



University
of Glasgow

<https://theses.gla.ac.uk/>

Theses Digitisation:

<https://www.gla.ac.uk/myglasgow/research/enlighten/theses/digitisation/>

This is a digitised version of the original print thesis.

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study,
without prior permission or charge

This work cannot be reproduced or quoted extensively from without first
obtaining permission in writing from the author

The content must not be changed in any way or sold commercially in any
format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author,
title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk/>
research-enlighten@glasgow.ac.uk

Regulation of *TamS1* gene expression during differentiation to the merozoite
in *Theileria annulata*

Mark Fox
Department of Veterinary Parasitology
University of Glasgow

A thesis submitted to the University of Glasgow for the Degree of Doctor of Philosophy

December 1997

ProQuest Number: 13815374

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

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



ProQuest 13815374

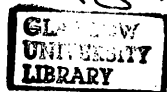
Published by ProQuest LLC (2018). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Thesis 11024
Copy 1



Acknowledgements.

I would like to thank Prof A Tait and Dr B. Shiels for the provision of the laboratory facilities and supervision. I must thank Dr. Swan, and Mrs Nybo for their lively debates. Special thanks goes to Dr Jane Kinnaird for her few well placed words of encouragement and support which meant much more than I will ever be able to explain.

I must thank Titch for sharing my lunch and saving my waist line from ever expanding. I am indebted to Brian, Fraser, Sue and Frank for technical comments and discussion. In addition they must be thanked for their tolerance and understanding of my lack of years, making sure that I never grew too big for my boots, and providing constructive criticism. Also for making life in the laboratory more enjoyable and accepting that I am no genius or rocket scientist. I also wish to thank Rowena for being Rowena, and for keeping me amused.

I would like to acknowledge the financial support of the Medical Research Council and the Cotton Club, Scott Street. I will always be grateful to the staff of the Cotton Club for their camaraderie and in particular to Sol, Jason, Stewart, Phil, and Will whose actions I can never forget or can thank enough.

On a more personal note, I would like to thank George Laird who taught me and demonstrated the meaning of the phrase "water off a ducks back". Also for accepting me as his friend warts and all. Special thanks must go to Linda Colman and Garn the gorilla, for excepting the number of times I would work well in to the small wee hours of the day, for providing endless cups of coffee and for caring. I must also thank my parents for going through this for a second time.

Declaration

The experimental work described in this thesis was carried out in the Department of Veterinary Parasitology, University of Glasgow, between October 1993 and December 1997. Except where stated otherwise, the results presented are my own.

Table of Contents	Page
Abbreviations	
List of Figures and Tables	
Summary	
1) Introduction	
1.1) <i>Theileria annulata</i>	1
1.2) Distribution	2
1.3) The Life Cycle of <i>T.annulata</i>	3
1.3.1) Bovine host	4
1.3.2) The Invertebrate Vector	8
1.4) Tropical Theileriosis	9
1.4.1) Pathogenesis	9
1.4.2) The immune response to <i>T.annulata</i> infection	11
1.4.3) Humoral Response	12
1.4.4) Cell mediated Response	14
1.4.5) Control Measures	16
1.4.6) Treatment of Tropical Theileriosis	16
1.4.7) Vector Control	16
1.4.8) Vaccination	17
1.4.8.1) Infection and treatment	17
1.4.8.2) Cell Culture Schizont Vaccine	17
1.4.8.3) Recombinant Vaccines	19
1.5) Stage Differentiation in Protozoan Parasite	24
1.5.1) Stage Differentiation in <i>Theileria</i>	39
1.6) Mechanisms of Gene Structure and Regulation in <i>Theileria</i>	34
1.6.1) Gene Structure	34
1.6.2) Initiation	37
1.6.3) RNA Processing and termination	41
1.6.4) Post-Transcriptional Gene Regulation	42
1.7) Objectives of this work	44

1.8) Summary of work presented	45
--------------------------------	----

2 The Cloning and Sequencing of the Regulatory Regions Controlling the Expression of the *TamS1* Gene

2.1) Introduction	47
2.1.2) Aims	50
2.2) Materials and Methods	51
2.2.1) Phage Stocks and Liquid Lysates	51
2.2.2) Restriction digests of DNA and agarose gel electrophoresis	53
2.2.3) Mapping of λ DASH clones by restriction digestion and southern blotting	53
2.2.4) Cloning and transformation λ DASH genomic clones	55
2.2.5) Preparation of plasmid DNA	57
2.2.5A) Alkaline lysis miniprep method	57
2.2.5B) Magic minipreps purification kit (Promega)	58
2.2.6) Generation of Exo III deletions via Erase-a-Base system (Promega)	58
2.2.7 DNA sequencing	59
2.2.7A) Manual DNA polymerase sequencing of double stranded DNA	59
2.2.7B) Automated DNA polymerase sequencing of double stranded DNA	60
2.2.8) Culture of <i>T. annulata</i> cell lines and a bovine B cell lymphosarcoma cell line	61
2.2.9) Isolation of RNA, gel electrophoresis and northern blotting	62
2.3) Results	64
2.3.1) Characterisation of the λ DASH clones	64
2.3.2) Subcloning the 5' and 3' regions of the <i>TamS1</i> gene	69
2.3.3) Initial sequencing of the 3kb Hind III + Sal I 5' <i>TamS1</i> intergenic fragment	72
2.3.4) Sequencing of the Sub-cloned 5' <i>TamS1</i> intergenic fragments.	74
2.3.5) Initial sequencing of the 4kb Hind III + Sal I 3' intergenic fragment	77
2.2.6) General analysis of sequence	78
2.3.7) Open reading frames	80
2.4) Discussion	89

3 Structural Analysis of the Upstream Region of the *TamS1* Gene.

3.1) Introduction	94
3.1.2) Aims	98
3.2) Materials and Methods	99
3.2.1) Nuclear run on analysis	99
3.2.2) 5' Rapid Amplification of cDNA Ends (5' RACE)	100
3.2.3) DNA amplification by Polymerase chain reaction (PCR)	101
3.2.4) Generation of genomic DNA mini-library and cloning of merozoite gene	102
3.3) Results	106
3.3.1) Isolation of gene clones encoding the major merozoite antigen in <i>T.parva</i>	100
3.3.2) Sequencing of the 3' untranslated of the <i>TamS1</i> related gene in <i>T.parva</i>	108
3.3.3) Sequence comparison of the <i>T.annulata</i> and <i>T.parva</i> contigs	111
3.3.4) Transcriptional regulation of the <i>TamS1</i> gene	112
3.3.5) Mapping the Transcriptional Initiation Site of the <i>TamS1</i> gene	114
3.3.6) Comparative Sequence Analysis of the 5' Sequence of Major Merozoite Surface Antigen Genes	116
3.3.7) Putative Promoter binding Sites in the 5' Untranslated Region of <i>TamS1</i>	120
3.3.8) Comparative Sequence Analysis of the upstream sequences of the <i>TamS1</i> , <i>TamR1</i> , Ta-ORF-1 and <i>SPAG-1</i> genes.	121
3.4) Discussion	127

