmic1-βgal-5'613, pmic1-βgal-5'583 and pmic1-βgal-5'520. Therefore this 120 bp region (between -503 and -383 bp) and the AGAGACC motif within it do not appear to influence expression (Fig. 5.6). The next construct pmic1-βgal-5'460, (350 bp of untranscribed and 110 bp of transcribed sequence) contained a further truncation of just 33 bp and gave a dramatic reduction (~61% on average from the eight replicas of each plasmid in the two transfection assays) in the number of blue parasites compared to pmic1-βgal-5'493, suggesting that the 33bp region between -383 and -350 contains sequences that are important for expression (Fig. 5.6). The final two constructs pmic1-βgal-5'434 (324 bp of untranscribed and 110 bp of transcribed sequence) and pmic1-βgal-5'410 gave similar numbers of blue parasites to pmic1-βgal5'460, demonstrating that sequence between -350 and -300 do not contribute to promoter activity (Fig. 5.6). Analyses of the 33 bp region between -383 and -350 did not reveal any common *cis*-acting motifs observed in other apicomplexan promoters. However TRANSFAC analysis (Heinemeyer *et al.*, 1998) identified multiple potential transcription factor binding sites for transcription factors including Dof, which can recognise 5'-AAAG-3' motifs and signal transducers and activators of transcription (STAT) which recognise 5'-TTCC-3' binding motifs. Also observed were binding sites for a CCAAT enhancing domain (CED) which recognises a 5'-GGAA-3' motif, a nuclear associated factor (NAF), which recognises 5'-GAAA-3' binding site, a heme associated protein (HAP), which binds to a 5'-CCAA-3' motif and E74-like factor 1 (ELF) which recognises a 5'-GTTT-3' motif.

5.3.2 Promoter mapping by 5' deletion analysis of the untranscribed region of *EtMIC2*

A similar system was employed to analyse the 5' upstream sequence of *EtMIC2* to establish the minimal sequence necessary to drive β-galactosidase expression. Pmic1-βgal-5'613 was used as a positive control for each transfection assay. Five constructs with increasingly larger truncations of the 5' upstream sequence of *EtMIC2* and the translational initiation codon fused to *lacZ* were made (Fig 5.7).

Constructs pmic2-βgal-5'1046 (952 bp of untranscribed and 94 bp of transcribed sequence) and pmic2-βgal-5'608 (514 bp of untranscribed and 94 bp of transcribed sequence) gave comparable numbers of blue parasites to each other when the mean values from eight replicas in two transfection experiments were compared (Fig 5.8) suggesting that sequences between -952 and -514 did not contribute to expression of *EtMIC2*. The next construct pmic2-βgal-5'245 (151 bp of untranscribed and 94 bp of transcribed sequence) gave an average of 23% fewer blue parasites in the transfection experiments demonstrating that sequences between -514 and -151 make some contribution to expression (Fig 5.8). Pmic2-βgal-5'131, (37 bp of untranscribed and 94 bp of transcribed sequence) which included both predicted transcription start sites one of which was within Inr motif, resulted in a dramatic decrease of ~70% (this is the mean value of eight replicas in two experiments), in the number of blue parasites when compared to pmic2-βgal-5'245 suggesting that sequences between -151 and -37 are important for expression (Fig 5.8). After transfection of sporozoites with the final construct in this group, pmic2-βgal-5'89, (89 bp of transcribed sequence only) which did not contain the potential transcription start site, no blue parasites were detected in any transfection experiment (Fig 5.8).

This set of transfection assays suggest that the minimum promoter sequence for *EtMIC2* lies between the transcriptional start sites and -37. Interestingly for *EtMIC2* a construct with only 37 bp of the 5' untranscribed sequence was sufficient for expression, whereas much more of the 5' untranscribed region of *EtMIC1* is required for its expression, suggesting that there is a qualitative difference in the ability of sequences surrounding the Inrs of *EtMIC1* and *EtMIC2* to drive expression.

The transfection assays described above clearly showed that, the regions between -514 and -37 bp upstream of the transcriptional start codon significantly influence the number of blue parasites obtained. To investigate these regions further another set of

deletion constructs was generated (Fig. 5.9). Constructs pmic2-βgal-5'558, 475, 401, 332 and 291 (464, 381, 307, 238 and 197 bp of untranscribed and 94 bp of transcribed sequence, respectively) gave comparable numbers of blue parasites to pmic2-βgal-5'608 when the mean values from two experiments were examined, demonstrating that the sequences between -514 and -197 are not important in expression (Fig 5.10). The next smallest truncated construct contained 151 bp of untranscribed and 94 bp of transcribed sequence showed a significant decrease (an average of ~50% from the eight replicas in the two experiments) in the number of blue parasites obtained when compared to pmic2-βgal-5'291 (Fig 5.10), suggesting this region (between -151 and -197) greatly influences expression. The next two constructs pmic2-βgal-5'208 and pmic2-βgal-5'179 did not show decreases in the number of blue parasites observed, suggesting that the region between -151 and -85 do not greatly influence expression. The next construct pmic2-βgal-5'152 did show a decrease in the number of blue parasites observed (~50%) indicating that the region between -151 and -37 does make a significant contribution to *EtMIC2* promoter activity (Fig 5.10).

5.3.3 Promoter mapping by 5' deletion analysis of the untranscribed region of *EtMIC3*

To determine the minimum sequence of the upstream sequence of *EtMIC3* required for promoter activity, truncation constructs similar to those for pmic1-β-gal were made using the same strategy. Four constructs, with 2268 bp, 629 bp, 127 bp or 45 bp (excluding the predicted Inr) of the 5' flanking sequence of *EtMIC3* and its translational initiation codon fused to *lacZ*, were made (Fig 5.11). However transfection of sporozoites with all constructs failed to produce blue parasites.

To investigate the possibility that expression of pmic3-βgal-5'1945 was stage specific, the transient transfection assays were repeated and sporozoites were allowed to develop through to the formation of mature schizonts (~50-60 h post incubation). Visualisation with X-gal again failed to detect blue parasites in the schizont however molecular constructs of *EtMIC1* and *EtMIC2* also resulted in complete loss of β-galactosidase expression in schizonts. If development were allowed to continue to the formation of merozoites, blue parasites were observed following electroporation with constructs of *EtMIC1* and *EtMIC2* however *EtMIC3* constructs failed to produce blue parasites.
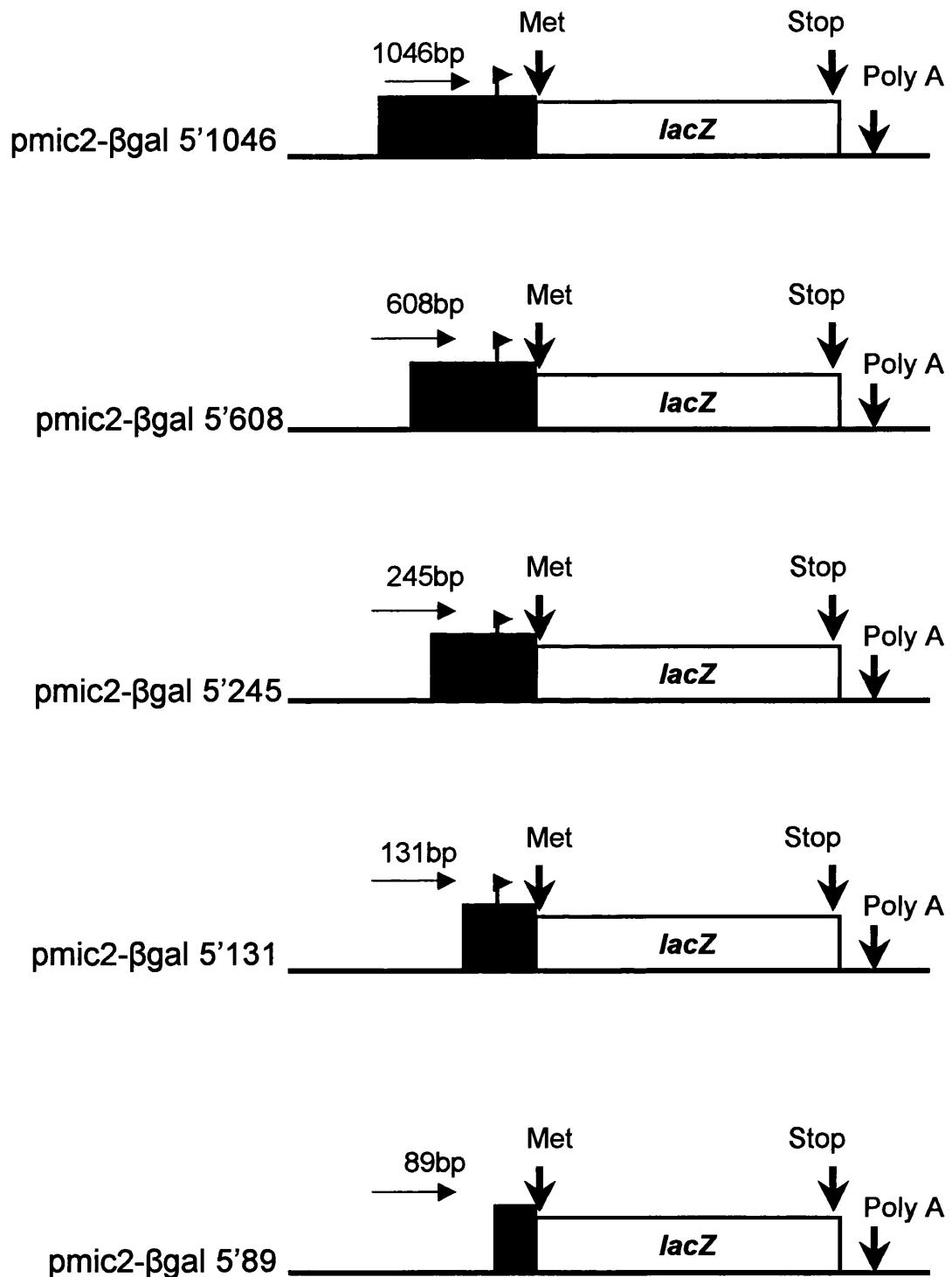
Fig. 5.7 Schematic diagram of initial set of *EtMIC2* plasmid constructs used in transient transfection experiments. Filled boxes represent 5' upstream sequences of the microneme gene, *EtMIC2*. The open box illustrates the coding region of the *lacZ* gene and the thin line designates pSV-βgal sequence (Promega), which includes the first SV40 poly(A) addition signal which is 99 bp beyond the authentic *lacZ* translational stop codon. Arrows show the orientation of upstream UTRs (3' at arrow head). Bold arrow heads show the positions of initiation codons (Met, from *EtMIC1*), termination codons (Stop, from pSV-βgal vector) and proposed polyadenylation signals (Poly A, from pSV-βgal vector). The flag represents the transcription start site in the upstream region of *EtMIC1*. Genetic elements are not drawn to scale.

133

Fig. 5.8 Transient transfection analyses of _E. tenella EtMIC2_ promoter.

Functional analysis of _EtMIC2_ 5' flanking regions by transient transfection of sporozoites of _E. tenella_. Various fragments from the _EtMIC2_ upstream regions was cloned upstream of a promoterless _lacZ_ gene in pSV-βgal. Sporozoites were electroporated with 100 µg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. The number of blue (transfected) sporozoites was counted. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

134
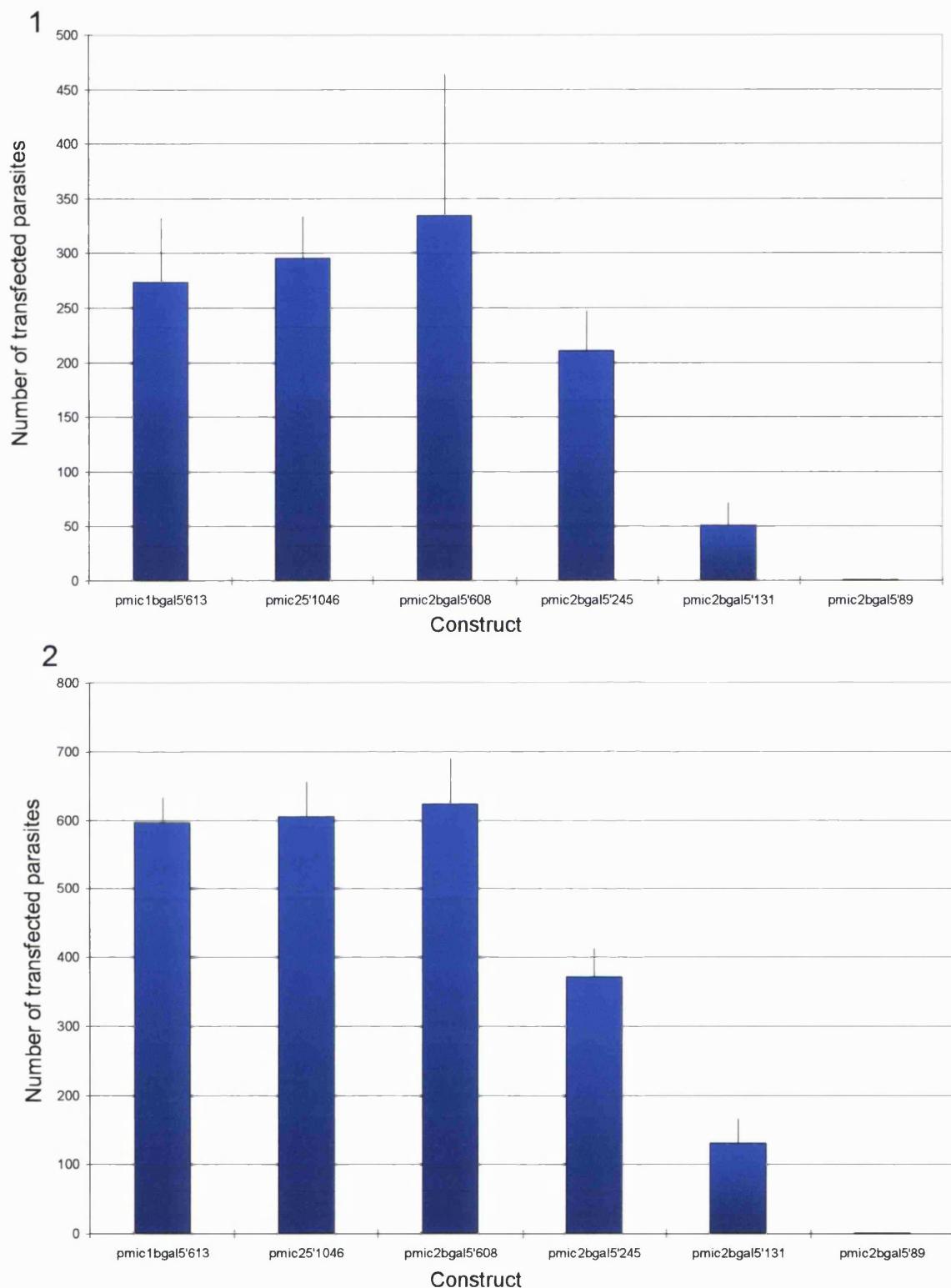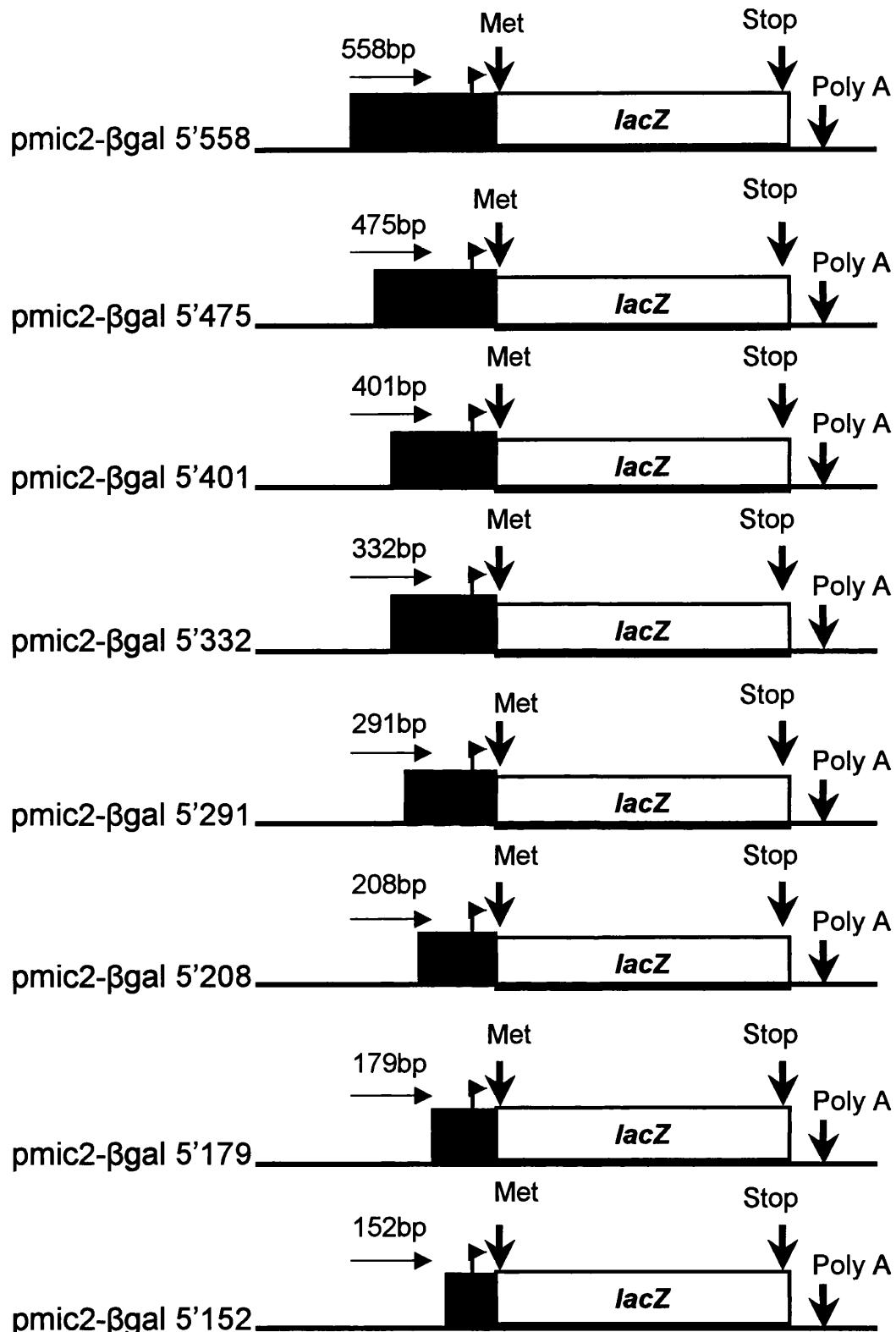
Fig. 5.9 Schematic diagram of *EtMIC2* plasmid constructs used in transient transfection experiments.

Filled boxes represent upstream flanking sequences from the microneme gene of *Eimeria*, *EtMIC2*. The open box illustrates the coding region of the *lacZ* gene and the thin line designates pSV-βgal sequence (Promega), which includes the first SV40 poly(A) addition signal which is 99 bp beyond the authentic *lacZ* translational stop codon. Arrows show the orientation of upstream UTRs (3' at arrow head). Bold arrow heads show the positions of initiation codons (Met, from *EtMIC1*), termination codons (Stop, from pSV-βgal vector) and proposed polyadenylation signals (Poly A, from pSV-βgal vector). The flag represents the transcription start site in the upstream region of *EtMIC2*. Genetic elements are not drawn to scale.

1

2
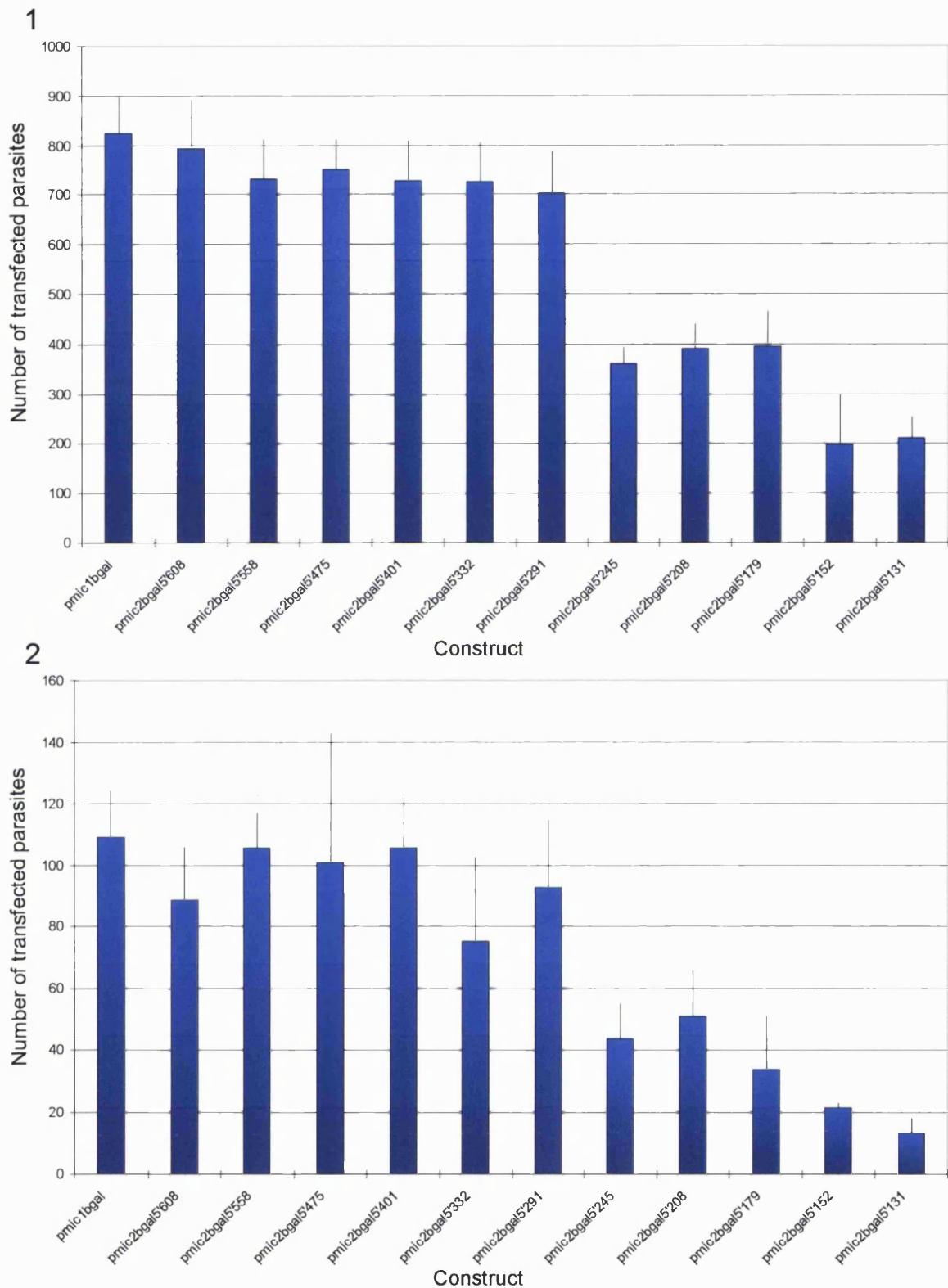
Fig. 5.10 Transient transfection analyses of *E. tenella* EtMIC2 promoter.
Functional analysis of *EtMIC2* 5' flanking regions by transient transfection of sporozoites of *E. tenella*. Various fragments from the *EtMIC2* upstream regions was cloned upstream of a promoterless *lacZ* gene in pSV-βgal. Sporozoites were electroporated with 100 μg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. The number of blue (transfected) sporozoites was counted. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

136

Fig. 5.11 Schematic diagram of *EtMIC3* plasmid constructs used in transient transfection experiments. Filled boxes represent upstream flanking sequences from the microneme gene, *EtMIC3*. The open box illustrates the coding region of the *lacZ* gene and the thin line designates pSV-βgal sequence (Promega), which includes the first SV40 poly(A) addition signal which is 99 bp beyond the authentic *lacZ* translational stop codon. Arrows show the orientation of upstream UTRs (3' at arrow head). Bold arrow heads show the positions of initiation codons (Met, from *EtMIC3*), termination codons (Stop, from pSV-βgal vector) and proposed polyadenylation signals (Poly A, from pSV-βgal vector). The flag represents the transcription start site in the upstream region of *EtMIC2*. Genetic elements are not drawn to scale.

5.4 Discussion

The transient transfection system used for these studies is "semi-quantitative" and attempts to develop an accurate chemiluminescence system were not successful. X-gal staining for β-galactosidase is not quantitative because the intensity of blue staining varied from parasite-to-parasite, possibly because the number of copies of plasmid that each parasite received up during electroporation varies. It was not possible to quantitate the intensity of X-gal staining by eye, so the numbers of parasites that stained blue after a specified staining period, usually three hours were scored arithmetically. A transient transfection system using green fluorescence protein (GFP) would make quantitation possible, for example by chemiluminescence or flow cytometry, but this has not yet been developed for E. tenella.

It might be expected that although a similar number of blue parasites should appear when using any functional promoter, the intensity of X-gal staining would vary according to the promoter strength. During the analysis, it was clear that the numbers of blue parasites varied significantly according to the length of 5' upstream sequence present. The most likely explanation is that as the promoter strength decreased, the lower limit of sensitivity for the detection of X-gal staining by eye was reached, particularly for those parasites in the population that had received a low dose of plasmid. To be confident with the data, all transfections were repeated at least twice with a minimum of eight replicas for each plasmid construct used.

As well as copy number, differences in expression levels between experiments can be attributed to differences in biological parameters. These include the age of the oocyst culture from which the sporozoites were prepared, the age of MBDK cells used and the length of time parasites were left before electroporation with plasmid DNA. No β-galactosidase activity was observed when sporozoites were incubated with constructs and not electroporated or if sporozoites were electroporated with the parent plasmid, pSV-βgal (Promega). β-galactosidase activity was not due to bacterial contamination as all buffers, media and samples were plated onto LB (Luria Bertani) agar, incubated overnight at 37°C and no colonies were detected.

As microneme proteins are expected to share similar trafficking mechanisms, are secreted simultaneously upon parasite invasion (Dubremetz et al., 1998b) and appear in a co-ordinated fashion at both mRNA and protein levels (Ryan et al., 2000) their expression may be regulated by similar processes at the transcriptional level. Therefore, it could be predicted that such genes would contain common, regulated cis-

acting elements in their promoter regions. Analysis of the upstream DNA sequence of the five microneme-encoding genes (EtMIC1-5), identified possible cis-acting elements, including Inr motifs, TATA boxes, polypurine rich elements and GAAC motifs sequences, all of which have been observed in the promoter regions of parasite genes and, in some cases, have been found to be critical for gene expression.

The current study used transient transfection assays in which the β-galactosidase gene was driven by the 5' upstream region of the microneme-encoding genes to define regions important for promoter activity and to identify important elements involved in expression. Deletion analyses of the gene encoding microneme protein EtMIC1 demonstrated that sequences upstream of position -383 (383 bp upstream of the transcription start site) are not required for a significant increase in expression levels of the reporter product, β-galactosidase and that 124 bp upstream of the transcription start site is the minimum required to drive expression. However the minimum may be less than this as analysis of the 87 bp region between positions -124 and -37 (which was not sufficient to promote expression) was not performed.

Deletion analyses of the upstream region of EtMIC1 also demonstrated a significant decrease in expression when a degenerate TATA box (-144) was removed thus implicating its involvement in gene expression. Also the potential Inr alone was insufficient for promoter activity, indicating that the DNA sequences between -124 and -37 must contain elements essential for gene expression. Within this region is a GAAC motif (-81), like those observed in E. histolytica as described in Chapter Four. In E. histolytica this motif controls the rate of gene expression and both it and a degenerate TATA box are involved in protein complex assembly at the core promoter of hgl5 (Singh and Rogers, 1998; Singh et al., 1997; Singh et al., 2002). Both truncation assays implicate an involvement of the above motifs in gene expression, however they are some distance upstream of the Inr motif. Thus it seems unlikely that they function in a similar way as described in Entamoeba genes.

It has been suggested that purine rich elements play a role in the transcription initiation and possible act like SP1 elements to determine the transcription start site by protein-protein interaction in a TATA-less promoter (Kadonaga et al., 1986; Soldati and Boothroyd, 1995). SP1recognise GC-rich sequences which are usually found in tandem repeats and in either orientation. The most important SP1-responsive element in gene expression is the one nearest the transcription start site (Kadonaga et al., 1986) as with the repeats of the GRA proteins. As there was only a single purine rich motif

and as its presence or absence did not affect expression it seemed highly unlikely that that this motif functions like SP1 responsive motifs. Truncation analyses of *EtMIC1* demonstrated that the presence or absence of the AGAGACC sequence was not significant in aiding the expression of *EtMIC1* unlike its function *SAG1, TUB1* or *GRA1, 2, 5* and *6*. Further upstream were two GGAGAGG motifs as observed in *NTPase 1* and *3* (Nakaar *et al.*, 1992) however due to time limitations it was not possible to investigate if they influence expression of β-galactosidase. Transfection analyses identified a 33 bp region which made a significant contribution to expression, however no common *cis*-acting motifs observed in other apicomplexan promoters were found in this region. TRANSFAC analysis (Heinemeyer *et al.*, 1998) identified multiple potential transcription factor binding sites including Dof, STAT, CED, NAF, HAP and ELF. The involvement that these factors may have in expression remains to be elucidated.

Transfection assays using the upstream region of *EtMIC2* suggest that 37 nucleotides can function as a minimum promoter. However the minimum promoter may be less than this as analysis of the region between positions -37 (which was sufficient to promote expression) and the transcription start site was not preformed. Analyses of this region did not identify any common *cis*-acting elements observed in other apicomplexan promoters. Interestingly for *EtMIC2*, a construct with only 37 bp of 5' untranscribed sequence was sufficient for expression, whereas much more of the 5' untranscribed region of *EtMIC1* is required for its expression, suggesting that there is a qualitative difference in the ability of sequences surrounding the Inrs of *EtMIC1* and *EtMIC2* to drive expression. Deletion analysis of the 5' flanking sequence of *EtMIC2* demonstrated that loss of the degenerate TATA box and the GAAC motif did not abolish expression, unlike its effect in the promoter region of *EtMIC1*.

Deletion analysis of the upstream sequences of *EtMIC3* resulted in a failure to drive expression of the *lacZ* gene when sporozoites were electroporated and allowed to invade a monolayer of MBDK cells for approximately 24 hours. As a result it was not possible to investigate the importance of the GAAC motif or the TATA box. The failure to detect β-galactosidase could be due to sensitivity of the system or lack of expression at this life cycle stage, however expression was not detected in schizonts either. Constructs of *EtMIC3* failed to drive expression of β-galactosidase in the schizont, however molecular constructs of *EtMIC1* and *EtMIC2* also did not promote expression in schizonts, demonstrating that expression of *EtMIC1* and *EtMIC2* are

stage specific. If development was allowed to continue to the production and release of merozoites, expression of β-galactosidase was observed in merozoites. These results correlate with the results of ultrastructural studies where it is clear that micronemes are formed afresh during each successive stage of the life cycle (Chobotar *et al.*, 1975; Dubremetz, 1975; Dubremetz, 1979). Earlier RT-PCR results are in agreement with this scenario, also microneme proteins EtMIC2 and EtMIC5 gradually disappear during early schizogony and reappear as invasive merozoites mature, suggesting that microneme protein expression is co-ordinated and occurs only when micronemes are being assembled in readiness for the next round of host cell invasion (Brown *et al.*, 2000b; Tomley *et al.*, 1996). All of these results agree with the detection of β-galactosidase activity during sporozoite and merozoites maturation. This observation of down regulation in gene expression during the conversion from one life cycle stage to another has been observed for other genes in other protozoan species. For example, *SAG1* and *SAG2* of *T. gondii* are abundantly expressed in the tachyzoite stage and down regulated during the conversion to slow growing stages, such as the bradyzoite stage of the parasite. Also the high levels of gene expression from the *NTP3* promoter is reduced when tachyzoites differentiate into the bradyzoites (Nakaar *et al.*, 1992).

