Expression of microneme genes in *Eimeria tenella*

by

Rachel M. Ryan

A thesis submitted (in partial fulfilment) for the degree of Doctor of Philosophy

Division of Molecular Parasitology
Institute of Animal Health
Compton
Newbury
Berkshire
RG20 7NN

&

Cell Biology Department
University College London
Gower Street
London
WC1E 6BT

September 2002
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

Abstract ............................................................................................................................. 3
Table of Contents ............................................................................................................. 4
Figures .............................................................................................................................. 7
Tables ............................................................................................................................... 8
Abbreviations .................................................................................................................. 9
Parasites ........................................................................................................................... 11
Bacteria ............................................................................................................................ 11
Acknowledgements ......................................................................................................... 12
1.1 Coccidiosis and its control ...................................................................................... 13
  1.1.1 The life-cycle of *E. tenella* ............................................................................. 14
  1.1.2 Genomic organisation ....................................................................................... 16
1.2 Invasion by apicomplexan parasites ....................................................................... 18
  1.2.1 Parasite invasion and motility ......................................................................... 21
  1.2.2 The parasitophorous vacuole ......................................................................... 25
1.3 Apical organelles of apicomplexan parasites .......................................................... 29
  1.3.1 Dense granules ............................................................................................... 29
  1.3.2 Rhoptries ....................................................................................................... 30
  1.3.3 Micronemes .................................................................................................... 31
1.4 The apicomplexan secretory system ....................................................................... 31
  1.4.1 Trafficking to the ER and Golgi apparatus .................................................... 32
  1.4.2 Membrane docking and fusion proteins ....................................................... 35
  1.4.3 Targeting to the dense granules ..................................................................... 35
  1.4.4 Post secretory targeting of GRA proteins .................................................... 36
  1.4.5 Targeting of proteins to the rhoptries and beyond ....................................... 37
  1.4.6 Targeting to the micronemes ........................................................................ 38
1.5 Micronemes ............................................................................................................. 42
  1.5.1 Microneme protein adhesive domains ............................................................ 42
    1.5.1.1 TSP domains ............................................................................................ 43
    1.5.1.2 I domains ............................................................................................... 45
    1.5.1.3 EGF-like domains .................................................................................. 45
    1.5.1.4 Apple domains ....................................................................................... 46
  1.5.2 Function in recognition and attachment .......................................................... 50
  1.5.3 Secretion ........................................................................................................ 52
  1.5.4 Processing ....................................................................................................... 54
  1.5.5 Transmembrane and cytoplasmic tail ............................................................. 55
1.6 Project aims .............................................................................................................. 56
2.1 General parasitology methods ................................................................................. 57
  2.1.1 Parasites ......................................................................................................... 57
  2.1.2 Recovery of oocysts from infected caeca ...................................................... 57
  2.1.3 Sporulation .................................................................................................... 58
  2.1.4 Recovery and purification of sporulated oocysts ............................................ 58
  2.1.5 Preparation of sporozoites ............................................................................ 58
2.2 General molecular biology methods ....................................................................... 59
  2.2.1 Preparation of plasmid and cosmid DNA ...................................................... 59
    2.2.1.1 Small scale .............................................................................................. 59
    2.2.1.2 Large scale ............................................................................................. 59
  2.2.2 Polymerase chain reaction (PCR) ................................................................ 60
  2.2.3 Restriction endonuclease digestion ............................................................... 61
2.2.4 Gel purification of DNA fragments ............................................................... 61
2.2.5 Ligation of DNA ........................................................................................ 62
2.2.6 Transformation of DNA with E. coli cells ................................................. 62
2.2.7 DNA sequencing ......................................................................................... 63
  2.2.7.1 Automated sequencing reactions and precipitations ............................. 63
  2.2.7.3 Automated sequencing gels .................................................................. 63
  2.2.7.4 Manual sequencing reactions ............................................................... 64
  2.2.7.5 Manual sequencing gel ........................................................................ 64
2.2.8 DNA detection ........................................................................................... 65
  2.2.8.1 Southern blotting .................................................................................. 65
  2.2.8.2 Hybridisation of radiolabelled DNA probes ......................................... 65
2.2.9 Detection of mRNA ................................................................................... 66
  2.2.9.1 Total RNA extraction from a time-course of sporulating oocysts ....... 66
2.2.10 Protein analyses ....................................................................................... 66
  2.2.10.1 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) ..................................................................................................................... 66
  2.2.10.2 Gel staining and drying ........................................................................ 67
  2.2.10.3 Western blots and alkaline phosphatase detection ............................ 67
2.2.11 Transfection of E. tenella sporozoites .................................................... 68
  2.2.11.1 Preparation of cell monolayers ........................................................... 68
  2.2.11.2 Electroporation ...................................................................................... 68
  2.2.11.3 Fixing and visualisation for β-galactosidase ....................................... 69
  2.2.11.4 Preparation of parasite/cell lysates for assay of β-galactosidase.... 69
  2.2.11.5 Chemiluminescent assay for β-galactosidase using a luminometer 69
3.1 Introduction .................................................................................................... 71
  3.1.1 Background of the work .......................................................................... 73
3.2 Methods ........................................................................................................ 73
  3.2.1 Oocysts sampling ...................................................................................... 73
  3.2.2 Protein Analyses ....................................................................................... 74
    3.2.2.1 Extraction of protein from sporulating oocysts .................................... 74
  3.2.3 Pulse field gel electrophoresis (PFGE) .................................................... 74
    3.2.3.1 Preparation of chromosomal DNA blocks ............................................ 74
    3.2.3.2 Conditions of pulse field gel electrophoresis (PFGE) ......................... 75
  3.2.4 Detection of mRNA.................................................................................. 75
    3.2.4.1 Reverse transcription-polymerase chain reaction (RT-PCR)............ 75
3.3 Results ........................................................................................................... 76
  3.3.1 Synchrony of development and morphology of oocysts during sporulation .............................................................. 76
  3.3.2 Appearance of microneme proteins .......................................................... 80
  3.3.3 Chromosomal localisation of genes encoding microneme proteins ....... 83
  3.3.4 Expression of microneme-specific mRNAs during sporulation ............. 84
3.4 Discussion ...................................................................................................... 91
4.1 Introduction ................................................................................................... 94
  4.2 Methods ........................................................................................................ 98
    4.2.1 Rapid amplification of cDNA ends (RACE) ........................................... 98
    4.2.2 Primer extension ...................................................................................... 100
4.3 Results .......................................................................................................... 100
  4.3.1 Isolation of genomic clones ...................................................................... 100
  4.3.2 Transcription start site mapping of EtMIC1-5 ........................................ 103
  4.3.3 Sequence analysis of genomic DNA (5' untranscribed regions) .......... 104
4.4 Discussion ............................................................................................................ 113
5.1 Introduction ............................................................................................................ 120
  5.2.1 Molecular constructs ..................................................................................... 121
5.3 Results ..................................................................................................................... 123
  5.3.1 Promoter mapping by 5’ deletion analysis of the untranscribed region of EtMIC1 ...................................................................................................................... 123
  5.3.2 Promoter mapping by 5’ deletion analysis of the untranscribed region of EtMIC2 ........................................................................................................................................ 131
  5.3.3 Promoter mapping by 5’ deletion analysis of the untranscribed region of EtMIC3 and EtMIC4 ........................................................................................................................................ 132
5.4 Discussion .............................................................................................................. 138
6.1 Introduction ............................................................................................................ 142
  6.2 Methods ............................................................................................................ 144
  6.2.1 Site directed mutagenesis .............................................................................. 144
6.3 Results ..................................................................................................................... 148
  6.3.1 Mutational analysis of the Inr of EtMIC1 ................................................... 148
  6.3.2 Mutational analysis of the Inr of EtMIC2 ................................................... 151
  6.3.3 Mutational analysis of potential transcription factor binding sites 155
6.4 Discussion .............................................................................................................. 158
7.1 Summary of results .............................................................................................. 163
7.2 Future Work ........................................................................................................... 168
References ..................................................................................................................... 176
Appendix A ................................................................................................................... 218
Appendix B ................................................................................................................... 224
Appendix C ................................................................................................................... 234
Appendix D ................................................................................................................... 241
### Figures

#### Chapter One

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 1.1</td>
<td>The life-cycle of <em>E. tenella</em>.</td>
<td>16</td>
</tr>
<tr>
<td>Fig. 1.2</td>
<td>Diagrammatic representation of an eimerian sporozoite (A) and merozoite (B) reproduced from Ball and Pittilo (1990).</td>
<td>18</td>
</tr>
<tr>
<td>Fig. 1.3</td>
<td>Invasion by <em>E. tenella</em> sporozoites.</td>
<td>27</td>
</tr>
<tr>
<td>Fig. 1.4</td>
<td>The organelles of <em>E. tenella</em> sporozoites.</td>
<td>28</td>
</tr>
<tr>
<td>Fig. 1.5</td>
<td>Current model for protein targeting in the early secretory pathway</td>
<td>40</td>
</tr>
<tr>
<td>Fig. 1.6</td>
<td>Current model for protein targeting from the golgi to the secretory organelles, micronemes, rhoptries and dense granules</td>
<td>41</td>
</tr>
<tr>
<td>Fig. 1.7</td>
<td>Schematic representation of some of the micronemal proteins in apicomplexan parasites</td>
<td>48</td>
</tr>
</tbody>
</table>

#### Chapter Three

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 3.1</td>
<td>Synchrony of oocysts during sporulation.</td>
<td>80</td>
</tr>
<tr>
<td>Fig. 3.2</td>
<td>Morphology of <em>E. tenella</em> oocysts during sporulation.</td>
<td>81</td>
</tr>
<tr>
<td>Fig. 3.3</td>
<td>Total lysate analysis of oocysts proteins during sporulation.</td>
<td>83</td>
</tr>
<tr>
<td>Fig. 3.4</td>
<td>Detection of microneme proteins during oocyst sporulation.</td>
<td>84</td>
</tr>
<tr>
<td>Fig. 3.5</td>
<td>Chromosomal localisations of <em>EtMIC1-5</em>.</td>
<td>87</td>
</tr>
<tr>
<td>Fig. 3.6</td>
<td>Fidelity of total RNA.</td>
<td>89</td>
</tr>
<tr>
<td>Fig. 3.7</td>
<td>Detection of microneme-specific RNA during oocyst sporulation.</td>
<td>90</td>
</tr>
<tr>
<td>Fig. 3.8</td>
<td>Detection of microneme-specific RNA from sporozoites.</td>
<td>91</td>
</tr>
</tbody>
</table>