4 Identification and Characterisation of an Essential Element in the Promoter of the 30kDa Merozoite Surface Antigen

4.1) Introduction	133
4.1.2) Aims	136
4.2) Materials and Methods	138
4.2.1) Isolation of Macroschizont or Differentiating Macroschizont Nuclear Extracts	138

4.2.2) Indirect Immunofluorescence Assay	138
4.2.3) SDS-polyacrylamide gel electrophoresis	138
4.2.4) Preparation of labelled DNA probes for retardation assay and South-western assays	140
4.2.5) Gel Mobility Shift Assay	141
4.2.6) Preparation of Single Stranded Plasmid DNA	143
4.2.7) Preparation of Radiolabelled Probe for UV Crosslinking	143
4.2.8) UV Crosslinking Procedure	144
4.2.9) South-Western Blotting	144
4.3) Results	146
4.3.1) Analysis of Nuclei from Macroschizont cell lines	146
4.3.2) Preparation of nuclear extract	148
4.3.3) Subcloning the 5' untranslated region of the <i>TamS1</i> gene	152
4.3.4) Gel Shift Analysis of the 5' Untranslated Fragments of the <i>TamS1</i> Gene	154
4.3.5) Evidence for Parasite Associated Mobility Shifts	155
4.3.6) Specificity of Binding to <i>TamS1</i> pE Probe	160
4.3.7) Defining the Position of Binding Sites Within the <i>TamS1</i> pE Probe	160
4.3.8) Conformation of DNA-Protein interactions with Oligonucleotides GATA, CAT1 and TAT1	163
4.3.9) Effect of Poly d(I-C) concentration on Protein Binding to CAT1	169
4.3.10) Analysis of Protein Binding to CAT1	169
4.3.11) CAT1 Binding Elements	175
4.3.12) Generation of Mutations that Abolish Specific Binding to CAT1	177
4.3.13) Identification of Transcription Factors	181
4.3.14) UV Crosslinking of Transcription Factors to CAT1 Sequence	181
4.3.15) Identification of DNA-Binding Proteins by South-Western Blotting	185
4.4) Discussion	190

5 Comparative Analysis of Gene Expression in Enhanced and Diminished Macroschizont Cell Lines

5.1 Introduction	197
5.1.2 Aims	201
5.2 Material and Methods	202
5.3 Results	204
5.3.1 Restriction fragment polymorphism (RFLP) analysis of the macroschizont enhanced cell line D7 and diminished cell line D7/B12	204
5.3.2 To compare the ability of enhanced and diminished cell lines to regulate parasite gene expression	206
5.3.3 Mobility shift analysis of nuclear extracts derived from D7 and D7/B12 cell lines	210
5.3.4 Sequence comparison of the 5' untranslated region of the D7 and D7/B12 alleles of <i>TamS1</i> .	212
5.4 Discussion	215
6 Discussion	220
7 References	226
Appendix I	254
Appendix II	266
Appendix III	274

Abbreviations

ATP	Adenosine triphosphate
BoLA	bovine leucocyte antigen
bp	base pair
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
dH ₂ O	distilled water
dATP	2'-deoxyadenosine 5'-triphosphate
dCTP	2'-deoxycytidine 5'-triphosphate
dGTP	2'-deoxyguanosine 5'-triphosphate
dTTP	2'-deoxythymidine 5'-triphosphate
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetra-acetic acid
EMSA	electrophoretic mobility shift assay
IFA	immuno-fluorescence assay
kb	kilobase
kDa	kilodalton
mcAb	monoclonal antibody
OD	Optical density
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute medium
SDS	sodium dodecyl sulphate
<i>spp.</i>	species
Tris	Tris (hydroxymethyl) aminomethane

Figures

Figure 1.1	Life cycle of <i>Theileria annulata</i>	5
Figure 2.1	Agarose electrophoresis analysis of genomic clone isolated from λ Dash genomic library	66
Figure 2.2	Autoradiograph of Southern blot of Lambda Dash clone TamS1 1.1 probed with TA37d	66
Figure 2.3	Restriction enzyme map of Tams1 1.1 bacteriophage clone	67
Figure 2.4	Autoradiograph of bacteriophage DNA TamS1 1.1 cut with HindIII, Sal I and Hind III x Sal I	68
Figure 2.5	A diagrammatic representation of the sub-cloning of the 5' and 3' <i>TamS1</i> regions into the vector pGem3zf	70
Figure 2.6	Southern blot analysis of genomic sub-clones following digestion with either Hind III + Sal I or EcoRI restriction enzymes and probing with cDNA probes as indicated.	71
Figure 2.7	Agarose electrophoresis analysis of the 5' genomic clone of the 30kDa molecule from λ Dash genomic library.	73
Figure 2.8	A diagrammatic representation of the sub-cloning 5' TamS1 into plasmids 5'Tams 1.0, 5'Tams1 0.8 and Tams1 1.2	75
Figure 2.9	A diagrammatic representation of the sub-cloning of the 3' <i>TamS1</i> into plasmids Tam13 and Tam33	79
Figure 2.10	The nucleotide and derived amino acid sequence of the <i>TamS1</i> gene	81
Figure 2.11	The nucleotide and predicted amino acid sequence of open reading frame 1	82
Figure 2.12	The nucleotide and predicted amino acid sequence of open reading frame 2	83
Figure 2.13	DNA sequence of open reading frame 3 plus 400bp of 3' intergenic sequence	84
Figure 2.14	Open reading frames and selected restriction enzyme sites in the <i>TamS1</i> genomic clone	86
Figure 2.15	Distribution of nucleotide composition across the ~8kb genomic clone	87