Sequence analyses identified potential Inrs in the upstream regions of *EtMIC1, 2* and *3*. Deletion constructs used in transient transfection assays demonstrated that the potential Inr of *EtMIC2* was sufficient for basal expression. Due to the prevalence and importance of these motifs in gene expression we decided to investigate which nucleotides are of most significance in the consensus sequence. To do, this transient transfection assays were performed in conjunction with molecular constructs, which contained single point mutations or deletions in the potential Inr motif.

Chapter Six
Mutational analysis of untranscribed regions of five microneme genes using transient transfection in *Eimeria tenella*

6.1 Introduction

The results of experiments described in Chapters Three, Four and Five indicate that differential expression of microneme-specific mRNAs occurs at the transcriptional level. 5' RACE analyses mapped the transcription start sites of *EtMIC1, 2* and *3* to adenosine nucleotides which were surrounded by sequence which conformed to Inr motifs, which have been identified in a number of apicomplexan genes (Nakaar *et al.*, 1992; Soldati and Boothroyd, 1995; Shiels *et al.*, 2000; Elmendorf *et al.*, 2001). Transient transfection assays in Chapter Five, demonstrated that the Inr alone was not sufficient for promoter activity and that a minimum of 124 nucleotides (upstream of the translational start site) is required for gene expression. Deletion analyses demonstrated that the potential Inr of *EtMIC2* is sufficient for promoter activity, in addition to 37 bp of untranscribed sequence.

Conservation of the Inr element at the transcription start site of *EtMIC1, 2* and *3* and its similarity to other protozoan Inr elements led us to question whether this apparent Inr functions as a core promoter element. The principal method for identifying the DNA sequence elements that control transcription in eukaryotes has been to mutagenise native DNA sequences near the start of transcription (either by deletion or single/multiple mutational analysis). Changes in gene expression of the altered templates can be tested either by reintroduction into the cell or by introduction into an experimental system such as an *in vitro* transcription system. The accumulation of the RNA, protein or a reporter product (β-galactosidase in this case) can be measured after a period of time and from these measurements an inference can be made about the effect of sequence changes on the strength of the promoter.

Previous mutational analysis of the weak consensus Inr sequence in mammalian TATA-containing genes demonstrated that mutations in the region resulted in use of an alternative transcription start site and/or a reduction in promoter strength (Talkington and Leder, 1982; Dierks *et al.*, 1983; Concino *et al.*, 1984). Also accurate initiation occasionally was observed following specific mutagenesis of the TATA box (Hen *et al.*, 1982; Dierks *et al.*, 1983; Jones *et al.*, 1988). Previous functional

mutational analysis of mammalian and trichomonad Inrs revealed that certain nucleotides within the motif are critical for its efficient function, such as the presence of a TCA^{+1} motif and the pyrimidine content surrounding it (Liston and Johnson, 1999; Javahery *et al.*, 1994; Lo and Smale, 1996). It has been demonstrated that a tyrosine nucleotide is preferred at position +3 (O'Shea-Greenfeild and Smale, 1992). Moreover if a large number of pyrimidines are present and surrounds the start site, then low levels of Inr activity can be imparted in the absence of either the adenosine nucleotide at +1 or the tyrosine nucleotide at +3 (O'Shea-Greenfeild and Smale, 1992). Overall the most critical nucleotides, within the Inr consensus $(PyPyA^{+1}PyT/APyPy)$, for determining the strength of the Inr are the nucleotides at positions +1 (A), +3 (preferably T or A) and -1 and -2 (C or T) (Javahery *et al.*, 1994; Lo and Smale, 1996).

To determine whether these conserved sequences play an essential, functional role and whether the critical nucleotides within the Inrs of microneme-coding genes, are similar to those observed in other Inrs of apicomplexan genes and higher eukaryotes, we carried out detailed mutational analyses of some of the conserved nucleotides and assayed their effects in transient transfection studies. However an enormous number of mutation and deletion combinations exist in the simple 7 base pair Inr consensus and due to time constraints only a limited number of mutant combinations could be produced.

Also transient transfection analyses of the upstream region of *EtMIC1* demonstrated that a 33 bp region between -383 and -350 contains sequences that are important for expression, as demonstrated with constructs pmic1-βgal-5'493 and pmic1-βgal-5'460. TRANSFAC analyses (Heinemeyer *et al.*, 1998) of this region identified potential binding sites for activator or repressor proteins which are typically present and serve to regulate the efficiency and specificity of transcription. Mutations were made in some of the conserved, core nucleotides within the potential binding sites and the mutant constructs were assayed in transient transfection studies, to gain an insight into the importance of the potential transcription factors.

6.2 Methods

Sporozoites were purified as detailed in section 2.1.2 - 2.1.4 and counted using a Fuchs Rosenthal counting chamber. Sporozoites were washed once in incomplete cytomix and then resuspended in complete cytomix at a concentration of 1.4 x 10^7 per ml and were stored on ice prior to transfection. All solutions, wool and tubes used were sterilised before use.

6.2.1 Site directed mutagenesis

Site-directed PCR-based mutagenesis was performed essentially as described by Hutchinson (1978). The DNA to be mutagenised was part of the 5' upstream regions of both *EtMIC1* and *2*. Each PCR contained 50 ng of cosmid DNA, 12.5 pmol of each of two suitable primers (see Appendix C), 10 nmol dNTP's (100 mM dNTP mixture of each), 2.5 U of *Taq* DNA polymerase and made up to 50 µl with 1x PCR buffer (both from Amersham Pharmacia Biotech). All DNA templates were denatured for 3 min at 94°C and amplified through 30 cycles of 30 s at 94°C; 50 s at 62°C; 1 min 40 s at 72°C and finally extended at 72°C for 10 min. PCRs were carried out using primers with synthetic *Kpn*I restriction sites included. The amplified products of *EtMIC1* (685 bp) and *EtMIC2* (1045 bp) were digested using 5 U of the restriction endonuclease *Kpn*I with 50 µl of 1x React 4 buffer (both from Gibco BRL) and incubated for 1 h at 37°C. The digested products were then electrophoresed through a 1% agarose gel and visualised with ethidium bromide. Bands of the correct size were cut out and gel extracted using the QIAgen gel extraction kit, according to the manufacturer's instructions (see section 2.2.4) and ligated into pUC18 (Amersham Pharmacia Biotech) which had also been digested with *Kpn*I and alkaline phosphatase treated to prevent religation (see section 2.2.3 and 2.2.5).

Complementary primers used to produce the required mutation or deletion were designed with the desired deletion or base substitution in the centre with at least 15 nucleotides on either side and a G or C base at the 3' end. Two PCRs were carried out each containing a single mutant primer (MUTA-1A or MUTA-1B) and a primer complementary to the vector cloning site (pUC18 reverse or –20). The reagents used for the PCR reactions can be seen in Table 6.1. All DNA templates were denatured for 3 min at 94°C and amplified through 25 cycles of 15 s at 94°C; 30 s at 50°C; 2 min at 72°C and finally extended at 72°C for 10 min. The amplified products were

electrophoresed through an ethidium bromide-stained, low melting agarose gel (SeaPlaque GTG; Flowgen) and DNA products of the correct size, were excised from it and extracted using a QIAgen Gel Extraction kit, as outlined in section 2.2.4.

A final PCR to produce the full length, mutated product was carried out using the two amplified fragments obtained from the initial two PCRs as template DNA and the M13 reverse and -20 primers only. The same volumes for PCR were used as described in Table 6.1 and all DNA templates were denatured for 3 min at 94°C and amplified through 30 cycles of 30 s at 94°C; 50 s at 62°C; 1 min 40 s at 72°C and finally extended at 72°C for 10 min. The final EtMIC1 and EtMIC2 mutated products were KpnI digested, electrophoresed through an ethidium-bromide stained, low melting agarose gel (SeaPlaque GTG; Flowgen) excised and purified using a QIAgen Gel Extraction kit. The product was then ligated into KpnI digested and alkaline phosphatase-treated pSV-β-gal vector and transformed into competent XL1-Blue E. coli cells as described in section 2.2.3 to 2.2.6. A schematic diagram of the protocol used to produce mutated products is given in Fig. 6.1.

DNA for each molecular construct was prepared using a QIAgen mini prep kit following the manufacturers instructions. All constructs were digested with KpnI to check for the presence of the required insert and were sequenced to determine if the flanking sequences were in the correct orientation when compared to the lacZ coding region, to conclude if the gene was in the correct frame and to ensure that the desired mutation was present (see section 2.2.7).

Caesium chloride prepared DNA of each construct was resuspended at a final concentration of 100 μg under ethanol (100 μl DNA, 250 μl 95% ethanol and 10 μl 3 M sodium acetate, pH 4.8) and stored at -20°C. When required for transfection into E. tenella sporozoites, the DNA was centrifuged at 13,000g for 5 min, the supernatant removed and the pellet washed in 70% ethanol. Finally the dry DNA pellet was resuspended in 100 μl of complete cytomix and used in transient transfection assays (see section 2.2.11).

PCR number 1		PCR number 2	
Template DNA	50 ng	Template DNA	50 ng
MUTA-1A (2.5 pmol/μl)	2.5 μl	MUTA-1B	2.5 μl
M13 reverse primer	2.5 μl	M13 - 20 primer	2.5 μl
100 mM dNTP set	1 μl	100 mM dNTP set	1 μl
10x *Pfu* polymerase buffer	5 μl	10x *Pfu* polymerase buffer	5 μl
Pfu polymerase (2-3 U/μl)	0.5 μl	*Pfu* polymerase	0.5 μl
H$_2$O to a final volume of	50 μl	H$_2$O to a final volume of	50 μl

Table 6.1 Template DNA was mutated by PCR. The contents of the mutation PCRs are given above.

Fig. 6.1 Schematic representation of the procedure used to produce mutated molecular constructs.
The DNA to be mutated (red blocks, with the area to be mutated shown as a thick black line) was digested out of the pSV-βgal vector using the restriction endonuclease *KpnI* and ligated into *KpnI* digested pUC18 (blue circle or line). Two primers containing the desired mutation or deletion in the middle of them were designed to be complementary to this region, going in both directions. Two PCRs were performed, one containing the –20 primer of pUC18 and MUTA-1B and the other contained the reverse primer and MUTA-1A. The amplified products were gel purified and a second fusion PCR was performed using both products and the primers complementary to pUC18 only (-20 and reverse). The final product was digested with *KpnI* and cloned into a *KpnI* digested pSV-βgal plasmid and used in transient transfection analysis.

6.3 Results

6.3.1 Mutational analysis of the Inr of *EtMIC1*

5'RACE analysis identified a single potential transcription start site for *EtMIC1* within a sequence which conformed to an Inr motif. Deletion analyses of the gene encoding microneme protein EtMIC1 revealed that 234 nucleotides of 5' upstream sequence are required for promoter activity and that the potential Inr alone was not sufficient for promoting expression. To investigate whether the predicted Inr-like sequence is an important regulatory element for *EtMIC1* expression, single point mutations or small clusters of point mutations or deletions were introduced into the Inr sequence within the construct pmic1-βgal-5'613. As differences between the constructs all lie within the Inr motifs only, changes in the number of blue parasites are presumed to reflect relative promoter strength and result from the mutations introduced. A list of the mutations investigated is given in Table 6.1.

To investigate the requirement for an adenosine nucleotide at the A^{+1} position within the predicted Inr motif of *EtMIC1*, this nucleotide was mutated to either a guanosine or cytosine nucleotide (Table 6.1). When the mutated molecular construct, Mic1mut2 ($TCTCG^{+1}CAG$; the mutation nucleotide is shown in bold type), was introduced to the transient transfection system there was modest reduction in the number of blue parasites, to ~85% (the mean value from eight replicas, from two independent experiments) when compared to wild type, pmic1-βgal-5'613 (Fig. 6.2). When Mic1mut3 ($TCTCC^{+1}CAG$) was used there was a larger reduction in the number of blue parasites, down to ~62% when compared to the wild type construct (Fig. 6.2). These decreases demonstrate a preference for an adenosine nucleotide at the transcription start site. Both mutants indicate that an adenosine nucleotide at the transcription start site is preferred and show that another purine (G) is tolerated better than a pyrimidine (C).

Sequence analyses from a variety of efficient promoters containing well conserved Inr motifs identified a $TCA^{+1}NT/A$ consensus as the minimal sequence conserved around a transcription start site (Concino *et al.*, 1984; Du *et al.*, 1993; Liston and Johnson, 1999; Quon *et al.*, 1994). Analyses of many protozoan Inrs (Table 4.1, Chapter Four) revealed the presence of a TCA^{+1} motif and as the Inr of *EtMIC1* contained a TCA^{+1}, it was desirable, to see the effect of its removal. When the mutated construct which

Construct	Sequence 5' – 3'	% of blue parasites compared to WT
WT	ATC GTC TCA^{+1} CAG TAG	100
Mic1mut2	ATC GTC TC**G**$^{+1}$ CAG TAG	85
Mic1mut3	ATC GTC TC**C**$^{+1}$ CAG TAG	62
Mic1mut4	ATC GTC - - -$^{+1}$ CAG TAG	28
Mic1mut5	ATC GTC TC -$^{+1}$ - AG TAG	81
Mic1mut6	ATC GTC **TT**A^{+1} CAG TAG	70
Mic1mut7	ATC GTC **TA**A^{+1} CAG TAG	52

Table 6.1 Analysis of mutated *EtMIC1* promoter.
Alignment of mutational sequences surrounding the initiator region of *EtMIC1*. All constructs contained 613 bp of the 5' upstream sequence of *EtMIC1*, but only the relevant nucleotide sequences between -8 and +6 are shown. Pmic1βgal5'613 is the wild type *lacZ* construct that was used as the positive control and this was also used as the background in which all the mutations were made. Mutants used in the transient transfection assays are outlined (Mic1mut2-7). The point mutations introduced are shown in bold type and the deletions are represented by a single dash for each base removed. Percentage data are representative of at least two independent experiments.
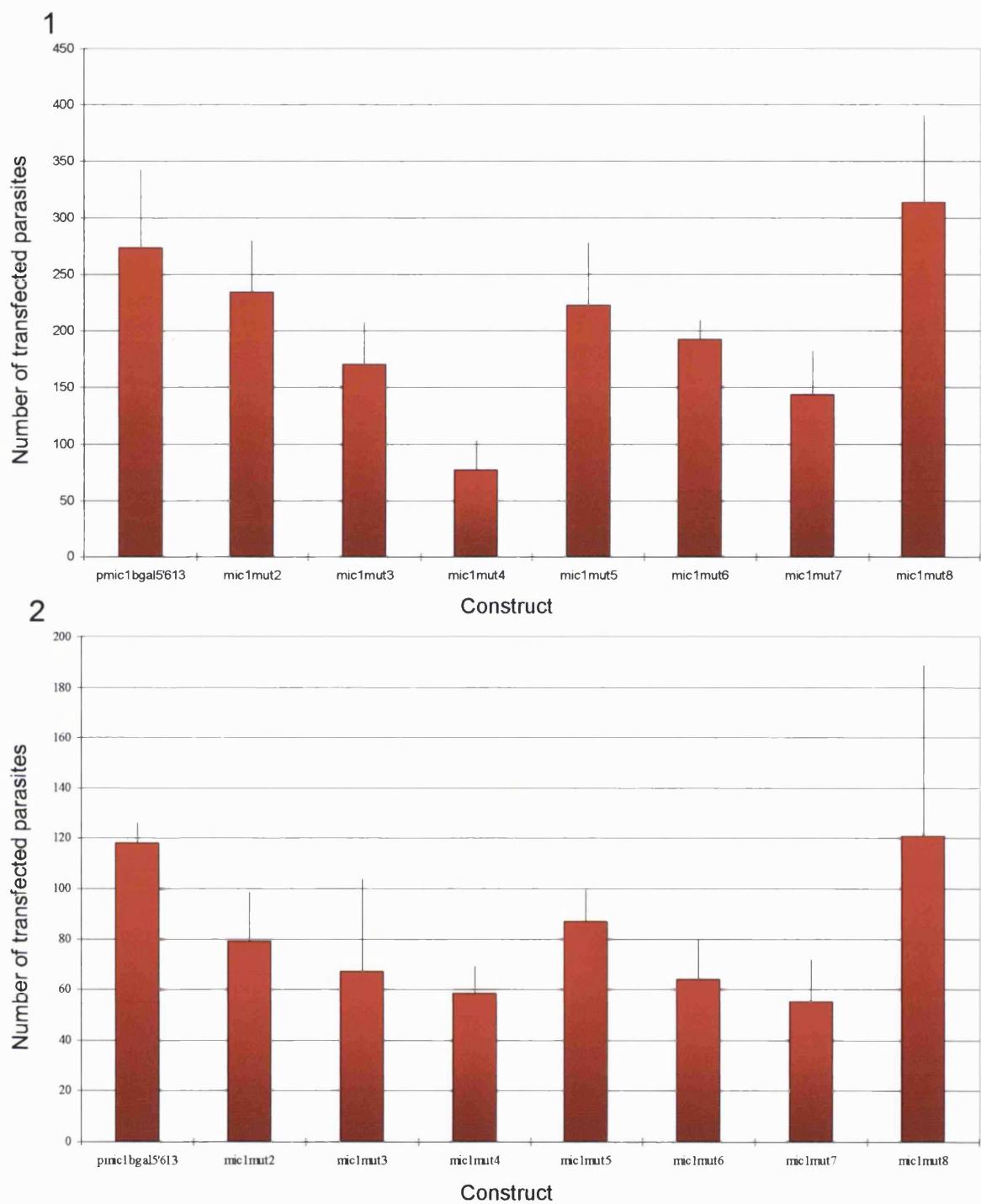
Fig. 6.2 Transient transfection analyses of mutated constructs of *E. tenella EtMIC1* promoter.

Functional analysis of *EtMIC1* 5' flanking regions by transient transfection of sporozoites of *E. tenella*. Various mutated fragments from the *EtMIC1* upstream regions were cloned upstream of a promoterless *lacZ* gene in pSV-βgal. Sporozoites were electroporated with 100 µg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

150

had its TCA^{+1} motif deleted was introduced into the transient transfection system (Mic1mut4; TC---$^{+1}$CAG), the number of blue parasites observed was significantly reduced to ~28% (Fig. 6.2), when compared with the non-mutated construct, suggesting a preference for a TCA^{+1} motif surrounding the transcription start site. The next mutated construct investigated was Mic1mut5, which had a deletion of both A^{+1} and C^{+2} (TCTC-$^{+1}$-AG), resulted in a reduction of only ~19% in the number of blue parasites observed (Fig. 6.2). Removal of both the transcription start site and the following nucleotide would have expected to yield a greater reduction, however further analysis of the new mutated construct revealed the production of a new TCA motif. Due to the low reduction in blue parasite numbers it is highly likely that the transcription stat site of *EtMIC1* moved from position A^{+1} which was deleted to position A^{+3}. The results of this mutation and Mic1mut4 demonstrate the requirement for a TCA^{+1} motif surrounding the transcription start site, and suggest that this motif is significant in gene expression and possibly increased transcription from the other transcription start site at -3.

It has also been observed that the pyrimidine content, which surrounds the transcription start site, is important in determining the transcription start site location and in aiding expression, in particular positions -1 and -2. The cytosine nucleotide at position -1 was converted to a thymine nucleotide and when this mutation (Mic1mut6; TATTA^{+1}CAG) was introduced into the transfection system a modest reduction of ~30% in the number of blue parasites was found (Fig. 6.2). However when the cytosine nucleotide was converted to an adenosine nucleotide (Mic1mut7; TATAA^{+1}CAG) the negative effect on the number of blue parasites was greater, with a reduction to ~48% (Fig.6.2), when compared to the non-mutated construct. These results demonstrate the preference for another pyrimidine nucleotide at the -1 position, however if this nucleotide is mutated to a purine nucleotide (in this case an adenosine nucleotide) expression is reduced to a greater extent.

6.3.1 Mutational analysis of the Inr of *EtMIC2*

The transcription start sites for *EtMIC2* were preferentially mapped to two adenosine nucleotides positioned 94 (has been designated +1) and 97 bp upstream of the ATG start codon, and further analyses of the sequence surrounding A^{+1} revealed an Inr-like motif. Deletion analyses of the 5' flanking region of *EtMIC2* demonstrated that the minimum sequence necessary for promoter activity was 131 bp of the upstream

sequence and the Inr is sufficient for basal level expression. Therefore to investigate the importance of the sequences surrounding both potential transcription start sites as important *cis*-acting elements, single point mutations or small clusters of point mutations or deletions, similar to those introduced to the Inr of *EtMIC1* were made to the transcription start sites of construct pmic2-βgal-5'1046. A list of the mutations investigated is given in Table 6.2.

The sequence surrounding A^{-3} did not conform to an Inr consensus as it only contained two pyrimidines. However A^{+1} had an excellent fit to the Inr consensus, as the motif surrounding the potential transcription start site contained exclusively pyrimidines. To investigate the importance of A^{-3} as a transcription start site, constructs which contained a deletion of the adenosine and the preceding guanine (G^{-4}) were generated. When this mutated construct was introduced into the transfection system (Mic2mut1; $GAA\text{--}^{-3}TTA^{+1}$) the number of blue parasites observed was reduced by ~66%, when compared to the wild type (Fig 6.3), demonstrating that both the adenosine at the preceding guanine nucleotides are significantly important in promoter activity. Removal of both purines at positions A^{-5} and A^{-6} (Mic2mut2; $G\text{--}GA^{-3}TTA^{+1}C$) also reduced the number of blue parasites but to a lesser extent (down to ~56%) when compared to pmic2-βgal-5'1046. Deletion of the second potential transcription start site, A^{-3}, (Mic2mut5; $GAAG\text{-}^{-3}TTA^{+1}C$) also reduced the number of blue parasites observed, to ~50%, thus demonstrating its importance as another transcription start site. Another mutant construct (Mic2mut6; $GAAAA^{-3}ATA^{+1}C$), which had both nucleotides at positions at T^{-2} and G^{-4} mutated to adenosine nucleotides also resulted in a reduction in the number of blue parasites observed (down to ~58%) compared to the wild type (Fig. 6.3), demonstrating that these nucleotides are important in aiding high promoter activity.

Collectively these results suggest that the nucleotide sequence between T^{-1} and A^{-6} influences transcriptional activity, but in the absence of a 100% reduction cannot be regarded as essential which is most likely due to the presence of a another transcription start site within the Inr motif. Together with the 5'RACE data these studies further suggest that the A^{-3} is a second transcription start site but it should be noted also that it does not conform to an Inr consensus and does not contain a TCA motif.

Construct	Sequence (5' - 3')	% of blue parasites compared to WT
WT	GAA GA^{-3}T <u>TA^{+1}C TCC</u> TTG CTT	100
Mic2mut1	GAA - - $^{-3}$T <u>TA^{+1}C TCC</u> TTG CTT	34
Mic2mut2	G - - GA^{-3}T <u>TA^{+1}C TCC</u> TTG CTT	56
Mic2mut5	GAA G -$^{-3}$T <u>TA^{+1}C TCC</u> TTG CTT	50
Mic2mut6	GAA AA^{-3}A <u>TA^{+1}C TCC</u> TTG CTT	58
Mic2mut8	GAA GA^{-3}T <u>T -$^{+1}$- - - - - -$^{+7}$</u> G CTT	26
WT	TAG TG<u>C^{-98} AAT^{-95}</u> TGT GCT AGC	100
Mic2mut9	TAG TG -$^{-98}$ - - -$^{-95}$ TGT GCT AGC	35

Table 6.2 Analysis of mutated *EtMIC2* promoter.
Alignment of mutational sequence surrounding the potential transcription start sites of *EtMIC2*. All constructs contained 1046 bp of the 5' upstream sequence of *EtMIC2*, but only relevant nucleotide sequence, between G^{-7} and G^{+11} (Inr) or between T^{-103} and C^{-86} (CAAT) are shown. Pmic2βgal5'1046 was used as a positive control and is the background in which all of the above mutant constructs were made. The potential transcription start sites at A^{+1} and A^{-3} are indicated and the Inr motif with the sequence TTA^{+1}CTCC, 5'–3', is underlined. For ease of description the transcription start site nearest the ATG start codon is regarded as +1. Below this are all the mutants used in the transient transfection assays. The point mutations introduced into the mutant constructs, Mic2mut1, 2, 5, 6, 8 and 9, are shown with the mutations shown in bold type the deletions are represented by a single dash for each base that has been removed. Percentage data are representative of at least two independent experiments, with a minimum of eight replicas representing each result.

Fig. 6.3 Transient transfection analyses of mutated constructs of *E. tenella* *EtMIC2* promoter.

Functional analysis of *EtMIC2* 5' flanking regions by transient transfection of sporozoites of *E. tenella*. Various fragments from the *EtMIC2* upstream regions were cloned upstream of a promoterless *lacZ* gene in pSV-βgal. Sporozoites were electroporated with 100 µg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

154

Construct Mic2mut8 (GAAGA^{-3}TT-$^{+1}$------$^{+7}$G) which contained a 7 bp deletion of part of the Inr-like motif from position A^{+1} to T^{+7} was used to investigate the importance of the potential Inr. When this mutated molecular construct was introduced into the transient transfection system a dramatic and significant reduction in the number of blue parasites was observed (down to ~26%), suggesting that the transcription start site within the Inr was the major transcriptional start site (A^{+1}) while the transcription start site A^{-3} plays a minor role.

Initial deletion analysis of the upstream region of *EtMIC2* demonstrated that loss of 114 bp between -151 and -37 resulted in a slight decrease in promoter activity. This region includes a consensus CCAAT box located at position -98. To investigate if the CCAAT box is an important *cis*-acting element of this region, it was deleted to produce mutant Mic2mut9 (TG-$^{-98}$---$^{-95}$TGT). When this mutated molecular construct was used in the transient transfection system, a large reduction in the number of blue parasites was observed (down to ~35%) when compared to the non-mutated construct, pmic2-βgal-5'1046 (Fig. 6.3). This result demonstrates that the CCAAT does make a significant contribution to promoter activity.

6.3.3 Mutational analysis of potential transcription factor binding sites

Deletion analysis of the 5' flanking region of *EtMIC1* identified a 33 bp region, between -383 and -350, which was significant in aiding the expression of the reporter product β-galactosidase. As discussed in Chapter Five sequence analysis of the region did not reveal any common *cis*-acting regulatory elements found in other apicomplexan genes, however TRANSFAC analysis (Heinemeyer *et al.*, 1998) identified potential transcription factor binding sites. To investigate if any of the potential transcription factor binding sites were necessary for efficient gene expression, constructs which contained deletions of the core recognition site for each transcription factor were made, using the same method as outlined in section 6.2 and their effects were revealed by β-galactosidase activity in transient transfection assays.

One of the potential transcription factor binding sites identified was for a DOF binding protein which recognise 5'-AAAG-3' motifs. When the mutated molecular construct Mic1-D, which contained a deletion of the sequence, was introduced to the sporozoites transient transfection system there was a reduction in the number of blue parasites observed (~50%), when compared to the wild type construct pmic1-β-gal-5'613 (Fig. 6.4).

Construct	Sequence (5' - 3')	% of blue parasites compared to WT
WT	TAAACCATGCTTGGAGGAAACTTTGCTTTT	100
Mic1-C	TAAACCATGCTTGGA - - - -ACTTTGCTTTT	100
Mic1-N	TAAACCATGCTTGGAG - - - - CTTTGCTTTT	100
Mic1-D	TAAACCATGCTTGGAGGAAA - - - - GCTTTT	50
Mic1-S	TAAACCATGCTTGGA - - - - ACTTTGCTTTT	55
Mic1-H	TAAACCATGC - - - - AGGAAACTTTGCTTTT	100
Mic1-E	T - - - - CATGCTTGGAGGAAACTTTGCTTTT	100

Table 6.3 Analysis of mutated *EtMIC1* promoter (series 2).
Alignment of mutational sequences surrounding various potential protein binding regions of *EtMIC1*. All constructs contained 613 bp of the *EtMIC1* promoter, but only relevant nucleotides, between -383 and -350, are shown. Pmic1βgal5'613 is the wild type *lacZ* construct which was used as the positive control and the background in which all the mutations were made. Below this are all the mutants used in the transient transfection assays. The deletions introduced are represented by a single blue dash for each base pair removed. Percentage data are representative of at least two independent experiments.
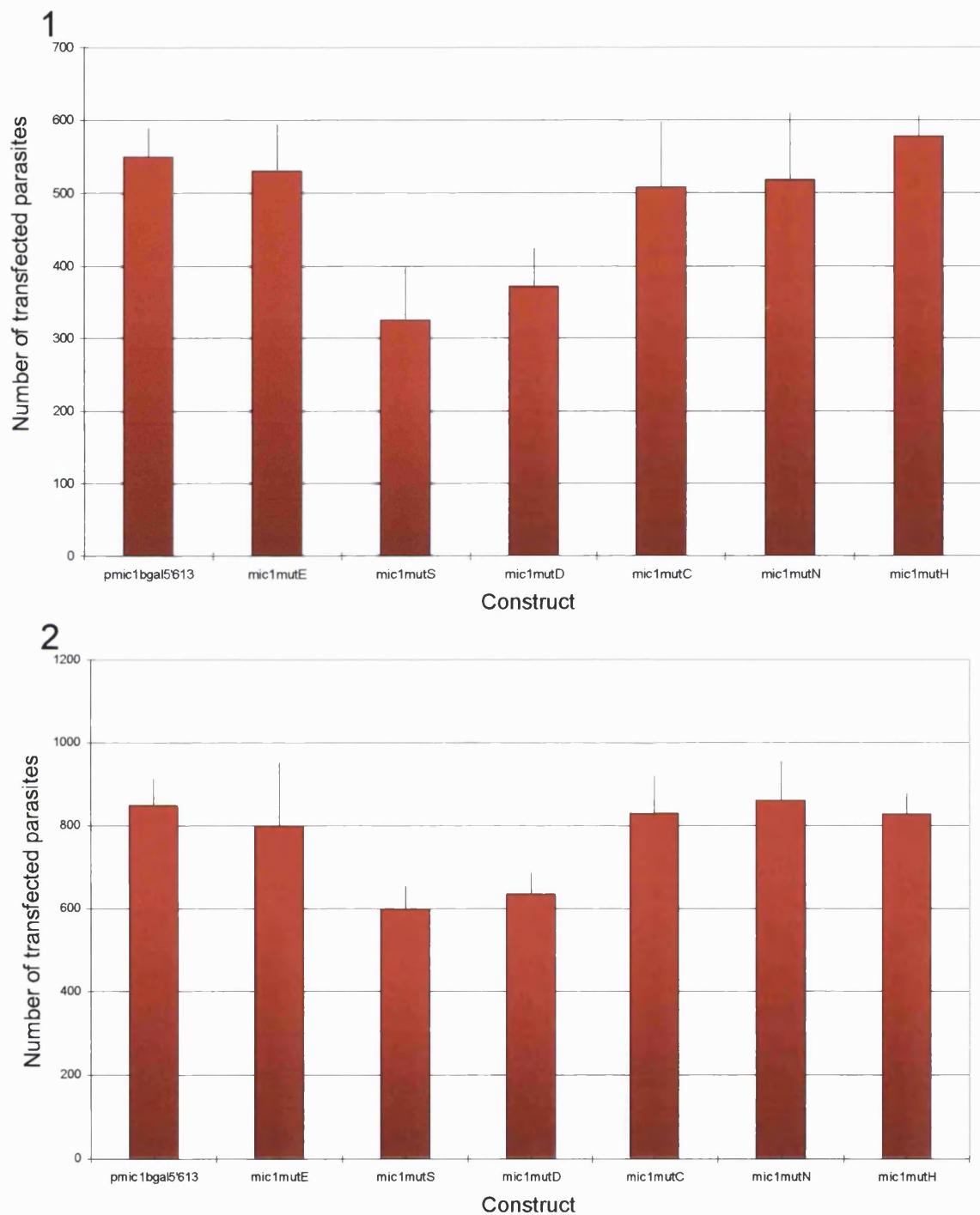
Fig. 6.4 Transient transfection analyses of mutated constructs of *E. tenella* *EtMIC1* promoter (series 2).