#### Chapter Four

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 4.1</td>
<td><em>EtMIC1</em> genomic DNA amplification to clone the 5' upstream region.</td>
<td>104</td>
</tr>
<tr>
<td>Fig. 4.2</td>
<td>Mapping <em>EtMIC1, 2 and 3</em> transcription start sites using 5' RACE PCR analysis of <em>E. tenella</em> total RNA.</td>
<td>108</td>
</tr>
<tr>
<td>Fig. 4.3.1</td>
<td>Schematic of the information available for the microneme encoding genes <em>EtMIC1</em> to 3 following sequence and 5’RACE analysis.</td>
<td>114</td>
</tr>
<tr>
<td>Fig. 4.3.2</td>
<td>Schematic of the information available for the microneme encoding genes <em>EtMIC4</em> to 5 following sequence and 5’RACE analysis.</td>
<td>115</td>
</tr>
</tbody>
</table>

#### Chapter Five

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fig. 5.1</td>
<td>Schematic diagram of the protocol used to produce molecular constructs for using the 5' flanking region of <em>EtMIC1</em>.</td>
<td>126</td>
</tr>
<tr>
<td>Fig. 5.2</td>
<td>Visualisation of β-galactosidase activity in developing sporozoites of <em>E. tenella</em> (Wis F96 line).</td>
<td>129</td>
</tr>
<tr>
<td>Fig. 5.3</td>
<td>Schematic diagram of initial set of <em>EtMIC1</em> plasmid constructs used in transfection experiments.</td>
<td>130</td>
</tr>
<tr>
<td>Fig. 5.4</td>
<td>Transient transfection analyses of <em>E. tenella</em> <em>EtMIC1</em> promoter.</td>
<td>131</td>
</tr>
<tr>
<td>Fig. 5.5</td>
<td>Schematic diagram of <em>EtMIC1</em> plasmid constructs used in transfection experiments.</td>
<td>132</td>
</tr>
<tr>
<td>Fig. 5.6</td>
<td>Transient transfection analyses of <em>E. tenella</em> <em>EtMIC1</em> promoter.</td>
<td>133</td>
</tr>
</tbody>
</table>
Fig. 5.7  Schematic diagram of initial set of EtMIC2 plasmid constructs used in transfection experiments.
Fig. 5.8  Transient transfection analyses of *E. tenella* EtMIC2 promoter.
Fig. 5.9  Schematic diagram of EtMIC2 plasmid constructs used in transfection experiments.
Fig. 5.10 Transient transfection analyses of *E. tenella* EtMIC2 promoter.
Fig. 5.11 Schematic diagram of EtMIC3 plasmid constructs used in transient transfection experiments.

**Chapter Six**

Fig. 6.1  A schematic representation of the procedure used to obtain mutated constructs.
Fig. 6.2  Transient transfection analyses of mutated constructs of *E. tenella* EtMIC1 promoter.
Fig. 6.3  Transient transfection analyses of mutated constructs of *E. tenella* EtMIC2 promoter.
Fig. 6.4  Transient transfection analyses of mutated constructs of *E. tenella* EtMIC1 promoter (series 2).

**Tables**

**Chapter One**
Table 1.1  Taxonomic arrangement of coccidial parasites.  

**Chapter Three**
Table 3.1  Chromosome localisations of EtMIC1-5.  
Table 3.2  Oligonucleotide primers for RT-PCR analysis.

**Chapter Four**
Table 4.1  Results of transcription start site mapping by 5'RACE  
Table 4.2  Oligonucleotide primers for 5'RACE analysis.  
Table 4.3  Prevalence of Inr consensus in natural Pol II promoters.  
Table 4.4  Repeat sequences in the upstream regions of Apicomplexan genes.  
Table 4.5  Repeat sequences in the upstream regions of Apicomplexan genes.

**Chapter Six**
Table 6.1  Template DNA was mutated by PCR.  
Table 6.1  Analysis of mutated EtMIC1 promoter.  
Table 6.2  Analysis of mutated EtMIC2 promoter.  
Table 6.3  Analysis of mutated EtMIC1 promoter (series 2).
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Adenosine</td>
</tr>
<tr>
<td>ADP</td>
<td>Adenosine 5'-diphosphate</td>
</tr>
<tr>
<td>AP</td>
<td>Adaptor proteins</td>
</tr>
<tr>
<td>APS</td>
<td>Ammonium persulphate</td>
</tr>
<tr>
<td>ARF</td>
<td>ADP ribosylation factor</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine 5'-triphosphate</td>
</tr>
<tr>
<td>BAP</td>
<td><em>E. coli</em> alkaline phosphatase</td>
</tr>
<tr>
<td>BCIP</td>
<td>Bromo chloro indolyl phosphate</td>
</tr>
<tr>
<td>BDM</td>
<td>Butanedione monoxime</td>
</tr>
<tr>
<td>BFA</td>
<td>Brefeldin A</td>
</tr>
<tr>
<td>BLA</td>
<td><em>E. coli</em> β-lactamase</td>
</tr>
<tr>
<td>β-gal</td>
<td>β-galactosidase reporter gene</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>bp</td>
<td>Base pairs</td>
</tr>
<tr>
<td>C</td>
<td>Cytosine</td>
</tr>
<tr>
<td>CAT</td>
<td>Chloroamphenicol acetyl transferase reporter gene</td>
</tr>
<tr>
<td>CBB</td>
<td>Coomassie Brilliant Blue</td>
</tr>
<tr>
<td>cbEGF</td>
<td>Calcium binding epidermal growth factor</td>
</tr>
<tr>
<td>CD</td>
<td>Cytochalasin D</td>
</tr>
<tr>
<td>cDNA</td>
<td>DNA complementary to RNA</td>
</tr>
<tr>
<td>CS</td>
<td>Circumporozoite protein</td>
</tr>
<tr>
<td>CTRP</td>
<td>CS and TRAP related protein</td>
</tr>
<tr>
<td>dATP</td>
<td>Deoxyadenosine 5'-triphosphate</td>
</tr>
<tr>
<td>DBP</td>
<td>Duffy-binding protein</td>
</tr>
<tr>
<td>DEPC</td>
<td>Diethyl pyrocarbonate-treated deionised water</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethylsulphoxide</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxynuclease</td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EBA-175</td>
<td>Erythrocyte-binding antigen-175 Kda</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EGF</td>
<td>Epidermal growth factor</td>
</tr>
<tr>
<td>EM</td>
<td>Electron microscopy</td>
</tr>
<tr>
<td>EMSA</td>
<td>Electrophoretic mobility shift assay</td>
</tr>
<tr>
<td>ER</td>
<td>Endoplasmic reticulum</td>
</tr>
<tr>
<td>EtMICX</td>
<td><em>E. tenella</em> microneme (X refers to an number)</td>
</tr>
<tr>
<td>FCS</td>
<td>Foetal calf serum</td>
</tr>
<tr>
<td>G</td>
<td>Guanine</td>
</tr>
<tr>
<td>GAG</td>
<td>Glycosaminoglycans</td>
</tr>
<tr>
<td>GCG</td>
<td>Genetics Computer Group</td>
</tr>
<tr>
<td>GFP</td>
<td>Green fluorescent protein</td>
</tr>
<tr>
<td>GRA</td>
<td>Dense granule protein</td>
</tr>
<tr>
<td>GST</td>
<td>Glutathione-S-transferase</td>
</tr>
<tr>
<td>HAMs</td>
<td>HAMs nutrient mixture-F12</td>
</tr>
<tr>
<td>HSPG</td>
<td>Heparin sulphate proteoglycans</td>
</tr>
<tr>
<td>IBP</td>
<td>Initiator binding protein</td>
</tr>
<tr>
<td>I domain</td>
<td>Integrin insertion domain</td>
</tr>
<tr>
<td>ISG</td>
<td>Immature secretory granule</td>
</tr>
</tbody>
</table>
Inr  Initiator region
IPTG  Isopropyl β-D-thiogalactopyranoside
IVN  Intravacuolar network
Kb  Kilobase(s) or 1000 base pairs
kDa  Kilodalton(s)
LB (broth)  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  Rhotry associated protein
Rhs  Right hand side
RNA  Ribonucleic acid
RNase  Ribonuclease
ROP  Rhotry 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′-tetramethylethylendiamine
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<sub>m</sub>  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-galactopyranoside
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
ParacoxS, 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 sporozoysts 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
### Table 1.1 Taxonomic arrangement of coccidial parasites.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Class</th>
<th>Subclass</th>
<th>Order</th>
<th>Suborder</th>
<th>Family</th>
<th>Subfamily</th>
<th>Genus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apicomplexa</td>
<td>Sporozoisa</td>
<td>Coccidiasa</td>
<td>Eucoccidiordia</td>
<td>Eimeriorina</td>
<td>Eimeridae</td>
<td>Cryptosporidiidae</td>
<td>Caryospora</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Eimeria</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Isospora</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cryptosporidium</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Sarcocystis</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Frenkelia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Besnoitia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Hammondia</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Neospora</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Toxoplasma</td>
</tr>
</tbody>
</table>
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 tufA, 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-
terminus 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  
PCR  Posterior precocoidal ring  
C   Conoid  
OM  Outer membrane of pellicle  
IMC  Inner membrane complex  
PR  Polar ring  
IM  Inner membrane of pellicle  
MN  Microneme  
MT  Microtubule  
R   Rhoptery  
RF  Refractile body  

PT  Protein granule  
MP  Micropore  
MV  Multimembranous vacuole  
V  Vacuole  
G  Golgi body  
N  Nucleus  
PG  Polysaccharide granule  
MI  Mitochondrion  
L  Lipid  
VT  Vesicle with thick wall  
ER  Granular endoplasmic reticulum
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.*, 2002).
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). Immuno-electron 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. yoelii*, 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., 1999b).
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 auxotrophic 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
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.
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 (Wemer-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 RAPI 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 coatamer proteins (α- β- β'- γ- δ- ε- ζ-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-acetylgalactosamine 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-Schintlter 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 Δgra5 *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 \(\Delta gra2\) 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 \(\alpha\)-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 \textit{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 \(\text{Ca}^{2+}\) binding domain known as an EF-hand and it has been postulated that this protein may function as a \(\text{Ca}^{2+}\) buffer, modulating the \(\text{Ca}^{2+}\) 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 \textit{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, \textit{Plasmodium} ROPs are first targeted to a pre-rhoptry organelle (Shaw et al., 1998) and such structures have been observed in \textit{Plasmodium} (Lingelbach and Joiner, 1998).