Figure 2.16	Analysis of RNA levels during a differentiation time course of macroschizont infected cell line D7 and Piroplasm RNA by Northern blotting	88
Figure 3.1	Southern blot analysis of <i>T.parva</i> genomic DNA following restriction digestion and hybridisation with (A) 5' <i>TamS1</i> cDNA and (B) 3' <i>TamS1</i> cDNA probe.	107
Figure 3.2	Agarose electrophoresis analysis of the TpmS13.2 genomic clone isolated from <i>T.parva</i> genomic mini-library.	107
Figure 3.3	Southern blot analysis of genomic clone isolated from the <i>T.parva</i> genomic library.	110
Figure 3.4	<i>TamS1</i> and <i>TamR1</i> gene define independent transcription units.	113
Figure 3.5	Southern blot analysis of 5' RACE products derived from <i>TamS1</i> mRNA	115
Figure 3.6	Comparison of <i>TamS1</i> like genes isolated form different <i>Theileria</i> species.	118
Figure 3.7	Comparison of <i>TamS1</i> like genes isolated from different <i>Theileria</i> species over the first 178bp of 5' region.	119
Figure 3.8	Factor binding sequence motifs identified in the 5' untranslated region of <i>TamS1</i> by GCG patternfind program	122
Figure 3.9	Comparison of the intergenic regions of <i>TamS1</i> , <i>TamR1</i> , <i>SAPG-1</i> and ORF-1 of the first 1kb 5' to the ATG start codon.	125
Figure 3.10	Comparison of the intergenic regions of the TamS1, TamR1 and Spag-1 genes over the first 100bp upstream of the transcription initiation site of each genes	126
Figure 4.1	DAPI staining of fixed slide preparations of the cloned cell line D7 after cultures were incubated at 37°C or at 41°C for 6 days.	149
Figure 4.2	Immunofluorescence reactivity of monoclonal antibody 5E1 on fixed slide preparations of the cloned cell line D7 after cultures were incubated at 37°C or at 41°C for 6 days.	150
Figure 4.3	Immunofluorescence reactivity of anti-TamR1 serum on fixed slide preparations of the cloned cell line D7 after cultures were incubated at 37°C or at 41°C for 6 days.	151

Figure 4.4.	Coomassie blue stained SDS-polyacrylamide gel of total cell extract and material from different points of the nuclear isolation procedure.	153
Figure 4.5	The promoter region of the 30kDa molecule	156
Figure 4.6	Analysis of protein concentration upon DNA binding complexes	157
Figure 4.7	Gel mobility shift analysis of parasite specific DNA complexes.	159
Figure 4.8	Competition EMSA with probe TamS1pD	161
Figure 4.9	Analysis of the position of protein binding within the TamS1pD probe	164
Figure 4.10	EMSA of the CAT1 sequence element with nuclear extracts	166
Figure 4.11	EMSA of the GATA sequence element with nuclear extracts	167
Figure 4.12	EMSA of the TAT1 sequence element with nuclear extracts	168
Figure 4.13	Binding of radiolabelled CAT1 in the presence of poly d(I-C)	170
Figure 4.14	Titration of probe concentration in CAT1 EMSA	172
Figure 4.15	Titration of extract concentration in CAT1 EMSA	173
Figure 4.16	Competition experiments showing specificity of CAT1-extract complex formation	174
Figure 4.17	The area depicted represents part of the TamS1pD fragment which is proximal to the RNA initiation site of the <i>TamS1</i> gene.	176
Figure 4.18	Binding activity of the CAT1, 2, 3 and 4 probes in parasite enriched extracts prepared from (A) D7 cells cultured at 37°C or (B) 41°C for 6 days.	178
Figure 4.19	EMSA competition experiments of CAT1 complex formation	179
Figure 4.20	Base substitutions of CAT1	182
Figure 4.21	EMSA and base substitution analysis of CAT1	183
Figure 4.22	Competition experiments using base substitution mutants to demonstrate specificity of the CAT1 element.	184
Figure 4.23	UV crosslinking of mobility shift reaction	186
Figure 4.24	Competition analysis of UV crosslinking of mobility shift reactions.	187
Figure 4.25	South-Western analysis of CAT1, TamS1pD and GATA	189
Figure 5.1	Southern blot analysis of genomic DNA isolated from D7 cells and D7/B12 cells digested with a variety of restriction enzymes and hybridised with <i>TamS1</i> cDNA probe.	205

Figure 5.2	Analysis of RNA levels during a differentiation time course of cell lines D7 and D7/B12 by Northern blotting.	207
Figure 5.3	Analysis of RNA levels during a differentiation time course of cell lines D7 and D7/B12	209
Figure 5.4	Interaction of the CAT1 sequence element with parasite enriched nuclear extracts prepared from differentiation time course D7 and D7/B12 macroschizont infected cell lines.	211
Figure 5.5	Sequence comparison of the 5' untranslated sequence of <i>TamSI</i> from the cell lines D7 and D7/B12	214

Tables

Table 2.1	Nucleotide composition of different regions of the 8.0kb genomic fragment	87
Table 3.1	The oligonucleotide primers used for PCR and 5' RACE system	103
Table 4.1	Oligonucleotide primers used to generate gel shift assay probes by annealing or PCR amplification of TamS1 genomic DNA	142

Summary

Much of the pathogenicity of the bovine disease tropical theileriosis, caused by the protozoan parasite *Theileria annulata*, results from the parasite differentiating from one life cycle stage to another. Given the observation, by Shiels *et al.*, (1994) that the timing of differentiation from macroschizont to merozoite can be altered and that this correlates with changes in the expression of the 30kDa merozoite surface polypeptide (*TamS1*); the aim of this study was to identify the molecular mechanisms controlling the expression of the *TamS1* gene. Furthermore, the identification of the factors controlling *TamS1* expression could lead to the identification of a factor involved in the timing of differentiation events. Therefore, it was intended to clone and sequence the regulatory regions controlling the expression of the *TamS1* gene and to develop a gel retardation mobility assay in order to analyse transcription factors interacting with *TamS1* promoter elements particularly with respect to factors binding during induction of differentiation.