Functional analysis of *EtMIC1* 5' flanking regions by transient transfection of sporozoites of *E. tenella*. Various fragments from the *EtMIC1* upstream regions were cloned upstream of a promoterless *lacZ* gene in pSV-βgal. Sporozoites were electroporated with 100 μg of plasmid, allowed to invade MDBK cells and assayed for reporter activity 24 h post transfection. Each graph shown represents a single experiment and each data point plotted represents the mean of a minimum of eight replicas. The error bars represent the standard deviation from the mean values.

Another potential transcription factor binding site identified was for STAT binding proteins which recognise a 5'-TTCC-3' motif. When the mutated molecular construct Mic1-S, which contained a deletion of the sequence, was introduced into the transient transfection system there was a reduction in the number of blue parasites observed (reduced to ~55%), when compared to the non-mutated construct pmic1-β-gal-5'613 (Fig. 6.4). Another potential transcription factor identified was CED, which contained an identical binding site as the STAT binding site. However its mutated molecular construct when used in transient transfection did not reduce the number of blue parasites observed, in contrast to the STAT mutated construct. As all the mutated constructs were sequenced to confirm the mutations it seems unlikely that the either the STAT or the CED mutated construct are incorrect and thus responsible for the differences observed between the two constructs. The other possibility is that the transfection data is incorrect, however the fact that two separate transfections were carried out, each with eight replicas shown similar data. It is possible that either the STAT or the CED mutated DNA was somehow contaminated and this could be responsible for the differences in blue parasites observed. Other potential transcription factor binding sites identified were for NAF, HAP and ELF binding proteins which recognise a 5'-GAAA-3', 5'-CCAA-3' and 5'-GTTT-3' motifs respectively (Table 6.4). All deletion constructs failed to reduce the number of blue parasites observed when compared to the wild type construct (Fig. 6.4).

6.4 Discussion

Examination of the 5' untranscribed regions of *EtMIC1* and *2* has revealed the presence of highly conserved Inr elements surrounding some of the transcription start sites. To investigate which nucleotides surrounding the start sites and within the potential Inr elements are most important for efficient transcription initiation, we introduced single point mutations, small clusters of point mutations or deletions into the sequences surrounding the start sites of constructs pmic1-βgal-5'613 or pmic2-βgal-5'1046. The mutated 5' untranscribed regions of either *EtMIC1* or *2* were fused to the *lacZ* gene of pSV-βgal and these constructs were then used to transiently transfect *E. tenella* sporozoites and the resulting β-galactosidase (as visualised by staining with X-gal) activities were measured. We demonstrated that this motif is a promoter element which is both structurally and functionally similar to metazoan Inrs (Nakaar *et al.,* 1992; Soldati and Boothroyd, 1995; Sheils *et al.,* 2000; Elmendorf *et*

al., 2001). Specific conserved nucleotides comprising the core of this promoter element are shown to be necessary for accurate selection of the start site of transcription and efficient expression indicating it functions as a *bona fide* Inr.

Mutational analyses of the Inr in *EtMIC1* demonstrated a preference for an adenosine nucleotide at the transcription start site within the motif and indicated that another purine (Mic1mut2; TCTCG^{+1}CAG) is tolerated better than a pyrimidine (Mic1mut3; TCTCC^{+1}CAG). Previous mutational analyses demonstrated that an Inr can enhance promoter strength even if it shifted slightly upstream or downstream from the start site (O'Shea-Greenfield and Smale, 1992; Zenzie-Gregory *et al.* 1992), thus the transcription start site can be located five nucleotides upstream or downstream of the original transcription start site. Also previous mutational analysis of the weak Inr consensus sequence in a variety of mammalian protein-encoding genes demonstrated that mutations in the region resulted in use of an alternative transcription start site and/or a reduction in promoter strength (Corder *et al.*, 1980; Talkington and Leder, 1982; Dierks *et al.*, 1983; Concino *et al.*, 1984). Therefore in these mutation constructs the transcription start site may have moved to position A^{+3}, two nucleotides downstream of the original transcription start site. Had this nucleotide not been present the mutation might have been more effective as it has been demonstrated in other Inr elements that mutating an adenosine nucleotide at position +1 to a guanine, cytosine or thymine nucleotide, resulted in activity being almost abolished (Liston and Johnson, 1999; Smale, 1994).

Sequence analyses from a variety of efficient promoters containing well conserved Inr elements identified a TCA^{+1}NT/A motif as the minimal sequence conserved around a transcription start site (Concino *et al.*, 1984; Du *et al.*, 1993; Liston and Johnson, 1999; Quon *et al.*, 1994). Analyses of many protozoan Inrs (Table 4.1, Chapter Four) revealed the presence of a TCA^{+1} motif and it was desirable to see the effect of its removal. When the TCA^{+1} motif of the *EtMIC1* Inr was deleted (Mic1mut4; TC---$^{+1}$CAG), there was a large reduction in promoter activity, demonstrating a preference for a TCA^{+1} motif surrounding the transcription start site. The deletion of both the A^{+1} and the C^{+2} in the *EtMIC1* Inr motif (Mic1mut5; TCTC-$^{+1}$-AG) resulted in only a small reduction in the number of blue parasites observed. The deletion of both the transcription start site and the following nucleotide would have expected to result in a greater reduction, however this deletion produced a new TCA motif and the

transcription start site most likely moved to A^{+3}. These results demonstrate the requirement for a TCA^{+1} motif surrounding the transcription start site and indicate its significant importance in gene expression. These mutations again re-iterate the possibility of the transcription start site moving slightly upstream or downstream of the original start site. To confirm this is the case, primer extension or 5'RACE analyses could be performed on the mutant constructs to establish the possibly new transcription start site.

It was understood from previous experiments (O'Shea-Greenfeild and Smale, 1992) that the pyrimidine content at position -1 is of importance in the activity of the Inr. The requirement for pyrimidines at the -1 position was investigated in this study through mutation the C^{-1} in the Inr element of *EtMIC1* to either a thymine (Mic1mut6; $TATTA^{+1}CAG$) or adenosine (Mic1mut7; $TATAA^{+1}CAG$) nucleotide. Both mutations caused a reduction in gene expression suggesting that there is a preference for a cytosine nucleotide at position −1, but also demonstrated that a pyrimidine is preferred at position -1, as if this nucleotide was mutated to a purine (in this case an adenosine residue) expression was reduced to a greater extent. These results again demonstrate that importance of a TCA motif surrounding the transcription start site.

Analysis of multiple Inr-like elements of *G. lamblia* identified a string of consecutive adenosine nucleotides in the Inr motif, either before or after the transcription start site, as important in defining the start site location and efficiency of gene expression (Elmendorf *et al.*, 2001). Mutation of the sequence, before the Inr element of *EtMIC2*, to a string of adenosine nucleotides (Mic2mut6; $GAAAA^{-3}ATA^{+1}C$), resulted in decreased promoter activity. Therefore a string of adenosine nucleotides do not increase expression in eimerian Inrs. It may be possible that the string of adenosine nucleotides in *Giardian* Inrs are required for the binding of transcription factors to a transcription initiation complex, however this would indicate that both apicomplexan parasites transcribe their Inr containing genes differently.

Initial deletion analysis of the upstream region of *EtMIC2* demonstrated that loss of 114 bp, which included a consensus CCAAT box, between -151 and -37 upstream of the ATG, resulted in a decrease in promoter activity. Deletion of the CCAAT box (Mic2mut9; $TG-^{-98}---^{-95}TGT$) resulted in a large reduction in the number of blue parasites observed, demonstrating that the CCAAT box does make a significant contribution to promoter activity. It has been demonstrated in previous experiments

that the nature of the sequence at -25 appears to influence both the relative activity of an Inr and the degree to which the Inr influences promoter strength and start site placement (Wiley et al., 1992; Zenzie-Gregory et al., 1993). Many strong promoters appear to function regardless of the upstream sequence (Zenzie-Gregory et al., 1993), however, weaker Inrs may have little influence on core promoter activity and on start site placement when the promoter contains a strong consensus TATA box (Lo and Smale, 1996). The sequence 25 bp upstream of the Inr of EtMIC2 is only about 50% AT rich and at must therefore be regarded as a strong Inr. In the region between the Inr and the CCAAT box there are some AT rich areas which could serve as degenerate TATA boxes. This could mean that EtMIC2 could be a TATA-containing promoter and thus explain why the CCAAT box influences promoter activity.

Possibly the specific sequences found in these elements possess a reduced affinity for the putative Inr-binding protein. There is no evidence to date of proteins that interact with protozoan Inrs but the identification of sequence-specific, Inr-binding proteins is difficult. Purification of Eimeria nuclear protein that specifically recognises a functional but not a non-functional Inr should advance our understanding of transcription in protozoan parasites, since this protein is likely to be a homologue of protozoan Inr-binding proteins. If a transcription factor was found to interact with the CCAAT box of EtMIC2, it would confirm its importance in expression of EtMIC2.

Dof proteins belong to a family of transcription factors which to date have only been observed in plants (Yanagisawa, 1995; Yanagisawa and Schmidt, 1999). It was demonstrated by binding site selection and EMSAs that four different maize Dof proteins recognise a core motif (A/T)AAAG, however all have different preferences for flanking sequences. Further analysis of the Dof-DNA interaction in the context of native promoters in vivo is required to elucidate the exact means through which sequence recognition is achieved. It has been demonstrated that StDof1 (stomata Dof) binds to a novel cis-acting element implicating involvement of Dof proteins in cell-specific gene expression. Block mutagenesis of the TAAAG motif in the KST1 promoter (which encodes for a K^+ influx channel of guard cells in potatoes) dramatically reduced gene expression (Plesch et al., 2001). This is the first example of a potential Dof binding site within the 5' flanking region of any gene found outside the plant kingdom.

Signal transducers and activators of transcription (Stat) proteins were initially identified in mammals and to date seven have been observed (Stat1-6, including two Stat5 genes). Stat homologs have been defined in simpler eukaryotes including *Dictyostelium, Caenorhabditis elegans* and *Drosophila* however none have been observed in yeast (Kisseleva *et al.*, 2002; Miyoshi *et al.*, 2001). To date such transcription factors have not been identified in apicomplexa. The proteins recognise the core motif 5'-TTCC-3'. Numerous studies have described a strong interaction between Stats and other transcription factors. For example Stat1 was shown to interact with NF-κB, Sp1 and USF-1 (Look *et al.*, 1995; Muhlethaler-Mottet *et al.*, 1998; Ohmori *et al.*, 1997). In most cases the Stat proteins are most closely related to mammalian Stat3 and 5 and appear to regulate developmental processes (Kisseleva *et al.*, 2002). Again this is the first example of a potential Stat transcription factor binding site within apicomplexa. In both cases further experiments are required to confirm that these regions do contribute to promoter activity and in particular the core motifs identified. DNase footprinting analyses could be used to identify DNA-binding sites for proteins by their protection against nicking by DNase. Such studies may identify an area of protection surrounding regions identified as aiding gene expression through 5' deletion and mutation assays. Electrophoretic mobility shift assays could be used to identify protein(s), which can specifically recognise such areas of each of the microneme encoding genes.

Chapter Seven
Final Discussion

7.1 Summary of results

Host cell invasion is tightly associated with the secretion of three classes of protein from the apical tip of obligate intracellular parasites of the phylum Apicomplexa: rhoptries, micronemes and dense granules. Microneme proteins have been shown to contain important ligands for binding (Adams *et al.*, 1990; Sim *et al.*, 1992; Muller *et al.*, 1993; Robson *et al.*, 1995) and to be required for invasion of host cells (Barnwell and Galinski, 1989; Sim *et al.*, 1990; Sharma *et al.*, 1996) and are therefore essential for the establishment of intracellular parasitism. To date very little is known about microneme organelle biogenesis and the regulation of microneme gene expression.

This series of studies began with an assessment of oocyst sporulation in *E. tenella* as a means to examine the expression of genes encoding microneme proteins (Chapter Three). Dramatic morphological changes occur within the oocyst during the sporulation process and culminate in the production of sporozoites, the first invasive life cycle stage. Examination of sporulating oocysts during the time when sporocysts and sporozoites are formed revealed that sporulation is highly synchronous and around 90% of the oocysts proceeded to full sporulation by 22.5 h (the remaining 10% were most likely unfertilised macrogametes) (Chapter Three). The time-course of sporulation has proved to be a convenient, synchronous system and useful in the analysis of the relative abundance of specific products of gene expression.

The system revealed a high degree of co-ordination in the timing of expression of microneme proteins and all five microneme proteins examined (EtMIC1-5) were present from 22.5 h onwards, when the sporozoites were fully formed. This situation is similar to the timing of the appearance of EtMIC2 and EtMIC5 during first generation schizogony, when the proteins are detected only from the time at which daughter merozoites are forming (Tomley *et al.*, 1996; Brown, 2000). Examination of this set of data suggests that microneme proteins are made around the time when microneme organelle biogenesis occurs. Moreover the fact that all five microneme proteins appeared with such a high level of synchronicity, suggests that their expression is highly regulated (Chapter Three). The sporulating oocyst system also demonstrated that microneme specific mRNAs were first detected 6 or 12 h into sporulation,

indicating a regulation of expression between the unsporulated and the sporulating oocyst. It is not known whether this temporal regulation is due to differences in transcription between the life cycle stages or due to post-transcriptional effects such as differential mRNA turnover or mRNA stability. Each microneme protein was not detected until 10-12 h after their specific mRNAs were first observed, indicating that post-transcriptional factors are important in the regulation of microneme protein expression. Whether this is due to post-transcriptional effects during sporulation also remains to be elucidated. One of the major advantages this system has over other apicomplexan parasites is that with *Eimeria* it is possible to access large numbers of unsporulated diploid oocysts that develop through to haploid invasive sporozoites with a good level of synchrony. The system has since been used to investigate the expression of an aspartyl proteinase (Jean *et al.*, 2001) and a cyclin dependent kinase (Jane Kinnard and Brian Shiels, University of Glasgow, unpublished) of *E. tenella*. As rhoptry proteins are expected to be trafficked in a similar way to each other and are secreted upon invasion (somewhat later than microneme proteins) it would be expected that they might appear in a co-ordinated manner through oocyst development. Currently, two-dimensional electrophoresis of purified rhoptry organelles is being carried out to identify rhoptry proteins and once anti-sera have been raised against these, the sporulation system could also be used analyse the expression of rhoptry proteins. This system of sporulation can be used to study various aspects of parasite differentiation. SDS-PAGE analyses of total protein from different time points during sporulation of *E. tenella* identified numerous changes in protein profiles and many of these changes corresponded to major events such as sporoblast development, sporozoite formation and meiosis and mitosis.

It was not possible to further investigate the transcription kinetics of the microneme genes, as metabolic labelling of *E. tenella* oocysts is not possible and no *in vitro* transcription system is available. As transcriptional control was likely to be involved for the co-ordinated level of mRNA expression, an analysis of the flanking sequences of some of the microneme genes was undertaken.

Very little is known about protozoan parasite gene expression compared to other eukaryotes. Analyses of eukaryotic promoters and their interacting proteins have been mostly confined to higher eukaryotes. An initial step in defining transcriptional control was the identification of *cis*-regulatory elements, to gain an insight into the regulation of gene expression in microneme encoding genes of apicomplexa. Chapter Four

details the analysis of the 5' flanking genomic sequence and the transcription start site mapping of microneme-coding genes. Sequence analysis identified possible *cis*-acting elements, including Inrs, which have been observed in other apicomplexan parasites and shown to be important in gene expression (Table 4.1, Chapter Four).

The potential, major transcription start sites for *EtMIC1, 2* and *3* (Chapter Four) were mapped to adenosine nucleotides, within Inr consensus sequence. Further analysis of the upstream region of *EtMIC1* revealed a tripartite architecture (a degenerate TATA box, an Inr and a GAAC motif) similar to that observed in *E. histolytica*, however the elements in the promoter region of EtMIC1 were distributed over a much larger area. In *E. histolytica* the GAAC motif has been shown to control the rate of gene expression and is able to direct a new transcription start site, independent of both the TATA box and the Inr (Singh and Rogers, 1998; Singh *et al.*, 1997) and can mediate transcriptional activation of some upstream regulatory elements (Singh *et al.*, 2002). Mutations within the GAAC element and the TATA box in the promoter region of various genes in *E. histolytica* resulted in decreased reporter gene expression (Prurdy *et al.*, 1996; Singh *et al.*, 1997) indicating both play an important role in gene expression. Deletion assays demonstrated that loss of the degenerate TATA in *EtMIC1* reduced gene expression and loss of the GAAC motif completely abolished expression, indicating that the potential Inr alone was insufficient (Chapter Five). In *E. histolytica* the TATA box and GAAC elements are necessary for nuclear protein assembly, however it is not known if these high-order complexes represent sequential protein assembly of a single protein (Singh *et al.*, 2002). Truncation assays of the *E. histolytica* hgl5 implicate an involvement of the above motifs in gene expression, however they are some distance upstream of the Inr motif of *EtMIC1*. Due to the large area it seems unlikely that they function in a similar way as described in *Entamoeba* genes. The Inr of *EtMIC1* alone was insufficient for basal expression.

Polypurine motifs have been observed in the promoter regions of various protozoan genes (Barahona *et al.*, 1988; Brunk and Sadler, 1990; Nakaar *et al.*, 1992; Lanzer *et al.*, 1993a; Lanzer *et al.*, 1993b; Soldati and Boothroyd, 1995; Mercier *et al.*, 1996) and have been shown to be critical for efficient expression in some cases. The presence or absence of similar polypurine sequence (AGAGACC) in the upstream region of *EtMIC1* did not affect expression, unlike its function *SAG1, TUB1* or *GRA1, 2, 5* and *6*. These motifs have been postulated to be similar to SP1 motifs of higher eukaryotes. SP1 motif are usually found numerous times in the promoter region of a

gene and their loss results in loss of expression (Kadonaga *et al.*, 1986) however, it seems unlikely that they serve a similar function in *EtMIC1*, as there is a single repeat and its presence or absence did not alter expression (Chapter Five).

The potential transcription start site of *EtMIC1* was mapped to an adenosine nucleotide within the sequence $TCA^{+1}CAGT$ (where A^{+1} is the predicted transcription start site) which is similar to an Inr motif (Javahery *et al.*, 1994; Smale and Baltimore, 1989). Inrs have been found in an increasingly large number of eukaryotic and protozoan genes (Elmendorf *et al.*, 2001; Nakaar *et al.*, 1992; Shiels *et al.*, 2000; Singh *et al.*, 1997; Soldati and Boothroyd, 1995). Deletion analyses demonstrated that the Inr alone was not sufficient for basal expression but it was not known how important it was for efficient expression. Mutational analyses demonstrated a preference for an adenosine nucleotide at the +1 position of the *EtMIC1* Inr elements, since replacement or deletion of this nucleotide reduced expression and indicated that another purine is tolerated better than a pyrimidine. Sequence analyses from a variety of efficient promoters containing well conserved Inrs have identified a $TCA^{+1}NT/A$ motif as the minimal sequence conserved around a transcription start site (Concino *et al.*, 1984; Du *et al.*, 1988; Quon *et al.*, 1994; Liston and Johnson, 1999) and many protozoan Inrs (Table 4.1, Chapter Four) contain a TCA^{+1} motif. Deletion and mutational analyses of the TCA^{+1} motif in *EtMIC1* resulted in a significant reduction in expression, indicating its importance in the Inr motif. Also there is a preference for a cytosine residue at position -1. However if this cytosine residue is mutated to a purine residue (in this case an adenosine residue) expression is reduced to a greater extent.

Deletion analysis demonstrated that the Inr of *EtMIC2* (with the addition of 37 bp of untranscribed sequence) was sufficient for basal level expression. Loss of the degenerate TATA box and GAAC motif did not alter expression. The major and minor transcription start sites of *EtMIC2* were mapped to adenosine nucleotides within the sequence $A^{-3}TTA^{+1}CTCC$, the second of which is within a sequence that has an excellent match with the predicted Inr consensus (Javahery *et al.*, 1994; Smale and Baltimore, 1989). Deletion analyses of the 5' flanking region of *EtMIC2* demonstrated that the minimum sequence necessary for promoter activity was 131 bp and the Inr alone is sufficient for basal expression. Mutational analyses demonstrated that the nucleotide sequence between T^{-1} and A^{+5} influences transcriptional activity, but in the absence of a 100% reduction cannot be regarded as essential and this is most likely

due to the presence of a second transcription start site. Together with the 5'RACE data these studies further suggest that the A^{-3} is a second transcription start site but it should be noted also that it does not conform to an Inr consensus and does not contain a TCA motif. Deletion of most of the Inr motif surrounding the transcription start site at A^{+1} (i.e. deletion from position A^{+1} to T^{+7}) resulted in a dramatic and significant reduction in the number of blue parasites was observed suggesting that the second transcription start site within the Inr was the major start site (A^{+1}) while the transcription start site A^{-3} is the minor. Deletion and mutational analyses of the upstream region of EtMIC2 demonstrated that loss of a consensus CAAT box located 195 bp upstream of the ATG start codon resulted in a decrease in promoter activity. Mutational analysis of the CAAT box demonstrated that it makes a significant contribution to promoter activity indicating it is an important cis-acting element.

Molecular constructs containing the upstream sequences of either EtMIC3 or EtMIC4 failed to drive expression of the lacZ gene in sporozoites, possibly due to sensitivity of the system or lack of expression at this life cycle stage. Constructs of both EtMIC3 and EtMIC4 also failed to drive expression in the schizont but those of EtMIC1 and EtMIC2 similarly failed to promote expression in schizonts. However if development were allowed to continue to the production and release of merozoites, expression was observed in this later life cycle stage, suggesting that expression of EtMIC1 and EtMIC2 is stage specific.

These results correlate with ultrastructural data where it is clear that micronemes are formed afresh during each successive stage of the life cycle, e.g. in first generation schizogony the micronemes of the invading sporozoites gradually disappear (Chobotar et al., 1975) and reappear late in schizogony when daughter merozoites separate from the residuum (Dubremetz, 1975, Dubremetz, 1979). Further, microneme proteins EtMIC2 and EtMIC5 gradually disappear during early schizogony and are detected later in mature merozoites, suggesting that microneme protein expression occurs only when micronemes are being assembled in readiness for the next round of host cell invasion (Tomley, 1996; Brown, 2000b). All of these results agree with the detection of β-galactosidase activity during sporozoite and merozoite maturation. Such down regulation in gene expression during the conversion from one life cycle stage to another has been observed in SAG1 and SAG2 of T. gondii which are expressed in the tachyzoite stage but down regulated during its conversion to the bradyzoite stage of

the parasite. Also the high levels of gene expression from the *NTP3* promoter is reduced when tachyzoites differentiate into the bradyzoites (Nakaar *et al.*, 1992). Transcription factor binding site mutation indicates that transcription factors Dof and STAT may be important in aiding gene expression of *EtMIC1*. Dof proteins belong to a family of transcription factors which recognise a core motif (A/T)AAAG and have only been observed in plants to date (Yanagisawa, 1995; Yanagisawa and Schmidt, 1999). Further analysis of the Dof-DNA interaction in the context of native promoters *in vivo* is required to elucidate the exact means through which sequence recognition is achieved. This is the first example of a potential Dof binding site within the 5' flanking region of any gene found outside the plant kingdom. Stat proteins recognise the core motif 5'-TTCC-3' which has been defined in simpler eukaryotes including *Dictyostelium, Caenorhabditis elegans* and *Drosophila* however none have been observed in yeast (Kisseleva *et al.*, 2002; Miyoshi *et al.*, 2001). To date such transcription factors have not been identified in apicomplexan parasites. Again this is the first example of a potential Stat transcription factor within apicomplexa. Further experiments are required to confirm that these regions contribute to promoter activity and in particular the core motifs identified.

7.2 Future Work

To date a DNA binding protein(s) implicated in Inr function remains to be identified. Previous studies have demonstrated that the binding of both human and Drosophila TFIID to Inr elements relies on the nucleotides within the Inr consensus (Kaufman and Smale, 1994; Purnell *et al.*, 1994). Thus TFIID complex is an attractive candidate as the protein that functionally recognises an Inr consensus. However since the sequence requirements have not yet been established for the binding of two other proteins implicated in basal Inr function, RNA pol II (Carcamo *et al.*, 1991) and TFII-I (Roy *et al.*, 1991) it remains possible that one or both of these proteins may also contribute to the activity of consensus Inr elements. Purification of *Eimeria* nuclear protein that specifically recognises a functional but not a non-functional Inr should advance our understanding of transcription in protozoan parasites, since this protein is likely to be a homologue of other protozoan Inr-binding proteins.

To determine whether the *Eimeria* Inr-like element is a core promoter element, it would be of interest to establish whether these motifs are interchangeable between different *E. tenella* genes, other apicomplexan genes and mammalian Inrs. A further

aspect of the work would be to determine whether this element in *E. tenella* can be replaced *in vivo* by a heterologous Inr such as the Inr of *SAG1* of *T. gondii*, or *vice versa*, and examine if the two elements perform homologous functions.

From the structural analysis of the *E. tenella* microneme gene Inrs, it has been established that mutation of specific nucleotides within the motif can severely inhibit reporter gene expression. If the microneme Inrs play a role in transcription start site selection, specific mutations would be predicted to have a direct effect on its transcriptional performance. This can be demonstrated by analysing transfectants expressing the different mutations. Transfectants can be selected, poly(A)$^+$ RNA extracted and the transcription start site can be determined using primer extension or 5' RACE.

DNase footprinting analyses could be used to identify DNA-binding sites for proteins by their protection against nicking by DNase. Such studies may identify a region of protection surrounding the potential Inr motifs (of *EtMIC1, 2, 3* as identified by 5'RACE and *EtMIC4* and *5* observed from sequence data) as well as regions identified as aiding gene expression through 5' deletion assays. Electrophoretic mobility shift assays could be used to identify protein(s), which can specifically recognise Inr sequences and other elements in the 5' flanking regions of each of the microneme encoding genes. This method could also be used to determine if any of the DNA binding proteins involved in transcription could distinguish between wild-type and mutant Inrs.

Factors which influence Inr promoter strength include a strong preference for an adenosine nucleotide at the transcription start site, a high number of pyrimidines surrounding this, especially at position -1, -2, +4 and +5, the nature of the sequence ~25 bp upstream of the transcription start site (weak Inrs may be inactive if the –25 sequence is G/C rich and will therefore possesses lower affinity for TBP) and the presence or absence of upstream enhancer sequences. Only some of these factors were investigated in this thesis, therefore many other mutations could be used to investigate the strength of the Inr and its surrounding sequence.

It would also be of interest to determine if the predicted Inrs of both *EtMIC4* and *5* are functional. To investigate this it would be necessary to demonstrate if transcription initiates within the Inr elements. Once the transcription stat site of *EtMIC4* was determined it would also be possible to determine if the consensus TATA and CAAT boxes of *EtMIC4* play a part in transcription initiation and if they interact with each

other or with the predicted Inr or if the predicted Inr acts alone. It would be of interest to assess the effects of these mutations *in vivo,* to investigate their effects on the transcription of *EtMIC1* and *2,* and its knock-on effects (i.e. invasion, motility) however this is not possible at present.

With the advent of the *E. tenella* genome sequencing project, it should be possible to identify Inr binding protein(s) and nuclear proteins which are involved in transctiption.

References

Abrahamsen, M., Clark, T. and White, M. (1994). Developmental regulation of an *Eimeria bovis* mRNA encoding refractile body-associated proteins. *Molecular and Biochemical Parasitology* **68**(1), 25-34.

Achbarou, A., Mercereau-Puijalon, O., Sadak, A., Fortier, B., Leriche, M. A., Camus, D. and Dubremetz, J. F. (1991a). Differential targeting of dense granule proteins in the parasitophorous vacuole of *Toxoplasma gondii*. *Parasitology* **103**(3), 321-329.

Achbarou, A., Autheman, J. Fortier, B. Camus, D. and Dubremetz, J. (1991b). Characterization of microneme proteins of *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **47**(2), 223-33.

Adams, J., Torii, M. Ward, G. Wellems, T. Aikawa, M. and Miller, L. (1990). The Duffy receptor family of *Plasmodium knowlesi* is located within the micronemes of invasive malaria merozoites. *Cell* **63**(1), 141-153.

Adams, J. S., B. Dolan, S. Fang, X. Kaslow, D. and Miller, L. (1992). A family of erythrocyte binding proteins of malaria parasites. *Proceedings in the National Academy of Sciences USA* **89**(15), 7085-9.

Aikawa, M. (1971). Parasitological review: *Plasmodium*: The fine structure of malarial parasites. *Experimental Parasitology* **30**, 284-320.

Aikawa, M., Komata, Y., Asai, T. and Midorikawa, O. (1977). Transmission and scanning electron microscopy of host cell entry by *Toxoplasma gondii*. *American Journal of Pathology* **87**, 285-296.

Aikawa, M., Johnson, J. and Rabbege, J. (1978). Erythrocyte entry by malarial parasites. A moving junction between erythrocyte and parasite. *Journal of Cell Biology* **77**(1), 72-82.

Aikawa, M., Rabbege, J. and Epstein, N. (1981). Freeze-fracture study on the erythrocyte membrane during malarial parasite invasion. *Journal of Cell Biology* **91**(1), 55-62.

Ajioka, J. W. (1998). *Toxoplasma gondii*: ESTs and gene discovery. *International Journal for Parasitology* **28**(7), 1025-1031.

Alano, P., Silvestrini, F. and Roca, L. (1996). Structure and polymorphism of the upstream region of the pfg27/25 gene, transcriptionally regulated in gametocytogenesis of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **79**, 207-217.

Anderson, S., Chow, H. and Loh, D. (1988). A conserved sequence in the T-cell receptor beta-chain promoter region. *Proceedings in the National Academy of Sciences USA* **85**(10) 3551-3554.

Anon. P. (1999). EU facts and forecast. *Poultry World* **September issue**.

Asai, T., Miura, S., Sibley, L. D., Okabayashi, H., and Takeuchi, T. (1995). Biochemical and molecular characterisation of nucleoside triphosphate hydrolase isozymes from the parasitic protozoan *Toxoplasma gondii*. *Journal of Biological Chemistry* **270**(19), 11391-11397.

Baba, E., Uno, H., Sadano, N., Fukata, T., Sasai, K. and Arakawa, A. (1996). *Eimeria tenella*: Role of carbohydrates on sporozoite at the penetration into cultured cells. *Experimental Parasitology* **83**(1), 67-72.

Baglia, F. A., Jameson, B. A. and Walsh, P. N. (1990). Localisation of the high molecular weight kininogens binding site in the heavy chain of human factor XI to amino acids phenylanine 56 through serine 86. *Journal of Biological Chemistry* **265**(7), 4149-4154.

Baglia, F. A., Jameson, B. A. and Walsh, P. N. (1991). Identification and chemical synthesis of a substrate-binding site for factor-IX on the heavy chain of Factor-XIA. *Thrombosis and Haemostasis* **65**(6), 693.

Baglia, F. A., Jameson, B. A. and Walsh, P. N. (1993a). The A1 and A4 domains of Factor-XI are involved in the surface mediated activation of Factor-XIIA. *Thrombosis and Haemostasis* **69**(6), 1236.

Baglia, F. A., Jameson, B. A. and Walsh, P. N. (1993b). Identification and characterisation of a binding site for factor-XIIA in the apple-4 domain of coagulation factor-XI. *Journal of Biological Chemistry* **268**(6), 3838-3844.