ROP2 of \textit{T. gondii} contains four tyrosine residues, one of which forms part of the sequence, SEYEQL which is reminiscent of the consensus sequence \(\text{YXX}\phi\) (X, any amino acid, \(\phi\) 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* µ1 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
YXX\dot{\phi} 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\textsuperscript{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 Ca²⁺ 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
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-acetylgalactosamine) 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-acetylglactosamine 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.
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 undergo 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 $1 \times 10^3$ sporulated oocysts of the Wey or Wis strains or $1 \times 10^6$ 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 \(0.25 \times 10^6\) 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 \(\sim 5\) 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 \(2 \times 10^6\) 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 SS 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 SS plus centrifuge. The supernatants were decanted into 20 ml universals and volumes made up to 8 ml with milliQ water. Finally, 8 g of CsCl₂ (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
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 MgSO₄, 10 mM MgCl₂, 2.5 mM KCl, in a 500 ml baffled flask and re-incubated at 37°C, with shaking until the culture reached an A₆₀₀ 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® DNA polymerase, FS, PE Applied Biosystems), 2.0 µl of BD HalfTERM (Genpak Ltd.) and milliQ H2O 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 H2O, 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 H2O 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'-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-35S]dATP (10 μCi/μl), 2.0 μl Thermo sequenase DNA polymerase and sterile H2O 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® (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 H2O 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 α32P dCTP (10 mCi ml−1/3000 Ci mmol−1; Amersham Pharmacia Biotech) using a Prime-it® 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 x 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 H₂O (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 byinsertion 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/cm². 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 $6 \times 10^5$ ml⁻¹, and dispensed into 24 well plates containing sterile coverslips at 0.6 ml per well (giving $3.6 \times 10^4$ cells per well). Plates were incubated at 41°C with 5% CO₂ 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 $1 \times 10^7$ parasites in 700 μl of complete cytomix (Van der Hoff, 1992). For each electroporation, $1 \times 10^7$ 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) and streptomycin (20 µg/ml), dispensed into wells containing a monolayer of MDBK and incubated at 41°C with 5% CO₂, 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 MgCl₂, 0.1 M Na₂HPO₄/NaH₂PO₄ 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.
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 cimerian 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 endoplasmic 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 schizonts 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 schizonts 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 schizonts 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.
such as 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 other 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 chromatography, 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 x 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 x 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 \textit{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 \textit{Eimeria} is asynchronous (Ferguson \textit{et al.}, 1978a; Ferguson \textit{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 \textit{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 \textit{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.
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, 10μm.

**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.
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 ($10^9$) 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.
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).
<table>
<thead>
<tr>
<th>Location of microneme gene</th>
<th>Chromosome number</th>
<th>Size (Mbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td>1</td>
<td>1.05</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>1.20</td>
</tr>
<tr>
<td><em>EtMIC3</em></td>
<td>3 (9 in Wey)</td>
<td>2.0</td>
</tr>
<tr>
<td>-</td>
<td>4</td>
<td>2.2</td>
</tr>
<tr>
<td><em>EtMIC4</em></td>
<td>5/6</td>
<td>3/3.5</td>
</tr>
<tr>
<td>-</td>
<td>7</td>
<td>3.5</td>
</tr>
<tr>
<td>-</td>
<td>8</td>
<td>3.7</td>
</tr>
<tr>
<td><em>EtMIC2 and EtMIC5</em></td>
<td>9</td>
<td>4.5</td>
</tr>
<tr>
<td>-</td>
<td>10</td>
<td>4.5</td>
</tr>
<tr>
<td>-</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>-</td>
<td>12</td>
<td>5.5</td>
</tr>
<tr>
<td><em>EtMIC1</em></td>
<td>13</td>
<td>&gt;6</td>
</tr>
<tr>
<td>-</td>
<td>14</td>
<td>&gt;6</td>
</tr>
</tbody>
</table>

Table 3.1 Chromosome localisations of *EtMIC1-5.*
Fig. 3.6 Fidelity of total RNA.
Oocysts (25 x 10⁶) 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 *EiACTIN* and for each *EiMIC* 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.
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 *EtMIC*1-5 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.
<table>
<thead>
<tr>
<th>Gene</th>
<th>Primers</th>
<th>Sequence (5’ to 3’)</th>
<th>Product size</th>
</tr>
</thead>
<tbody>
<tr>
<td>EtMIC1</td>
<td>Mic23</td>
<td>TTGGTCATGACTGACGGGC</td>
<td>1209 bp</td>
</tr>
<tr>
<td></td>
<td>Tsp5D</td>
<td>GTGCAAGCTTACATAGAATTCTTATTTTGATTCG</td>
<td></td>
</tr>
<tr>
<td>EtMIC2</td>
<td>Mic2rr5.1</td>
<td>GAGCGAAGCAGGACTCTATTG</td>
<td>832 bp</td>
</tr>
<tr>
<td></td>
<td>Mic2rr3.1</td>
<td>ACTCTGCGTGAACTCTCTTCC</td>
<td></td>
</tr>
<tr>
<td>EtMIC3</td>
<td>Mic3g</td>
<td>TGTCGCTGTCAATGACCGCTTGAA</td>
<td>451 bp</td>
</tr>
<tr>
<td></td>
<td>Mic3b</td>
<td>GAGGCGCGCGGCGGACGGCTTGTA</td>
<td></td>
</tr>
<tr>
<td>EtMIC4</td>
<td>Mic4rr5.1</td>
<td>CCACGCTCTTGTGCCAAAAC</td>
<td>1127 bp</td>
</tr>
<tr>
<td></td>
<td>Mic4rr3.1</td>
<td>GAGGGTGTTGTGTGCACAC</td>
<td></td>
</tr>
<tr>
<td>EtMIC5</td>
<td>Pjb6</td>
<td>TTCCGTCAGGGCGTTGTGATAC</td>
<td>400 bp</td>
</tr>
<tr>
<td></td>
<td>Pjb7</td>
<td>ACTTCGTAGCCGAGGGCTG</td>
<td></td>
</tr>
<tr>
<td>EtACTIN</td>
<td>Act1</td>
<td>CTGTGAGAAGAACCGGGTGTCTC</td>
<td>350 bp</td>
</tr>
<tr>
<td></td>
<td>Act8rr</td>
<td>CGTGGAAAAATGCCGGACGAAGAG</td>
<td></td>
</tr>
</tbody>
</table>

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, RAP74 and a smaller subunit, RAP38. 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 TFIIF 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 TFIIID 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” TFIIID to the promoter, controlling the position of the pre-initiation complex, as the components of TFIIID 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+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+1NA/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+1NT-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 TFIIID 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, Theileria and Toxoplasma revealed the presence of RNA pol II and transcription factors TFIIID and TFIIIA.

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
primes 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^32P]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 ^32P 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 Sau3A 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 Sau3A cloning site).

Since genomic DNA had been prepared by partial digestion with Sau3A 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' untranscribed 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 \textit{EtMIC1-5}

In Chapter Three it was shown by RT-PCR that transcripts for \textit{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 \textit{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 \textasciitilde130 bp for each of \textit{EtMIC1} and \textit{EtMIC2} and of \textasciitilde100 bp for \textit{EtMIC3} (Fig 4.2.2). These PCR products were cloned into pGEM-T-Easy vector and approximately 15 clones of each were sequenced to identify the transcription start sites of \textit{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 \textit{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\textsuperscript{+1}CAGT, the transcription start site is designated A\textsuperscript{+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 \textit{EtMIC2} (5'-A\textsuperscript{3(+1)}TTA\textsuperscript{+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 \textit{EtMIC3} (CTA\textsuperscript{+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 \textit{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 than 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 ≥ 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
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.
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^5^ (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.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Sequence (5’ – 3’)</th>
<th>Bp to ATG start codon</th>
</tr>
</thead>
<tbody>
<tr>
<td>EtMIC1</td>
<td>TCTCA^3^CAGT</td>
<td>110</td>
</tr>
<tr>
<td>EtMIC2</td>
<td>GA^3+(+1)TTA^1CTCCT</td>
<td>94 and 97</td>
</tr>
<tr>
<td>EtMIC3</td>
<td>ACCTA^2^TTTTT</td>
<td>56</td>
</tr>
<tr>
<td>Gene</td>
<td>Primers</td>
<td>Sequence (5’ to 3’)</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>EtMIC1</td>
<td>Sp1</td>
<td>GCACCTTCTGGATCTGATGATGGCC</td>
</tr>
<tr>
<td></td>
<td>Sp2</td>
<td>CTACCAACATGACATCCAGAGGCC</td>
</tr>
<tr>
<td></td>
<td>Sp3</td>
<td>GCCAGAGCTGATGTTGCGCCGTCGG</td>
</tr>
<tr>
<td>EtMIC2</td>
<td>Sp1</td>
<td>GTTCATTTTGATGTTCTCATCCACCG</td>
</tr>
<tr>
<td></td>
<td>Sp2</td>
<td>GCCCTGCCGAGTTACCTTCATCTAG</td>
</tr>
<tr>
<td></td>
<td>Sp3</td>
<td>GTTCCGCGCATCTCTCCACACTGAG</td>
</tr>
<tr>
<td>EtMIC3</td>
<td>Sp1</td>
<td>GCCGAGCTCAAAATTGTAGAGCCG</td>
</tr>
<tr>
<td></td>
<td>Sp2</td>
<td>CGAGCTAATGGCTAAAAAAGAGTC</td>
</tr>
<tr>
<td></td>
<td>Sp3</td>
<td>CACAGCAGCGCAGTCGCGAGGG</td>
</tr>
<tr>
<td>EtMIC4</td>
<td>Sp1</td>
<td>GTTTGAATGGCGCTCGATCGAGTCC</td>
</tr>
<tr>
<td></td>
<td>Sp2</td>
<td>GCACTGACAAATGCGCTCCGGAGGG</td>
</tr>
<tr>
<td></td>
<td>Sp3</td>
<td>CACTGCTACAAATCGAGGCGG</td>
</tr>
<tr>
<td>EtMIC5</td>
<td>Sp1</td>
<td>CGCTGCTCATTTTCCCTGCTTGG</td>
</tr>
<tr>
<td></td>
<td>Sp2</td>
<td>GAACACCAAAATGGCATCAGCTCC</td>
</tr>
<tr>
<td></td>
<td>Sp3</td>
<td>GATGAGGCTGCAGGTGCGCGCTCC</td>
</tr>
<tr>
<td></td>
<td>Oligo d(T) primer</td>
<td>GACCACCGCTATGCTGACTTTTTTTTTTTTTTTTTTTTTTTT</td>
</tr>
<tr>
<td></td>
<td>PCR anchor primer</td>
<td>GACCACCGCTATGCTGACTGAGG</td>
</tr>
</tbody>
</table>

**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 GTGTGTG 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^+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^+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 *TamSl* 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.
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.
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 5', 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 a2 tubulin*, *a-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 NTPase1 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 SAGl play a role in the transcription initiation process (Soldati and Boothroyd, 1995). It has also been suggested that they act like SPI 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 SPI and are usually found in tandem repeats and in either orientation. The purine rich repeats of SAGl stimulated expression when introduced into a heterologous promoter (TUBl) 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 SPI-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 TamSl 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'-TTTGTAGG-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 TamSl 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.
### 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. lambia* 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.