Isolation and sequencing of the 5' and 3' regions flanking the *TamS1* protein coding sequence, identified three additional open reading frames. Attempts to identify each gene or functional motifs within each gene by sequence comparisons with polypeptides in known databases was unsuccessful. Nuclear run analysis demonstrated that the *TamS1* gene is monocistronically transcribed with definite start and termination signals. Additionally, the orientation of the other open reading frames would suggest that all of the genes within the isolated contig are monocistronically transcribed. In an attempt to identify the promoter region involved in the regulation of the *TamS1* gene the RNA start site was mapped. Two strategies were used in an attempt to identify the promoter elements; firstly the 5' intergenic regions from related genes in other species of *Theileria* were cloned and sequenced to allow for comparisons to be made and the 5' region of the *TamS1* gene was compared with that of known promoter elements. However, neither of these approaches was successful in highlighting possible functional promoter elements.

Theileria annulata macroschizont infected cell lines can be induced to produce merozoites by culture at an elevated temperature (41°C). This *in vitro* system, enabled the isolation of host-free parasite nuclear extracts for use in the development of a gel retardation

mobility shift assay. This technique enable the identification of a promoter element which bound three specific mobility bands. Analysis of parasite nuclear extracts made from two types of clonal cell lines (enhanced and diminished differentiators) showed that two electrophoretic bands were only present during a specific stage of parasite differentiation corresponding either to commitment or a significant increase in expression of the *TamSI* gene. Determination of the size and number of factors interacting with the promoter element were attempted using two strategies, by UV crosslinking and South-Western blotting. However, neither of these approaches were successful and the possible reasons for this are explained.

1 Introduction

1.1 *Theileria annulata*

Theileria annulata, a protozoan parasite of cattle and water buffalo, was first described by Dschunkowsky and Luhs (1904) and is the causative agent of tropical theileriosis or Mediterranean Coast fever, a debilitating and frequently fatal disease of cattle. Initially the disease was described as a tropical bovine piroplasmosis and was associated with a haemo-protozoan parasite, subsequently named *Piroplasma annulatum* by Bettencourt in 1907 (Irvin, 1987). Discovery of a schizont stage of *P.annulatum* led to reclassification of these organisms, placing them in the genus *Theileria*.

Other apicomplexan parasites of major importance are *Plasmodium*, *Eimeria*, *Babesia*, *Sarcocystis*, and *Toxoplasma*. Members of the phylum Apicomplexa are all characterised by the possession of an apical complex (consisting, in the majority of cases, of the polar ring and conoid complex, forming the leading pole, during the budding process of a developing parasite) during at least one stage of their life cycle, with intracellular forms occurring in the mammalian host (Shaw and Tilney, 1992). Within the apicomplexan phylum *Theileria* is most closely related to the genus *Babesia*. These genera are distinguished by their morphology at the erythrocytic stages and their target cells. *Babesia* species infect only erythrocytes while *Theileria* species also invade leukocytes. At least 2 representatives of the *Babesia* species are exceptions to this rule, for example *B.equi* and *B.microti*, which unlike other Babesiidae multiply initially in lymphocytes before infecting the erythrocytes. In addition, the morphology of the schizonts found in the cytoplasm of the lymphocytes have been described as being "*Theileria*-like" (Moltman *et al.*, 1983). This and morphological similarities coupled by molecular phylogeny studies have brought into question the taxonomical placement of these parasites within Babesiidae as they continue to show increasing similarities to *Theileria* (Knowles, 1996; Ellis *et al.*, 1992). Further analysis at the genetic (molecular level) should have the ability to resolve existing uncertainty over *Theileria* species and subspecies designation. The phylogenetic relationship between *Theileria* and other apicomplexan parasites has been established and is based on the sequence comparison of the small sub-unit ribosomal RNA molecules (Barta *et al.*, 1991; Gajadhar *et al.*, 1991). The currently accepted classification of the *Theileria* genus is shown below: (Levine, 1988).

Sub-kingdom	Protozoa
Phylum	Apicomplexa
Class	Aconoidasida
Order	Piroplasmorida
Family	Theileriidae
Genus	Theileria

Theileria parasites are a group of tick transmitted apicomplexan parasites which infect a variety of wild and domestic animals throughout the world. The genus *Theileria* contains a number of species whose characteristics have been reviewed by Uilenberg (1981), Dolan (1989) and Morzaria and Nene (1990). Six species of *Theileria* are infectious for cattle and domestic buffalo: *T.annulata*, *T.parva*, *T.sergenti*, *T.mutans*, *T.taurotragi* and *T.velifera* (Irvin, 1987). These species are classified according to geographical distribution, the morphology of macroschizont and piroplasm stages, pathogenicity, vector species and by indirect immunofluorescence antibody tests (Kimber *et al.*, 1973). Clinically and economically, *T.annulata* and *T.parva* are the most serious of the six. *T.parva* is responsible for East Coast Fever, January disease and Corridor disease and is transmitted by *Rhipicephalus* species of tick; whereas *T.annulata*, the active agent of tropical theileriosis, is transmitted by *Hyalomma* spp. *T.sergenti* is also significant and is an important pathogen in cattle of East Asia. *T.taurotragi*, *T.mutans*, and *T.velifera* are rarely pathogenic. Other *Theileria* species of which infect ruminants, include *T.hirci* a pathogenic parasite for sheep and goats, and can cause significant economic losses in regions where it is endemic (Hooshmand-Rad *et al.*, 1973) *T.camelensis* a parasite infecting camels, and *T.ovis*, *T.recondita* and *T.separata* are non-pathogenic parasites of small ruminants (Dolan, 1989, Papadopoulos *et al.*, 1996).

1.2 Distribution

Tropical theileriosis is widespread, with endemic regions stretching from the Western Mediterranean, Southern Europe, Northern Africa to the Pacific, in a broad band encompassing, India, the Middle East, China, and parts of the former Soviet Union (Dolan, 1992). Recently, the disease has also been reported in West Africa (Jacquet *et al.*, 1994). It

has been estimated that there are more than 250 million cattle across the globe at risk from tropical theileriosis (Tait and Hall, 1990). With such a large number of animals at risk making an assessment of economic damage is very difficult. Furthermore, cattle recovering from infection suffer weight loss, fertility problems and diminished milk yields. Another aspect of the disease is that the risk of contraction is such that it prevents the introduction of high-yielding cattle, which are a fully susceptible stock, to endemic areas where preventive controls are not practised. This is of major economic importance as it prevents genetic improvement to livestock productivity

The distribution and seasonal occurrence of *T.annulata* is restricted by the geographical location of the transmission vector, ixodid ticks of the *Hyalomma* species and depending on the geographical location, different species of *Hyalomma* are important with respect to transmission. Susceptible domestic cattle (*Bos taurus* and *Bos indicus*) and Asian water buffalo (*Bubalus bubalis*) in areas where the tick vector occurs are, in principle, at risk of contracting tropical theileriosis. Two important vectors of *T.annulata*, are *H. anatolicum anatolicum* and *H. detritum* occurring in Southern Europe, Northern Africa, Asia, parts of the former Soviet Union and the Middle East (Jongejan and Uilenberg, 1994). Developing ticks require shelter, warmth and a relatively humid climate and, as transfer of the parasite to the bovine coincides with feeding activity of the tick following moulting, outbreaks of Theileriosis can be seasonally dependent. Thus, the majority of outbreaks in North Africa occur from June to September (Flach *et al.*, 1994; Bouattour *et al.*, 1994), and in India outbreaks are generally reported during the rainy summer months (Grewal *et al.*, 1994). In general, sporadic cases are encountered in various geographical regions all year round (Pipano, 1989).