Baglia, F. A. and Walsh, P. N. (1996). A binding site for thrombin in the apple 1 domain of factor XI. *Journal of Biological Chemistry* **271**(7), 3652-3658.

Baldi, D., Andrews, K., Waller, R., Roos, D., Howard, R., Crabb, B. and Cowman, A. (2000). RAP1 controls rhoptry targeting of RAP2 in the malaria parasite *Plasmodium falciparum*. *EMBO Journal* **19**(11), 2435-2443.

Ball, S. and Pittilo, R. (1988). Ultrastructural observations of macrogametogenesis in *Eimeria bakuensis* (Syne ovina) of sheep. *Parasitology Research* **74**(5) 431-435.

Bannister, L. H., Mitchell, G. H., Dennis, E. D. and Butcher, G. A. (1975). Structure and invasive behaviour of *Plasmodium knowlesi* merozoites *in vitro*. *Parasitology* **71**, 483-491.

Bannister, L., Mitchell, G. H., Butcher, G. A. and Dennis, E. D. (1986). Lamellar membrane associated with rhoptries in erythrocytic merozoites of *Plasmodium knowlesi*- A clue to the mechanism of invasion. *Parasitology* **92**(2), 291-303.

Bannister, L. and Dluzewski, A. (1990). The ultrastructure of red cell invasion in malaria infections: a review. *Blood Cells* **16**(2-3), 257-92.

Barahona, I., Soares, H., Cyrne, L., Penque, D., Denoulet, P. and Rodrigues-Pousada, C. (1988). Sequence of one α- and two β-tubulin genes of *Tetrahymena pyriformis*. Structural and functional relationships with other eukaryotic tubulin genes. *Journal of Molecular Biology* **202**, 365-382.

Barnes, D. A., Bonnin, A., Huang, J. X., Gousset, L., Wu, J., Gut, J., Doyle, P., Dubremetz, J. F., Ward, H. and Petersen, C. (1998). A novel multi-domain mucin-like glycoprotein of *Cryptosporidium parvum* mediates invasion. *Molecular and Biochemical Parasitology* **96**(1-2), 93-110.

Barnwell, J. and Galinski, M. (1989) The adhesion of malaria merozoite proteins to erythrocytes- a reflection of function. *Research in Immunology* **142**(8) 666-672.

Baruch, D., Pasloske, B., Singh, H. (1995). Cloning the *Plasmodium falciparum* gene encoding PfEMP-1 a malarial variant antigen and adherence receptor on the surface of parasitised human erythrocytes. *Cell* **82**, 77-87.

Beaupain, D., Eleouet, J. F. and Romeo, P. H. (1990). Initiation of transcription of the erythroid promoter of the porphobilinogen deaminase gene is regulated by a

cis-acting sequence around the cap site. *Nucleic Acids Research* **18**, 6509-6515.

Beckers, C. J., Dubremetz, J., Mercereau-Puijalon, O. and Joiner, K. (1994). The *Toxoplasma gondii* rhoptry protein Rop2 is inserted into the parasitophorous vacuole membrane, surrounding the intracellular parasite and is exposed to the host cell cytoplasm. *Journal of Cell Biology* **127**(4), 947-961.

Bermudes, D., Peck, K., Afifi-Afifi, M., Beckers, C. and Joiner, K. (1994). Tandemly repeated genes encode nucleoside triphosphate hydrolase isoforms secreted into the parasitophorous vacuole of *Toxoplasma gondii*. *Journal of Biological Chemistry* **269**(46), 29252-29260.

Blanchard, J. (1999). The non-photosynthetic plastid in malarial parasites and other apicomplexans is derived from outside the green plastid lineage *Journal of Eukaryotic Microbiology*.

Bohne, W., Wirsing, A. and Gross, U. (1997). Bradyzoite-specific gene expression in *Toxoplasma gondii* requires minimal genomic elements. *Molecular and Biochemical Parasitology* **85**, 89-98.

Borst, P., Bitter, W., McCulloch, R., Van-Leeuwen, F. and Rudenko, G. (1995). Antigenic variation in malaria. *Cell* **82**, 1-4.

Breathnach, R. and Chambon, P. (1981). Organisation and expression of eukaryotic split genes-coding for proteins. *Annual review of Biochemistry* **50**, 349-383.

Brecht, S., Carruthers, V., Ferguson, D., Giddings, O., Wang, S., Jakle, U., Harper, J., Sibley, L. and Soldati, D. (2001). The *Toxoplasma* micronemal protein MIC4 is an adhesin composed of six conserved apple domains. *Journal of Biological Chemistry* **276**(6) 4119-4127.

Brown, P. (2000a). Molecular and functional characterisation of the *Eimeria tenella* (Apicomplexa) microneme protein, EtMIC5, Imperial College London.

Brown, P., Billington, K., Bumstead, J., Clark, J. and Tomley, F. (2000b). A microneme protein from *Eimeria tenella* with homology to the Apple domains of coagulation factor XI and plasma pre-kallikrein. *Molecular and Biochemical Parasitology* **107**, 91-102.

Brown, P., Gill, A., Nugent, P., McVey, J. and Tomley, F. (2001). Domains of invasion organelle proteins from apicomplexan parasites are homologous with the Apple domains of blood coagulation factor XI and plasma pre-kallikrein and are members of the PAN module superfamily. *FEBS Letters* **497**(1) 31-38.

Brunk, C. and Sadler, L. (1990). Characterisation of the promoter region of *Tetrahymena* genes. *Nucleic Acids Research* **18**(2), 323-329.

Brydges, S. D., Sherman, G. D., Nockemann, S., Loyens, A., Daubener, W., Dubremetz, J. and Carruthers, V. B. (2000). Molecular characterization of TgMIC5, a proteolytically processed antigen secreted from the micronemes of *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **111**(1), 51-66.

Bull, P., Lowe, B., Kortok, M., Molyneux, C., Newbold, C. and Marsh, K. (1995). Parasite antigens on the infected red cell surface are targets for naturally acquired immunity to malaria. *Nature Medicine* **4**, 358-360.

Bumstead, J. M. & Tomley, F. M. (2000). Induction of secretion and surface capping of microneme proteins in *Eimeria tenella*. *Molecular and Biochemical Parasitology* **110**(2), 311-321.

Bunick, D., Zandomeni, R., Ackerman, S. and Weinmann, R. (1982). Mechanism of RNA polymerase II- specific initiation of transcription *in vitro* ATP requirement and uncapped runoff transcripts. *Cell* **29**(3), 877-886.

Buratowski, S., Hahn, S., Guarente, L. and Sharp, P. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**, 549-561.

Burg, J. L., Perelman, D., Kasper, L. H., Ware, P. L. & Boothroyd, J. C. (1988). Molecular analysis of the gene encoding the major surface antigen of Toxoplasma gondii. *Journal of Immunology* **141**(10), 3584-91.

Butcher, P. (1990). Weight matrix descriptions of four eukaryotic RNA polymerase II promoter elements derived from 502 unrelated promoter sequences. *Journal of Molecular Biology* **212**, 563-578.

Calzone, F., Callehan, R. and Gorovsky, M. (1988). Direct measurements of tubulin and bulk message distributions on polysomes of growing, starved and deciliated *Tetrahymena* using RNA gel blots of sucrose gradients containing acrylamide. *Nucleic Acids Research* **16**(20) 9597-9609

Camus, D. and Hadley, T. (1985). A *Plasmodium falciparum* antigen that binds to host erythrocytes and merozoites. *Science* **230**(4725), 553-556.

Canning, E. and Anwar, M. (1968). Studies on meiotic division in coccidial and malarial parasites. *Journal of Protozoology* **15**(2), 290-298.

Carcamo, J., Buckbinder, L. and Reinberg, D. (1991). The initiator directs the assembly of a transcription factor IID-dependent transcription complex. *Proceedings in the National Academy of Sciences USA* **88**, 8052-8056.

Carruthers, V. and Sibley, L. (1997). Sequential protein secretion from three distinct organelles of *Toxoplasma gondii* accompanies invasion of human fibroblasts. *European Journal of Cell Biology* **73**(2), 114-123.

Carruthers, V., Giddings, O. and Sibley, L. (1999a). Secretion of micronemal proteins is associated with *Toxoplasma* invasion of host cells. *Cellular Microbiology* **1**(3), 225-235.

Carruthers, V. B., Moreno, S. N. and Sibley, L. D. (1999b). Ethanol and acetaldehyde elevate intracellular $[Ca^{2+}]$ and stimulate microneme discharge in *Toxoplasma gondii*. *Biochemical Journal* **342**(2), 379-386.

Carruthers, V. and Sibley, L. (1999). Mobilization of intracellular calcium stimulates microneme discharge in *Toxoplasma gondii*. *Molecular Microbiology* **31**(2), 421-428.

Carruthers, V. B., Hakansson, S., Giddings, O. K. and Sibley, L. D. (2000). *Toxoplasma gondii* uses sulphated proteoglycans for substrate and host cell attachment. *Infection and Immunity* **68**(7), 4005-4011.

Carruthers, V. B. (2002). Host cell invasion by the opportunistic pathogen *Toxoplasma gondii*. *Acta Tropica* **81**(2), 111-122.

Cerami, C., Kwakyeberko, F. and Nussenzweig, V. (1992). Binding of malarial circumsporozoite protein to sulphatides [Gal(3-SO4)Beta-1-Cer] and cholestrol-3-sulphate and its dependence on sulphide bond formation between cysteines in region-II. *Molecular and Biochemical Parasitology* **54**(1), 1-12.

Cesbron-Delauw, M. F., Guy, B., Torpier, G., Pierce, R. J., Lenzen, G., Cesbron, J. Y., Charif, H., Lepage, P., Darcy, F., Lecocq, J. P., Capron, A. (1989). Molecular characterisation of a 23 kilodalton major antigen secreted by *Toxoplasma*

gondii. Proceedings of the National Academy of Sciences of the United States of America **86**(19), 7537-7541.

Cesbron-Delauw, M. F. (1994). Dense-granule organelles of Toxoplasma gondii: Their role in the host-parasite relationship. Parasitology Today **10**(8), 293-296.

Chakrabarti, D., Reddy, G., Dae, J., Almira, E., Laipis, P., Ferl, R., Yang, T., Rowe, T., and Schuster, S. (1994). Analysis of expressed sequence tags from Plasmodium falciparum. Molecular and Biochemical Parasitology **66**, 97-104.

Chapman, H. (1997). Biochemical, genetic and applied aspects of drug resistance in Eimeria parasites of the fowl. Avian Pathology **26**, 221-244.

Chaturvedi, S., Qi, H., Coleman, D., Rodriguez, A., Hanson, P., Striepen, B., Roos, D. and Joiner, K. (1999). Constitutive calcium-independent release of Toxoplasma gondii granules occurs through the NSF/SNAP/SNARE/Rab machinery. The Journal of Biological Chemistry **274**(4), 2424-2431.

Chen, Q., Fernandez, V., Sundstrom, A., Schlichtherle, M., Datta, S., Hagblom, P. and Wahlgren, M. (1998). Developmental selection of var gene expression in Plasmodium falciparum. Letters to Nature **394**, 392-395.

Cheng, Q., Cloonan, N., Fischer, K., Thompson, J., Waine, G., Lanzer, M. and Saul, A. (1998). Stevor and rif are Plasmodium falciparum multicopy gene families which potentially encode variant antigens. Molecular and Biochemical Parasitology **97**, 161-176.

Chitnis, C. E. and Miller, L. H. (1994). Identification of the erythrocyte binding domains of Plasmodium vivax and Plasmodium knowlesi proteins involved in erythrocyte invasion. Journal of Experimental Medicine **180**(2), 497-506.

Chobotar, B., Senaud, J., Ernst, J. and Scholtyseck, E. (1975). A fine structural study of asexual stages of the murine coccidium Eimeria ferrisi (Levine and Ivens 1965). Z Parasitenkd **45**, 291-306.

Chobotar, B., Senaud, J., Ernst, J. and Scholtyseck, E. (1975). The ultrastructure of macrogametes of Eimeria ferrisi (Levine and Ivens 1965) in Mus musculus. Z Parasitenkd **48**(2), 111-21.

Chobotar, B. and Scholtyseck, E. (1982). Ultrastructure. The biology of the coccidia. (Long, P. D., Ed.), University Park Press, Baltimore.

Chung, D. W., Fujiawa, W., McMullen, B. A. and Davie, E.W. (1986). Human plasma prekallikrein, a zymogen to a serine protease that contains 4 tandem repeats. *Biochemistry* **25**(9), 2410-2417.

Conaway, R. A. Conaway, J. W. (1993). General initiation-factors for RNA polymerase II. *Annual review of Biochemistry* **62**, 161-190.

Concino, M. F., Lee, R. F., Merrywater, J. P. and Weinmann, R. (1984). The adenovirus major late promoter TATA box and initiation site are both necessary for transcription *in vitro*. *Nucleic Acids Research* **12**, 7423-7433.

Cosson, P., Lefkir, Y., Demolliere, C. and Letourneur, F. (1998). New COP1-binding motifs involved in ER retrieval. *Embo J* **17**, 6863-6870.

Crabb, B. and Cowman, A. (1996). Characterisation of promoters and stable transfection by homologous and nonhomologous recombination in *Plasmodium falciparum*. *Proceedings in the National Academy of Sciences USA* **93**, 7289-7294.

Cruz, A. and Beverly, S. M. (1990). Gene replacement in parasitic protozoa. *Nature* **348**, 171-173.

Dame, J., Williams, J., McCutchan, T., Weber, J., Wirtz, R., Hockmeyer, W., Maloy, W., Haynes, J., Schneider, I., Roberts, D., Sanders, G., Reddy, E., Diggs, C. and Miller, L. (1984). Structure of the gene encoding the immunodominant surface-antigen on the sporozoite of the human malaria parasite Plasmodium falciparum. *Science* **225**(4662), 593-599.

Danforth, H., Entzeroth, R. and Chobotar, B. (1992). Scanning and transmission electron microscopy of host cell pathology associated with penetration by *Eimeria papillata* sporozoites. *Parasitology Research* **78**(7), 570-3.

Daszak, P., Ball, S. J., Pittilo, R. M. and Norton, C. C. (1993). Ultrastructural evidence for dense granule exocytosis by 1st generation merozoites of *Eimeria tenella in vivo*. *Parasitology Research* **79**(3), 256-258.

Daszak, P. (1999). Secretion of multilamellar whorls by *Eimeria tenella* zoites. *Journal of Parasitology* **85**(4), 742-6.

De Carvalho, L. and De Souza, W. (1989). Cytochemical localisation of plasma membrane enzymes markers during internalisation of tachyzoites of *Toxoplasma gondii* by macrophages. *Journal of Protozoology* **36**, 164-170.

Delbac, F., Sanger, A., Neuhaus, E., Stratmann, R., Ajioka, J., Toursel, C., Herm-Gotz, A., Tomavo, S., Soldati, T. and Soldati, D. (2001). *Toxoplasma gondii* myosins B/C: one gene, two tails, two localisations, and a role in parasite division. *The Journal of Cell Biology* **155**(4), 613-623.

Delwiche, C. and Palmer, J. (1997). *Plant Systematics Evolution* **11**, 51-86.

Dessens, J. T., Beetsma, A.L., Dimopoulos, G., Wenhelnik, K., Crisanti, A., Kafatos, F.C. and Sinden, R.E. (1999). CTRP is essential for mosquito infection by malaria ookinetes. *EMBO Journal* **18**(22), 6221-6227.

Di Cristina, M., Spaccapelo, R., Soldati, D., Bistoni, F. and Crisanti, A. (2000). Two conserved amino acid motifs mediate protein targeting to the micronemes of the apicomplexan parasite *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **20**(19), 7332-7341.

Dickeson, S., Walsh, J. and Santoro, S. (1998). Binding of the alpha 2 integrin I domain to extracellular matrix ligands: structural and mechanistic differences between collagen and laminin binding. *Cell Adhesion Communications* **5**(4), 273-81.

Dierks, P., Vanooyen, A., Cochran, M., Dobkin, C., Reiser, J and Weissmann, C. (1983). Three regions from the capsite are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell* **32**(2) 695-706.

Dluzewski, A., Fryer, P., Griffiths, S., Wilson, R. and Gratzer, W. (1989). Red cell membrane protein distribution during malarial invasion. *Journal of Cell Science* **92**(4), 691-9.

Dluzewski, A., Mitchell, G., Fryer, P., Griffiths, S., Wilson, R. and Gratzer, W. (1992). Origins of the parasitophorous vacuole membrane of the malaria parasite, *Plasmodium falciparum*, in human red blood cells. *Journal of Cell Science* **102**(3), 527-32.

Dobrowolski, J. & Sibley, L. (1996). *Toxoplasma* invasion of mammalian cells is powered by the actin cytoskeleton of the parasite. *Cell* **84**(6), 933-939.

Dobrowolski, J., Carruthers, V. and Sibley, L. (1997a). Participation of myosin in gliding motility and host cell invasion by *Toxoplasma gondii*. *Molecular Microbiology* **26**(1), 163-73.

Dobrowolski, J., Niesman, I. and Sibley, L. (1997b). Actin in the parasite *Toxoplasma gondii* is encoded by a single copy gene, ACT1 and exists primarily in a globular form. *Cell Motility Cytoskeleton* **37**(3), 253-262.

Dobrowolski, J. and Sibley, L. (1997). The role of the cytoskeleton in host cell invasion by *Toxoplasma gondii*. *Behring Inst Mitt* **99**, 90-96.

Docampo, R. and Moreno, S. N. (1996). The role of calcium in cell invasion by intracellular parasites. *Parasitology Today* **12**, 61-65.

Donald, R. G. and Roos, D. S. (1993). Stable molecular transformation of *Toxoplasma gondii*: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria. *Proceedings in the National Academy of Sciences USA* **90**(24), 11703-11707.

Downing, A., Knott, V., Werner, J., Cardy, C., Campbell, I. and Handford, P. (1996). Solution structure of a pair of calcium-binding epidermal growth factor- like domains: implications for the Marfan syndrome and other genetic disorders. *Cell* **85**(4), 597-605.

Du, H., Roy, A.L. and Roeder, R. (1993). Human transcription factor USF stimulates transcription through the initiator elements of the HIV-1 and the AD-ML promoters. *Embo Journal* **12**(2), 510-511.

Dubremetz, J. F. (1975). Genesis of merozoites in the coccidia, *Eimeria necatrix*. Ultrastructural study. *Journal of Protozoology* **22**(1), 71-84.

Dubremetz, J. & Torpier, G. (1978). Freeze fracture study of the pellicle of an *Eimerian* sporozoite (Protozoa, Coccidia). *Journal of Ultrastructural Research* **62**(2), 94-109.

Dubremetz, J. F. (1979). Genesis of merozoites in the coccidia, *Eimeria necatrix*. Ultrastructure study. *Journal of Protozoology* **1975**(22), 71-84.

Dubremetz, J. F., Rodriguez, C. and Fereira, E. (1985). *Toxoplasma gondii*: redistribution of monoclonal antibodies on tachyzoites during host cell invasion. *Experimental Parasitology* **59**, 24.

Dubremetz, J., Achbarou, A., Bermudes, D. and Joiner, K. (1993). Kinetics and pattern of organelle exocytosis during *Toxoplasma gondii*/host-cell interaction. *Parasitology Research* **79**(5), 402-408.

Dubremetz, J. (1998a). Host cell invasion by *Toxoplasma gondii*. *Trends in Microbiology* **6**(1), 27-30.

Dubremetz, J., Garcia-Reguet, N., Conseil, V. and Fourmaux, M. (1998b). Apical organelles and host-cell invasion by Apicomplexa. *International Journal for Parasitology* **28**(7), 1007-1013.

Dunn, P. P. J., Bumstead, J. M., and Tomley, F. M. (1996). Sequence, expression and localisation of calmodulin-domain protein kinases in *Eimeria tenella* and *Eimeria maxima*. *Parasitology* **113**(5), 439-448.

Dunn, P. P. J., Stephens, P. J. and Shirley, M. W. (1998). *Eimeria tenella*: two species of extrachromosomal DNA revealed by pulsed-field gel electrophoresis. *Parasitology Research* **84**(4), 272-275.

Ellis, J. and Revets, H. (1990). *Eimeria* species which infect the chicken contain virus-like RNA molecules. *Parasitology* **101**(2), 163-169.

Ellis, J. and Thurlby, T. (1991). Changes in the messenger RNA population during sporulation of *Eimeria maxima*. *Parasitology* **102**(1), 1-8.

Elmendorf, H. G., Singer, S.M., Pierce, J., Cowan, J. and Nash, T.E. (2001). Initiator and upstream elements in the α2-tublin promoter of *Giardia lamblia*. *Molecular and Biochemical Parasitology* **113**, 157-169.

Entzeroth, R. (1984). Electron microscope study of host-parasite interactions of *Sarcocystis muris* (Protozoa, Coccidia) in tissue culture and *in vivo*. *Zeitschrift Fur Parasitenkunde* **70**(1), 131-134.

Entzeroth, R. (1985). Invasion and early development of *Sarcocystis muris* (Apicomplexa, Sarcocystidae) in tissue cultures. *Journal of Protozoology* **32**(3), 446-53.

Entzeroth, R., Kerckhoff, H. and Konig, A. (1992). Microneme secretion in Coccidia: confocal laser scanning and electron microscope study of *Sarcocystis muris* in cell culture. *European Journal of Cell Biology* **59**(2), 405-413.

Entzeroth, R., Stefen, C., Chobotar, B. and Dubremetz, J. (1993). Evidence of secretion of an internal antigen of *Eimeria papillata* into cultured cells. *Archiv Fur Protistenkunde* **143**(4), 347-351.

Entzeroth, R., Zgrzebski, G. and Dubremetz, J. (1989). Secretion of trials during gliding motility of *Eimeria nieschulzi* (Apicomplexa, Coccidia) sporozoites visualized by a monoclonal antibody and immuno-gold-silver enhancement. *Parasitology Research* **76**(2), 174-5.

Eschenbacher, K. H., Klein, H., Sommer, I., Meyer, H., Entzeroth, R., Mellhorn, H. and Ruger, W. (1993). Characterisation of cDNA clones encoding a major microneme antigen of *Sarcocystis muris* (Apicomplexa) cyst merozoites. *Molecular and Biochemical Parasitology* **62**, 27-36.

Etzion, Z., Murray, M. C. and Perkins, M. E. (1991). Isolation and characterisation of rhoptries of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **47**(1), 51-61.

Ferguson, D., Birch-Andersen, A., Hutchinson, W. and Sim, J. (1978a). Light and electron microscopy on the sporulation of the oocysts of *Eimeria brunetti*. I. Development of the zygote and formation of the sporoblasts. *Acta Patholgica Microbiolica Scandinavica [B]* **86**(1), 1-11.

Ferguson, D., Birch-Andersen, A., Hutchinson, W. and Sim, J. (1978b). Light and electron microscopy on the sporulation of the oocysts of *Eimeria brunetti*. II. Development into the sporocyst and formation of the sporozoite. *Acta Patholgica Microbiolica Scandinavica [B]* **86**(1), 13-24.

Fernando, M. and Pasternak, J. (1991). *Eimeria* spp. of the domestic fowl: resolution of chromosomes by field inversion gel electrophoresis. *Experimental Parasitology* **72**(3), 306-10.

Fernando, M., Rose, M. and Millard, B. (1987). *Eimeria* spp. of domestic fowl: the migration of sporozoites intra- and extra-enterically. *Journal of Parasitology* **73**(3), 561-7.

Fichera, M. and Roos, D. (1997). A plastid organelle as a drug target in apicomplexan parasites. *Letters to Nature* **390**, 407-409.

Forney, J., Vaughan, D., Yang, S. and Healey, M. (1998). Actin-dependent motility in *Cryptosporidium parvum* sporozoites. *Journal Parasitology* **84**(5), 908-913.

Frevert, U., Sinnis, P., Cerami, C., Shreffler, W., Takacs, B. and Nussenzweig, V. (1993). Malaria circumsporozoite protein binds to heparan sulphate proteoglycans associated with the surface membrane of hepatocytes. *Journal of Experimental Medicine* **177**(5), 1287-1298.

Fromm, M. and Berg, P. (1983). Simian virus-40 early-region and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Molecular Cell Biology* **3**(6), 991-999.

Fujikawa, K., Chung, D. W., Hendrickson, L. E. and Davie, E. W. (1986). Amino acid sequence of human Factor-XI, a blood coagulation factor with 4 tandem repeats that are highly homologous with plasma prekallikrein. *Biochemistry* **25**(9), 2417-2424.

Garcia-Reguet, N., Lebrun, M., Fourmaux, M. N., Mercereau-Puijalon, O., Mann, T., Beckers, C. J. M., Samyn, B., Van Beeumen, J., Bout, D. and Dubremetz JF. (2000). The microneme protein MIC3 of *Toxoplasma gondii* is a secretory adhesin that binds to both the surface of the host cells and the surface of the parasite. *Cellular Microbiology* **2**(4), 353-364.

Gilchirst, C. and Petri Jr, W. (1999). Virulence factors of *Entamoeba histolytica*. *Current Opinion in Microbiology* **2**, 433-437.

Goonewardene, R., Daily, J., Kaslow, D., Sullivan, T. J., Duffy, P., Carter, R., Mendis, K. and Writh, D. (1993). Transfection of the malaria parasite and expression of firefly luciferase. *Proceedings in the National Academy of Sciences USA* **90**, 5234-5236.

Greif, G. and Entzeroth, R. (1996). *Eimeria tenella:* localisation of rhoptry antigens during parasite-host cell interactions by a rhoptry-specific monoclonal antibody in PCKC culture. *Applied Parasitology* **37**, 253-259.

Gross, U., Bohne, W., Soete, M. and Dubremetz, J. (1996). Developmental differentiation between tachyzoites and bradyzoites of *Toxoplasma gondii.* *Parasitology Today* **12**, 30-33.

Gurnett, A., Dulski, P., Darkin-Rattray, S., Carrington, M and Schmatz, D. (1995). Selective labelling of intracellular parasite proteins by using ricin. *Proceedings in the National Academy of Sciences USA* **92**, 2388-2392.

Gustafson, P. V., Agar, H.D. and Cramer, D.I. (1954). An electron microscope study of *Toxoplasma. American Journal Tropical Medical Hygiene* **3**, 1008-1022.

Hager, K., Striepen, B., Tilney, L. and Roos, D. (1999). The nuclear envelope serves as an intermediary between the ER and Golgi complex in the intracellular parasite *Toxoplasma gondii. Journal of Cell Science* **112**, 2631-2638.

Hakansson, S., Morisaki, H., Heuser, J. and Sibley, L. (1999). Time-lapse video microscopy of gliding motility in *Toxoplasma gondii* reveals a novel, biphasic mechanism of cell locomotion. *Molecular Biology of the Cell* **10**(11), 3539-47.

Hampsey, M. (1998). Molecular genetics of the RNA polymerase II general transcriptional machinery. *Microbiology and Molecular Biology Reviews* **62**(2), 465-503.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *Journal of Molecular Biology* **166**, 557-580.

Hariharan, N. and Perry, R. P. (1990). Functional dissection of a mouse ribosomal protein promoter; significance of the polypyrimidine initiator and an element in the TATA-box region. *Proceedings in the National Academy of Sciences USA* **87**, 1526-1530.

Harris, C., Salgo, M., Tanowitz, H. and Wittner, M. (1988). *In vitro* assessment of antimicrobial agents against *Toxoplasma gondii. Journal of Infectious Diseases* **157**, 14-22.

Harter, C. and Wieland, F. (1996). The secretory pathway: mechanisms of protein sorting and transport. *Biochimica et Biophysica Acta* **1286**, 75-93.

Hayward, R., DeRisi, J., Alfadhli, S., Kaslow, D., Brown, P. and Rathod, P. (2000). Shotgun DNA microarrays and stage-specific gene expression in *Plasmodium falciparum* malaria. *Molecular Microbiology* **35**, 6-14.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A., Kel, O., Ignatieva, E., Ananko, E., Podkolodnaya, O., Kolpakov, F., Podkolodny, N and Kolchanov, N. (1998). Databases on transcription regulation: TRANSFAC, TRRD and COMPEL. *Nucleic Acids Research* **26**, 364-370.

Heintzelman, M. and Schwartzman, J. (1997). A novel class of unconventional myosins from *Toxoplasma gondii*. *Journal of Molecular Biology* **271**(1), 139-46.

Heintzelman, M. and Schwartzman, J. (1999). Characterisation of myosin-A and myosin-C: two class XIV unconventional myosins from *Toxoplasma gondii*. *Cell Motility Cytoskeleton* **44**, 58-67.

Heintzelman, M. and Schwartzman, J. (2001). Myosin diversity in Apicomplexa. *Journal of Parasitology* **87**, 429-432.

Hen, R., Sassonecors, P., Corden, J., Gaus, M. and Chambon, P. (1982). Sequences upstream from the TATA box are required *in vivo* and *in vitro* for efficient transcription from the adenovirus serotype-2 major late promoter. *Proceedings in the National Academy of Sciences USA* **79**(23) 7132-7136.

Hernandez, R., Mattei, D., Sterekers, Y., Peterson, D., Wellers, T. and Scherf, A. (1997). Expressed var genes are found in *Plasmodium falciparum* chromosomes. *Molecular and Cellular Biology* **17**, 604-611.

Hettmann, C., Herm, A., Geiter, A., Frank, B., Schwarz, E., Soldati, T. and Soldati, D. (2000). A dibasic motif in the tail of a class XIV apicomplexan myosin is an essential determinant of plasma membrane localization. *Molecular and Cellular Biology* **11**(4), 1385-1400.

Ho, D. H., Badellino, K., Baglia, F. and Walsh, P. (1998). A binding site for heparin in the apple 3 domain of factor XI. *Journal of Biological Chemistry* **273**(26), 16382-90.

Hobman, T. C., Zhao, B. P., Chan, H. N. and Farquhar, M. G. (1998). Immunoisolation and characterization of a subdomain of the endoplasmic reticulum that concentrates proteins involved in COPII vesicle biogenesis. *Molecular Biology of the Cell* **96**, 1265-1278.

Holberton, D. & Marshall, J. (1995). Analysis of consensus sequence patterns in *Giardia* cytoskeleton gene promoters. *Nucleic Acids Research* **23**, 2945-2953.

Hong, W. J. (1998). Protein transport from the endoplasmic reticulum to the Golgi apparatus. *Journal of Cell Science* **111**(19), 2831-2839.

Hoppe, H. and Joiner, K. (2000). Cytoplasmic tail motifs mediate endoplasmic reticulum localisation and export of transmembrane reporters in the protozoan parasite *Toxoplasma gondii*. *Cellular Microbiology* **2**(6), 569-578.

Hoppe, H., Ngo, H., Yang, M. and Joiner, K. (2000). Targeting to rhoptry organelles of *Toxoplasma gondii* involves evolutionarily conserved mechanisms. *Nature Cell Biology* **2**(7), 449-456.

Horrocks, P. and Kilbey, B. (1996). Physical and functional mapping of the *Plasmodium falciparum* proliferating cell nuclear antigen genes transcription start site. *Molecular and Biochemical Parasitology* **82**, 207-215.

Howard, R. F., Jacobson, K. C., Rickel, E. and Thurman, J. (1998). Analysis of inhibitory epitopes in the *Plasmodium falciparum* rhoptry protein RAP-1 including identification of a second inhibitory epitope. *Infection and Immunity* **66**(1), 380-386.

Huston, C., Houpt, E., Mann, B., Hahn, C. and Petri Jr, W. (2000). Caspase 3-dependent killing of host cells by the parasite *Entamoeba histolytica*. *Cellular Microbiology* **2**, 617-625.

Innis, M. and Gelfand, D. (1990). Optimisation of PCRs. In *PCR protocols: A guide to methods and applications*. (Innis, M. A., Gelfand, D.H., Sninsky, J.J. and White, T.J., ed.), pp. 3-12. Academic Press, London.