<table>
<thead>
<tr>
<th>Pol II promoter</th>
<th>Inr region</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>T-cell receptor, variable chain β (Vβ)</td>
<td>ACTCTCAGTTTCT</td>
<td>Anderson et al., 1988</td>
</tr>
<tr>
<td>Lymphocyte function associated antigen 1 (LFA 1α)</td>
<td>ATGATCAGTTTCC</td>
<td>Nueda et al., 1993</td>
</tr>
<tr>
<td>Mouse acetylcholinesterase (ACHE)</td>
<td>GCTGTCAGTTTG</td>
<td>Li et al., 1993</td>
</tr>
<tr>
<td>Human and rabbit acetylcholinesterase (BCHE)</td>
<td>TTTGTCAGTAACA</td>
<td>Jbilo et al., 1994</td>
</tr>
<tr>
<td>Human amyloid β-protein precursor (APP)</td>
<td>TCCGTCAGTTTCC</td>
<td>Quitschke et al., 1996</td>
</tr>
<tr>
<td><em>G. lambia</em>, α-giardin (α-gar)</td>
<td>CATTTAGAAAATG</td>
<td>Elmendorf et al., 2001</td>
</tr>
<tr>
<td><em>G. lambia</em>, δ-giardin (δ-gar)</td>
<td>CCAGAAAAAAATG</td>
<td>Elmendorf et al., 2001</td>
</tr>
<tr>
<td><em>G. lambia</em>, α2-tubulin (α2-tub)</td>
<td>GAAATRAAAATG</td>
<td>Elmendorf et al., 2001</td>
</tr>
<tr>
<td><em>T. gondii</em>, surface antigen 1 (SAG1)</td>
<td>GTAGTCAGTTTG</td>
<td>Soldati &amp; Boothroyd, 1995</td>
</tr>
<tr>
<td><em>T. gondii</em>, nucleotide triphosphate hydrolase 1 (NTP1)</td>
<td>CGTTTCAGTTTTT</td>
<td>Nakaar et al., 1992</td>
</tr>
<tr>
<td><em>T. gondii</em>, nucleotide triphosphate hydrolase 3 (NTP3)</td>
<td>CGTTTCAGTTTTT</td>
<td>Nakaar et al., 1992</td>
</tr>
<tr>
<td><em>T. annulata</em>, merozoite surface antigen 1 (TamS1)</td>
<td>CACTTCAGTTTATA</td>
<td>Sheils et al., 2000</td>
</tr>
<tr>
<td><em>E. histolytica</em>, Gal-GalNAc-specific lectin (hgl5)</td>
<td>AGAAAGAGAAAAT</td>
<td>Singh et al., 1997</td>
</tr>
<tr>
<td><em>E. tenella</em>, microneme protein 1 (EtMIC1)</td>
<td>CGTCTCACAGTAG</td>
<td></td>
</tr>
<tr>
<td><em>E. tenella</em>, microneme protein 2 (EtMIC2)</td>
<td>AAGATTACTCTTT</td>
<td></td>
</tr>
<tr>
<td><em>E. tenella</em>, microneme protein 3 (EtMIC3)</td>
<td>CACCTTATTTTCC</td>
<td></td>
</tr>
<tr>
<td>Consensus</td>
<td>PyPyANT/PyPyPy</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Sequence</td>
<td>Direct/Indirect</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>GRA1</td>
<td>CGACACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGACG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>CGTGACG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>TGAGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGACG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>CGAGACG</td>
<td>I</td>
</tr>
<tr>
<td>GRA2</td>
<td>TGCGACG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>AGAGACG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>AGAGATG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>AGAGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AGAGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AGCGACG</td>
<td>I</td>
</tr>
<tr>
<td>GRA5</td>
<td>AGACACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AGAGACG</td>
<td>D</td>
</tr>
<tr>
<td>GRA6</td>
<td>AGACACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGCGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AGCGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AGCGACG</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>TGCGACG</td>
<td>D</td>
</tr>
<tr>
<td>TUB1</td>
<td>GGAGATG</td>
<td>D</td>
</tr>
<tr>
<td>SAG1</td>
<td>TGAGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGACCG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGCG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGCG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGCG</td>
<td>D</td>
</tr>
<tr>
<td>EIMIC1</td>
<td>AGAGACG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGACG</td>
<td>I</td>
</tr>
<tr>
<td>EIMIC2</td>
<td>AGAGACG</td>
<td>I</td>
</tr>
<tr>
<td>EIMIC3</td>
<td>TGAGAGG</td>
<td>D</td>
</tr>
<tr>
<td>EIMIC4</td>
<td>GGTCTCA</td>
<td>I</td>
</tr>
<tr>
<td></td>
<td>CGTCTCT</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>AGAGAGG</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>TGAGAGC</td>
<td>D</td>
</tr>
<tr>
<td></td>
<td>GGTCTCA</td>
<td>I</td>
</tr>
<tr>
<td>EIMIC5</td>
<td>CGAGAGG</td>
<td>D</td>
</tr>
</tbody>
</table>

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 x 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 *KpnI* 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 *KpnI* 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
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 *KpnI* 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.
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 EtMICl plasmid constructs used in transient transfection experiments.

Filled boxes represent 5’ upstream sequences of the microneme gene, EtMICl. The open box illustrates coding regions of the lacZ gene and the thin line designates pSV-β-gal 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 EtMICl), termination codons (Stop, from pSV-β-gal vector) and proposed polyadenylation signals (Poly A, from pSV-β-gal vector). The flag represents the transcription start site. Genetic elements are not drawn to scale.
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.
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.
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.
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.
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-ßgal. 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.
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.
Functional analysis of *E*.*tenella* *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-β-gal. 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.
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). SP1 recognise 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$^+$ 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$^+$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 EndMIC1 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 $1.4 \times 10^7$ 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 KpnI restriction sites included. The amplified products of EtMIC1 (685 bp) and EtMIC2 (1045 bp) were digested using 5 U of the restriction endonuclease KpnI 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 KpnI 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-IA or MUTA-IB) 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).
<table>
<thead>
<tr>
<th>PCR number 1</th>
<th>PCR number 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Template DNA</td>
<td>Template DNA</td>
</tr>
<tr>
<td>50 ng</td>
<td>50 ng</td>
</tr>
<tr>
<td>MUTA-1A (2.5 pmol/µl)</td>
<td>MUTA-1B</td>
</tr>
<tr>
<td>2.5 µl</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>M13 reverse primer</td>
<td>M13 - 20 primer</td>
</tr>
<tr>
<td>2.5 µl</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>100 mM dNTP set</td>
<td>100 mM dNTP set</td>
</tr>
<tr>
<td>2.5 µl</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>1 µl</td>
<td>1 µl</td>
</tr>
<tr>
<td>10x <em>Pfu</em> polymerase buffer</td>
<td>10x <em>Pfu</em> polymerase buffer</td>
</tr>
<tr>
<td>5 µl</td>
<td>5 µl</td>
</tr>
<tr>
<td><em>Pfu</em> polymerase (2-3 U/µl)</td>
<td><em>Pfu</em> polymerase</td>
</tr>
<tr>
<td>0.5 µl</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>H₂O to a final volume of</td>
<td>H₂O to a final volume of</td>
</tr>
<tr>
<td>50 µl</td>
<td>50 µl</td>
</tr>
</tbody>
</table>

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
<table>
<thead>
<tr>
<th>Construct</th>
<th>Sequence 5' – 3'</th>
<th>% of blue parasites compared to WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>ATC GTC TCA*1 CAG TAG</td>
<td>100</td>
</tr>
<tr>
<td>Mic1mut2</td>
<td>ATC GTC TCG*1 CAG TAG</td>
<td>85</td>
</tr>
<tr>
<td>Mic1mut3</td>
<td>ATC GTC TCC*1 CAG TAG</td>
<td>62</td>
</tr>
<tr>
<td>Mic1mut4</td>
<td>ATC GTC - - - *1 CAG TAG</td>
<td>28</td>
</tr>
<tr>
<td>Mic1mut5</td>
<td>ATC GTC TC - *1 - AG TAG</td>
<td>81</td>
</tr>
<tr>
<td>Mic1mut6</td>
<td>ATC GTC TTA*1 CAG TAG</td>
<td>70</td>
</tr>
<tr>
<td>Mic1mut7</td>
<td>ATC GTC TAA*1 CAG TAG</td>
<td>52</td>
</tr>
</tbody>
</table>