1.3 The Life Cycle of *T.annulata*

All apicomplexan parasites undergo a progressive series of differentiation events which together comprise the Life-Cycle of the parasite. The parasite is dependent on the differentiation from one stage to the next, as these events are necessary for the generation of forms with specific function. In general, apicomplexan differentiation is either followed by a phase of asexual replication or the generation of a non-dividing invasive phase. As apicomplexans are unicellular organisms of haploid genotype, the generation of the sexual

phase of the Life-Cycle can also be considered to be a differentiation event. It can be seen that stage differentiation is a fundamental event of apicomplexan parasite biology. *T.annulata* has a complicated Life-Cycle generated by a progressive series of differentiation events within the vertebrate and invertebrate hosts. There are three main phases of asexual multiplication; sporogony, after sexual recombination in the vertebrate host, and schizogony and merogony in the vertebrate host. A simplified diagrammatic form of the Life-Cycle is shown in Figure 1.1.

1.3.1 Bovine Host

Infection of the bovine host with *T.annulata* is initiated when sporozoites, contained within the salivary glands of the tick vector, are deposited during a blood meal. The sporozoites rapidly invade leukocytes (Jura *et al.*, 1983), preferentially targeting cells displaying major histocompatibility complex (MHC) class II molecules; primarily B cells and monocytes/macrophages (Glass *et al.*, 1989). Sporozoites gain entry into lymphocytes within 5-60 minutes of inoculation by receptor mediated endocytosis, and as many as 15 sporozoites can be internalised within an individual leukocyte (Jura *et al.*, 1983). Sporozoite-lymphocyte recognition and binding is a temperature independent event which can occur at 0-2°C; all the subsequent stages are temperature dependent (Jura, 1984; Shaw *et al.*, 1991). Entry of the sporozoite occurs in a sequential manner with the internalisation step being characterised by "zippering" of the host cell and parasite membranes (Fawcett *et al.*, 1982). Subsequent to invasion, the apposed membranes of the host and parasite separate and at the same time the rhoptries and microspheres discharge their contents (Fawcett *et al.*, 1982; Shaw *et al.*, 1991). A thick layer of fuzzy material of 10-15 nm then appears on the surface of the sporozoite (Shaw *et al.*, 1991; Fawcett *et al.*, 1982) and the hosts cell membrane is broken down. The parasite finally comes to lie in the cytoplasm of the host cell, usually close to the Golgi apparatus, and its surface is surrounded by an orderly array which closely associates with the thick layer of fuzzy material (Fawcett *et al.*, 1984, Fawcett *et al.*, 1982; Shaw *et al.*, 1991).

Following invasion, the sporozoite develops into a uninucleate trophozoite, a transient feeding stage (Jura *et al.*, 1983). The trophozoite enlarges by ingesting the cytoplasm of the host cell and undergoes a series of nuclear divisions to produce the