Jackson, M., Nilsson, T. and Peterson, P. (1990). Identification of a consensus motif for retention of transmembrane proteins in the endoplasmic reticulum. *Embo Journal* **9**, 3153-3162.

Jacobs, D., Dubremetz, J. F., Loyens, A., Bosman, F. and Saman, E. (1998). Identification and heterologous expression of a new dense granule protein (GRA7) from *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **91**(2), 237-249.

Jacobs, L. (1967). *Toxoplasma* and toxoplasmosis. *Advances in Parasitology* **5**, 1-45.

Jakobsen, P. H., Heegaard, P. M. H., Koch, C., Wasniowska, K., Lemnge, M. M., Jensen, J. B. and Sim, B. K. L. (1998). Identification of an erythrocyte binding peptide from the erythrocyte binding antigen, EBA-175, which blocks parasite multiplication and induces peptide-blocking antibodies. *Infection and Immunity* **66**(9), 4203-4207.

Jamieson, J. (1998). The Golgi complex: perspectives and prospectives. *Biochimica et Biophysica Acta* **1404**, 3-8.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B. (1994). DNA sequence requirements for transcriptional initiator activity in mammalian cells. *Molecular and Cellular Biology* **14**(1), 116-127.

Jbilo, O., Toutant, J., Vatsis, K., Chatonnet, A. and Lockridge, O. (1994). Promoter and transcription start site of human and rabbit butyrylcholinesterase genes. *Journal of Biological Chemistry* **269**(33) 20829-20837.

Jean, L., Péry, P., Dunn, P., Bumstead, J., Billington, K., Ryan, R. and Tomley, F. (2001). Genomic organisation and developmentally regulated expression of an apicomplexan aspartyl proteinase. *Gene* **262**(1-2), 129-136.

Jeffers, T. (1975). Attenuation of *Eimeria tenella* through selection for precociousness. *Journal of Parasitology* **61**(6), 1083-90.

Jensen, J. B. and Edgar, S. A. (1976). Possible secretory function of the rhoptries of *Eimeria magana* during penetration of cultured cells. *Journal of Parasitology* **62**(6), 988-992.

Joe, A., Hamer, D. H., Kelley, M. A., Pereira, M. E., Keusch, G. T., Tzipori, S. andWard, H. D. (1994). Role of a Gal/GalNAc-specific sporozoite surface lectin in *Cryptosporidium parvum*-host cell interaction. *Journal of Eukaryotic Microbiology* **41**(5), 44S.

Joe, A., Verdon, R., Tzipori, S., Keusch, G. T. and Ward, H. D. (1998). Attachment of *Cryptosporidium parvum* sporozoites to human intestinal epithelial cells. *Infection and Immunity* **66**(7), 3429-3432.

John, N. M., Zea, M. E. R., Kawano, T., Omata, Y., Saito, A., Toyoda, Y. and Mikami, T. (1999). Identification of carbohydrates on *Eimeria stiedai* sporozoites and their role in the invasion of cultured cells *in vitro*. *Veterinary Parasitology* **81**(2), 99-105.

Joiner, K. A. (1991). Cell attachment and entry by *Toxoplasma gondii*. *Behring Inst Mitt* **88**, 20-26.

Joiner, K. A., Fuhrman, S. A., Miettinen, H. M., Kasper, L. H. and Mellman, I. (1990). *Toxoplasma gondii*- Fusion competence of parasitophorous vacuoles in FC-receptor transfected fibroblasts. *Science* **249**(4969), 641-646.

Jones, K., Luciw, P. and Duchange, N. (1988). Structural arrangements of transcription control domains within the 5'-untranslated leader regions of the HIV-1 and HIV-2 promoters. *Genes and Development* **2**(9) 1101-1114.

Jones, T. C., Yeh, S., and Hirsch, J.G. (1972). The interaction between *Toxoplasma gondii* and mammalian cells. *Journal of Experimental Medicine* **136**, 1157-1172.

Joyner, L. and Long, P. (1974). The specific characters of the *Eimeria* with special reference to the coccidia of the fowl. *Avian Pathology* **3**(3), 145-157.

Kadonaga, J., Jones, K. and Tjian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by SP1. *Trends in Biochemical Science* **11**, 20-23.

Kamata, T., Liddington, R. C. and Takada, Y. (1999). Interaction between collagen and the alpha(2) I-domain of integrin alpha(2)beta(1). Critical role of conserved residues in the metal ion- dependent adhesion site (MIDAS) region. *Journal of Biological Chemistry* **274**(45), 32108-11.

Kappe, S., Bruderer, T., Gantt, S., Fujioka, H., Nussenzweig, V. and Menard, R. (1999). Conservation of a gliding motility and cell invasion machinery in apicomplexan parasites. *Journal of Cell Biology* **147**(5), 937-944.

Karsten, V., Qi, H., Beckers, C., Reddy, A., Dubremetz, J., Webster, P. and Joiner, K. (1998). The protozoan parasite *Toxoplasma gondii* targets proteins to dense granules and the vacuolar space using both conserved and unusual mechanisms. *Journal of Cell Biology* **141**(6), 1323-1333.

Kasper, L. H., Crabb, J. H. and Pfefferkorn, E.R. (1983). Purification of a major membrane proteins of *Toxoplasma gondii* by immuno-adsorption with a monoclonal antibody. *Journal of Immunology* **147**, 3501-3506.

Kasper, L. H. (1989). Identification of stage specific antigen of *Toxoplasma gondii*. *Immunity* **57**, 668-672.

Kaufmann, J. and Smale, S. (1994). Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes and Development* **8**, 821-829.

Kawazoe, U., Chapman, H. and Shaw, M. (1991). Sensitivity of field isolates of *Eimeria acervulina* to salinomycin, maduramicin and a mixture of clopidol and methyl benzoquate in the chicken. *Avain Pathology* **2**(3), 439-446.

Kawazoe, U., Tomley, F., and Frazier, J. (1992). Fractionation and antigenic characterization of organelles of *Eimeria tenella* sporozoites. *Parasitology* **104**(1), 1-9.

Kelleher, M. and Tomley, F. M. (1998). Transient expression of beta-galactosidase in differentiating sporozoites of *Eimeria tenella*. *Molecular and Biochemical Parasitology* **97**(1-2), 21-31.

Khan, S. M., Jarra, W., and Preiser, P. R. (2001). The 235 Kda rhoptry protein of *Plasmodium yoelii*: function at the junction. *Molecular and Biochemical Parasitology* **117**(1), 1-10.

Kim, K., Soldati, D. and Boothroyd, J. (1993). Gene replacement in *Toxoplasma gondii* with chloramphenicol acetyltransferase as selectable marker. *Science* **262**(5135), 911-914.

King, C. (1981). Cell surface interaction of the protozoan *Gregarina* with concanavalin A beads - implications for models of gregarine gliding. *Cell Biology International Repeats* **5**(3), 297-305.

King, C. (1988). Cell motility of *Sporozoan* protozoa. *Parasitology Today* **4**, 315-319.

Kisseleva, T., Bhattacharya, S., Schroeder-Braunstein, J. and Schindler, C. (2002). Signalling through the JAK/STAT pathway, recent advances and future challenges. *Gene* **285**, 1-24.

Klausner, R. D., Donaldson, J. D. and Lippinocott-Schwartz, J. (1992). Brefaldin A: Insights into the control of membrane traffic and organelle structure. *Journal of Cell Biology* **116**, 1071-1080.

Klein, H., Loschner, B., Zyto, N., Portner, M. and Montag, T. (1998). Expression, purification, and biochemical characterization of a recombinant lectin of

Sarcocystis muris (Apicomplexa) cyst merozoites. *Glycoconjugate Journal* **15**(2), 147-153.

Kohler, S., Delwiche, C. F., Denny, P. W., Tilney, L. G., Webster, P., Wilson, R. J. M., Palmer, J. D. and Roos, D. S. (1997). A plastid of probable green algal origin in apicomplexan parasites. *Science* **275**(5305), 1485-1489.

Laban, A., Tobin, J. F., Delafaille, M. A. C. and Wirth, D. F. (1990). Stable expression of the bacterialneor gene in *Leishmania enrietti*. *Nature* **343**, 572-574.

Labruyere, E., Lingnau, M., Mercier, C., and Sibley, L. D. (1999). Differential membrane targeting of the secretory proteins GRA4 and GRA6 within the parasitophorous vacuole formed by *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **102**(2), 311-324.

Ladda, R., Aikawa, M. and Sprinz, H. (1969). Penetration of entrocytes by merozoites of mammalian and avian malarial parasites. *Journal of Parasitology* **55**, 633-644.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-5.

Lanzer, M., de Bruin, D. and Ravetch, J. (1992a). A sequence element associated with the *Plasmodium falciparum* KAHRP gene is the site for developmentally regulated protein-DNA interactions. *Nucleic Acids Research* **20**(12), 3051-3056.

Lanzer, M., de Bruin, D. and Ravetch, J. (1992b). Transcription mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates. *Embo Journal* **11**(5), 1949-1955.

Lanzer, M., Wertheimer, S., de Bruin, D. and Ravetch, J. (1993). *Plasmodium*: Control of gene expression in malaria parasites. *Experimental Parasitology* **77**, 121-128.

Lawn, A. and Rose, M. (1982). Mucosal transport of *Eimeria tenella* in the cecum of the chicken. *Journal of Parasitology* **68**(6), 1117-23.

Lecordier, L., Mercier, C., Torpier, G., Tourvielle, B., Darcy, F. Liu, J., Meas, P., Tartar, A., Capron, A. and Cesbron-Delauw, M. F. (1993). Molecular structure of a *Toxoplasma gondii* dense granule antigen (GRA-5) associated with the

parasitophorous vacuole membrane. *Molecular and Biochemical Parasitology* **59**(1), 143-154.

Lecordier, L., Moleonborodowsky, I., Dubremetz, J. F., Tourvielle, B., Mercier, C., Deslee, D., Capron, A., Cesbron-Delauw, M. F. (1995). Characterisation of a dense granule antigen of *Toxoplasma gondii* (GRA-6) associated to the network of the parasitophorous vacuole. *Molecular and Biochemical Parasitology* **70**(1-2), 85-94.

Lee, E. H. and Alizzi, S. A. (1981). Selective killing of macrophages in the peritoneal-cavity by carrageenan and its effect on normal infection of *Eimeria tenella* in chickens. *Avian Diseases* **25**(2), 503-512.

Lee, S., Fernando, M. A. and Nagy, E. (1996). DsDNA associated with virus-like particles in *Eimeria* spp of the domestic fowl. *Parasitology Research* **82**(6), 518-523.

Lee, M. and Van der Ploeg, L. (1990). *Science* **250**, 1583-1587.

Leriche, M. A. and Dubremetz, J. (1991). Characterisation of the protein contents of rhoptry and dense granules of *Toxoplasma gondii* tachyzoites subcellular fractionation monoclonal antibodies. *Molecular and Biochemical Parasitology* **45**(2), 249-260.

Letourneur, F., Gaynor, E., Hennecke, S., Demolliere, C., Duden, R., Emr, S., Riezman, H. and Cosson, P. (1994). Coatomer is essential for retrieval of dilysine-tagged protein to the endoplasmic reticulum. *Cell* **79**, 1199-1207.

Li, Y., Camp, S., Rachinsky, T., Bongiorno, C. and Taylor, P. (1993). Promoter elements and transcriptional control of the mouse acetylcholinesterase gene. *Journal of Biological Chemistry* **268**(5) 3563-3572.

Liendo, A. and Joiner, K. (2000). *Toxoplasma gondii*: conserved protein machinery in an unusual secretory pathway? *Microbes and Infection* **2**(2), 137-144.

Lingelbach, K. and Joiner, K. A. (1998). The parasitophorous vacuole membrane surrounding *Plasmodium* and *Toxoplasma*: an unusual compartment in infected cells. *Journal of Cell Science* **111**, 1467-1475.

Liston, D. and Johnson, P. (1999). Analysis of a ubiquitous promoter element in a primitive eukaryote: early evolution of the initiator element. *Molecular and Cellular Biology* **19**(3), 2380-2388.

Lo, K. and Smale, S. (1996). Generality of a functional initiator consensus sequence. *Gene* **182**, 13-22.

Long, P. L. and Joyner, L. P. (1984). Problems in the identification of species of *Eimeria*. *Journal of Protozoology* **31**(4), 535-541.

Look, D., Pelletier, M., Tidwell, R., Roswit, W. and Holtzman, M. (1995). Stat1 depends on transcriptional synergy with Sp1. *Journal of Biological Chemistry* **270**, 30264-30267.

Lovett, J. L., Howe, D.K. and Sibley, L.D. (2000). Molecular characterisation of a thrombospondin-related anonymous protein homologue in *Neospora caninum*. *Molecular and Biochemical Parasitology* **107**, 33-43.

Lu, H. G., Zhong, L., Chang, K. P. and Docampo, R. (1997). Intracellular calcium pool content and signalling and expression of a calcium pump are linked to virulence in *Leishmania mexicana* amazonensis amastigotes. *Journal of Biological Chemistry* **272**, 9464-9473.

Lycke, E., Carlberg, K. and Norrby, R. (1975). Interactions between *Toxoplasma gondii* and its host cells: functions of the penetration-enhancing factor of *Toxoplasma*. *Infection and Immunity* **11**, 853-861.

Mallabiabarrena, A., Jimenez, M., Rico., M. and Alarcon, B. (1995). A tyrosine-containing motif mediates ER retention of CD3-epsilon and adopts a helix-turn structure. *Embo Journal* **14**, 2257-2268.

Matsumoto, A., Bessho, H., Uehira, K. and Suda, T. (1991). Morphological studies of the association of mitochondria with *Chlamydial inclusions* and the fusion of *Chlamydial inclusions*. *Journal of Electron Microscopy* **40**(5), 356-363.

Matuschewski, K. M. M., Pinder, J. C., Nussenzweig, V. and Kappe, S. H. I. (2001). Identification of the class XIV myosins Pb-MyoA and Py-MyoA and expression in *Plasmodium* sporozoites. *Molecular and Biochemical Parasitology* **112**(1), 157-161.

Matuschewski, K., Nunes, A., Nussenzweig, V. and Menard, R. (1999). Two adhesion motifs in TRAP contribute independently to host cell invasion by *Plasmodium* sporozoites. *In abstracts of Molecular Parasitology Meeting, Woods Hole, MA*, 105.

McAndrew, M., Read, M., Sims, P. and Hyde, J. (1993) Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely AT-rich human malaria parasite *Plasmodium falciparum*. *Gene* **124**, 165-171.

McConkey, G., Rogers, M. and McCutchan, M. (1997). Inhibition of *Plasmodium falciparum* protein synthesis-Targeting the plastid-like organelle eiththiostrepton. *Journal of Biological Chemistry* **272**, 2046-2049.

McCormick, C. J., Tuckwell, D. S., Crisanti, A., Humphries, M. J. and Hollingdale, M. R. (1999). Identification of heparin as a ligand for the A-domain of *Plasmodium falciparum* thrombospondin-related adhesion protein. *Molecular and Biochemical Parasitology* **100**(1), 111-124.

McDonald, V. and Rose, M. E. (1987). *Eimeria tenella* and *E. necatrix*: a third generation of schizogony is an obligatory part of the developmental cycle. *Journal of Parasitology* **73**(3), 617-22.

McFadden, G. and Roos, D. (1999). Apicomplexan plastids as drug targets. *Trends in Microbiology* **7**(8), 328-333.

McFadden, G. and Waller, R. (1997). Plastids in parasites of humans. *BioEssays* **19**(11), 1033-1040.

McFadden, G., Waller, R., Reith, M. and Lang-Unnasch, N. (1997). Plastids in apicomplexan parasites. *Plant Systematics and Evolution* **11**, 261-287.

McLaren, D., Bannister, L., Trigg, P. and Butcher, G. (1977). A freeze-fracture study on the parasite-erythrocyte interrelationship in *Plasmodium knowlesi* infections. *Bull World Health Organ* **55**(2-3), 199-203.

McMullen, B. A., Fujikawa, K. and Davie, E. W. (1991). Location of the disulfide bonds in human coagulation factor XI: the presence of tandem apple domains. *Biochemistry* **30**(8), 2056-60.

Means, A. L. and Farnham, P. J. (1990). Transcription initiation from the dihydrofolate reductase promoter positioned by H1P1 binding at the initiation site. *Molecular Cell Biology* **10**, 653-661.

Meissner, M., Reiss, M., Viebig, N., Carruthers, V. B., Toursel, C., Tomavo, S., Ajioka, J. W. and Soldati, D. (2002). A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorters. *Journal of Cell Science* **115**(3), 563-574.

Meissner, M., Reiss, M., Viebig, N., Carruthers, V., Toursel, C., Tomavo, S., Ajioka, J and Soldati, D. (2001). A family of transmembrane microneme proteins of *Toxoplasma gondii* contain EGF-like domains and function as escorters. *Journal of Cell Science* **115**(3), 563-574.

Mercier, C., Lecordier, L., Darcy, F., Deslee, D., Murry, A., Tourveillie, B. Maes, P., Vapron, A. and Cesbron-Delauw, M-F. (1993). Molecular characterisation of a dense granule antigen (GRA-2) associated with the network of the parasitophorous vacuole in *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **58**(1), 71-82.

Mercier, C., Lefebvre-Van Hende, S., Garber, G., Lecordier, L., Capron, A. and Cesbron-Delauw, M. (1996). Common *cis*-acting elements critical for the expression of several genes of *Toxoplasma gondii*. *Molecular Microbiology* **21**(2), 421-8.

Mercier, C., Howe, D., Mordue, D., Lingnau, M. and Sibley, L. (1998). Targeted disruption of the GRA2 locus in *Toxoplasma gondii* decreases acute virulence in mice. *Infection and Immunity* **66**(9), 4176-82.

Mercier, C., Dubremetz, Z., Rauscher, B., Lecordier, L., Linguau, M., Sibley, L., Cesbron-Delauw, M-F. (2001a). Biogenesis of the *Toxoplasma* parasitophorous vacuole: both the dense granules proteins GRA2 and GRA6 play a key role in the tubular organization of the network. *Molecular Biology of the Cell* **13**(7), 2397-2409.

Mercier, C., Rauscherb, B., Lecordierb, L., Desléeb, D., Dubremetz, J-F and Cesbron-Delauw, M-F. (2001b). Lack of expression of the dense granule protein GRA5 does not affect the development of *Toxoplasma* tachyzoites. *Molecular and Biochemical Parasitology* **116**(2), 247-251.

Mevelec, M. N., Chardes, T., Mercereaupuijalon, O., Bourguin, I., Achbarou, A., Dubremetz, J. F. and Bout, D. (1992). Molecular cloning of GRA4, a *Toxoplasma gondii* dense granule protein, recognised by mucosal IgA antibodies. *Molecular and Biochemical Parasitology* **56**(2), 227-238.

Michishita, M., Videm, V. and Arnaout, M. A. (1993). A novel divalent cation-binding site in the A-domain of the beta-2-integrin-CR3 (CD11B/CD18) is essential for ligand binding. *Cell* **72**(6), 857-867.

Miller, L., Hudson, D. and Haynes, J. (1988). Identification of *Plasmodium knowlesi* erythrocyte binding proteins. *Molecular and Biochemical Parasitology* **31**(3), 217-22.

Miller, L. H., Aikawa, M., Johnson, J.G. and Shiroishi, T. (1979). Interaction between cytochalasin B-treated malarial parasites and erythrocytes. Attachment and junction formation. *Journal of Experimental Medicine* **149**, 172-184.

Mineo, J. R., McLeod, R., Mack, D., Smith, J., Khan, I. A., Ely, K. H. and Kasper, L. H. (1993). Antibodies to *Toxoplasma gondii* major surface protein (SAG-1, P30) inhibit infection of host cells and are produced in murine intestine after peroral infection. *Journal of Immunology* **150**(9), 3951-3964.

Miyoshi, K. e. a. (2001). Structure of the mouse Stat3/5 locus: evolution from Drosophila to zebrafish to mouse. *Genomics* **71**, 150-155.

Mordue, D. G., Desai, N., Dustin, M. and Sibley, L. D. (1999a). Invasion by *Toxoplasma gondii* establishes a moving junction that selectively excludes host cell plasma membrane proteins on the basis of their membrane anchoring. *Journal of Experimental Medicine* **190**(12), 1783-1792.

Mordue, D. G., Hakansson, S., Niesman, I. and Sibley, L. D. (1999b). *Toxoplasma gondii* resides in a vacuole that avoids fusion with host cell endocytic and exocytic vesicular trafficking pathways. *Experimental Parasitology* **92**, 87-99.

Moreno, S. N., Silva, J., Vercesi, A. E. and Docampo, R. (1994). Cytosolic free calcium elevation in *Trypanosoma cruzi* is required for cell invasion. *Journal of Experimental Medicine* **180**, 1535-1540.

Moreno, S. N. J. and Zhong, L. (1996). Acidocalcisomes in *Toxoplasma gondii* tachyzoites. *Biochemical Journal* **313**, 655-659.

Morisaki, J., Heuser, J. and Sibley, L. (1995). Invasion of *Toxoplasma gondii* occurs by active penetration of the host cell. *Journal of Cell Science* **108**(6), 2457-2464.

Morita, T., Asano, K. and Saeki, H. (1994). Invasion of *Babesia gibsoni* into mouse erythrocytes. *Journal of Protozoology Research* **5**(3) 129-135.

Muhich, M. and Boothroyd, J. (1988). Polycistronic transcripts in trypanosomes and their accumulation during heat shock: evidence for a precursor role in mRNA synthesis. *Molecular Cell Biology* **8**, 3837-3846.

Muhlethaler-Mottet, A., Di-Berardino, W., Otten, L. and Mach, B. (1998). Activation of the MHC class II transactivator CIITA by interferon-gamma requires cooperative interaction between Stat1 and USF-1. *Immunity* **8**, 157-166.

Muller, H., Reckmann, I., Hollingdale, M., Bujard, H., Robson, K. and Crisanti, A. (1993). Thrombospondin related anonymous protein (TRAP) of *Plasmodium falciparum* binds specifically to sulphated glycoconjugates and to HepG2 hepatoma cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. *Embo Journal* **12**(7), 2881-2889.

Mullis, K. and Faloona, F. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. *Methods Enzymology* **155**, 335-350.

Nagel, S. and Boothroyd, J. (1988). The alpha- and beta-tubulins of *Toxoplasma gondii* are encoded by single copy genes containing multiple introns. *Molecular and Biochemical Parasitology* **29**, 261-273.

Naitza, S., Spano, F., Robson, K. J. H. and Crisanti, A. (1988). The thrombospondin related family of apicomplexan parasites: the gears of the cell invasion machinery. *Parasitology Today* **14**, 479-484.

Nakaar, V., Bermudes, D., Ran Peck, K. and Joiner, K. (1992). Upstream elements required for expression of nucleotide triphosphate hydrolase genes of *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **92**, 229-239.

Nichlos, B. A. and O'Connor, G. R. (1981). Penetration of mouse peritoneal macrophages by the protozoan *Toxoplasma gondii*. *Laboratory Investigations* **44**, 324-335.

Nichlos, B. A., Chiappino, M. and O'Connor, G. R. (1983). Secretion from the rhoptries of *Toxoplasma gondii* during host cell invasion. *Journal of Ultrastructure Research* **83**(1) 85-98.

Nichlos, M. E., Rubinstein, P., Barnwekk, J., Decordoba, S. R. and Rosenfeild, R. E. (1987). A new human Duffy blood group specificity defined by a murine monoclonal antibody: Immunogenetics and association with susceptibility to *Plasmodium vivax*. *Journal of Experimental Medicine* **166**(3), 776-785.

Nichols, B. and Pelham, H. (1998). SNAREs and membrane fusion in the Golgi apparatus. *Biochimica et Biophysica Acta* **1404**, 9-31.

Nishimura, N., Bannykh, S., Slabough, S., Matteson, J., Altschuler, Y., Hahn, W. (1999). A di-acidic (DXE) code directs concentration of cargo during export from the endoplasmic reticulum. *Journal of Biological Chemistry* **274**, 15937-15946.

Norton, C. C. and Chard, M. J. (1983). The oocyst sporulation time of *Eimeria* species from the fowl. *Parasitology* **86**(2), 193-8.

Nueda, A., Lopez-Cabrera, M., Vara, A. and Corbi, A. (1993). Characterisation of the CD11a (alpha L, LFA-alpha) integrin gene promoter. *Journal of Biological Chemistry* **268**(26) 19305-19311.

Nussenzweig, V. and Nussenzweig, R. S. (1985). Circumsporozoite proteins of malaria parasites. *Cell* **42**, 401-403.

Odenthal-Schinttler, M., Tomavo, S., Becker, D., Dubremetz, J-F and Schwarz, R. (1993). Evidence for N-linked glycosylation in *Toxoplasma gondii*. *Biochemical Journal* **291**, 713-721.

Ohmori, Y., Schreiber, R. and Hamilton, T. (1997). Synergy between interferon-gamma and tumor necrosis factor-alpha in transcriptional activation is mediated by cooperation between signal transducer and activator of transcription 1 and nuclear factor kappaB. *Journal of Biological Chemistry* **272**, 14899-14907.

Opitz, C., Di Cristina, M., Reiss, M., Ruppert, T., Crisanti, A. and Soldati, D. (2002). Intramembrane cleavage of microneme proteins at the surface of the apicomplexan parasite *Toxoplasma gondii*. *Embo Journal* **21**(7), 1577-1585.

Ortega-Barria, E. and Boothroyd, J. C. (1999). A *Toxoplasma* lectin-like activity specific for sulphated polysaccharides is involved in host cell infection. *Journal of Biological Chemistry* **274**(3), 1267-1276.

O'Shea-Greenfield, A. and Smale, S. T. (1992). Roles of TATA and initiator elements in determining the start site location and direction of RNA polymerase II transcription. *Journal of Biological Chemistry* **267**(2), 1391-1402.

Ossario, O. N., Schwartzman, J. and Boothroyd, J. (1992). A *Toxoplasma gondii* rhoptry protein associated with host cell penetration has unusual charge asymmetry. *Molecular and Biochemical Parasitology* **50**, 1-16.

Pancake, S. J., Holt, G., Mellouk, S. and Hoffman, S. (1992). Malaria sporozoites and circumsporozoite proteins bind specifically to sulphated glycoconjugates. *Journal of Cell Biology* **117**(6), 1351-1357.

Parvin, J. and Sharp, P. (1993). DNA topology and a minimal set of basal factors for transcription by RNA polymerase II. *Cell* **73**(3), 533-540.

Parvin, J. D., Timmers, H. T. and Sharp, P. A. (1992). Promoter specificity of basal transcription factors. *Molecular Biology of the cell* **68**, 1135-1144.

Pelham, H. (1982). A regulatory upstream promoter element in the Drosophila HSP 70 heat-shock gene. *Cell* **30**(2), 517-528.

Perkins, M. E. (1991). Approaches to study merozoite invasion of erythrocytes. *Research in Immunology* **142**(8), 662-665.

Perrin, L. H., Merkli, B., Gabra, M. S., Stocker, J. W. Chizzlini, C. and Richle, R. (1985). Immunisation with a *Plasmodium falciparum* merozoite surface antigen induces a partial immunity in monkeys. *Journal of Clinical Investigation* **75**(5), 1718-1721.

Petri Jr, W., Smith, R., Schesinger, P., Murphy, C. and Ravdin, J. (1987). Isolation of the galactose-binding lectin that mediates the *in vitro* adherence of *Entamoeba histolytica*. *Journal of Clinical Investigation* **80**, 1238-1244.

Pinder, J. C., Fowler, R.E., Dluzewski, A.R., Bannister, L.H., Lavin, G.H., Mitchell, R.J.M. and Gratzer, W.B. (1998). Actomyosin motor in the merozoites of the malaria parasite, *Plasmodium falciparum*: implications for red cell invasion. *Journal of Cell Science* **111**, 1831-1839.

Pizarro-Cerda, J., Meresse, S., Parton, R. G., van der Goot, G., Sola-Landa, A., Lopez-Goni, I., Moreno, E. and Gorvel, J. P. (1998). *Brucella abortus* transits through the autophagic pathway and replicates in the endoplasmic reticulum of nonprofessional phagocytes. *Infection and Immunity* **66**(12), 5711-5724.

Plesch, G., Ehrhardt, T. and Mueller-Roeber, B. (2001). Involvement of TAAG elements suggests a role for Dof transcription factors in guard cell-specific gene expression. *The Plant Journal* **28**(4), 455-464.

Porchet-Hennere, E. and Nicolas, G. (1983). Are rhoptries of Coccidia really extrusomes? *Journal of Ultrastructural Research* **84**(2), 194-203.

Poupel, O. and Tardieux, I. (1999). *Toxoplasma gondii* motility and host cell invasiveness are drastically impaired by jasplakinolide, a cyclic peptide stabilizing F-actin. *Microbes and Infection* **1**(9), 653-662.

Preston, T. M. and King, C. A. (1992). Evidence for the expression of actomyosin in the infective stage of the sporozoan protist *Eimeria*. *Cell Biology International Reports* **16**(4), 377-381.

Prurdy, J., Pho, L., Mann, B. and Petri Jr., W. (1996). Upstream regulatory elements controlling expression of the *Entamoeba histolytica* lectin. *Molecular and Biochemical Parasitology* **78**, 91-103.

Purnell, B., Emanuel, P. and Gilmour, D. (1994). TFIID sequence recognition of the initiator and sequences farther downstream in Drosophila class II genes. *Genes and Development* **8**, 830-842.

Quitschke, W., Matthew, J., Kraus, R. and Vostrov, A. (1996). The initiator element and proximal upstream sequences affect transcriptional activity and start site selection in the amyloid beta-protein precursor promoter. *Journal of Biological Chemistry* **271**(36) 22231-22239.

Quon, D. V., Delgadillo, M.G., Khachi, A., Smale, S.T. and Johnson, P.J. (1994). Similarity between a ubiquitous promoter element in an ancient eukaryote and mammalian initiator elements. *Proceedings in the National Academy of Sciences USA* **91**, 4579-4583.

Rabenau, K., Sohrabi, A., Tripathy, A., Reitter, C., Ajioka, J., Tomley, F. and Carruthers, V. (2001). TgM2AP participates in *Toxoplasma gondii* invasion of

host cells and is tightly associated with the adhesive protein TgMIC2. *Molecular Microbiology* **41**(3), 537-547.

Reiss, M., Viebig, N., Brecht, S., Fourmaux, M-N., Soete, M., Di Cristina, M., Dubremetz, J F. and Soldati, D. (2001). Identification and characterisation of an escorter for two secretory adhesins in *Toxoplasma gondii. The Journal of Cell Biology* **152**(3), 563-578.

Reithinger, R. (2002). RTS,S/AS02 and the quest for the Holy Grail. *Trends in Parasitology* **18**(5), 202.

Ridley, R. G., Takacs, B., Etlinger, H. and Scaife, J. G. (1990). A rhoptry antigen of *Plasmodium falciparum* is protective in saimiri monkeys. *Parasitology* **101**(2), 187-192.

Roberts, W., Speer, C. and Hammond, D. (1970). Electron and light microscope studies of the oocyst walls, sporocysts, and excysting sporozoites of *Eimeria callospermophili* and *E. larimerensis. Journal of Parasitology* **56**(5), 918-26.

Roberts, D., Craig, A., Berendt, A., Pinches, R., Nash, G., Marsh, K. and Newbold, C. (1992). Rapid switching to multiple antigenic and adhesive phenotypes in malaria. *Nature* **357**, 689-692.

Robson, K., Frevert, U., Reckmann, I., Cowan, G., Beier, J., Scragg, I., Takehara, K., Bishop, D., Pradel, G. and Sinden, R. (1995). Thrombospondin-related adhesive protein (TRAP) of *Plasmodium falciparum*: expression during sporozoite ontogeny and binding to human hepatocytes. *Embo Journal* **14**(16), 3883-3894.