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.
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 $\sim$28% (Fig. 6.2), when compared with the non-mutated construct, suggesting a preference for a TCA$^{+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 $\sim$19% in the number of blue parasites observed (Fig. 6.2). Removal of both the transcription start site and the following nucleotide would have expected to yield a greater reduction, however further analysis of the new mutated construct revealed the production of a new TCA motif. Due to the low reduction in blue parasite numbers it is highly likely that the transcription start site of EtMIC1 moved from position A$^{+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 $\sim$30% in the number of blue parasites was found (Fig. 6.2). However when the cytosine nucleotide was converted to an adenosine nucleotide (Mic1mut7; TATAA$^{+1}$CAG) the negative effect on the number of blue parasites was greater, with a reduction to $\sim$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 EtMICl 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\textsuperscript{3} did not conform to an Inr consensus as it only contained two pyrimidines. However A\textsuperscript{41} 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\textsuperscript{3} as a transcription start site, constructs which contained a deletion of the adenosine and the preceding guanine (G\textsuperscript{4}) were generated. When this mutated construct was introduced into the transfection system (Mic2mut1; GAA\textsuperscript{4}-TTA\textsuperscript{41}) 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\textsuperscript{5} and A\textsuperscript{6} (Mic2mut2; G--GA\textsuperscript{3}-TTA\textsuperscript{41}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\textsuperscript{3}, (Mic2mut5; GAAG\textsuperscript{3}-TTA\textsuperscript{41}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\textsuperscript{3}-ATA\textsuperscript{41}C), which had both nucleotides at positions at T\textsuperscript{2} and G\textsuperscript{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\textsuperscript{1} and A\textsuperscript{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 another transcription start site within the Inr motif. Together with the 5'RACE data these studies further suggest that the A\textsuperscript{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.
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\textsuperscript{7} and G\textsuperscript{+11} (Inr) or between T\textsuperscript{103} and C\textsuperscript{-86} (CAAT) are shown. Pmic2\textgreek{p}gal5'1046 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\textsuperscript{+1} and A\textsuperscript{3} are indicated and the Inr motif with the sequence TTA\textsuperscript{+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.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Sequence (5' - 3')</th>
<th>% of blue parasites compared to WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>GAA GA\textsuperscript{T} TA\textsuperscript{+1}C TCC TTG CTT</td>
<td>100</td>
</tr>
<tr>
<td>Mic2mut1</td>
<td>GAA - - \textsuperscript{T} TA\textsuperscript{+1}C TCC TTG CTT</td>
<td>34</td>
</tr>
<tr>
<td>Mic2mut2</td>
<td>G - - GA\textsuperscript{T} TA\textsuperscript{+1}C TCC TTG CTT</td>
<td>56</td>
</tr>
<tr>
<td>Mic2mut5</td>
<td>GAA G - \textsuperscript{T} TA\textsuperscript{+1}C TCC TTG CTT</td>
<td>50</td>
</tr>
<tr>
<td>Mic2mut6</td>
<td>GAA AA\textsuperscript{A} TA\textsuperscript{+1}C TCC TTG CTT</td>
<td>58</td>
</tr>
<tr>
<td>Mic2mut8</td>
<td>GAA GA\textsuperscript{T} T - - - - - - \textsuperscript{+7} G CTT</td>
<td>26</td>
</tr>
<tr>
<td>WT</td>
<td>TAG TGC\textsuperscript{98} AAT\textsuperscript{-95} TGT GCT AGC</td>
<td>100</td>
</tr>
<tr>
<td>Mic2mut9</td>
<td>TAG TG - -\textsuperscript{98} - -\textsuperscript{-95} TGT GCT AGC</td>
<td>35</td>
</tr>
</tbody>
</table>
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.
Construct Mic2mut8 (GAAGA⁻³TT⁺¹---⁺⁷G) 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 ---⁻⁹⁵TGT). 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).
Table 6.3 Analysis of mutated *EtMICl* promoter (series 2).
Alignment of mutational sequences surrounding various potential protein binding regions of *EtMICl*. All constructs contained 613 bp of the *EtMICl* 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.

<table>
<thead>
<tr>
<th>Construct</th>
<th>Sequence (5′ - 3′)</th>
<th>% of blue parasites compared to WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>TAAACCATGCTTGGAGGAAACTTTGCTTTT</td>
<td>100</td>
</tr>
<tr>
<td>Mic1-C</td>
<td>TAAACCATGCTTGGAA - -ACTTTGCTTTT</td>
<td>100</td>
</tr>
<tr>
<td>Mic1-N</td>
<td>TAAACCATGCTTGGAG - -CTTTGCTTTT</td>
<td>100</td>
</tr>
<tr>
<td>Mic1-D</td>
<td>TAAACCATGCTTGGAGGAAA - -GCTTTT</td>
<td>50</td>
</tr>
<tr>
<td>Mic1-S</td>
<td>TAAACCATGCTTGGAA - -ACTTTGCTTTT</td>
<td>55</td>
</tr>
<tr>
<td>Mic1-H</td>
<td>TAAACCATGC - -AGGAAACTTTGCTTTT</td>
<td>100</td>
</tr>
<tr>
<td>Mic1-E</td>
<td>T - - - CATGCTTGGAGGAAACTTTGCTTTT</td>
<td>100</td>
</tr>
</tbody>
</table>
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
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; TCTG*¹*CAG) is tolerated better than a pyrimidine (Mic1mut3; TCTC*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; Zenie-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*¹²* in the *EtMIC1* Inr motif (Mic1mut5; TCTC*¹*CAG) 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\textsuperscript{+3}. These results demonstrate the requirement for a TCA\textsuperscript{+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\textsuperscript{-1} in the Inr element of EtMIC1 to either a thymine (Mic1mut6; TATTA\textsuperscript{+1}CAG) or adenosine (Mic1mut7; TATAA\textsuperscript{+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\textsuperscript{-1}ATA\textsuperscript{+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\textsuperscript{-98}---\textsuperscript{-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 KSTI
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 hg15 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 \textit{EtMIC1}, as there is a single repeat and its presence or absence did not alter expression (Chapter Five).

The potential transcription start site of \textit{EtMIC1} was mapped to an adenosine nucleotide within the sequence TCA$^{-1}$CACGT (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 \textit{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 \textit{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 \textit{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 \textit{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 \textit{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 \textit{E. tenella} can be replaced \textit{in vivo} by a heterologous Inr such as the Inr of \textit{SAG1} of \textit{T. gondii}, or \textit{vice versa}, and examine if the two elements perform homologous functions.

From the structural analysis of the \textit{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 \textit{EtMIC1}, 2, 3 as identified by 5'RACE and \textit{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 \textit{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 \textit{EtMIC4} was determined it would also be possible to determine if the consensus TATA and CAAT boxes of \textit{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 transcription.
References


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*.


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


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


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


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)


Appendix A

Alkaline phosphate buffer (x 1)
Tris 100 mM
NaCl 100 mM
MgCl₂ 10 mM
Adjusted to pH 9.5 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
H₂O 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
H₂O 1 ml
Filter sterilised and stored in aliquots at −20°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
MgCl₂ 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$_4$Fe(CN)$_6$·3H$_2$O, 3.3 mM
- K$_3$Fe(CN)$_6$, 3.3 mM
- Na$_2$HPO$_4$, 60 mM
- NaH$_2$PO$_4$, 40 mM

**Beta-galactosidase staining solution**
- 2% Xgal in dimethyl formamide
- Yellow solution

**Bjerrum/Schafer-Niehran transfer buffer**
- Tris 48 mM
- Glycine 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 brilliant blue (CBB) stain**
- 0.25 g
- Coomassie brilliant blue R-250 dye (Sigma) 90 ml
- Methanol:H$_2$O (1:1v/v) 10 ml
- Glacial acetic acid
- Filtered through No.1 Whatman paper.

**Cytomix (incomplete):**
- K$_2$HPO$_4$/KH$_2$PO$_4$ 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$_2$O 33.3 %
DTT  (DL-Dithiothreitol; Sigma)
2.35M in 400mM KAc, pH6.0

Hybridisation solution
A: Phosphate buffer  \( \text{Na}_2\text{HPO}_4 \)  134 g/l
     85\% (v/v) \( \text{H}_3\text{PO}_4 \)  4 ml/l
     \( \text{H}_2\text{O} \)
B: SDS solution  10\% \( \text{NaDodSO}_4 \)

IPTG stock solution (Sigma)
IPTG  23.2mg
\( \text{H}_2\text{O} \)  1.0ml
Filter sterilised 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
\( \text{H}_2\text{O} \)
Autoclaved for sterilisation.

LB-broth
Bacto-tryptone  10 g/l
Bacto-yeast-extract  5 g/l
NaCl  10 g/l
\( \text{H}_2\text{O} \)
Autoclaved for sterilisation.

Loading dye
Formamide  98\%
EDTA  10 mM
Xylene cyanol  0.1\%
Bromophenol blue  0.1\%

Lysis buffer
100 mM \( \text{K}_2\text{HPO}_4 \)  9.15 ml
100 mM \( \text{KH}_2\text{PO}_4 \)  0.85 ml
1 M DTT  10 \( \mu \)l

MgKCl
\( \text{MgSO}_4 \)  1 M
\( \text{MgCl}_2 \)  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-lauroylsarcosine 1 %

PBS (pH 7.4, 7.6 or 8.0)
Na$_2$HPO$_4$ 1.44g/l
KH$_2$PO$_4$ 0.24g/l
KCl 0.20g/l
NaCl 8.00g/l
H$_2$O
Adjusted to the required pH with HCl and autoclaved for sterilisation.

Protease inhibitor cocktail (Sigma P2714)
Contains 4-(2-aminoethyl)benzenesulfonyl fluoride,
trans-epoxysuccinyl-$\gamma$-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: Citric 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$_2$O

SDS-PAGE (10% acrylamide gel)
H$_2$O 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)**

- H₂O: 0.68ml
- 30% acrylamide mix: 0.17ml
- 1.0M Tris (pH 6.8): 0.13ml
- 10% SDS: 0.01ml
- 10% ammonium persulphate: 0.01ml
- TEMED: 0.001ml

**SOC medium**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB medium</td>
<td>10 ml</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>10 mM</td>
</tr>
<tr>
<td>MgSO₄</td>
<td>10 mM</td>
</tr>
<tr>
<td>KCl</td>
<td>2.5 mM</td>
</tr>
<tr>
<td>Glucose</td>
<td>20 mM</td>
</tr>
</tbody>
</table>

**Southern blotting solutions**

**Denaturing buffer:**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>1.7 M</td>
</tr>
<tr>
<td>NaOH</td>
<td>0.5 M</td>
</tr>
<tr>
<td>Ammonium acetate</td>
<td>10 M</td>
</tr>
</tbody>
</table>

**Neutralising buffer:**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaOH</td>
<td>5 mM</td>
</tr>
</tbody>
</table>

**2 x SSC**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>300 mM</td>
</tr>
<tr>
<td>Tri-NaCitrate.2H₂O</td>
<td>3 mM</td>
</tr>
<tr>
<td>Adjust to pH 7.0.</td>
<td></td>
</tr>
</tbody>
</table>

**TBE buffer x10 (Tris-Borate-EDTA)**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris base</td>
<td>108 g/l</td>
</tr>
<tr>
<td>Boric acid</td>
<td>55 g/l</td>
</tr>
<tr>
<td>0.5M EDTA (pH 8.0)</td>
<td>20 ml/l</td>
</tr>
<tr>
<td>H₂O</td>
<td></td>
</tr>
</tbody>
</table>

**TFB (transformation buffer)**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>K-MES (pH 6.2-6.3)</td>
<td>10 mM</td>
</tr>
<tr>
<td>RbCl or KCl</td>
<td>100 mM</td>
</tr>
<tr>
<td>MnCl₂</td>
<td>45 mM</td>
</tr>
<tr>
<td>CaCl₂</td>
<td>10 mM</td>
</tr>
<tr>
<td>HACoCl₂</td>
<td>3 mM</td>
</tr>
</tbody>
</table>

**T4 Polynucleotide kinase buffer, 10x (Promega)**

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tris-HCl</td>
<td>500 mM</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>100 mM</td>
</tr>
<tr>
<td>DTT</td>
<td>50 mM</td>
</tr>
<tr>
<td>Spermidine</td>
<td>1 mM</td>
</tr>
<tr>
<td>Adjusted to pH 7.5</td>
<td></td>
</tr>
</tbody>
</table>
X-Gal stock solution
X-Gal (Sigma) 20 mg
Dimethyl formamide 1.0 ml

Bacterial strains
XL1-Blue MRF': Obtained from Stratagene, has a genotype of $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA1 lac [F' proAB lacI$^\delta$ $\Delta M15$ Tn10 (Tet')]. Used for the propagation of high efficiency cloning and antibiotic selection to allow expression of F' episome.