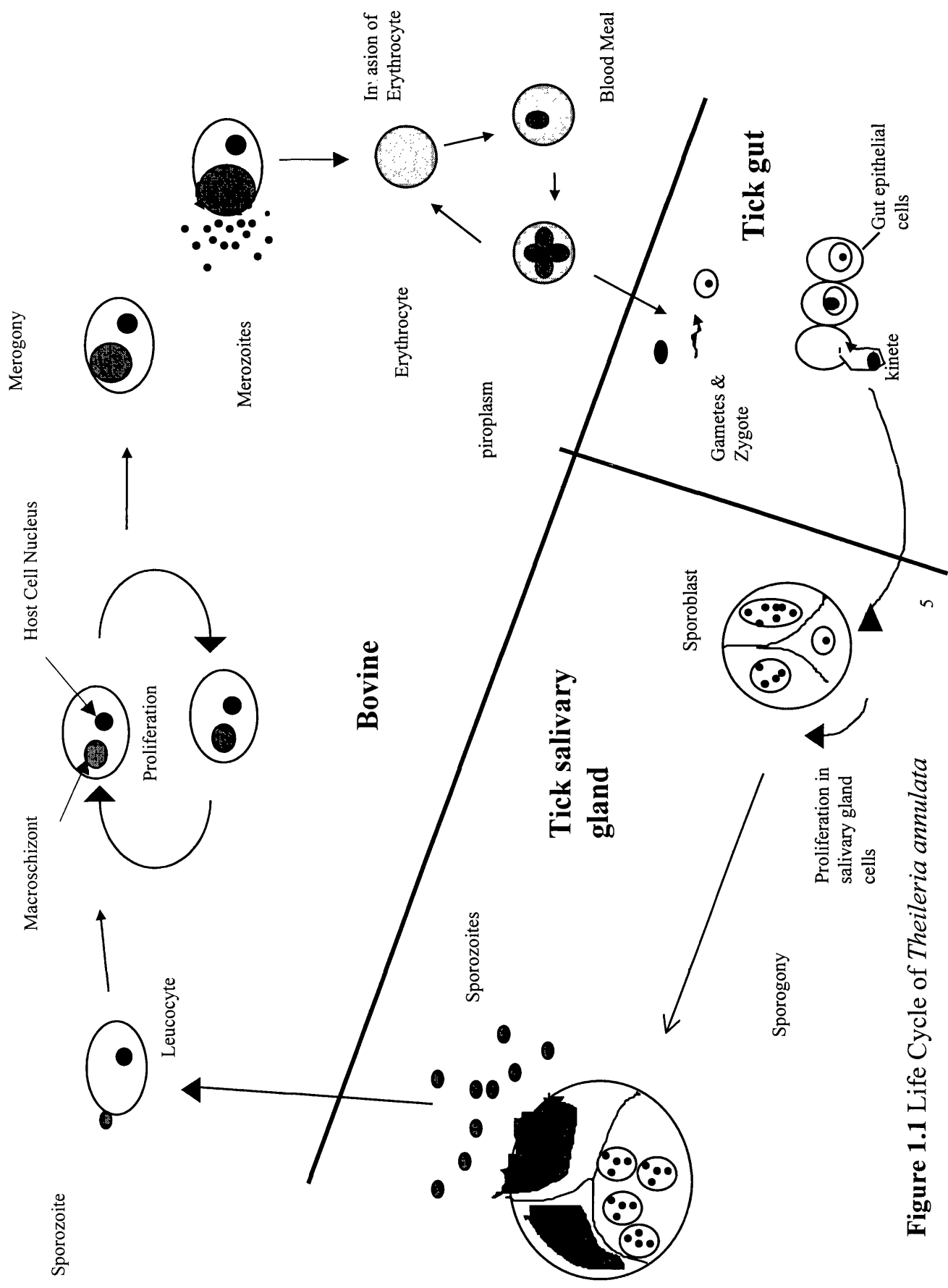


Figure 1.1 Life Cycle of *Theileria annulata*

multinucleated macroschizont stage. The schizont syncytium contains on average, 15-20 nuclei and is associated with host cell immortalization. Induction of host cell blastogenesis which leads to the rapid clonal expansion of parasitised cells and, *in vitro*, synchronous division of host and parasite, is achieved through attachment of the parasite to the host spindle apparatus in prophase and metaphase (Carrington *et al.*, 1995). During this period, the DNA is replicated and daughter parasites are separated by cytokinesis of the host cell during anaphase (Irvin, *et al.*, 1982). Division of the parasite is sometimes unequal, resulting in cells that may have lost the parasite which cease to divide after a few days. The process of synchronous division of host cell and macroschizont is the first phase of asexual multiplication in the bovine host. Recently a cdc2-related kinase gene was isolated from both *T.annulata* and *T.parva* which probably functions in all dividing stages of the parasite and is likely to play an important role in the regulation of nuclear division (Kinnaid *et al.*, 1996).

During the macroschizont stage of the parasite's life-cycle, it lies free within the host cell cytoplasm, surrounded by a plasma membrane with no obvious outer surface coat (Shaw and Tilney, 1992). The nuclei are surrounded by a typical nuclear envelope with nuclear pores. Parasite nuclei and mitochondria are scattered randomly throughout the schizont cytoplasm. Apart from numerous ribosomes and occasionally some small membrane-bounded vesicles, the schizont cytoplasm contains very few other organelles. No smooth or rough endoplasmic reticula or golgi apparatus have been observed in the parasite's cytoplasm (Shaw and Tilney, 1992). *Theileria* differs from other related protozoa such as *Toxoplasma* and *Plasmodium* in that it is not retained within a parasitophorous vacuole (Mehlhorn and Schein, 1984).

Later in the infection, after a finite number of cell divisions, enlarged macroschizonts undergo merogony, differentiating into microschizonts. This event is characterised by a series of structural and organisational changes within the schizont. These changes include the development of smooth and rough endoplasmic reticulum, and the formation of an external coat on the outer surface of the schizont plasma membrane which is found on the surface of mature merozoites. Nuclei migrate to the periphery of the schizont, by an unknown mechanism, and uninucleate merozoites start to form. Rhoptries appear in

the schizont cytoplasm frequently in small clusters, associated with the schizont nuclei and fibrous structures. Ultimately the fibrous material connects the rhoptries and the nuclear envelope to a peg which inwardly projects from the schizont plasma membrane, and mitochondria become closely associated with the outer membrane of the nuclear envelope. Merozoites bud from the surface of the schizont in a synchronous manner, liberating themselves from the host by breaking down the host cell membrane (Shaw and Tilney, 1992).

Upon release into the bloodstream, the merozoites rapidly enter erythrocytes and up to 90% of erythrocytes can become infected in extreme cases (Mehlhorn and Schein, 1984). Very little is known about the process of erythrocyte invasion by merozoites but it has been suggested that this process is mediated by ligand receptor interactions as demonstrated in malaria (Kawamoto *et al.*, 1990). More recently it has been proposed by Shaw and Tilney (1995) that merozoite invasion occurs via a similar process to that described for sporozoites (Shaw *et al.*, 1991). Following invasion the erythrocyte membrane enveloping the merozoite is disintegrated and the rhoptry contents are discharged. The parasite lies free within the cytoplasm of the erythrocyte and, unlike the macroschizont, host cell microtubules do not associate with the surface of the merozoite. Two forms of piroplasms have been observed free within the erythrocyte matrix: a) spherical form and b) a slender comma shaped form (Mehlhorn and Schein, 1984). The frequency with which these two different forms are observed varies for different *Theileria* species. For example in *T.annulata* both forms occur in approximately equal numbers whereas in *T.parva* 80% of all piroplasms are comma shaped. The two forms are believed to have differing roles; the ovoid form representing gamonts which are pre-adapted for the generation of gametes in the gut of the tick, while the comma forms are thought to undergo further division (Mehlhorn and Schein, 1984). Division of the comma form is thought to occur by binary fission and this nuclear division is associated with cellular division so no multi-nucleate schizont like stages can occur. Although true schizonts do not occur, in some species of *Theileria* four merozoites occasionally form a tetrad or "Maltese-cross" arrangement. Division of the merozoites often leads to the destruction of the host cell (Mehlhorn and Schein, 1984). The merozoites released from the erythrocytes are identical to those released from schizont-infected lymphocytes, which has led to the suggestion that these forms could be responsible for the

re-invasion of erythrocytes (Conrad *et al.*, 1985). When a tick feeds on an infected host, piroplasms and infected erythrocytes are taken up and, thus transmission of parasite between host and vector occurs.