Robson, K., Hall, J., Jennings, M., Harris, T., Marsh, K., Newbold, C., Tate, V. and Weatherall, D. (1988). A highly conserved amino-acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. *Nature* **335**(6185), 79-82.

Robson, K. J., Naitza, S., Barker, G., Sinden, R. E. and Crisanti, A. (1997). Cloning and expression of the thrombospondin related adhesive protein gene of *Plasmodium berghei. Molecular and Biochemical Parasitology* **84**(1), 1-12.

Rothman, J. and Orci, L. (1992). Molecular dissection of the secretory pathway. *Nature* **355**, 409-415.

Roy, A., Meisterernst, M., Pognonec, P. and Roeder, G. (1991). Cooperative interaction of an initiator-binding transcription initiation-factor and the helix loop helix activator USF. *Nature* **354**(6350), 245-248.

Rubio, J., Thompson, J. and Cowman, A. (1996). The var genes of *Plasmodium falciparum* are located in the subtelomeric region of most chromosomes. *Embo Journal* **15**, 4069-4077.

Russell, D. G. and Sinden, R. E. (1981). The role of the cytoskeleton in the motility of coccidian sporozoites. *Journal of Cell Science* **50**, 345-59.

Russell, D. G. and Sinden, R. E. (1982). Three-dimensional study of the intact cytoskeleton of coccidian sporozoites. *International Journal for Parasitology* **12**(2-3), 221-226.

Ryan, R., Shirley, M. and Tomley, F. (2000). Mapping and expression of microneme genes in *Eimeria tenella*. *International Journal for Parasitology* **30**, 1493-1499.

Ryming, F. W. and Remington, J. S. (1978). Effect of cytochalasin D on *Toxoplasma gondii* cell entry. *Infection and Immunity* **20**(3), 739-743.

Saffer, L., Mercereau-Puijalon, O., Dubremetz, J. and Schwartzman, J. (1992). Localisation of a *T. gondii* rhoptry protein by immunoelectronmicroscopy during and after host cell penetration. *Journal of Protozoology* **39**, 526-530.

Saliba, K., and Kirk, K. (2001). Nutrient acquisition by intracellular apicomplexan parasites: staying in for dinner. *International Journal for Parasitology* **31**(12), 1321-1330.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular cloning. A laboratory manual*, Cold Spring Harbour Laboratory Press.

Sam-Yellowe, T. Y. and Perkins, M. E. (1991). Interaction of the 140/130/110 Kda rhoptry protein complex of *Plasmodium falciparum* with the erythrocyte-membrane and liposomes. *Experimental Parasitology* **73**(2), 161-171.

Sawadogo, M. and Roeder, R. (1984). Energy requirement for specific transcription initiation by the human RNA polymerase II system. *Journal of Biological Chemistry* **259**(8), 5321-5326.

Schmatz, D., Crane, M. and Murray, P. (1984). Purification of *Eimeria* sporozoites by DE-52 anion exchange chromatography. *Journal of Protozoology* **31**(1), 181-183.

Schmatz, D., Crane, M. and Murray, P. (1986). *Eimeria tenella*: parasite-specific incorporation of ^{3}H-uracil as a quantitative measure of intracellular development. *Journal of Protozoology* **33**, 109-114.

Schofield, L., Bushell, G. R. Cooper, J. A., Saul, A. J., Upcroft, J. A. and Kidson, C. (1986). A rhoptry antigen of *Plasmodium falciparum* contains conserved and variable epitopes recognised by inhibitory monoclonal antibodies. *Molecular and Biochemical Parasitology* **18**(2), 183-195.

Schutze, M., Peterson, P. and Jackson, M. (1994). An N-terminal double-arginine motif maintains type II membrane proteins in the endoplasmic reticulum. *Embo Journal* **13**, 1696-1705.

Schwab, J. C., Beckers, C. J. and Joiner, K. A. (1994). The parasitophorous vacuole membrane surrounding intracellular *Toxoplasma gondii* functions as a molecular sieve. *Proceedings in the National Academy of Sciences USA* **91**(2), 509-513.

Schwab, J. C., Afifi Afifi, M., Pizzorno, G., Handschumacher, R. E. and Joiner, K. A. (1995). *Toxoplasma gondii* tachyzoites possess an unusual plasma membrane adenosine transporter. *Molecular and Biochemical Parasitology* **70**(1-2), 59-69.

Schwartzman, J. D. and Pfefferkorn, E. R. (1982). *Toxoplasma gondii*: purine synthesis and salvage in mutant host cells and parasites. *Experimental Parasitology* **53**, 77-86.

Schwartzman, J. (1986). Inhibition of a penetration enhancing factor of *Toxoplasma gondii* by a monoclonal antibodies specific for rhoptries. *Infection and Immunity* **51**(3), 760-764.

Schwarz, R. and Tomavo, S. (1993). The current status of the glycobiology of *T. gondii*: glycosylphosphatidylinositols, N- and O-linked glycans. *Research in Immunology* **144**, 24-31.

Sharma, P., Bharadwaji, A., Bhasin, V. K., Sailaja, V. N. and Chauhan, V. S. (1996). Antibodies to a conserved-motif peptide sequence of the *Plasmodium falciparum* thrombospondin-related anonymous protein and circumsporozoite protein recognise a 78-kilodalton protein in the asexual blood stages of the parasite and inhibit merozoite invasion *in vitro*. *Infection and Immunity* **64**, 2172-2179.

Shaw, M. K. and Tilney, L. G. (1995). The entry of *Thieleria parva* merozoites into bovine erythrocytes occurs by a process similar to sporozoite invasion of lymphocytes. *Parasitology* **111**(4), 455-461.

Shaw, M. K. (1997). The same but different: the biology of *Thieleria* sporozoite entry into bovine cells. *International Journal for Parasitology* **27**(5), 457-74.

Shaw, M. K., Roos, D. and Tilney, L. (1998). Acidic compartments and rhoptry formation in *Toxoplasma gondii*. *Parasitology* **117**(5) 435-443.

Shaw, M. K. and Tilney, L. G. (1999). Induction of an acrosomal process in *Toxoplasma gondii*: Visualisation of actin filaments in a protozoan parasite. *Proceedings in the National Academy of Sciences USA* **96**, 9095-9099.

Shiels, B., Fox, M., McKeller, S., Kinnard, J. and Swan, D. (2000). An upstream element of the TamS1 gene is a site of DNA-protein interactions during differentiation to the merozoite in *Thieleria annulata*. *Journal of Cell Science* **113**(12), 2243-2252.

Shirley, M., Bushell, A., Bushell, J., McDonald, V. and Roberts, B. (1995a). A live attenuated vaccine for the control of avian coccidiosis: trials in broiler breeders and replacement layer flocks in the United Kingdom. *Veterinary Research* **137**(18), 453-7.

Shirley, M. W. (1994). The genome of *Eimeria tenella*: further studies on its molecular organisation. *Parasitology Research* **80**(5), 366-373.

Shirley, M. W. (1995b). *Eimeria* species and strains of chickens. In *Biotechnology - Guidelines on techniques in coccidiosis research* (Eckert J, B. R., Shirley MW and Coudert P, ed.), pp. 1-24. The European Commission DGXII, Luxembourg.

Shirley , M. W. and Harvey, D. (1999). *Eimeria tenella*: Infection with a single sporocyst gives a clonal population. *Parasitology* **122**(6) 523-528.

Shirley, M. W. (2000). The genome of *Eimeria* spp., with special reference to *Eimeria tenella*- a coccidium from the chicken. *International Journal for Parasitology* **30**(4), 485-93.

Shirley, M. W., Kemp, D. J., Pallister, J. and Prowse, S. J. (1990). A molecular karyotype of *Eimeria tenella* as revealed by contour-clamped homogeneous electric field gel electrophoresis. *Molecular and Biochemical Parasitology* **38**(2), 169-173.

Sibley, L., Krahenbuhl, J. L., Adams, G. and Weidner, E. (1986). *Toxoplasma* modifies macrophage phagosomes by secretion of a vesicular network rich in surface proteins. *Journal of Cell Biology* **103**(3), 867-874.

Sibley, L., Niesman, I., Asai, T. and Takeuchi, T. (1994a). *Toxoplasma gondii-* secretion of a potent nucleoside triphosphate hydrolase into the parasitophorous vacuole. *Experimental Parasitology* **79**(3), 301-311.

Sibley, L. D., Messina, M. and Niesman, I. R. (1994b). Stable DNA transformation in the obligate intracellular parasite *Toxoplasma gondii* by complementation of tryptophan auxotrophy. *Proceedings in the National Academy of Sciences USA* **91**(12), 5508-5512.

Sibley, L. D., Niesman, I. R., Parmley, S. F. and Cesbrondelauw, M. F. (1995). Regulated secretion of multilamellar vesicles leads to formation of a tubulovesicular network in host cells vacuoles occupied by *Toxoplasma gondii*. *Journal of Cell Science* **108**(4), 1669-1677.

Sim, B. K. (1990). Sequence conservation of a functional domain of erythrocyte binding antigen 175 in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* **41**(2), 293-295.

Sim, B., Toyoshima, T., Haynes, J. and Aikawa, M. (1992). Localization of the 175-kilodalton erythrocyte binding antigen in micronemes of *Plasmodium falciparum* merozoites. *Molecular and Biochemical Parasitology* **51**(1), 157-159.

Sim, B. K., Chitnis, C. E., Wasniowska, K., Hadley, T. J. and Miller, L. H. (1994). Receptor and ligand domains for invasion of erythrocytes by *Plasmodium falciparum*. *Science* **264**(5167), 1941-1944.

Sinai, A. P., Webster, P. and Joiner, K. A. (1997). Association of host cell endoplasmic reticulum and mitochondria with the *Toxoplasma gondii* parasitophorous vacuole membrane: a high affinity interaction. *Journal of Cell Science* **110**(17), 2117-2128.

Singer, V., Yee, J. and Nash, T. (1998). Episomal and integrated maintenance of foreign DNA in *Giardia lamblia*. *Molecular and Biochemical Parasitology* **92**, 59-69.

Singer, V. L., Wobbe, C. R. and Struhl, K. (1990). A wide variety of DNA sequences can functionally replace a yeast TATA element for transcriptional activation. *Genes and Development* **4**, 636-645.

Singh, U., Rogers, J.B., Mann, B.J. and Petri, Jr. W.A. (1997). Transcription initiation is controlled by three core promoter elements in the hgl5 gene of the protozoan parasite *Entamoeba histolytica*. *Proceedings in the National Academy of Sciences USA* **94**, 8812-8817.

Singh, U. and Rogers, J. (1998). The novel core promoter element GAAC in the hgl5 gene of *Entamoeba histolytica* is able to direct a transcription start site independent of TATA or initiator regions. *The Journal of Biological Chemistry* **273**(34), 21663-21668.

Singh, U., Gilchrist, C., Schaenman, J., Rogers, J., Hockensmith, J., Mann, B. and Petri Jr, W. (2002). Context-dependent roles of the *Entamoeba histolytica* core promoter element GAAC in transcriptional activation and proteins complex assembly. *Molecular and Biochemical Parasitology* **120**, 107-116.

Sinnis, P. and Kim Lee Sim, B. (1997). Cell invasion by the vertebrate stages of *Plasmodium*. *Trends in Microbiology* **50**(2), 52-58.

Smale, S. and Baltimore, D. (1989). The "initiator" as a transcriptional control element. *Cell* **57**, 103-113.

Smale, S. T. (1994). Core promoter architecture for eukaryotic protein-coding genes. In *Transcription: mechanisms and regulation*. (R. C. Conaway and Conaway, J. W., ed.), pp. 63-81. Raven Press, New York.

Smale, S. T. (1997). Transcription initiation from TATA-less promoters within eukaryotic protein coding genes. *Biochimica et Biophysica Acta* **1351**, 73-88.

Smith, J., Chitnis, C., Craig, A., Roberts, D., Hudson-Taylor, P., Perterson, D., Pinches, R., Newbold, C. and Miller, L. (1995). Switches in expression of *Plasmodium falciparum* var genes correlated with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* **82**, 101-110.

Soldati, D. and Boothroyd, J. C. (1993). Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. *Science* **260**(5106), 349-352.

Soldati, D. and Boothroyd, J. C. (1995). A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Molecular and Cellular Biology* **15**(1), 87-93.

Soldati, D., Kim, K., Kampmeier, J., Dubremetz, J. F. and Boothroyd, J. C. (1995). Complementation of a *Toxoplasma gondii* ROP1 knock-out mutant using phleomycin selection. *Molecular and Biochemical Parasitology* **74**(1), 87-97.

Soldati, D., Dubremetz, J-F. and Leburn, M. (2001). Microneme proteins: structural and functional requirements to promote adhesion and invasion by the apicomplexan parasite *Toxoplasma gondii*. *International Journal for Parasitology* **31**, 1293-1302.

Soldati, D., Lassen, A., Dubremetz, J-F. and Boothroyd, J. (1998). Processing of Toxoplasma ROP1 protein in nascent rhoptries. *Molecular and Biochemical Parasitology* **96**, 37-48.

Sonda, S., Fuchs, N., Gottstein, B. and Hemphill, A. (2000). Molecular characterisation of a novel microneme antigen in *Neospora caninum*. *Molecular and Biochemical Parasitology* **108**, 39-51.

Spaccapelo, R., Naitza, S. and Robson, K.J. (1997). Thrombospondin-related adhesive protein (TRAP) of *Plasmodium falciparum* and parasite motility. *Lancet* **350**, 355.

Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research* **10**(15), 4731-4751.

Stedmann, T. and Joiner, K. (2000). *Advances in cell and molecular biology of membranes and organelles.* En route to the vacuole: tracing the secretory pathway of *Toxoplasma gondii*. (Gordon, S., Ed.), 5, JAI Press, Greenwich.

Stewart, M. J. and Vanderberg, J. P. (1988). Malaria sporozoites leave behind trails of circumsporozoite protein during gliding motility. *Journal of Protozoology* **35**(3), 389-393.

Stewart, M. J. and Vanderberg, J. P. (1991). Malaria sporozoites release circumsporozoite protein from their apical end and translocate it along their surface. *Journal of Protozoology* **38**(4), 411-421.

Striepen, B., Soldati, D., Garcia-Reguet, N., Dubremetz, J. F. and Roos, D. (2001). Targeting of soluble proteins to the rhoptries and micronemes in *Toxoplasma gondii*. *Molecular and Biochemical Parasitology* **113**(1), 45-53.

Strout, R. G., Alroy, J., Lukacs, N. W., Ward, H. D. and Pereira, M. E. (1994). Developmentally regulated lectins in *Eimeria* species and their role in avain coccidiosis. *Journal of Parasitology* **80**(6), 946-951.

Su, X. and Wellems, T. (1994). Sequence, transcript characterisation and polymorphisms of a *Plasmodium falciparum* gene belonging to the heat-shock protein (HSP) 90 family. *Gene* **151**, 225-230.

Sultan, A. A., Thathy, V., Frevert, U., Robson, K. J., Crisanti, A., Nussenzweig, V., Nussenzweig, R. S. and Menard, R. (1997). TRAP is necessary for gliding motility and infectivity of *Plasmodium* sporozoites. *Cell* **90**(3), 511-522.

Sun, G. and Tai, J. (1999). Identification and characterisation of a ran gene promoter in the protozoan pathogen *Giardia lamblia*. *Journal of Biological Chemistry* **274**, 19699-19706.

Suss-Toby, E., Zimmerberg, J. and Ward, G. E. (1996). *Toxoplasma* invasion: the parasitophorous vacuole is formed from host cell plasma membrane and pinches off via a fission pore. *Proceedings in the National Academy of Sciences USA* **93**(16), 8413-8418.

Sutton, C., Shirley, M. and Wisher, M. (1989). Characterization of coccidial proteins by two-dimensional sodium dodecyl sulphate-polyacrylamide gel electrophoresis. *Parasitology* **99**(2), 175-187.

Swanson, M. S. and Isberg, R. R. (1995). Association of *Legionella pneumophila* with the macrophage endoplasmic reticulum. *Infection and Immunity* **63**(9), 3609-3620.

Tabor, S. and Richardson, C. (1995). A Single Residue in DNA Polymerases of the *Escherichia coli* DNA Polymerase I Family is Critical for Distinguishing Between Deoxy- and Dideoxyribonucleotides. *Proceedings in the National Academy of Sciences USA* **92**, 6339-6343.

Talkington, C. A. and Leder, P. (1982). Rescuing the *in vitro* function of a globin pseudogene promoter. *Nature* **298**(5870) 192-195.

Templeton, T. J. and Kaslow, D. C. (1997). Cloning and cross-species comparison of the thrombospondin-related anonymous protein (TRAP) gene from *Plasmodium knowlesi*, *Plasmodium vivax* and *Plasmodium gallinaceum*. *Molecular and Biochemical Parasitology* **84**(1), 13-24.

Tetley, L., Brown, S., McDonald, V. and Coombs, G. (1998). Ultrastructural analysis of the sporozoite of *Cryptosporidium parvum*. *Microbiology* **144**(12), 3249-3255.

Thastrup, O., Dawson, A., Scharff, O., Foder, B., Cullen, P., Frobak, B., Bferrum, P.J., Christensen, S. and Hanley, M. (1989). Thapsigargin, a novel molecular probe for studing intracellular calcium release and storage. *Agents actions*. **27**, 17-23.

Tomavo, S., Fortier, B., Soete, M., Ansel, C., Camus, D. and Dubremetz, J. (1991). Characterisation of bradyzoite-specific antigens of *Toxoplasma gondii*. *Infection and Immunity* **59**(10) 3750-3753.

Tomley, F. M., Clarke, L. E., Kawazoe, U., Dijkema, R. and Kok, J. J. (1991). Sequence of the gene encoding an immunodominant microneme protein of *Eimeria tenella*. *Molecular and Biochemical Parasitology* **49**(2), 277-288.

Tomley, F. (1994a). Antigenic diversity of the asexual developmental stages of *Eimeria tenella*. *Parasite Immunology* **16**(8), 407-413.

Tomley, F. M. (1994b). Characterization of rhoptry proteins of *Eimeria tenella* sporozoites: antigenic diversity of rhoptry epitopes within species of the genus *Eimeria* and among three asexual generations of a single species, *E. tenella*. *Infection and Immunity* **62**(10), 4656-4658.

Tomley, F. M., Bumstead, J. M., Billington, K. J., Dunn, P. P. (1996). Molecular cloning and characterization of a novel acidic microneme protein (Etmic-2)

from the apicomplexan protozoan parasite, *Eimeria tenella. Molecular and Biochemical Parasitology* **79**(2), 195-206.

Tomley, F. and Soldati, D. (2000). Mix and Match Modules: Structure and Function of Microneme Proteins in Apicomplexan Parasites. *Parasitology Today* **17**(2), 81-88.

Tomley, F. M., Billington, K. J., Bumstead, J. M., Clark, J. D. and Monaghan, P. (2001). EtMIC4: a microneme protein from *Eimeria tenella* that contains tandem arrays of epidermal growth factor-like repeats and thrombospondin type-1 repeats. *International Journal for Parasitology* **31**(12), 1303-1310.

Torii, M., Adams, J. H., Miller, L. H. and Aikawa, M. (1989). Release of merozoite dense granules during erythrocyte invasion by *Plasmodium knowlesi. Infection and Immunity* **57**(10), 3230-3233.

Trottein, F., Triglia, T. and Cowman, A. F. (1995). Molecular cloning of a gene from *Plasmodium falciparum* that codes for a protein sharing motifs found in adhesive molecules from mammals and *Plasmodia. Molecular and Biochemical Parasitology* **74**(2), 129-141.

Tschudi, C. and Ullu, E. (1988). Polygene transcriptis are precursors to calmodulin mRNAs in trypanosomes. *Embo J* **7**, 455-463.

Tyree, C., George, C., Liradevito, L., Wampler, S., Dahmus, M., Zawel, L. and Kadonaga, J. (1993). Identification of a minimal set of proteins that is sufficient for accurate initiation of transcription by RNA polymerase II. *Genes and development* **7**(7A), 1254-1265.

Tyzzer, E. E. (1929). Coccidiosis in gallinaceous birds. *American Journal of Hyginene* **10**, 269-383.

Van den Boogaart et al. (1998). *Coccidiosis poultry vaccine.*

Van der Hoff, J., Moorman, F. and Lamers, W. (1992). Electroporation in "intracellular" buffers increases cell survival. *Nucleic Acids Research* **20**, 2902.

Van Dijk, M. R., Janse, C. J. and Waters, A. P. (1996). Expression of a *Plasmodium* gene introduced into subtelomeric regions of *Plasmodium berghei* chromosomes. *Science* **271**, 662-664.

Van Dijk, M. R., Waters, A. P. and Janse, C. J. (1995). Stable transfection of malaria parasite blood stages. *Science* **268**(5215), 1358-62.

Van Doornink, W. M. and Becker, E. R. (1957). Transport of sporozoites of *Eimeria necatrix* in macrophages. *Journal of Parasitology* **43**, 40-44.

Vasconcelos, E., Ferreira, S., deCarvalho, T., deSouza, W., Kettlun, A., Mancilla, M., Valenzuela, M. and Verjovski-Almeida, S. (1996). Partial purification and immunohistochemical localization of ATP diphosphohydrolase from *Schistosoma mansoni* - Immunological cross-reactivities with potato apyrase and *Toxoplasma gondii* nucleoside triphosphate hydrolase. *Journal of Biological Chemistry* **271**(36), 22139-22145.

Vervelde, L., Vermeulen, A., and Jeurissen, S. (1995). *Eimeria tenella* - sporozoites rarely enter leukocytes in the cecal epithelium of the chicken (*Gallus domesticus*). *Experimental Parasitology* **81**(1), 29-38.

Vieira, M. C. F. and Moreno, S. N. J. (2000). Mobilisation of intracellular calcium upon attachment of *Toxoplasma gondii* tachyzoites to human fibroblasts is required for invasion. *Molecular and Biochemical Parasitology* **106**, 157-162.

Vines, R., Ramakrishnan, G., Rogers, J., Lockhart, L., Mann, B. and Petri Jr., W. (1998). Regulation of adherence and virulence by the *Entamoeba histolytica* lectin cytoplasmic domain, which contains a beat2 integrin motif. *Molecular and Cellular Biology* **9**, 2069-2979.

Waller, R. F., Keeling, P. J., Donald, R. G., Striepen, B., Handman, E., Lang-Unnasch, N., Cowman, A. F., Besra, G. S., Roos, D. S. and McFadden, G. I. (1998). Nuclear-encoded proteins target to the plastid in *Toxoplasma gondii* and *Plasmodium falciparum*. *Proceedings in the National Academy of Sciences USA* **95**(21), 12352-12357.

Wan, K. L., Carruthers, V. B., Sibley, L. D. and Ajioka, J. W. (1997). Molecular characterisation of an expressed sequence tag locus of *Toxoplasma gondii* encoding the micronemal protein MIC2. *Molecular and Biochemical Parasitology* **84**(2), 203-214.

Wang, C. C. and Simashkevich, P. M. (1981). Purine metabolism in the protozoan parasite *Eimeria tenella*. *Proceedings in the National Academy of Sciences USA* **78**(11), 6618-6622.

Wang, C. C. and Stotish, R. L. (1975). Changes of nucleic acids and proteins in the oocysts of *Eimeria tenella* during sporulation. *Journal of Protozoology* **22**, 438-443.

Ward, G. E., Miller, L. H. and Dvorak, J. A. (1993). The origin of parasitophorous vacuole membrane lipids in malaria-infected erythrocytes. *Journal of Cell Science* **106**(1), 237-248.

Wefald, F. C., Devlin, B. H. and Williams, R. S. (1990). Functional heterogeneity of mammalian TATA-box sequences revealed by interaction with a cell-specific enhancer. *Nature* **344**, 260-262.

Weiher, H., Konig, M. and Gruss, P. (1983). Multiple point mutations affecting the Simian virus 40 enhancer. *Science* **219**, 626-631.

Weiss, W. R., Mellouk, S., Houghten, R. A., Sedegah, M., Kumar, S., Good, M. F., Berzofsky, J. A., Miller, L. H. and Hoffman, S. L. (1990). Cytotoxic T-cells recognise a peptide from the circumsporozoite protein on malaria infected hepatocytes. *Journal of Experimental Medicine* **171**(3), 763-773.

Wengelnik, K., Spaccapelo, R., Naitza, S., Robson, K. J., Janse, C. J., Bistoni, F., Waters, A. P. and Crisanti, A. (1999). The A-domain and the thrombospondin-related motif of *Plasmodium falciparum* TRAP are implicated in the invasion process of mosquito salivary glands. *Embo Journal* **18**(19), 5195-5204.

Werner-Meier, R. and Entzeroth, R. (1997). Diffusion of microinjected markers across the parasitophorous vacuole membrane in cells infected with *Eimeria nieschulzi* (Coccidia, Apicomplexa). *Parasitology Research* **83**(6), 611-613.

Wiley, S., Kraus, R. and Mertz, J. (1992). Functional binding of the TATA box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. *Proceedings in the National Academy of Sciences USA* **89**, 5814-5818.

Williams, R. B. (1999). A compartmentalised model for the estimation of the cost of coccidiosis to the world's chicken production industry. *International Journal for Parasitology* **29**, 1209-1229.

Wilson, R. J. M., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W. and Williamson, D. H.

(1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *Journal of Molecular Biology* **261**(2), 155-172.

Wu, Y., Sifri, C. D., Lei, H. H., Su, X. Z. and Wellems, T. E. (1995). Transfection of *Plasmodium falciparum* within human red blood cells. *Proceedings in the National Academy of Sciences USA* **92**(4), 973-977.

Yanagisawa, S. (1995). A novel DNA binding domain that may form a single zinc finger motif. *Nucleic Acids Research* **23**, 3403-3410.

Yanagisawa, S. and Schmidt, R. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. *Plant Journal* **17**, 209-214.

Yuda, M., Sakaida, H., and Chinzei, Y. (1999a). Targeted disruption of the *Plasmodium berghei* CTRP gene reveals its essential role in malaria infection of the vector mosquito. *Journal of Experimental Medicine* **190**(11), 1711-1715.

Yuda, M., Sawai, T. and Chinzei, Y. (1999b). Structure and expression of an adhesive protein-like molecular of mosquito invasive-stage malarial parasite. *Journal of Experimental Medicine* **189**(12), 1947-1952.

Zawel, L. and Reinberg, D. (1993). Initiation of transcription by RNA polymerase II- a multistep process. *Proceedings in the National Academy of Sciences USA* **44**, 67-108.

Zenzie-Gregory, B., Khachi, A., Garraway, I.P. and Smale, S.T. (1993). Mechanism of initiator-mediated transcription: Evidence for a functional interaction between the TATA-binding protein and DNA in the absence of a specific recognition sequence. *Molecular and Cellular Biology* **13**(7), 3841-3849.

Zenzie-Gregory, B., O'Shea-Greenfeild, A. and Smale, S.T. (1992). Similar mechanisms for transcription initiation mediated through a TATA box or an initiatior element. *Molecular and Cellular Biology* **267**, 2823-2830

Appendix A

Alkaline phosphate buffer (x 1)

Tris	100 mM
NaCl	100 mM
MgCl$_2$	10 mM

Adjusted to pH9.5 with HCl.

Alkaline phosphate visualisation mixture

Alkaline phosphate buffer (x 1)	100 ml
BCIP	330 μl
NBT	660 μl

Antibiotic stock solutions
Ampicillin (1000x)

Ampicillin	100 mg
H$_2$O	1 ml

Filtered sterilised and stored at either –20°C or at 4°C for a maximum of 3 weeks.

Kanamycin (333x)

Kanamycin	10 mg
H$_2$O	1 ml

Filter sterilised and stored in aliquots at –20^{0}C.

Tetracycline (400x)

Tetracycline	5 mg
Ethanol 80% (v/v)	1 ml

Filter sterilized and stored at –20°C.

AMV Primer extension buffer, 2x (Promega)

Tris-HCl	100 mM
MgCl$_2$	20 mM
KCl	100 mM
DTT	20 mM
dNTP	2 mM
Spermidine	1 mM

Adjusted to pH 8.3

BCIP (5-bromo, 4-chloro, 3-indolyl phosphate)

BCIP (Sigma)	0.5 g
Di-methly formamide	10 ml

Beta-galactosidase solution:

$MgCl_2$,	1 mM
NaCl,	150 mM
$K_4Fe(CN)_6.3H_2O$,	3.3 mM
$K_3Fe(CN)_6$,	3.3 mM
Na_2HPO_4,	60 mM
NaH_2PO_4	40 mM

Beta-galactosidase staining solution

2% Xgal in dimethyl formamide	10 ml
Yellow solution	90 ml

Bjerrum/Schafer-Niehran transfer buffer

Tris	48 mM
Gylcine	39 mM
Methanol	20 %
Adjusted to pH 9.0	

cDNA synthesis buffer, 5x (Boehringer-Mannheim)

Tris-HCl	250 mM
$MgCl_2$	40 mM
KCl	150 mM
Dithiothreitol	5 mM
Adjusted to pH 8.5	

Coomassie brillant blue (CBB) stain

	0.25 g
Coomassie brillant blue R-250 dye (Sigma)	90 ml
Methanol:H_2O (1:1v/v)	10 ml
Glacial acetic acid	
Filtered through No.1 Whatman paper.	

Cytomix (incomplete):

K_2HPO_4/KH_2PO_4 Ph 7.6,	10 mM
KCL,	120 mM
$CaCl_2$,	0.15 ml
HEPES,	25 mM
EGTA,	2 mM
$MgCl_2$	5 mM

Cytomix (complete):

10ml incomplete cytomix.

ATP	2 mM
Gluthathione	5 mM

Destaining soln.

Methanol	50.0 %
Acetic acid	16.7 %
H_2O	33.3 %

DTT (DL-Dithiothreitol; Sigma)

2.35M in 400mM KAc, pH6.0

Hybridisation solution

A: Phosphate buffer	Na₂HPO₄	134 g/l
	85% (v/v) H₃PO₄	4 ml/l
	H₂O	
B: SDS solution	10% NaDodSO₄	

IPTG stock solution (Sigma)

IPTG	23.2mg
H₂O	1.0ml

Filter steilised and stored at –20°C

LB-agar

Bacto-tryptone	10 g/l
Bacto-yeast extract	5 g/l
NaCl	10 g/l
Agar	15 g/l
H₂O	

Autoclaved for sterilisation.

LB-broth

Bacto-tryptone	10 g/l
Bacto-yeast-extract	5 g/l
NaCl	10 g/l
H₂O	

Autoclaved for sterilisation.

Loading dye

Formamide	98 %
EDTA	10 mM
Xylene cyanol	0.1 %
Bromophenol blue	0.1 %

Lysis buffer

100 mM K₂HPO₄	9.15 ml
100 mM KH₂PO₄	0.85 ml
1 M DTT	10 µl

MgKCl

MgSO₄	1 M
MgCl₂	1 M
KCl	0.25 M

NBT (Sigma)

Nitro blue tetrazolium	0.5 g
Dimethyl formamide (70%)	10 ml

Lauroylsarcosine buffer

EDTA (pH 9.5)	0.5 M
Tris-HCl (pH 9.5)	10 mM
N-lauroylsarosine	1 %

PBS (pH 7.4, 7.6 or 8.0)

Na_2HPO_4	1.44g/l
KH_2PO_4	0.24g/l
KCl	0.20g/l
NaCl	8.00g/l
H_2O	

Adjusted to the required pH with HCl and autoclaved for sterilisation.

Protease inhibitor cocktail   (Sigma P2714)

Contains 4-(2-aminoethyl)benzenesulfonyl fluoride, trans-epoxysuccinyl-L-leucyl amido(4-guanidino), butane (E-64), bestatin, leupeptin, aprotinin and sodium EDTA.