XL1-Blue MR: Obtained from Stratagene has a genotype of $\Delta(mcrA)183 \Delta(mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA96$ gyrA1 lac and was used for the propagation 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.
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
The *EtMCI* genomic clone was identified by PCR using primers T3/r7 and miclrr5. The genomic clones were sequenced using primers miclrr5, miclrr5.7, miclrr5.8 and miclrr5 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)N/A/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 histolytica* 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 GTGTGTG (-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 AAATCGATGG TTTTACAAGGC TCCGGGCTAA ACGGCCAGAC AGAGTGGTAA

mic1rr5.10

-2324 ATGACGCAGCT GCATTTCGAG GGCCCTCCACG CAGCTATGTA ATAAAGGGCAA

-2274 GTCTCAGTTC GACGAACGGGT GATATCTGGG AGGGCAGGTTT ACGCGCCACT

-2224 AGTGCTTGGCT TCAAAGAACCA CGCGCATGGG AGAAGGCCCTG GTTGCCATAG

-2174 GTAATTGAAA GCAGACGGGAG GCTGGAATAA GCTACCATT TCTCTGTATG

-2124 TAGAGAAACCC CTCCTCGTGG TAATGCGAAGCCA CTAAAGAGT AACGCCCTTG

-2074 AGACCGACAC ATAGTTGTTG TTATGTATTC TAAACACGGG CGGTTCTGT

-2024 CTGCCGTTTCA TAATAGAGCT GAGGCAGACT CAGTACAACA GGCTCCCGAT

-1974 GAGGGGGCTGG GACGGAGGCCA GCTGGAAGCC TATGCGCCGT AGCTACCCGT

-1924 GACACAAATTC AAAACGGTGA CCGGGATTTCC AGTCGATGTC TTGTAACTTC

-1874 CACCACGCGCC AGGGGACGGGT AGTCCGACGG TAAATGACAG CTCGCTACCA

-1824 TTGACTGGAC GCAGAGGATT TAAATGAAGA AGGAAATTCG TACAAAATTC

-1774 AACTAAGCAG TTTAGAGACT TGTTCTGATG TTGTGCGAGA GAATCCGGTA

-1724 GGGAGAGGGG TACCTCGGCT TCAGCGCTCG TGCTGAGCAGG

-1674 CGGTTTGGAG AACGACAGGC ACAATGTAAC AAGCAGAGCC TACGGCACAG

-1624 GTGAGGGGGCC TGACGTTGCA AATTTGTTCC ATTTCTGGAA AGCAGGGCGC

mic1rr5.8

-1574 TGCTTTATATA ACGCAACCAA TTAGTTTGTC GGGAGAGTGC CTGGGACCTT

-1524 CTATCTACCT GCCGCTCTTA GTTGCGGACT ACAGGTAACC AGCGAGCAGC

-1474 AAAGGGCCTTT ACCTTTCTGG GCGGAGTTTT ACCAGGAGGG TGCGGCGATCC

-1424 AGTAAGAAGC CACCTCGGTT GCCTCCAAAA ATGTTGCAAG TCCCTTCTTG

-1374 TCAGCTGAAA CCGTTTTGTT GAGTCAAGGA AAGCAGCCAC CAGAACTGTA

-1324 ATTACTAAGG AGGGGCAAAT TGGAAGCAAA CAATACGGCT GACGAATTGT

-1274 GCTCACACGG TTCTTAGCTG GTCATAGAAAG GAGAGGGAGG TGACGGAACCT

-1224 GTGGAGACAA AGAAATTTGA GACGGACGGGT TAAACCTCAG TGTAGCTGGA

-1174 GGGAAGTGGAA AGAAGACAGA AGGAACAAAT GAGAGTGTTT TGAACACCTC

-1124 AAGCTGCTAC TGATAGCTCC ACAGACCCAG TCAGCAAGGG TGATAGAGG

-1074 AGCCAGCTAG AGAGTCCACA GGATTCATTG TGAATGATGC GTCAATACT

mic1rr5.7

-1024 GGAGCGTTCC CTATAAGAGG GACGCGGAGG CCCCCTGGTC ACAGTTGTG

-974 CACGCCCAAC CATTCAATC CAGTAAAAT TACAGGAAA 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 histolytica 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 ACTGPCAACA GTTGTGAGGG
-2191 TCAATTTGGT CAAAACTGTGT ACCATTAGCT GACCTGGCTT GTAGGTGCGG
-2141 GTTACGAATG GGACTGGAGA CTTCTGCGGA GACCGAAAAT AACACTGTGC
-2091 AGAGTAACGA TTGCGCTAAG CAAAACCTTGC AAAAGCCGTGC ATGGAACTGT
-2041 GTTACGAATG GGACTGGAGA CTTCTGCGGA GACCGAAAAT AACACTGTGC
-1991 AGGAACATGG TGGTACCACC AGTTTCAAGCA AGTTGTCCTCC
-1941 AGGGACCAGC AGACAGAAAG CAGCTTGCGA AATTGATTATG CAACACTTAA
-1741 AAAAGCGCCA GCTTTGGAGA GACCTGGGAGA AACTCTTGGGA
-1691 GATTACAGTC CTAGATCATTG AAAGGATAGT GCATTGCACT
-1641 AGCTGAGATC CTGGCTACGC TCTTGAGAGC AGGGTTTGGGA GCCGATCTTG
-1591 AGGTGAAACAG AGATCTTGCTG GTGTAGTACCC TCCAGGACAG TAGCCAAAGAG
mic3rr5ah ←
-1541 TAAACAAACCG TTGCACTGTA GCTGAACGCG ACCGTAGGAG AGAGATTAGC
-1491 AGGAGGAGAAA AGAGGGCAAGA AGATTAACGG TCTGCCATTT AGCAACAGTT
-1441 GCTTAAGCAT CAGACTAATA AGGTGCTAAG CTACAGTACT TAGGCGGTGT
-1391 AGCTTGCTGA CGGCCATCGTT CTAGTTTAAGC AGGCGAATAC CCCCCGACAC
-1341 AGCTGAGATC CTGGCTACGC TCTTGAGAGC AGGGTTTGGGA GCCGATCTTG
-1291 GATATTCCAT TATATGGTAG AGAAAGACAG CAGAGGAAAT TCTGCTGAAA
-1241 ATAGATTCGA TCTGCGCAGT TTTAATCTAC TCCGCTTTCT AGGAGACGCC
-1191 ACTGGCAAGT GTGTGAAGCG CCCTTTACTG TAAGACAGGG
-1141 GTGGTAACGC GGGAAACGCC AGTCACTCGG AAAGTGTCGCC GGGATGCGCG
-1091 GAGGGACGCAG CGGAGTCGCT AAGTCTTCTC AAGTCAAGAG
-1041 ACTCTGCAGTC TTTTTTCTTC TGTGGCCCTT GCCAGGCCCTG
mic3rr5ah ←
-991 GCTTTTTGCA GAAATAATATG AGCCTGGCAC TCTGCAAGAGA AAAGCAAAAT
-941 TGGTTGCCATT GACGCGGCAG GCCCATCCTA TCTACAGGGA CATCTCTTTT
mic3rr5ah ←
-891 ACACAGTCAA GCTGTGATGT CCTGCTTCTAA CTCCTGAAGCA AGACGGCCAGA
-841 GTGGCAGAAG ATATTAAACAG ACGCGTATGC TTTGCTTTCGT ACAGATTCGTT
-791 CATACTGAAT GAGCAGATCC AAGAAGAAAG TCCCTTTTGC AGAGCAACAA
-741 CAGCCGACCG TGGCTGATGT TGGCTGAGGT CAGCCCATCT ATGTGATGCT
-691 TAACCTTTCA TATGTTATCC ACCGAGCAT TATGTTATCC TCGATATTTA
-641 TAGATTGCC TAAAGCTTAA ATGGATAGGA CGGATACCTA CGTGAACAGCA
-591 TAGCGATTGC GGGAAATGCGG CAGTTGACGC GTTCCTGCGT GACAATCGG

223
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, mic3rr5af and mic3rr5ah and are shown in red type, with an arrow indicating the direction of sequencing. A degenerate TATA box (-195) and CAA T box (-211) are shown boxed and underlined, respectively. The initiator element, with bases confirming to the PyPyA(+1)NATPyPy 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 histolytica 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).
The *EtMIC4* genomic sequence was identified by sequence analysis through the *E. tenella* genome sequencing project. A consensus TATA (-1452) and CAAAT (-1467) boxes are shown boxed in and underlined, respectively. Also the predicted transcription start site is between -1437 to -1440, and the potential Inr

Fig. B.4 Upstream genomic sequence of *EtMIC4*
element which conforms to the PyPyA-N/A/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. histolytica 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, TamSl, 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), TGAGAC (-1586, direct orientation) and GGTCTCA (-2426, indirect orientation) and all are shown with a blue background.
Fig. B.5 Upstream genomic sequence of EtMIC5