1.3.2 The Invertebrate Vector

Infected erythrocytes are lysed within the gut of the tick releasing the piroplasms which then undergo further development (Mehlhorn and Schein, 1984). The piroplasms differentiate into a form known as ray bodies, which are morphologically similar to ‘strahlenkörper’ in *Babesia* (Schein, 1975), and spherical forms. This process is initiated when the piroplasms are released from their erythrocytes during lysis inside the intestine. The ray bodies produce uninucleated gamete-like stages which are considered to be microgamonts. In addition the spherical forms, which do not divide, are considered to be macrogametes. The haploid microgametes fuse with haploid macrogametes to form diploid zygotes (Gauer *et al.*, 1995) which enters tick gut epithelial cells. This and the succeeding kinete are believed to be the only diploid stages during the entire Life-Cycle of the parasite. Morzaria *et al.*, (1992) showed that recombinant parasites of mixed parental genotypes can be obtained from ticks fed on animals infected with two distinct parental stocks confirming sexual reproduction in *T.parva*. At this particular stage in the parasite Life-Cycle the *Theileria* species *T.annulata* and *T.parva* appear to diverge in their respective development. The zygotes of *T.parva* undergo a two-step meiosis in the gut epithelium of their vector tick (Gauer *et al.*, 1995) which appears to be a step for amplifying the parasite in the host. 12 to 30 days after feeding transformation of the stationary ovoid or spherical zygote is initiated leading to a club-shaped motile stage called a kinete. The kinete penetrates the gut epithelial cells of the tick as it migrates to the salivary glands via the haemolymph (Schein, 1975) where it remains dormant in the cytoplasm of the salivary gland cells. Since few kinete reveal diploid DNA contents in *T.annulata*, it seems likely that a postzygotic meiosis occurs after differentiation of zygotes into kinetes (Gauer *et al.*, 1995). The two species of *Theileria* again show similarities in their development as the tick moults and re-attaches to a bovine host. An initial feeding phase is followed by a phase of intensive multiplication steps yielding thousands of small nuclei. The sporoblast then enlarges dramatically to give a multinucleate syncytium, and uninucleate sporozoites form by cytoplasmic fission (Fawcett

et al., 1982). The sporozoites are released into the salivary gland and infection of the bovine host then occurs during the next blood meal, completing the Life-Cycle of *T.annulata*.

Interestingly, *Theileria* parasites undergo a number of highly similar developmental processes during their Life-Cycle (Shaw and Tilney, 1992). There is a remarkable similarity in the structure of the merozoites, sporozoites and piroplasms and in the process of their formation. Additionally invasion of the bovine host cells by *T.parva* sporozoites and merozoites occurs in a morphologically similar manner. Shaw and Tilney (1992) suggest that these repeating patterns in the life-cycle of *Theileria* are unlikely to have evolved independently, and they put forward a hypothesis that a developmental pattern of cellularisation was evolved only once, but has been used repeatedly by the parasite throughout its life-cycle. Therefore, it seems logical to conclude that the parasite may use the same cassette of genes to control these processes. It is possible to speculate that a replacement of only a few genes within each stage could account for minor differences in each process. For example, sporozoites and merozoites would only need to change the molecular composition of their surface coat to account for the selective invasion of different host cells. Similarly, the only difference between sporozoites and merozoites is the presence of microspheres in the sporozoite. Thus, Shaw and Tilney (1997) have proposed that the parasite may undergo a number of seemingly disparate developmental processes numerous times throughout its Life-Cycle without the requirement of an extensive genome.

1.4 Tropical Theileriosis

1.4.1 Pathogenesis

The severity of the disease in cattle infected with *T.annulata* is dependent upon the susceptibility of an animal, the virulence of the parasite strain and the number of sporozoites inoculated. Clinical symptoms of the disease are identified with the destruction of lymphoid cells as a result of the intracellular schizont differentiating into merozoites and haemolytic anaemia associated with merozoite division and differentiation of merozoites to the piroplasm stage within the erythrocyte.

The first symptoms of an acute infection are detectable following an incubation period of between 9 and 24 days after inoculation. The infected animal suffers a massive

proliferation of parasitized lymphocytes, causing swelling of the lymph nodes draining the sites of infection. A high fever of 41°C develops, which usually stays until recovery or death, coinciding with the onset of hypertrophy of parasitized and non-parasitized cells in the lymphoid tissues. Schizonts are detected in the bloodstream, liver and spleen of the animal. The animal loses conditioning and is anorexic. Additional clinical signs include diarrhoea, accelerated pulse and breathing, swelling of the eyelids, drooling from the mouth, a reduction of milk production, and a cessation of rumination (Barnett, 1977). During the course of the infection there is a progressive leucocytosis, followed by leucopenia (Preston *et al.*, 1992) and the red blood count also drops and associated haemolytic anaemia. It has been demonstrated that up to 90% of erythrocytes can be infected by the parasites, and although direct lysis of infected cells by the parasite adds to the anaemia, it is believed that the removal of infected erythrocytes by the spleen and liver is the direct cause (Hooshmand-Rad, 1976; Barnett, 1977; Uilenberg, 1981). If the erythrocyte count recovers and the animal continues to feed then it has a good chance of recovery; if the erythrocyte count remains low, however, and leads to severe anaemia, the animal will die approximately 20 days post infection.

Until recently it has not been clear how disease progression is related to parasite development. It has been proposed that *Theileria* infection of leukocyte cells induces the production of novel metalloproteinases, including matrix metalloproteinases (Baylis *et al.*, 1995). The matrix metalloproteinases are thought to play an important role in the formation of ulcerative lesions of the abomasum and cause digestion of connective tissue. Matrix metalloproteinases also activate the production of tumour necrosis factor alpha (TNF- α) (Adamson and Hall 1996) which has been shown to induce cachexia, pyrexia and leucopenia, symptoms similar to those observed in tropical theileriosis (Gearing *et al.*, 1995, Beutler and Cerami, 1986, Ulich *et al.*, 1987). Further induction of TNF- α synthesis in macrophages and interferon- γ (IFN- γ) in lymphocytes has been demonstrated *in vitro* by *T.annulata* macroschizont infected cells producing IFN- α (Preston *et al.*, 1993). Therefore, *T.annulata* infection may stimulate the production of metalloproteinases and a cascade of cytokine production through several pathways, and the action of these mediators could account for some of the clinical symptoms and tissue destruction associated with tropical theileriosis demonstrating a link between the pathogenicity of the parasite and the immune

response of the host. Furthermore, cytokines may not only underlie certain clinical symptoms due to IFN- γ induced synthesis of macrophage-derived NO, but could also induce the pathological lesions associated with NO production under other circumstances (Visser *et al.*, 1995).