RNA extraction solutions (Gentra Systems, provided by Flowgen)
Cell lysis buffer: citric acid, ethylenediaminetetraacetic acid, SDS
Protein-DNA precipitation buffer: Citirc acid, sodium chloride
(concentrations not supplied by Gentra)

React 4 (Gibco BRL)

Tris-HCl	20 mM
$MgCl_2$	5 mM
KCl	50 mM
Adjusted to pH 7.4	

Reaction buffer, 10x (Boehringer Mannheim)

Tris-HCl	100 mM
$MgCl_2$	15 mM
KCl	500 mM
Adjusted to pH 8.3	

Running buffer x 5 (protein gels)

Tris base	0.125 mM
Glycine	0.96 mM
Sodium dodecyl sulphate	0.5 %
H_2O	

SDS-PAGE   (10% acrylamide gel)

H_2O	1.90ml
30% acrylamide mix (BioRad)	1.70ml
1.5M Tris (pH8.8)	1.30ml
10% SDS	0.05ml
10% ammonium persulphate	0.05ml
TEMED	0.002ml

SDS-PAGE   (5% stacking gel)

H_2O	0.68ml
30% acrylamide mix	0.17ml
1.0M Tris (pH6.8)	0.13ml
10% SDS	0,01ml
10% ammonium persulphate	0.01ml
TEMED	0.001ml

SOC medium

LB medium	10 ml
$MgCl_2$	10 mM
$MgSO_4$	10 mM
KCl	2.5 mM
Glucose	20 mM

Southern blotting solutions
Denaturing buffer:

NaCl	1.7 M
NaOH	0.5 M
Ammonium acetate	10 M

Neutralising buffer:

NaOH	5 mM

2 x SSC

NaCl	300 mM
Tri-NaCitrate.$2H_2O$	3 mM

Adjust to pH 7.0.

TBE buffer x10 (Tris-Borate-EDTA)

Tris base	108 g/l
Boric acid	55 g/l
0.5M EDTA (pH8.0)	20 ml/l
H_2O	

TFB (transformation buffer)

K-MES (pH 6.2-6.3)	10 mM
RbCl or KCl	100 mM
$MnCl_2$	45 mM
$CaCl_2$	10 mM
$HACoCl_2$	3 mM

T4 Polynucleotide kniase buffer, 10x (Promega)

Tris-HCl	500 mM
$MgCl_2$	100 mM
DTT	50 mM
Spermidine	1 mM

Adjusted to pH 7.5

X-Gal stock solution

X-Gal (Sigma)	20 mg
Dimethyl formamide	1.0 ml

Bacterial strains

XL1-Blue MRF': Obtained form stratagene, has a genotype of Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA1 lac* [F' *proAB lacI^q* ZΔM15 *Tn10* (Tet^r)]^c. Used for the propagation of high efficiency cloning and antibiotic selection to allow expression of F' episome.

XL1-Blue MR: Obtained form stratagene has a genotype of Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA96 gyrA1 lac* and was used for the propogation of SuperCos1 vector (Stratagene).

Antibodies

Micronemes were prepared from freshly excysted, purified sporozoites by sonication and sucrose density gradient ultra-centrifugation as described (Tomley, 1997). Microneme proteins were separated by one and two dimensional SDS-PAGE, visualised by staining in aqueous CBB and harvested by electro-elution. Hyperimmune sera were prepared in rabbits against five microneme proteins, designated EtMIC1-5.

```
-1191  TTGCCCTTAA  AACCAAATGC  TTTCCCTTTT  TTTTCAAGGT  ACTATTGGAA
                        mic2rr5.5 ➤
-1141  TTTTTGGTTT  GACACAAAGT  CCAGTCACAC  GTTAAGTGGA  ACCAGACTTA
-1091  TAACCAGCTT  GGCGAACCGG  CCCCCGTTT   TTGGTGGTGG  GTTACGCTCA
-1041  ATTGCAGTTC  CCAAGCTAGC  TTGTTACTTC  ACCACCAAAA  TGACCACTTA
 -991  TGTTTTAAAA  GGAGGCTTGT  GTCGCTCTCT  TGATGGAACT  GGCAGTGGGC
 -941  TAGTGCCAAC  GAGCGAACGG  TGCCCTTCCA  TTGGTCGTGT  GTTCGACAAT
 -891  CCCTGCAGTT  GCTCCCGGTC  CAGCGTCTAG  CGGTCTGACA  GCTATGCGCA
 -841  ATATTCATGC  AGTTATTATA  CCCATCTGAA  TCCTTTGCAC  GTGCTTTAGC
 -791  GTTTATAATT  CAGAGTATTT  GCTTCTGGCG  CACATTTGTC  TCATCCCAGT
 -741  CGTGTGCCGA  CATCAATGCC  ACAGCGGCCG  CTCAATGCAG  CATTTGTATC
 -691  AGGCTATTAT  TAAGTGAAGC  ACCCCCTGAA  TTTTTCTTAT  GCGCATCATG
 -641  TAAAGCAGGA  TGGAGCCCCG  CACTATAAAT  AGCCTTTCTT  TCTCGGCACG
 -591  GCGATCATTT  TATCAAGTGC  TTCGGAATGT  TACATGGCTG  CCACTAGAAA
 -541  TAATTAACAG  CAAACGAATG  CGGTTAATGA  CTGTTAAATA  AGTGATGATA
                                          mic2rr5.6 ◄
 -491  GTGCTTTTTC  TGCTTTGTGC  CACAGGAAGA  CGGTTAAGTA  CAGCTGCTCC
 -441  AGCGGCCCAA  CGTGTGTGTC  TTGGAGCCAA  CCAGGAGCCA  ATCAGCACGT
 -391  AGAATAAAAA  GCGCAACGAA  TTCTCGTATG  CTTACCAACA  TTTCATAATA
 -341  GTCTCTTGCT  TACTGTGAGC  TTTTATACCA  GGAACTGCTT  AGAGTTGGCT
 -291  TGCAAGCCCG  TACGTTCAGT  GAAGCTGCTG  TTGCCATCGT  GGTCAGATGG
 -241  GGTCTCCGGT  GCGCGCCTCC  CAGAAAGCA   AGAAAGGCTC  CCCAAAAAT
 -191  TAATGGCTTC  ATACTGGCTC  GACAATGTAC  AATACATTAT  AGTCATATTA
 -141  CAGAATGCTA  AAAGGAGAAA  TAAAAACGCG  TGATTCCTTA  GTGCAATTGT
  -91  GCTAGCCACT  GACTGTGCCT  CTAGCGGACC  GCAAACCTGC  ACAACCTGT
                                                      +1  +1
  -41  GTGCTTTCCA  ACTCAAGGTC  CCATCATTTG  CAGCAACGAA  GATTACTCCT
  +10  TGCTTTTTGG  CAATTGTTTG  GCTTTTAAAA  AGTATTTTTT  CCTTTGTTGC
  +60  ATTGCATAAC  CTCATTTCTC  TTTGTATTCA  CATTCAAAAT G
```

Fig. B.2 Upstream genomic sequence of *EtMIC2*

The *EtMIC2* genomic clone was identified by PCR using primers T3/T7 and mic2rr5. The genomic clones were sequenced using primers mic2rr5, mic2rr5.6 and mic2rr5.5 and are shown in red type, with an arrow indicating the direction of sequencing. Degenerate CAAT boxes (at +20 and -98) and a single degenerate TATA box (at −788) are shown underlined and boxed in, respectively. The initiator element with bases

```
-924   CTAGAGAATG ATCAACTGGA AGTTCGGCAT GAGTTCAAGG ACATGACGTA

-874   GAGCTCGCAA TTCGTACTGC GAAGGCGTGT CAAATTAGTC AGAAGTTGAA

-824   TCCATACAAA TACCTACACA ACTTTCCACA AAATAACTAA GCTCAATATG

-774   CGTAGCCGCC CTTAATAGCC GGTAGCGCTC TCGTTAAATG CTCCCCAAAG

-724   CGCTGAGTTC ATCAGGGCTG CCATCCAACG CCAAAACAAG CTGGAAACGC

-674   CAGCACGAGA TAGGAGAGGA AAGCTTAAGC AAGACTACAG TGAACGCGTA
                        mic1rr5◄─────

-624   TACACATGAT GAGTCTTGCC CGGAGCCTTG CTAAAACAAA CGAGTTCGCT

-574   ACGTCGATTG CATCATATAC ACCTCATACT AACCAGTCCA GATAATGAAC

-524   ATTCGCAATC AAACTTACAA CCAAACTTTT GAAGCATTGT GCGAACAAGA

-474   ACTAATTTGC CTCTCGAAGG ATCTGAATGC AAGCAAACAT AACTATTCCT

-424   TGTGTGCGAA TCGACATATG CACACAGCGC AGAGACCGTC ACTTTAAACC

-374   ATGCTTGGAG GAAACTTTGC TTTTCCATGT CTAATGGCAA AAAGCAAGTT

-324   GGAATAACAC TCCTCAATAT CTAGCTACGG ATATTCTACA CTATCAGAGC

-274   CCGATACGTG TGTGAAGGCG AGCGAAGCTG TTCACTCGTG GCTTCAGCTG

-224   GAAAAATAAT CGACAATGCG CTAGACGCGG GAGTCGAAGC GTGTCTGTAT

-174   GCTGTCGTGA GTAAACCCAC AGGAAATCGC TTAATAAGAG CTGAAGAAAG

-124   CAATAACAAT ATACGTGATC TGCCTGACCT TGCAGATCCC CGGGAACTAC

 -74   CCACGTCATT GCGCGGCCAG CCGCGAGTAC GCACTAGCAG GACCCTTGCC
                                +1

 -24   GAGCCATCCG ACTTCAATCG TCTCACAGTA GAATTCTGGT TAGGTTTCAA

 +27   AACATATTGT TTTGTTAGGT CATTTCTTA ATCTCCTATT CTTCATCTTT

 +77   TTACCAGTT CCGCTTTCTT GCATCATTCC CGCAATG
```

Fig. B.1 Upstream genomic sequence of *EtMIC1*

The *EtMIC1* genomic clone was identified by PCR using primers T3/T7 and mic1rr5. The genomic clones were sequenced using primers mic1rr5, mic1rr5.7, mic1rr5.8 and mic1rr5.10 and are shown in red type, with an arrow indicating the direction of sequencing. Multiple degenerate CAAT boxes (between –6 and –310 bp) and a single degenerate TATA box (at –144) are shown underlined and boxed in, respectively. The initiator element with bases confirming to the PyPyA(+1)NA/TPyPy (Smale and Baltimore, 1989; Javahery *et al.*, 1994) is shown boxed in with a double line and the predicted transcription start site (A) is shown as +1. The sequence is numbered from the beginning of the newly identified transcription start site and the cDNA sequence previously identified from clone cDNA10 by Karen Billington is shown in blue type (Tomley *et. al*, 1991). The translation initiating codon is in bold type. A GAAC motif observed in *Entamoeba histoltica* is shown with a turquoise background (-81). The GGAGAGG motif which has been found in the upstream regions of *NTPase 1* and *3* of *T. gondii* (Nakaar *et al.*, 1992) is shown with a pink background at positions -1244 and -662. Motifs, similar to the A/TGAGAGC motif found in the upstream region of *SAG1* (Soldati and Boothroyd, 1995) and the AGAGACGC motif found in the promoter region of *NTPase 1* and *3* of *T. gondii* (Nakaar *et al.*, 1992), are shown with a blue background at positions –394 (direct orientation) and –6 (inverted orientation). The CACACA (-410) or GTGTGT (-1141) motifs defined in the upstream region of *T. annulata* gene, *TamS1*, are shown with a yellow background (Shiels *et al.*, 2000).

Appendix B

```
-2374   AAATCGATTG TTTTACAAGC TCCGGGCTAA ACGGCCAGAC AGAGTGTTAA
                     mic1rr5.10 ——▶
-2324   ATGACCGACT GCATTTGCAG GGCCTCCACG CAGCTATGTA ATAAGGCGAA
-2274   GTCTCAGTTC GACGAACGGT GATATCTGGG AGGGCAGTTT ACGCGGCACT
-2224   AGTCGTTGCT TCAAAGAACA CGCGCATTGG AGAAGCCTTG GTTGCCAATG
-2174   GTAATTGAAA GCGAGCGAAG GCTGCAAATA GCTACCATTT TCCTTGTATG
-2124   TAGAGAAAGC CCCTTCGTCG TAATGCGAAG CCACTAAGAT AACGCCCTTG
-2074   AGACCGACAC ATAGTTGTGT TTATGTTATC TAAACAACGA GCGTTCTGTT
-2024   CTGCCGTTCA TAATTGAGCT GAGGCGATCA CAGTACAACA GGCTCCCGAT
-1974   GAGCGGCTGG GAACGAACGA GCTGGACTCC ATGTGGCCTG GCAGCGCTCT
-1924   GACCAAAATC AAAAACGTGA CGGGAGTTCC AGTCGTGATC TTGTACTTTC
-1874   CACCACGCCG AGGCGCAGCC AGGTGCCGTG ACAATGCAAC CGTTACCCAA
-1824   TGCACTGAGC GCGCAGGATT GTAATGAAAA AGGAAAATCG TACAAAATTC
-1774   AACTAAGCAA TTTGAAGACT TGTCTGCTAG TTGTGCGCAG GAATCCGTGG
-1724   GGGAGAAGAA GCTCTCCGAG TCACTTCCTC TGCTCCTCGC TGCAGAGCGA
-1674   CGGTTTTGAG AACAGCAGCG ACAATGTAAC AAGCAGAAGC TAAGCGCAGA
-1624   GTGAGGGACC TGACGTGGCA AATTGTGTCC ATTTCTGGAA AGCCGGCGCT
                mic1rr5.8 ◀——
-1574   TGCTTATATA ACGCAACCAA TTAGTTTTGC CGGAGAGTGC CTGGGACCTT
-1524   CTATCTACTC GCGCTGCTTA GTTGGGCAGT ACAGGGTAAC AGGCAGCAGC
-1474   AAAGGGCCTT ACCTTTCTGG GGCCAGTTTT ACCAGGGACC TTGCGCATCC
-1424   AGTAAGAAGC CACCTGCGTT GCGCTTCAAA ATGGTGCAAG TCCCTTTCGT
-1374   TCACTGAAAA CCTTTTGTCG TGAGTCAAGA AAACAGCACC CACTTATGTA
-1324   ATTACTAAGG AGGGGCAAAT TCGGAGCAAA CAATACGCTG CAGCAATTGT
-1274   GCTCACACGC TTCTTAGCTG GTACTAGAAT GGAGAGGACG TGACCGAACT
-1224   GTGGAGCAAA AGAAATTGAC TGAACCCCGT TAAAACCTCA TGTAGCTGAA
-1174   GGGAAGTGGA AGAAGACAGA AGGAACAAAT GGAGTGTGTT TGAACACCTC
-1124   AAGCTCATAC TGAGTAGCTC ACAGACCCAG TCGCAAAGGT TTGATAGAGG
-1074   AGCCAGCTAG AGAGTCCACA GGATTCATTC TGAATGATGC CGTCAATACT
                mic1rr5.7 ◀——
-1024   GGAAGCGTTC CTAATAAGAG GACGCCGAGG CCCCTGGGTC CACTGTTGAC
 -974   CACGCCCCAC CATTCAAATC CAGTAAAAAT TACAAGAAAA ACAGCAACAA
```

confirming to the PyPyA(+1)NA/TPyPy (Smale and Baltimore, 1989; Javahery *et al.*, 1994) is boxed in with a double line and the predicted transcription start site (A) is shown as +1. The sequence is numbered from the beginning of the newly identified transcription start site and the cDNA sequence previously identified by Karen Billington is shown in blue type (Tomley *et. al*, 1991). The translation initiating codon is in bold type. The GAAC motif defined in *Entamoeba histoltica* is shown with a turquoise background (-310). Two GTGTGT motifs which were observed in the upstream region of *T. annulata* gene, *TamS1*, are shown with a yellow background at positions −905 and -430 (Shiels *et al.*, 2000). A motif, similar to the A/TGAGAGC motif found in the upstream region of *SAG1* (Soldati and Boothroyd, 1995) and the AGAGACGC motif found in the promoter region of *NTPase 1* and *3* of *T. gondii* (Nakaar *et al.*, 1992), is shown with a blue background at position −968 (inverted orientation).

```
         mic3rr5af  ──▶
-2241 ACTGTGTGAC TCATGCACTA GTTTTTTGTT ACTGTCAACA GTTGTGAGGG
-2191 TCAATTTGGT CAAACTGTGT ACCATTAGCT GACCTGGCTT GTAGGTCGGC
-2141 GTTACGAATG GGACTGGAGA CTCTGTCGCA GACCGAAATT AACACTGTGC
-2091 AGAGTAACGA TTGCCGTAAG CAAACTTGTC AAAGCCGTGC ATGGAACTGT
-2041 GCTGCGAACT ATTTGGAATT TTTAT GTGTG T GGCGTTTCT AGTTTGCAAA
-1991 AGGAACATGG TGGTACCACC AGTTTCAGCA GCAGCTCCAC AGTTGTCCCC
-1941 AGGGACCACG AGACAGAAGG CAGCTTCAGG ATTTATTATG CAAACTCAAG
-1741 AAAAGCGCCA GCTTGGACCA GACTTGCAGG GGATCTTGCT GAACTTTGGA
-1691 GATTACAGTC CTAGAATCTT AAGGACATGT ATTGAGTTAT TCCTTGGGTT
-1641 ACTACAGTCA TCCGGCATCC TCT TGAGAGG AGGGTTTGGA GGCGCATTGT
-1591 AGGTAAACAG GAGTCTGGCT GTGTAGTAGC TTCAGCAGTA TAGCCAAGAG
                                   mic3rr5ae ◀──
-1541 TAAACAACCG TTGCACTGTA GCTGAACTGC ACCGTAGTGG AGAGTATTGC
-1491 AGGAGAAGAA ACGAGGCAAA AAGTAAACGG TCTGCCATTG ATGCAACAGT
-1441 GCTTAAGCAT CAGACTAATA AGGTGGTTCA CTACAGTACT TAGGCGGTGT
-1391 AGCTTGTTCA GGCCAGTAGC CCCGTCTATC GGTAGGGAAC CGCGCCAATC
-1341 ACGGCGTGGA TATGCCAATG AGCAAAGTAG CGGTTGGTCC GATTCCAAAG
-1291 GTAATTCGTG TTATATTGAG AGAAGCACAG ACGACGAAAT TGCCTTGAAA
-1241 ATAGATGGAT TCTGGCGGAC TTTACTCTAC TTCGCTTCTT ACGGACGCCC
-1191 ACTGCGAAGT GTGGAAGTGC ACCGTTGGCG CGTCTTACTA GTAAGCAGGT
-1141 GTTGTAAGCA GGAGAACGCC AGTCACTCGG AAGTGTCGCC GGGATGCGCG
-1091 GAGGCCAGCT CTGGAGTGCA GGGAATCGCT AGGTCCTCCA AGCTACAGAG
-1041 ACTCTGCACT ATTTATTTCA TGTGCCTTTT GCGTTCGCCT GGCAGTTCTG
         mic3rr5ah  ──▶
 -991 GCCTTTGCGA GAAATGAATG AGCTGGCGAC TCTGCAGAGA AAAGCAAATT
 -941 TGTTGCCATT GACGGGCGGC GCCCATCTTA TCTACAGGGA CATCTCCTTT
                                           mic3rr5ab ◀──
 -891 ACACAGTCAA GCTGTGATGT CCTGTTCTAA CTCCTGAAGC AGACCGCCAA
 -841 GTGCCGAAGA TTATTAACGA ATGCGGTAGT CTTGGTTGTC ACAGATTGTT
 -791 CATACTGAAT GAGCAGATCC AAGAAGAAAG TCCCCTTTGC AGAGCAACAA
 -741 CAGCCAACAG TGCGTGATGT TGCGTAGGGT CAGCCACCTT AGTGTAGCTT
 -691 TAACTTTTCA TATGGTATCC ACCGGACCAT TATGTAGTTC TCAGTATTTA
 -641 CTAGATTGCC TAAGACTAAT ATGGATAGGA CGGAATCCTA CTGAACAGCA
 -591 TACGATGGAC GGGAATGCCG CAGTTGAGCC GGTCCTGCGT GACAATCGCG
```

```
-541   GAACAGTTAT ACCATGACTT GAGCTGAATG AGAACAAGCT TATCGGAAAT

-491   GGTTAGGAAC CTTTCGGGTA TCTTTACGCA AGTAAGGCTC TCACATCACG

-441   GCACAATGGG CAGCAGTGCA GGCACTGGCA GCTGGAAGCT TTTGCTCTAG

-391   AGCAACACTA AACAAACACA GACAATTGTA TACGCTTCTA TTTGAGCACA

-341   CCTATCATGC TGTCACAAGG TATGATTGTG TTCATGAGGT AGGGATCGTC

                                        mic3rr5aa ←

-291   GCTCTATTGC CTGCGGAACT CGAGCAACCT GTCACAGCAT CGGATTTGCG

-241   GGAAAAGAAC CGATTATCTT GAGCGTGCAT CAAATGATTT GAAAAATTAA

-191   TAATGAAACG AATGCAGTCT GTCATTTCTA GTTTATACAT GAAGCAGAAG

-141   CAGCCAGCTA GTCGCATCAC TGGGATAGCG CACAGCTTCG TTGCTCGATT

 -91   GCTAGCGGCG AGCCGAGTCA CACCACCATC TTTCTAGACG TTGTTCTGAC

                                             +1

 -41   ATGTGTTTTT TCCAGCTTCA CACCCGACGT TCCTCAC CTA TTTT TCCCTC

 +12   CATGTTGTTT CACATTTAAC CACAAGCCTT TCTTACTTCC CCAGGATG
```

Fig. B.3 Upstream genomic sequence of *EtMIC3*

The *EtMIC3* genomic clone was identified by PCR using primers T3/T7 and mic3rr5. The genomic clones were sequenced using primers mic3rr5, mic3rr5aa, mic3rr5ab, mic3rr5ae, mic3rr5af and mic3rr5ah and are shown in red type, with an arrow indicating the direction of sequencing. A degenerate TATA box (-195) and CAAT box (-211) are shown boxed in and underlined, respectively. The initiator element, with bases confirming to the PyPyA(+1)NA/TPyPy consensus (Smale and Baltimore, 1989; Javahery *et al.*, 1994), is boxed in with a double line and the predicted transcription start site (A) is shown as +1. The sequence is numbered from the beginning of the newly identified transcription start site. No cDNA sequence was available. The translation initiating codon is shown in bold type. The GAAC motif defined in *Entamoeba histoltica* is shown with a turquoise background at positions –235 and -276. A region, similar to the A/TGAGAGC motif which has been found in the upstream region of *SAG1* (Soldati et al., 1995) and the AGAGACGC motif found in the promoter region of *NTPase 1* and *3* of *T. gondii* (Nakaar *et al.*, 1992), is shown with a blue background blue at position -1617. A GTGTGT motif defined in the upstream region of *T. annulata* gene, *TamS1*, is shown with a yellow background -2016 (Shiels *et al.*, 2000).

```
-3090 TGGATCCGTA GGAAAACAAA CAACCCGAGA TGAAGCGGCA ACGGCATTTC
-3040 ATGACAGCTG CGTTAGCACC AAAATATCAG CAGAAGCAGC ATCAGCGGCG
-2990 CCAGAAGCAT CTGCTCCCGT TGCAGGAACA ATAGCTGTTT TTTTGGGACC
-2940 TTACACTGCA CGACCACAGC AAAGTTGAAA AGTTTCAGCA TAGCAGCAAG
-2890 CCAAAGTGCG CTCATAGATT GCACAAGCGA ATACATAACT GCGTATTCGT
-2840 TTGTATTTTA GTTGAATAGT AGAAATATGA TCAGAAATCT CCAAGTGTGA
-2790 AGCAGACCAA ACAGTGAAGA GCTCTGCAGA TATCTTGCGC AGGATGAGGC
-2740 CTCACCAAGC AGGACATCCA CCCAGTGAGC TGGCTACCTG TCCACTGCCT
-2690 GAACTATTCT TTTCGTGGAG GAGCTGTAGC TACTGATTTC CTTTTTGTTG
-2640 TCGCTTCGTA CTAATGCTGC ATTTCGTTGT TTGTGCTGCG GTCGTTGTAG
-2590 GGATGAAACT TTCTGCACTG GATACTATGC CATCAACTCG TCTTTCTTAT
-2540 GAAAAAGAG TAAACTAAAA TTTTACTGTT TACTGAAACT CGCTTGTATG
-2490 CAGGAGGCCA ATCGTGCCGA GGGGGCGCGA AAGGTGAATC GTGGCCATTT
-2440 CAGATGCCAC GTTGGGTCTC ACGCGGTGCT CATGCTTTGA AGTAAGTTAA
-2390 TGGAATTATG CCTTCGGCAA TGGACGAAAC ACTAACTTCC GGGGATGAAA
-2340 TACTGGGGCG CCATAAACGA GGTTATCATG TGACCAGTGC TCTCTTTGTT
-2290 CATGCCCTCC AGCCAAAGCA CTTCCAAACT GAATGCAGCT TTCCTAAAAG
-2240 GTGTAGCAGC ATAGTAGCTA TTGTGAAAGC CGCTTCCGCT TGCTCCAATA
-2190 CGAGGCAGCT TCTGAGCAGC AACGAGCACG AGCACTGCTG TTCAGCTACG
-2140 GCAAAGAAA ATGATTCTCA CGAAACTGAA CTGTTGAGAA ATAGAAATGA
-2090 AACAAAGCAC TGAAACAGAG AAAACGCGCA AAGCAGGCCG TGCTCGTGTG
-2040 CTCGTGGGAG TCAAGGGCAT GCGTACCGGG TTTGAGGGTT GCAGTTGCTG
-1990 GTTAGATGTT TACTCTCATA TATCTTGGAC ATAGTTTTCG TTTCTAAATT
-1940 GTTCATTCAA CTACCCAAGT AGGGAAATCG AAAGGCGATC CTAAAGGAAA
-1890 GCGATGCACA CAGCTTTCAC ATTGTGGCGT CGCGGCCGTA AACCAGATCA
-1840 CGCTTCCATG CAATAAACAC AAAAAGCACG GAGAGAATGC TCACTGAGCC
-1790 TCGAAACTGC ATGCAAAATA ACACGACAAC TCTGCGCTCG GTGTACAGGC
-1740 ACAGTGAACT GGCGAGTAGC AGGAAGAACT GGCCAGCGTT CGGGAAAATT
-1690 CGGCCCCGTT ATCTTCTATT GGAATTAGTA TTGCGAGCAA ATTATTTCTT
-1640 TTTTAATTTC CCCGCTAGAA ACAAATTAAG AAAAGAAGCA TCTTCATCGT
-1590 CATTTGAGAG CTCGGGCGAT CTGATGGTCG ATTGGGAACA TGAACGGGCA
-1540 GTGCTCGCTC CCACGTGAGG AGAGCGTGCT AATTCTCCCT GCTTCGCGCT
-1490 AGCCAAACTT GAGAAATAAG TTTCAATTAG TTTACAATTT AATATGGGCT
```

```
      (+1)

-1440  TCAGAGAGGC GACAACAGCG AGGGAAACAG CAGCCATCGA CGCTGGTCCG

-1390  CCCGCAGCTG GCCGCACACA GGAGGCATGG CGTAGCTAGC TCTGGCCGAT

-1340  CTCGTCTGGT TTTAACTGCA TTCATTAATT TTCCTGATCC CTATTTGGGA

-1290  CTTATTCCTC AATTTTTCCT CATCAACCAC CACTTCCTTT TTCTTATTTG

-1240  GGCTTCTTGT GCTCACCTCT CGTGTCGCTT TTCAGAATCG GGCGCCACCA

-1190  GCTAACCGCC CTTCACAGAG TCGTTCCACA TGCTGGCTCA TGAGCCAAAG

-1140  GGTTATGTTT CTTCCTTCTG TTGATGACAA TTCCTGCTTC TGTGTAGCTA

-1090  GTATTTAAGC ACAGTTTGGC CGTTCCTGCG GCAACAATGC TGCAACGCAA

-1040  CCCGCGGTGG GCGCTTTGCG CGACCCTTGC CGCACTCTAC AGCGGCACGC

 -990  GCATCCAGAG TAAGCGAAAG CACATACAGA TCAGGCTCCA TCCTCTTTTA

 -940  CAGCTAATGT TAGGGCATTC CAGTCTCATT CCAGTGCATT TCACGTTTGA

 -890  AGGGGCACTG TTGTGTCTGA ATTAATGCTG CCCAGCGGCA GATGTTCACG

 -840  TGGCCAAGGG CTCACTCAGC GTGCCGTTTT GTTGTGAAGG AACCGTAGTT

 -790  GGCACGCTTC ATGTCTTCGT GCTTCTCCGA GTGCTTGTGG CGCGTCTCTT

 -740  GAGCTGAGTT GCCGCCAGGC ATGTGTAGTC ATAGGCGTGC TGAGTATGTT

 -690  TGCTGCTGTT TTGCTGCCGC TTTGCACAAA CCACTGTTTG CTGGTTTTGG

 -640  TTTTGTTGCT ATGCAGGCGC ATCAGGTGCT GAATCTCAGG ATATCAGCCA

 -590  ATGTGAGTCA GCTTGGACTC CGTGGACAAC TTGCAACCCG AGTACAGGCT

 -540  TGCAAGAGAG ACATAACTCG CAGTGCGAGA CGTGGGTGGA AGTGAAAGAA

 -490  TGCCAGAAGC TCTCAGGATG TACGTGTTTC GCGGTCTTAA CCGACCCTGA

 -440  ACCCTCAAGT CGTTTCTGTA GCTGGCTTCT GTTGTTTCGC GTCTCAGGCT

 -390  AGATTGTGTG CGGGGCGTAT TATACTGTGT TCGTTTTTTG TCTTCTCAGG

 -340  TGGTGCTTTT TCGGCGTGGT CACCCAGCGA CGTTGCATGC GTCCCCGGTC

 -290  AGTTCCAGAG TCGATCTCGG GAAAACTGTC CGGAGGTGAT GGAAGTCAGG

 -240  TCGTGTAAGC ACGAGTCAAA AGAGTGTAAC AAACTATGGA CCCCGTGGAC

 -190  AGTGTGCGAC ACCAACAACT TGCAAGAAAG ATACAACTCT GCGTGCGGGC

 -140  CTGTTGAAGT GCGCGAATGC CAAATGTCAG AAGAAGAAGT CAAGAAGTGT

  -90  GAGTTTTGGA ATTCTGCAAA TTTGAAAGTT GTGCCGTTGC CAAATAGGTG

  -40  CCATTCATTC TTCCGTATGT GTGTTACACG AATAAAGGGA ATG
```