The EtMIC5 genomic clone was identified PCR using primers T3/T7 and mic5r5. The genomic clones were sequenced using primers mic5r5, mic5r5.6, mic5r5.7, mic5r5.8 and mic5r5.9 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; Javaheiy 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 CACAAG motif which has been observed in the upstream region of T. annulata gene, TamSI, is shown with a yellow background at position -740 (Shiels et al., 2000).
## Appendix C

<table>
<thead>
<tr>
<th>Name</th>
<th>Size (bp)</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3</td>
<td>16</td>
<td>TTC CCT TTA GTG AGG G</td>
</tr>
<tr>
<td>T7</td>
<td>19</td>
<td>TAA TAC GAC TCA CTA TAG GG</td>
</tr>
<tr>
<td>SP6</td>
<td>20</td>
<td>TTC TAT AGT GTC ACC TAA AT</td>
</tr>
<tr>
<td>Mic1rr5</td>
<td>20</td>
<td>GTT TTA GCA AGG CTC CGG GC</td>
</tr>
<tr>
<td>Mic1rr5.7</td>
<td>20</td>
<td>GTC CTC TTA TTA GGA ACG TC</td>
</tr>
<tr>
<td>Mic1rr5.8</td>
<td>23</td>
<td>CGT TAT ATA AGC AAG CGC CGG CT</td>
</tr>
<tr>
<td>Mic1rr5.10</td>
<td>20</td>
<td>GCA TTT GCA GGG CCT CCA ACG</td>
</tr>
<tr>
<td>Mic2rr5</td>
<td>20</td>
<td>CGA CGA AGA ATA TTC ACA TG</td>
</tr>
<tr>
<td>Mic2rr5.5</td>
<td>21</td>
<td>GAC ACA AAG TCC AG TCA CAC</td>
</tr>
<tr>
<td>Mic2rr5.6</td>
<td>21</td>
<td>GAG CAG CTG TAC TTA ACC GTC</td>
</tr>
<tr>
<td>Mic2rr3</td>
<td>20</td>
<td>GGC TGC TGA CCA GTC CTT GCT</td>
</tr>
<tr>
<td>Mic2rr3a</td>
<td>20</td>
<td>GAT GGC TCT GGG AGA AAG TA</td>
</tr>
<tr>
<td>Mic2rr3.4</td>
<td>20</td>
<td>GCA CACATG CCA ACC ACA GC</td>
</tr>
<tr>
<td>Mic3rr5</td>
<td>21</td>
<td>CAG GGC GTT TGT CAC GAA TGC</td>
</tr>
<tr>
<td>Mic3rr5aa</td>
<td>20</td>
<td>GCG TAG CTG TGC CAG GTT GC</td>
</tr>
<tr>
<td>Mic3rr5ab</td>
<td>21</td>
<td>CTG CCG CAC TGG GGC GTC TGC</td>
</tr>
<tr>
<td>Mic3rr5ae</td>
<td>21</td>
<td>CTC TCC ACT ACG GTG CAG TTC</td>
</tr>
<tr>
<td>Mic3rr5af</td>
<td>20</td>
<td>CTG TGT GAC TCA TGC ACT AG</td>
</tr>
<tr>
<td>Mic3rr5ah</td>
<td>20</td>
<td>GCC TTT GCG AGA AAT GAA TG</td>
</tr>
<tr>
<td>Mic3rr3</td>
<td>20</td>
<td>GCT TGC ACA ATT TGT CGA GG</td>
</tr>
<tr>
<td>Mic3rr3b</td>
<td>21</td>
<td>CGA CGA AGA ATA TTC ACA TGA</td>
</tr>
<tr>
<td>Mic3rr3.9</td>
<td>22</td>
<td>GTG CCT AGC GCC ATT CTG CAT G</td>
</tr>
<tr>
<td>Mic3rr3.11</td>
<td>20</td>
<td>GGC TTT GTC GTA TAC CCT AC</td>
</tr>
<tr>
<td>Mic4rr5</td>
<td>20</td>
<td>CTC GGA AAC CCG GTT TGC AC</td>
</tr>
<tr>
<td>Mic4rr5a</td>
<td>20</td>
<td>GCA AGG TCG CTG GGT GAC CA</td>
</tr>
<tr>
<td>Mic4rr5.10</td>
<td>18</td>
<td>CGG ATC AAA GGC GAG TAC</td>
</tr>
<tr>
<td>Mic4rr5.11</td>
<td>21</td>
<td>CAG CGG TAA GAT CCT TGA GAG</td>
</tr>
<tr>
<td>Mic4rr5.12</td>
<td>19</td>
<td>GTC TCA TGA GCG GAT ACA T</td>
</tr>
<tr>
<td>Mic4rr5.13</td>
<td>20</td>
<td>CGG GCC TCT TCG CTA TTA CG</td>
</tr>
<tr>
<td>Mic4rr3</td>
<td>20</td>
<td>TGC CGA GGA GCC TGA GCA AG</td>
</tr>
<tr>
<td>Mic4rr3.5</td>
<td>21</td>
<td>CAT GAA CGT CTG AGT GCT CCC</td>
</tr>
</tbody>
</table>

230
<table>
<thead>
<tr>
<th>Primer</th>
<th>Length</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mic4rr3.7</td>
<td>21</td>
<td>GCC TCC TTT CTT CAT CAC GCG</td>
</tr>
<tr>
<td>Mic4rr3.10</td>
<td>20</td>
<td>GTC ATT TCA GTC ATT TAT CA</td>
</tr>
<tr>
<td>Mic5rr5</td>
<td>19</td>
<td>CAG AAA GCC ACG CCA AGG G</td>
</tr>
<tr>
<td>Mic5rr5.6</td>
<td>20</td>
<td>CTT TGA GTC TGG ATG CAG AG</td>
</tr>
<tr>
<td>Mic5rr5.7</td>
<td>20</td>
<td>TGT GTG CAT GCC AAG AGG TA</td>
</tr>
<tr>
<td>Mic5rr5.8</td>
<td>22</td>
<td>GAG ACC TGC GGA ACC TCA CCT G</td>
</tr>
<tr>
<td>Mic5rr5.9</td>
<td>20</td>
<td>ACT TCA TTC CGG CTT TTT AT</td>
</tr>
<tr>
<td>Mic5rr3</td>
<td>19</td>
<td>GCC TCA CAT TAG CAC GCC T</td>
</tr>
<tr>
<td>Mic5rr3.5b</td>
<td>20</td>
<td>CCT GGG CAG CTT ATC TGC TT</td>
</tr>
<tr>
<td>Mic5rr3.6</td>
<td>21</td>
<td>CAC TGG AAT CTC GGG AGT TCA</td>
</tr>
<tr>
<td>Mic5rr3.10</td>
<td>19</td>
<td>CCT TCA CAT GCG TTG TCG G</td>
</tr>
<tr>
<td>Mic5rr3.11</td>
<td>20</td>
<td>CAG ATG TGG AAA CAA TTG CG</td>
</tr>
</tbody>
</table>

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.
<table>
<thead>
<tr>
<th>Name</th>
<th>Size bp</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>EtMIC1 molecular constructs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mic1R-kpn1</td>
<td>31</td>
<td>ggaaggggtac(+106)CATTGCAGGAATGATGCAAG(+95)</td>
</tr>
<tr>
<td>mic1F-kpn1</td>
<td>30</td>
<td>gccggttac(+12)CTGGTTAGGTTCAAAACATAT(+35)</td>
</tr>
<tr>
<td>mic1F-kpn2</td>
<td>31</td>
<td>ccgggtac(-37)CAGGACCTTGCGAGGCATCCG(-24)</td>
</tr>
<tr>
<td>mic1F-kpn3</td>
<td>30</td>
<td>gaaggtac(-124)CAATAACAAATACAGTGACCTG(-107)</td>
</tr>
<tr>
<td>mic1F-kpn4</td>
<td>31</td>
<td>ataggtac(-300)CTACGGATATTCATACATACAG(-278)</td>
</tr>
<tr>
<td>mic1F-kpn5</td>
<td>31</td>
<td>cttgggtac(-503)CAAACCTTTGAGCATTTGTGC(-482)</td>
</tr>
<tr>
<td>EtMIC1 internal molecular constructs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mic1F-kpn7</td>
<td>26</td>
<td>cccgggtac(-473)CTAATTTCCTCTGAG(-456)</td>
</tr>
<tr>
<td>Mic1F-kpn9</td>
<td>25</td>
<td>cccggta(-410)CATATGCACACAGCGCAG(-393)</td>
</tr>
<tr>
<td>Mic1F-kpn10</td>
<td>28</td>
<td>gacggta(-383)CTTTAAACCATGCTGAGGAG(-364)</td>
</tr>
<tr>
<td>Mic1F-kpn11</td>
<td>28</td>
<td>tgggttac(-349)CATGTCTAATGGCAAAAGGC(-330)</td>
</tr>
<tr>
<td>Mic1F-kpn12</td>
<td>29</td>
<td>agcggta(-324)GGAATAAACACTCTCTAATAC(-304)</td>
</tr>
<tr>
<td>EtMIC2 molecular constructs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mic2-rev-kpn1</td>
<td>33</td>
<td>cccgggtac(+100)CATTTTGGTGAATAACAAAAAGAG(+39)</td>
</tr>
<tr>
<td>mic2F-kpn1</td>
<td>36</td>
<td>cccggttac(+12)CTTTTGGCAATTGTTGGCTTTTTAAA(+12)</td>
</tr>
<tr>
<td>mic2F-kpn2</td>
<td>33</td>
<td>gtgggtac(-33)CAACTCAAGTCCATCATTTCCAG(-12)</td>
</tr>
<tr>
<td>mic2F-kpn3</td>
<td>33</td>
<td>atgggtac(-148)CATATTACAGAATGTATAAAAGGAGA(-125)</td>
</tr>
<tr>
<td>mic2F-kpn4</td>
<td>30</td>
<td>gtcgggtac(-510)CTTAAAATAAAGTGATAGTGCT(-487)</td>
</tr>
<tr>
<td>mic2F-kpn5</td>
<td>33</td>
<td>ggaaggtac(-949)CAGTGGCTAGTGCAACAGGAGAA(-925)</td>
</tr>
<tr>
<td>EtMIC2 internal molecular constructs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mic2F-kpn10</td>
<td>36</td>
<td>aacgggtac(-55)CTGCAACAACCATGTGTGCTTTCCAACCTC(-28)</td>
</tr>
<tr>
<td>Mic2F-kpn11</td>
<td>36</td>
<td>cccgggtac(-80)CTGACTGTGCTAGGGACCGAAGCAAC(-56)</td>
</tr>
<tr>
<td>Mic2F-kpn12</td>
<td>37</td>
<td>gtgggtac(-113)CTGATCCTTTAGTGCAATTTGTGCTAGCC(-85)</td>
</tr>
<tr>
<td>Mic2F-kpn13</td>
<td>39</td>
<td>ctcgggtac(-204)CTCCCCAAAAATAATGGGCTTACATGGCC(-185)</td>
</tr>
<tr>
<td>Mic2F-kpn14</td>
<td>38</td>
<td>gtgggtac(-235)CGGTGGCGGGCCTCCAGAAAAGAAGAAAG(-206)</td>
</tr>
<tr>
<td>Mic2F-kpn15</td>
<td>40</td>
<td>cccgggtac(-293)CTTAGAGTTGGCTTGCCACCCCCGATCCTGCAG(-273)</td>
</tr>
<tr>
<td>Mic2F-kpn16</td>
<td>41</td>
<td>gtgggtac(-380)CGCAACCGAATTTCTCGTATGCTTACAAACATTC(-348)</td>
</tr>
<tr>
<td>Mic2F-kpn17</td>
<td>40</td>
<td>gtgggtac(-461)CGGTTAAGTACAGCCTGCTGCCAGGAGGCAAC(-430)</td>
</tr>
<tr>
<td>EtMIC3 molecular constructs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mic3R-kpn1</td>
<td>37</td>
<td>Cccgggtac(+59)CATCCTGGGAAGTAAGAAAGGGCTTGTGC(+31)</td>
</tr>
<tr>
<td>mic3F-kpn1</td>
<td>31</td>
<td>Cccgggtac(+12)CATGTGTGGTTACAATTAACCAC(+34)</td>
</tr>
<tr>
<td>Primer</td>
<td>Length</td>
<td>Sequence (5' to 3')</td>
</tr>
<tr>
<td>------------</td>
<td>--------</td>
<td>-----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>mic3F-kpn3</td>
<td>30</td>
<td>Cccggtac(-71)CACACCACCATCTTAGACGTG(-49)</td>
</tr>
<tr>
<td>mic3F-kpn5</td>
<td>39</td>
<td>Cccggtac(-571)CAGTTGAGCCGGTCCTGCGTGAACATCGCG(-542)</td>
</tr>
<tr>
<td>mic3F-kpn7</td>
<td>36</td>
<td>Cccggtac(-2210)CCTGTCAACAGTTGAGGGTCAATT(-2184)</td>
</tr>
</tbody>
</table>