Fig. B.4 Upstream genomic sequence of *EtMIC4*

The *EtMIC4* genomic sequence was identified by sequence analysis through the *E. tenella* genome sequencing project. A consensus TATA (-1452) and CAAT box (-1467) are shown boxed in and underlined, respectively. Also the predicted transcription start site is between –1437 to –1440, and the potential Inr

element which conforms to the PyPyA^{+1}NA/TPyPy consensus (Smale and Baltimore, 1989; Javahery *et al.*, 1994) is shown boxed in with a double line and the predicted transcription start site (A) is shown as (+1). The sequence is numbered relative to the translation initiating codon which is shown in bold. No cDNA sequence was available. Numerous GAAC motifs which has been observed in *E. histoltica* are present and are shown with a turquoise background at positions -1549, -1555, -1715, -1735, -2690 and -2965. CACACA or GTGTGT motifs observed in the upstream region of *T. annulata* gene, *TamS1*, are shown with a yellow background at -22, -1376 and -1884, reversed (Shiels *et al.*, 2000). Also present are motifs similar to the repeats observed in *SAG1*, *TUB1*, *GRA1, 2, 5* and *6* and *NTPase 1* and *3* (Soldati and Boothroyd, 1995; Mercier *et al.*, 1996; Nakaar *et al.*, 1992), including a GCGAGACG motif (-516), GGTCTCA motif (at position –401, indirect orientation), AGAGAGC (-536, indirect orientation), CGTCTCT (-748, indirect orientation), AGAGAGG (-1438, direct orientation), TGAGAGAC (-1586, direct orientation) and GGTCTCA (-2426, indirect orientation) and all are shown with a blue background.

```
-2530 TGGACATAAA TTAGGCTCTT ATCACAAGGA AAGGGCTCCT TGCTCTGGGT
                             mic5rr5.9 ⟶
-2480 ATGCGTACGA AGTAAATTTT ACTTCATTCC GGCTTTTTAT CTCAGTTGGT
-2430 TGATCGGGCG CACTTGGATG CTCGTACATC ACTCTGAAAT GGAAAACGTG
-2380 GTAGAAGCAA CCCTGTGACC ACTCATAGCT GACAGCCTTG GCGACGTGTC
-2330 GTTAGTGAAG ATGCTGCTGA TACTTAGGTA GAGAAGGTGC GGACGCTGGT
-2280 GTGATCTCTG CGCTTGCTGT GGACAAAGCA TCGCCTGGGA CGGAATCATT
-2230 TAATGTTTGC CCCACAGGTT TGCGTATTTT AGGCATTCCT ATGGCAAGAC
-2180 AAGCAACACC AACTTGAAAT GTGTCCAATG CAACGAGTTA ACAGGAATTC
-2130 AGGGATAGAC GGGTAGGCAA AGTTGGGGAA GATGTTGGTG TACATTGATT
-2080 AATTCGACGT CTCTTTTGCC TCTACCACAG TATGGAAAAG ACGGGAAAAA
-2030 GTCCAGCGGT CTAGTTCTGT ACCACGATGA GGTACAGGAG CAGCCAACTC
-1980 AGAAACACAG GCTGTACTGC AACCCAGCGA GCACTTTCCA CGCGCCTAAC
-1930 ATACATTTAC TGTACTGACG ACCCATTGCT TGTTGAGACT AAGCTGAGAA
-1880 TGCACCAGGT GATGAAACAT CAATTGAATT GTTTCTCAGA TGTGACTGTA
-1830 ACGTTAGTAT CCACGCGGCC CTTTCCAGTT CCAGACCTAC CTTTTATACT
-1780 GGAGTGCAAA GAGTTCAAGT TCGATTGACA CAATTATAAC AATATGCAGG
                        mic5rr5.8 ⟵
-1730 CAAACTGACT TTCGCAGTTC AGGTGAGGTT CCGCAGGTCT CTCAGGAATG
-1680 TAAAAGGAAC AGCACTATTC CTTTTCCGTG AAGGAGAATG AGGACCAGAT
-1630 GCTTTCAGCA AATTGTTCGA GCAATACGAA CAGCTTGGGC AAAACATCTG
-1580 CCCCAGTCTC TCCATCACCT GGAAGCGAAG AAATGTCCTG TATTTGATTG
-1530 ACTATATACA CGCCTCCCAG AATCAATATT AGAAAGCAGG CAAGACTGGT
-1480 ACGAGAGGAA TTACTTTCCA AAAGTTTCTA GAGGCTAGGA AGTATCCCGG
                     mic5rr5.7 ⟶
-1430 GGCTAAAATC TGTGTGCATG CCAAGAGGTA GGTACACAGA CACAAGATAC
-1380 TTCTTACTCC CTCTGCTATG ACAGTAGAAT GTTGTAACAA CCTGACTCTT
-1330 GGGGAGCCAC ACACCATCGG AATAACCGGT GGTTCGGAGA CAAAGGTGC
-1280 TCCCGTAGAA GTCCTCAACA AGCCAACCAT TTCCCCTTTA AAAAAGCACG
-1230 TAATTCGCTG CAGAGTAAGC TAAAGAAGGA AGCTTCTCCG AACTACTCTG
          mic5rr5.6 ⟵
-1180 CATCCAGACT CAAAGTATTA ACGCTGGCCT GTGGCGCGTC TAGGGCTACT
-1130 AGTGCCCTGA TTCCTCGATC TACAAGCAAA TCTATGATAT AAATTGCATA
```

```
-1080 GAAAGCAGAA ATAATAGCTC AAAGCGCTTG CTTCCCTGAC GCTGACAACT
-1030 CGGCAGTAGA CTGTTCCGCA TAGCGCCTTC CAGCTGAGCA ATCGACACTG
                              (+1)
 -980 AAAATGCTCG AAAAAAGAGC ACTC[CTAGTC T]CTTATTGTA GCGTTTTTCT
 -930 GAGTTTTAAA TAAGTTGTTT AGTCTCTGTA CTCATGTGCT CTTATTTTTG
 -880 TGACTCCTCT CAACTTTTCG TAATCGGTCG CTGCCTGCTT CACTGGTGCG
 -830 AGCAGCAGCT AGCTGGCACG CGGCACCGCA GTGTCCTCAT CTCTGTTCAC
 -780 GGCGACCAGC TGTTGCATCT GGAGAATTGT TGCTTCGAAT CACACAGTTC
 -730 ATTTCACTAC TTCCAAGTCC TTTTTTCCTT TGAGCGGCTC AGCCTTTCAG
 -680 ATCGCTTTGG GTATTCGTGT GGCGAGCCGA ATACTGTCAC AAAGTTGGTG
                                           mic5rr5◄───
 -630 TGCTTGTTCG CCGATTATGC ACATCACTTT CTCCCTTGGC GTGGCTTTCT
 -580 GAGTAAGCAG TTTTTATCTC TTGGGGTTTT ATTTATTGCA AGACGACTGG
 -530 AAGCTGCCTT CAGGACTTAC CGTTCCCGGT AGCCACGTGC GAACGGTTTT
 -480 TCATTTTTTC CATTGACTCA CTTCCCAGTT TCACCCAATA GTTACCAGAG
 -430 CAAGCGTCGA ACGCGTGTTG ACCAGGAAAC TGTAGAAATT CACAGTTATT
 -380 TCTTTCCGTG TCAATATATC TCTAAGCAAT GGGGCGTATC CGGCGAGCAG
 -330 CGGCGGTAGC AGGCTTGGGT ATCACACTGT TGCCTCGGGA TGGAACAGAG
 -280 GCCTGGCTAG CACCAAACAG ACGCCATGCT GAAGTCAGTG GAACCTTTGA
 -230 GGCACCACCA TCCATGGTGG AGACACAGTG GGGCTCGCCT TACACTTACA
 -180 TCCCCCAGGT GACTGTTAAT GAGAAGGCGG CGGACTTCTT CTTCGCATCC
 -130 GTTCTCGAGT CCGTAGCGGA GTTGCCAGCT GAGGGCCAAG GAGGCGTGGA
  -80 GGCTGGGGAA GGCGCAGAGG CCGGGCAGGG CGGAGAAGCA GCGGGTCCCC
  -30 AGCCCAAGGC CTGCGAGGCG CTGAACGCGG ATG
```

Fig. B.5 Upstream genomic sequence of *EtMIC5*

The *EtMIC5* genomic clone was identified PCR using primers T3/T7 and mic5rr5. The genomic clones were sequenced using primers mic5rr5, mic5rr5.6, mic5rr5.7, mic5rr5.8 and mic5rr5.9 and are shown in red type, with an arrow indicating the direction of sequencing. A consensus TATA (-1093) and degenerate CAAT box (-1104) are boxed in and underlined, respectively. The sequence is numbered from the beginning of the genomic sequence, and the cDNA sequence previously identified from clone pPB7 by Phil Brown is shown in blue type. A sequence which conforms to the Inr consensus is shown boxed in with a double line (Smale and Baltimore, 1989; Javahery *et al.*, 1994). The translation initiating codon is shown in bold. A GGAGAGG motif which has been found in the upstream regions of *NTPase 1* and *3* of *T. gondii* (Nakaar *et al.*, 1992) is shown with a blue background at position -1479. A CACACA motif which has been observed in the upstream region of *T. annulata* gene, *TamS1*, is shown with a yellow background at position -740 (Shiels *et al.*, 2000).

Appendix C

Name	Size (bp)	Sequence (5'-3')
T3	16	TTC CCT TTA GTG AGG G
T7	19	TAA TAC GAC TCA CTA TAG GG
SP6	20	TTC TAT AGT GTC ACC TAA AT
Mic1rr5	20	GTT TTA GCA AGG CTC CGG GC
Mic1rr5.7	20	GTC CTC TTA TTA GGA ACG TC
Mic1rr5.8	23	CGT TAT ATA AGC AAG CGC CGG CT
Mic1rr5.10	20	GCA TTT GCA GGG CCT CCA ACG
Mic2rr5	20	CGA CGA AGA ATA TTC ACA TG
Mic2rr5.5	21	GAC ACA AAG TCC AG TCA CAC
Mic2rr5.6	21	GAG CAG CTG TAC TTA ACC GTC
Mic2rr3	20	GGC TGC TGA CCA GTT CGT GC
Mic2rr3a	20	GAT GGC TCT GGG AGA AAG TA
Mic2rr3.4	20	GCA CACATG CCA ACC ACA GC
Mic3rr5	21	CAG GGC GTT TGT CAC GAA TGC
Mic3rr5aa	20	GCG TAG CTG TGC CAG GTT GC
Mic3rr5ab	21	CTT CGG CAC TTG GCG GTC TGC
Mic3rr5ae	21	CTC TCC ACT ACG GTG CAG TTC
Mic3rr5af	20	CTG TGT GAC TCA TGC ACT AG
Mic3rr5ah	20	GCC TTT GCG AGA AAT GAA TG
Mic3rr3	20	GCT TGC ACA ATT TGT CGA GG
Mic3rr3b	21	CGA CGA AGA ATA TTC ACA TGA
Mic3rr3.9	22	GTG CCT ACG GCC ATT CTG CAT G
Mic3rr3.11	20	GGC CTT GTC GTA TAC CCT AC
Mic4rr5	20	CTC GGA AAC CCG GTT TGC AC
Mic4rr5a	20	GCA ACG TCG CTG GGT GAC CA
Mic4rr5.10	18	CGG ATC AAA GGC GAG TAC
Mic4rr5.11	21	CAG CGG TAA GAT CCT TGA GAG
Mic4rr5.12	19	GTC TCA TGA GCG GAT ACA T
Mic4rr5.13	20	CGG GCC TCT TCG CTA TTA CG
Mic4rr3	20	TGC CGA GGA GGC TGA GCA AG
Mic4rr3.5	21	CAT GAA CGT CTG AGT GCT CCC

Mic4rr3.7	21	GCC TCC TTT CTT CAT CAC GCG
Mic4rr3.10	20	GTC ATT TCA GTC ATT TAT CA
Mic5rr5	19	CAG AAA GCC ACG CCA AGG G
Mic5rr5.6	20	CTT TGA GTC TGG ATG CAG AG
Mic5rr5.7	20	TGT GTG CAT GCC AAG AGG TA
Mic5rr5.8	22	GAG ACC TGC GGA ACC TCA CCT G
Mic5rr5.9	20	ACT TCA TTC CGG CTT TTT AT
Mic5rr3	19	GCC TCA CAT TAG CAC GCC T
Mic5rr3.5b	20	CCT GGG CAG CTT ATC TGC TT
Mic5rr3.6	21	CAC TGG AAT CTC GGG AGT TCA
Mic5rr3.10	19	CCT TCA CAT GCG TTG TCG G
Mic5rr3.11	20	CAG ATG TGG AAA CAA TTG CG

Table C.1 Oligonucleotide primers used in the amplification of the 5' upstream region of *EtMIC2, EtMIC3, EtMIC4* and *EtMIC5* (Chapter Four). Primers were supplied by MWG-Biotech UK Ltd.

Name	Size bp	Sequence (5'-3')
EtMIC1 molecular constructs		
mic1R-kpn1	31	ggaaggggtac(+106)CATTGCGGGAATGATGCAAG(+95)
mic1F-kpn1	30	gccggtac(+12)CTGGTTAGGTTTCAAAACATAT(+35)
mic1F-kpn2	31	cgcggtac(-37)CAGGACCCTTGCCGAGCCATCCG(-24)
mic1F-kpn3	30	gaaggtac(-124)CAATAACAATATACGTGATCTG(-107)
mic1F-kpn4	31	ataggtac(-300)CTACGGATATTCTACACTATCAG(-278)
mic1F-kpn5	31	cttggtac(-503)CAAACTTTTGAAGCATTGTGCG(-482)
EtMIC1 internal molecular constructs		
Mic1F-kpn7	26	cccggtac(-473)CTAATTTGCCTCTCGAAG(-456)
Mic1F-kpn9	25	cccggta(-410)CATATGCACACAGCGCAG(-393)
Mic1F-kpn10	28	gacggtac(-383)CTTTAAACCATGCTTGGAGG(-364)
Mic1F-kpn11	28	tgcggtac(-349)CATGTCTAATGGCAAAAAGC(-330)
Mic1F-kpn12	29	agcggtac(-324)GGAATAACACTCCTCAATATC(-304)
EtMIC2 molecular constructs		
mic2-rev-kpn1	33	cccggtac(+100)CATTTTGGTGAATACAAAAGAG(+39)
mic2F-kpn1	36	cccggtac(+12)CTTTTTGGCAATTGTTTGGCTTTTAAAA(+12)
mic2F-kpn2	33	gtgggtac(-33)CAACTCAAGGTCCCATCATTTGCAG(-12)
mic2F-kpn3	33	attggtac(-148)CATATTACAGAATGCTAAAAGGAGA(-125)
mic2F-kpn4	30	gttggtac(-510)CTGTTAAATAAGTGATAGTGCT(-487)
mic2F-kpn5	33	ggaggtac(-949)CAGTGGGCTAGTGCCAACGAGCGAA(-925)
EtMIC2 internal molecular constructs		
Mic2F-kpn10	36	aacggtac(-55)CTGCACAACCATGTGTGCTTTCCAACTC(-28)
Mic2F-kpn11	36	cccggtac(-80)CTGACTGTGCTAGCGGACCGCAAAC(-56)
Mic2F-kpn12	37	gtgggtac(-113)CGTGATTCCTTAGTGCAATTGTGCTAGCC(-85)
Mic2F-kpn13	39	cgcggtac(-204)CTCCCCAAAAAATTAATGGCTTCATACTGGC(-185)
Mic2F-kpn14	38	gtgggtac(-235)CGGTGCGCGCCTCCCAGAAAAGCAAGAAAG(-206)
Mic2F-kpn15	40	cgcggtac(-293)CTTAGAGTTGGCTTGCACCCCCGTACGTTCAG(-273)
Mic2F-kpn16	41	gtgggtac(-380)CGCAACGAATTCTCGTATGCTTACCAACATTTC(-348)
Mic2F-kpn17	40	gtgggtac(-461)CGGTTAAGTACAGCTGCTCCAGCGGCCCAACG(-430)
EtMIC3 molecular constructs		
mic3R-kpn1	37	Cccggtac(+59)CATCCTGGGGAAGTAAGAAAGGCTTGTGC(+31)
mic3F-kpn1	31	Cccggtac(+12)CATGTTGTTTCACATTTAACCAC(+34)

mic3F-kpn3	30	Cccggtac(-71)CACACCACCATCTTAGACGTTG(-49)
mic3F-kpn5	39	Cccggtac(-571)CAGTTGAGCCGGTCCTGCGTGACAATCGCG(-542)
mic3F-kpn7	36	Cccggtac(-2210)CCTGTCAACAGTTGTGAGGGTCAATTTG(-2184)

Table C.2 Primers used for production of molecular constructs containing the 5' flanking regions of either *EtMIC1, EtMIC2* and *EtMIC3*. Capital letters in the oligonucleotide primers indicate nucleotides identical to the template used for PCR amplification, while letters in the lower case represent nucleotides not present in the template sequence (Chapter Five). The 5' and 3' locations of the genomic DNA are shown in brackets. Primers were supplied by MWG-Biotech UK Ltd.

Name	Size (bp)	Sequence (5'-3')
MUTA-1A	50	(-26)CCGAGCCATCCGACTTCAATCGTCTCTCAGTAGAATTCTGGTTA GGTTTC(+24)
MUTA-1B	50	(+24)GAAACCTAACCAGAATTCTACTGAGAGACGATTGAAGTCGGA TGGCTCGG(-26)
MUTA-2A	50	(-26)CCGAGCCATCCGACTTCAATCGTCTCGCAGTAGAATTCTGGTTA GGTTTC(+24)
MUTA-2B	50	(+24)GAAACCTAACCAGAATTCTACTGCGAGACGATTGAAGTCGGA TGGCTCGG(-26)
MUTA-3A	50	(-26)CCGAGCCATCCGACTTCAATCGTCTCCCAGTAGAATTCTGGTTA GGTTTC(+24)
MUTA-3B	50	(+24)GAAACCTAACCAGAATTCTACTGGGAGACGATTGAAGTCGAA TGGCTCGG(-26)
MUTA-4A	58	(-26)CCGAGCCATCCGACTTCAATCGTCCAGTAGAATTCTGGTTAGGT TTCAAAACATATTG(+25)
MUTA-4B	58	(+25)CAATATGTTTTGAAACCTAACCAGAATTCTACTGGACGATTG AAGTCGGATGGCTCCG(-26)
MUTA-5A	64	(-26)CCGAGCCATCCGACTTCAATCGTCTCAGTAGAATTCTGGTTAGG TTTCAAAACATATTGTTTTG(+113)
MUTA-5B	64	(+113)CAAAACAATATGTTTTGAAACCTAACCAGAATTCTAATGAG ACGATTGAAGTCGGATGGCTCGG(-26)
MUTA-6A	66	(-26)CCGAGCCATCCGACTTCAATCGTCTTACAGTAGAATTCTGGTTA GGTTTCAAAACATATTGTTTTG(+113)
MUTA-6B	66	(+113)CAAAACAATATGTTTTGAAACCTAACCAGAATTCTACTGTA AGACGATTGAAGTCGGATGGCTCGG(-26)
MUTA-7A	66	(-26)CCGAGCCATCCGACTTCAATCGTCTAACAGTAGAATTCTGGTTA GGTTTCAAAACATATTGTTTTG(+113)
MUTA-7B	66	(+113)CAAAACAATATGTTTTGAAACCTAACCAGAATTCTAATGTT AGACGATTGAAGTCGGATGGATAGCC(-26)

Table C.3 Oligonucleotide primers used in the construction of *EtMIC1* transcription start site mutant constructs (Chapter Six). The 5' and 3' locations of the genomic DNA are shown in brackets. Primers were supplied by MWG-Biotech UK Ltd.

| Name	Size (bp)	Sequence (5'-3')
Mic1-mutced-1A	49	(-389) CCGTCACTTTAAACCATGCTTGGAACTTTGCTT TTCCATGTCTAATGGC (-337)
Mic1-mutced-1B	49	(-337) GCCATTAGACATGGAAAAGCAAAGTTCCAAGCA TGGTTTAAAGTGACGG (-389)
Mic1mutnf at-1A	49	(-389) CCGTCACTTTAAACCATGCTTGGAGCTTTGCTT TTCCATGTCTAATGGC (-337)
Mic1mutnf at-1A	49	(-337) GCCATTAGACATGGAAAAGCAAAGCTCCAAGCA TGGTTTAAAGTGACGG (-389)
Mic1-stat-1A	49	(-389) CCGTCACTTTAAACCATGCTTGGAACTTTGCTT TTCCATGTCTAATGGC (-337)
Mic1-stat-1B	49	(-337) GCCATTAGACATGGAAAAGCAAAGTTCCAAGCA TGGTTTAAAGTGACGG (-389)
Mic1-dof-1A	49	(-389) GCCATTAGACATGGAAAAGCTTTCCTCCAAGCA TGGTTTAAAGTGACGG (-337)
Mic1-dof-1B	49	(-337) CCGTCACTTTAAACCATGCTTGGAGGAAAGCTT TTCCATGTCTAATGGC (-389)
Mic1-hap-1A	49	(-389) GCCATTAGACATGGAAAAGCAAAGTTTCCTGCA TGGTTTAAAGTGACGG (-337)
Mic1-hap-1B	49	(-337) CCGTCACTTTAAACCATGCAGGAAACTTTGCTT TTCCATGTCTAATGGC (-389)
Mic1-elf-1A	52	(-355) GCAAAGTTTCCTCCAAGCATGAAAGTGACGGTC TCTGCGCTGTGTGCATAT (410)
Mic1-elf-1A	52	(-410) CATATGCACACAGCGCAGAGACCGTCACTTTCA TGCTTGGAGGAAACTTTGC (-355)

Table C.4 Oligonucleotide primers used in the construction of *EtMIC1* transcription factor binding sites mutant constructs (Chapter Six). Primers were supplied by MWG-Biotech UK Ltd.

Name	Size (bp)	Sequence (5'-3')
Mic2-mut-1A	51	(-22)CCCATCATTTGCAGCAACGAATTACTCCTTGCTTTTTGGCAATTGTTTGGC(+31)
Mic2-mut-1B	51	(+31)GCCAAACAATTGCCAAAAAGCAAGGAGTAATTCGTTGCTGCAAATGATGGG(-22)
Mic2-mut-2A	51	(-22)CCCATCATTTGCAGCAACGGATTACTCCTTGCTTTTTGGCAATTGTTTGGC(+31)
Mic2-mut-2B	51	(+31)GCCAAACAATTGCCAAAAAGCAAGGAGTAATCCGTTGCTGCAAATGATGGG(-22)
Mic2-mut-5A	53	(-22)CCCATCATTTGCAGCAACGAAGTTTACTCCTTGCTTTTTGGCAATTGTTTGGC(+31)
Mic2-mut-5B	53	(+31)GCCAAACAATTGCCAAAAAGCAAGGAGTAAACTTCGTTGCTGCAAATGATGGG(-22)
Mic2-mut-6A	53	(-22)CCCATCATTTGCAGCAACGAAAAATACTCCTTGCTTTTTGGCAATTGTTTGGC(+31)
Mic2-mut-6B	53	(+31)GCCAAACAATTGCCAAAAAGCAAGGAGTATTTTCGTTGCTGCAAATGATGGG(-22)
Mic2-mut-8A	55	(-56)CCTGCACAACCCATCATTTGCAGCAACGAAGATTGCTTTTTGGCAATTGTTTGGC(+31)
Mic2-mut-8B	55	(+31)GCCAAACAATTGCCAAAAAGCAATCTTCGTTGCTGCAAATGATGGGTTGTGCAGG(-56)
Mic2-mut-9A	51	(-125)GAAATAAAAACGCGTGATTCCTTAGTGTGTGCTAGCCACTGACTGTGCCTC(-71)
Mic2-mut-9B	51	(-71)GAGGCACAGTCAGTGGCTAGCACACACTAAGGAATCACGCGTTTTTATTTC(-125)

Table C.5 Oligonucleotide primers used for in the construction of *EtMIC2* transcription start site mutant constructs (Chapter Six). The 5' and 3' locations of the genomic DNA are shown in brackets. Primers were supplied by MWG-Biotech UK Ltd.

pGEM®-T Easy Vector (3018bp)

Xmn I 2009
Sca I 1890
Amp'
f1 ori
Nae I 2710
lacZ
ori

T7 ↓	1 start
Apa I	14
Aat II	20
Sph I	26
Nco I	37
BstZ I	43
Not I	43
Sac II	49
EcoR I	52

Spe I	64
EcoR I	70
Not I	77
BstZ I	77
Pst I	88
Sal I	90
Nde I	97
Sac I	109
BstX I	118
Nsi I	127
↑ SP6	141

SuperCos 1 Cosmid Vector 7.6 kb

ampicillin
ori
NEO
SV40 ori
cos
cos
Xba I

EcoR I
Not I
T3 ↓
BamH I
T7 ↑
Not I
EcoR I
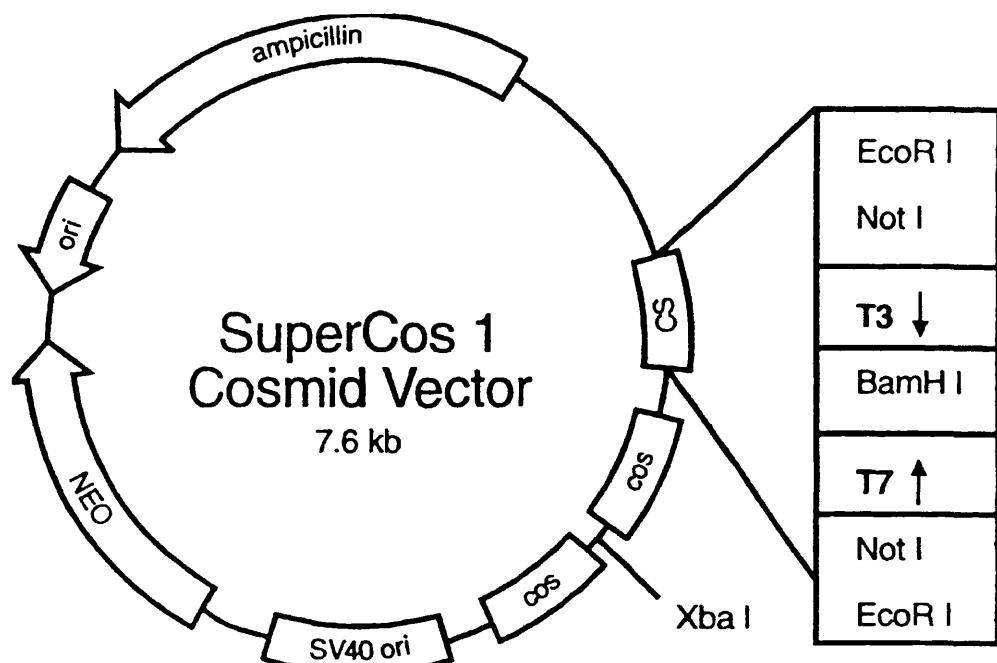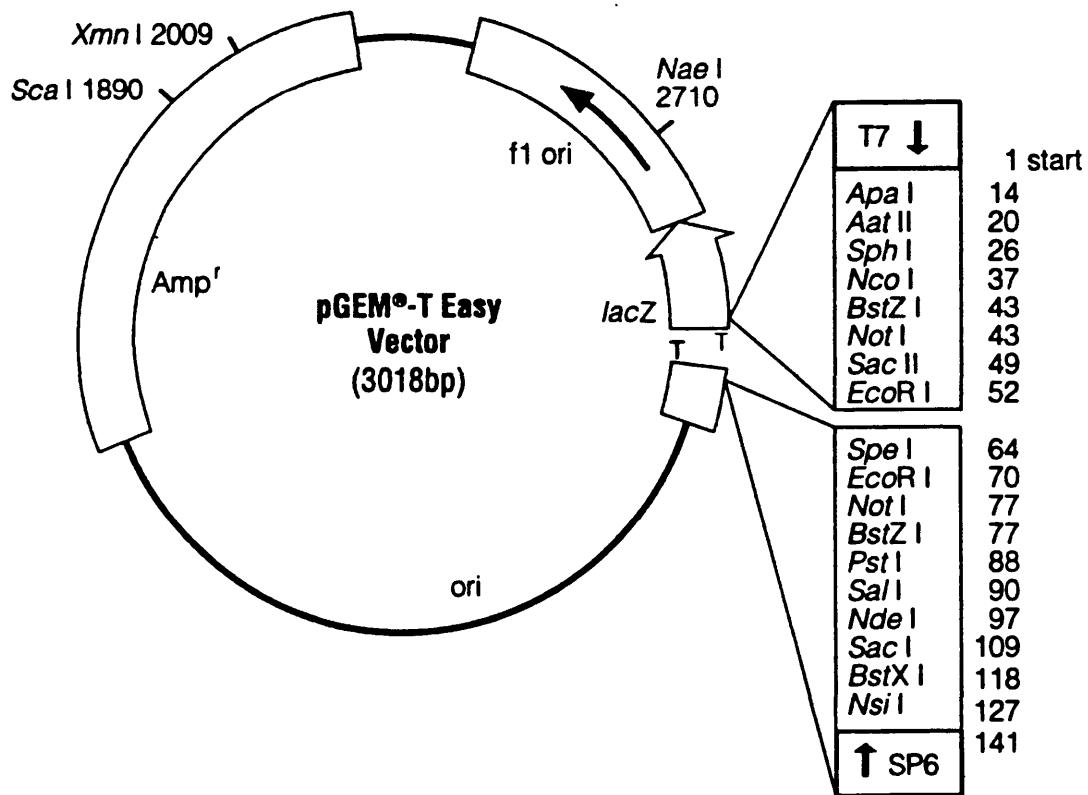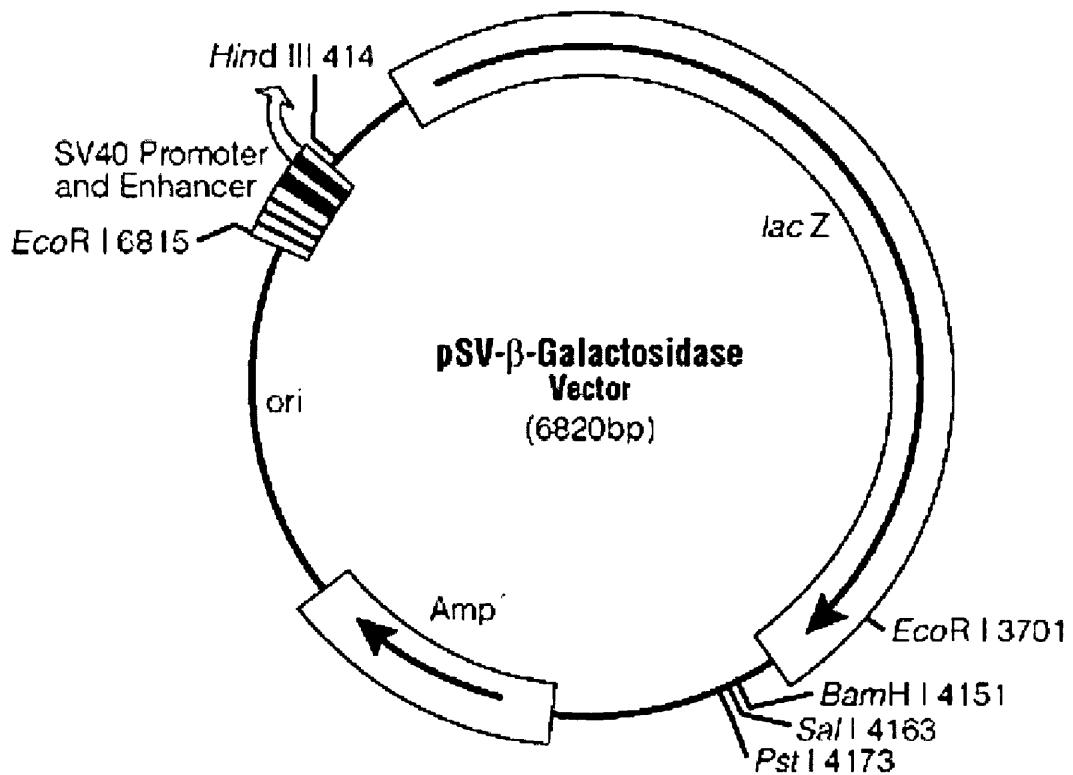
Plasmid maps

A: pGEM-T Easy vector, used for standard PCR cloning, contains an origin of replication, ampicillin resistance gene, multicloning site within the *LacZ* ORF and T7 and SP6 promoter binding sites.

B: SuperCos 1 cosmid vector, used to generate cosmid libraries, with a cloning capacity of between 30 and 42kb. Contains an ampicillin resistance ORF, neomycin resistance ORF, SV40 promoter, pUC origin of replication and a T3 and T7 promoter binding sites.

C: pSV-β-Galactosidase vector, used in transient transfection assays to analyse β-galactosidase expression. Contains an ampicillin resistance ORF, a *LacZ* ORF, and a SV40 early promoter and enhancer segment.