**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.
<table>
<thead>
<tr>
<th>Name</th>
<th>Size (bp)</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUTA-1A</td>
<td>50</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+24)</td>
</tr>
<tr>
<td>MUTA-1B</td>
<td>50</td>
<td>(+24) GAAACCTAACCAGAATTCTACTGAGAGACGATTTGAAGTCGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGCTCGG (−26)</td>
</tr>
<tr>
<td>MUTA-2A</td>
<td>50</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+24)</td>
</tr>
<tr>
<td>MUTA-2B</td>
<td>50</td>
<td>(+24) GAAACCTAACCAGAATTCTACTGAGAGACGATTTGAAGTCGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGCTCGG (−26)</td>
</tr>
<tr>
<td>MUTA-3A</td>
<td>50</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+24)</td>
</tr>
<tr>
<td>MUTA-3B</td>
<td>50</td>
<td>(+24) GAAACCTAACCAGAATTCTACTGAGAGACGATTTGAAGTCGGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TGGCTCGG (−26)</td>
</tr>
<tr>
<td>MUTA-4A</td>
<td>58</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+24)</td>
</tr>
<tr>
<td>MUTA-4B</td>
<td>58</td>
<td>(+25) CAATATGTTTTTGAACCTAACCAGAATTCTACTTGAGACGATTTGA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AAGTGATGGCTCGG (−26)</td>
</tr>
<tr>
<td>MUTA-5A</td>
<td>64</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+113)</td>
</tr>
<tr>
<td>MUTA-5B</td>
<td>64</td>
<td>(+113) CAAAACAATATGTTTTTGAACCTAACCAGAATTCTACTGAG</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACGATTGAAGTGATGCGATGGCTCGG (−26)</td>
</tr>
<tr>
<td>MUTA-6A</td>
<td>66</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+113)</td>
</tr>
<tr>
<td>MUTA-6B</td>
<td>66</td>
<td>(+113) CAAAACAATATGTTTTTGAACCTAACCAGAATTCTACTGTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AGACGATTGAAGTGATGCGATGGCTCGG (−26)</td>
</tr>
<tr>
<td>MUTA-7A</td>
<td>66</td>
<td>(−26) CCGAGCCATCCGACTTCAATCGTCCTCAGTAGAAATTCTGGTTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>GGTTC (+113)</td>
</tr>
<tr>
<td>MUTA-7B</td>
<td>66</td>
<td>(+113) CAAAACAATATGTTTTTGAACCTAACCAGAATTCTACTGTA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>AGACGATTGAAGTGATGCGATGGATAGCC (−26)</td>
</tr>
</tbody>
</table>

Table C.3 Oligonucleotide primers used in the construction of *EtMCI* 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.
<table>
<thead>
<tr>
<th>Name</th>
<th>Size (bp)</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mic1-_mutced-1A</td>
<td>49</td>
<td>(-389) CCGTCACCTTTAAACCATGCTTGAACTTTGTTCATGTCTAATGGC (-337)</td>
</tr>
<tr>
<td>Mic1-mutced-1B</td>
<td>49</td>
<td>(-337) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-389)</td>
</tr>
<tr>
<td>Mic1-mutnf at-1A</td>
<td>49</td>
<td>(-389) CCGTCACCTTTAAACCATGCTTGAACTTTGTTCATGTCTAATGGC (-337)</td>
</tr>
<tr>
<td>Mic1-mutnf at-1A</td>
<td>49</td>
<td>(-337) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-389)</td>
</tr>
<tr>
<td>Mic1-stat-1A</td>
<td>49</td>
<td>(-389) CCGTCACCTTTAAACCATGCTTGAACTTTGTTCATGTCTAATGGC (-337)</td>
</tr>
<tr>
<td>Mic1-stat-1B</td>
<td>49</td>
<td>(-337) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-389)</td>
</tr>
<tr>
<td>Mic1-dof-1A</td>
<td>49</td>
<td>(-389) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-337)</td>
</tr>
<tr>
<td>Mic1-dof-1B</td>
<td>49</td>
<td>(-337) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-389)</td>
</tr>
<tr>
<td>Mic1-hap-1A</td>
<td>49</td>
<td>(-389) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-337)</td>
</tr>
<tr>
<td>Mic1-hap-1B</td>
<td>49</td>
<td>(-337) GCCATTAGACATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-389)</td>
</tr>
<tr>
<td>Mic1-elf-1A</td>
<td>52</td>
<td>(-355) GCAAAGTTTCCTCAAGCATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-337)</td>
</tr>
<tr>
<td>Mic1-elf-1A</td>
<td>52</td>
<td>(-355) GCAAAGTTTCCTCAAGCATGGAAGACAAAGTTCTCAAGCTGGTTTAAAGTGACCG (-337)</td>
</tr>
</tbody>
</table>

Table C.4 Oligonucleotide primers used in the construction of _EtMICl_ transcription factor binding sites mutant constructs (Chapter Six). Primers were supplied by MWG-Biotech UK Ltd.
<table>
<thead>
<tr>
<th>Name</th>
<th>Size (bp)</th>
<th>Sequence (5'-3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mic2-mut-1A</td>
<td>51</td>
<td>(-22) CCCATCATTTTGCAAGCAACGAATTACTCCTTGCTTTTTTGGCAATTGTTTGGC (+31)</td>
</tr>
<tr>
<td>Mic2-mut-1B</td>
<td>51</td>
<td>(+31) GCCAAACAATTTGCCAAAAAGCAAGGAGTAATTCGTTGCTGCAAATGATGGG (-22)</td>
</tr>
<tr>
<td>Mic2-mut-2A</td>
<td>51</td>
<td>(-22) CCCATCATTTTGCAAGCAACGGATTACTCCTTGCTTTTTTGGCAATTGTTTGGC (+31)</td>
</tr>
<tr>
<td>Mic2-mut-2B</td>
<td>51</td>
<td>(-22) GCCAAACAATTTGCCAAAAAGCAAGGAGTAATTCGTTGCTGCAAATGATGGG (+31)</td>
</tr>
<tr>
<td>Mic2-mut-5A</td>
<td>53</td>
<td>(-22) CCCATCATTTTGCAAGCAACGAATTACTCCTTGCTTTTTTGGCAATTGTTTGGC (+31)</td>
</tr>
<tr>
<td>Mic2-mut-5B</td>
<td>53</td>
<td>(-22) GCCAAACAATTTGCCAAAAAGCAAGGAGTAATTCGTTGCTGCAAATGATGGG (+31)</td>
</tr>
<tr>
<td>Mic2-mut-6A</td>
<td>53</td>
<td>(-22) CCCATCATTTTGCAAGCAACGAATTACTCCTTGCTTTTTTGGCAATTGTTTGGC (+31)</td>
</tr>
<tr>
<td>Mic2-mut-6B</td>
<td>53</td>
<td>(-22) GCCAAACAATTTGCCAAAAAGCAAGGAGTAATTCGTTGCTGCAAATGATGGG (+31)</td>
</tr>
<tr>
<td>Mic2-mut-8A</td>
<td>55</td>
<td>(-56) CCTGCACAACCAATCTATTGTGCAAGCAACGAAGATCGTTCCTGGCTGGTGTTGCTGCAATTGATGGG (-22)</td>
</tr>
<tr>
<td>Mic2-mut-8B</td>
<td>55</td>
<td>(+31) GCCAAACAATTTGCCAAAAAGCAATCTCTCTCTGCTGCAATTGATGGG (-22)</td>
</tr>
<tr>
<td>Mic2-mut-9A</td>
<td>-125</td>
<td>(+31) GCCAAACAATTTGCCAAAAAGCAATCTCTCTCTGCTGCAATTGATGGG (-22)</td>
</tr>
<tr>
<td>Mic2-mut-9B</td>
<td>-125</td>
<td>(+31) GCCAAACAATTTGCCAAAAAGCAATCTCTCTCTGCTGCAATTGATGGG (-22)</td>
</tr>
</tbody>
</table>

Table C.5 Oligonucleotide primers used for 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.
**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.