The simplest test for diagnosis of tropical theileriosis is by examination of Giemsa stained blood or tissue smears for macroschizonts and/or piroplasms. Serological tests such as indirect fluorescent antibody test (IFAT) and enzyme-linked immunosorbent assay (ELISA) have been developed to detect circulating antibodies and production or release of parasite antigens, respectively. IFAT is limited as it is not suitable for large-scale surveys. An ELISA was developed by Kachani *et al.*, (1992) using sera from immunised cattle which exhibits low non-specific detection by normal sera and high post-infection values. The assay has provided an opportunity for epidemiological studies of *T.annulata* infection in Morocco (Kachani *et al.*, 1994). The most sensitive method for the detection of the parasite in carrier cattle, recently developed by d'Oliveira *et al.*, (1995), is the use of polymerase chain reaction (PCR) for specific amplification of *T.annulata* DNA from blood samples. Species specific primers for the *TamS1* major merozoite surface antigen were used to amplify a 372bp fragment specific to *T.annulata* which was successfully able to amplify DNA from stocks of *T.annulata* from distinct geographic regions. This assay has been shown to specifically diagnose *T.annulata* at parasitaemias as low as 0.00005% and provides an opportunity to determine whether animals that are translocated from regions where theileriosis is endemic are carriers of *T.annulata*. The sensitivity of the PCR will also facilitate monitoring of animals after vaccination with attenuated macroschizont infected cell cultures to determine whether vaccinated animals become carriers of *T.annulata* and thus a source of infection for *Hyalomma* ticks. DNA probes have also been used to detect *T.annulata* infection and have provided an opportunity to study genetic polymorphism of parasite populations within Tunisia (Ben-Miled *et al.*, 1994).

1.4.2 The Immune Response to *T.annulata* Infection

Following each Life-Cycle differentiation event within an infected animal, the host's immune system is exposed to a different set of antigenic determinants. The bovine immune response against *T.annulata* can be divided into humoral and cell mediated responses. Like

many protozoan pathogens cattle recovering from the infection develop immunity to further challenge. This immunity develops irrespective of the source of primary infection, whether from feeding ticks or artificial inoculation, with or without drug treatment, or from a vaccine derived from attenuated macroschizont infected cells cultured *in vitro*. Immunity lasts three years in the absence of further challenge. A complete understanding of the basis of immunity generated by a recovered animal would greatly aid in generating a rational approach to vaccine design. However, it seems from vaccine work that reasonable cross protection can be gained from attenuated vaccine.

Parasites can evade the immune system of the vertebrate host by several techniques (Borst, 1991); (1) invading cells or hiding in sites in the body where the immune system is less effective, (2) mimicry of host proteins, (3) suppressing the immune system of the host and (4) antigenic polymorphism. All of these mechanisms have been observed to have been used by the *Theileria* parasite during the life-cycle stages in the vertebrate host. The importance of antigenic polymorphism is discussed in section 1.4.8.3.

1.4.3 Cell Mediated Response

Animals recovering from a sporozoite challenge have been shown to generate cytotoxic T lymphocytes, which *in vitro* have been shown to lyse infected lymphoblasts (Singh *et al.*, 1977; Preston *et al.*, 1983). Cattle, upon a primary challenge of *T.annulata*, generate two peaks of cytotoxic cells (Preston *et al.*, 1983). The first peak which is restricted to target cells with BoLA antigens (bovine MHC Class 1 antigens) to the infected host and occurred two weeks post infection. These cells were shown to lyse schizont infected cells and were comprised of BoLA restricted cells, therefore, it was postulated that they were cytotoxic T cells. Cytotoxic cells were detected in the circulation and lymph nodes of cattle destined to recover. The second cell population, cytotoxic for schizont infected cells appeared four weeks post-challenge and manifested itself as both BoLA restricted and non-restricted cytotoxicity. The non-restricted cytotoxic response was interpreted as evidence for the activation of natural killer cells.

Investigations into the cell mediated immune response against *T.annulata* infection suggests that the macroschizont infected leukocyte is the main target. Macroschizont

infected cells have been shown to have increased MHC class II expression, allowing the infected cell to act as an antigen-presenting cell (APC), for CD4+ T cells *in vitro* (Glass *et al.*, 1990). In addition, infected cells *in vitro* have been shown to activate T cells, irrespective of their memory status (Campbell *et al.*, 1995). Therefore, infected cells may have an inherent ability to activate T cells, but inappropriate activation may prevent the host from mounting an effective immune response. Further avoidance of the hosts immune system may be achieved by the modification of BoLA class I antigens on the cell surface of *T.annulata* transformed cells (Oliver and Williams, 1996) since infection is controlled by the MHC class I restricted cytotoxic T cell killing of the infected cells.

In addition to cytotoxic T cell responses, work by Preston and co-workers has shown that cytokines and macrophage-mediated cytostasis of macroschizont-infected cells is a component of the immune response of cattle infected with *T.annulata* (Preston and Brown 1988; Preston *et al.*, 1993). *In vitro* studies have demonstrated that macroschizont infected cells produce TNF- α , and induce TNF- α synthesis in macrophages and interferon- γ (IFN- γ) in lymphocytes (Preston *et al.*, 1993). Also bovine recombinant TNF- α and INF- γ can suppress the transformation of cells infected with trophozoites into macroschizont-infected cells *in vitro* (Preston *et al.*, 1992). Therefore, in addition to cytotoxicity activation of macrophages, cytokines may prevent infection from developing by inhibiting schizont formation. So to summarise, there is evidence for a number of mechanisms which function at cell mediated and the humoral level.

A recent study by Campbell *et al.*, (1995) demonstrated that *T.annulata*-infected APC have the potential to evade the hosts immune system by supplying sufficient signals *in vitro* to activate T cells irrespective of specificity. Importantly, a similar phenomenon is seen *in vivo* with the rapid appearance of activated T cells within the draining lymph node. Subsequently T cell priming in recognised sites of the lymph node does not occur. The alteration of T cell function is most dramatically manifested by the loss of the T cell-dependent compartments of germinal centres, followed by complete germinal centre breakdown.

1.4.4 Humoral Response

Humoral responses to parasite infection have been observed but they do not appear to play a primary role in the development of a protective immune response. Antibodies are consistently observed reacting against the sporozoite, macroschizont and piroplasm stage of the parasite, but, unfortunately sera do not appear to recognise the surface of infected leukocytes in *T.annulata* (Shiels *et al.*, 1989) or *T.parva* (Creemers, 1982; Duffus *et al.*, 1978), or the surface of infected erythrocytes (Hall, 1988). Reaction of sera to the surface of an infected cell would be the first step of antibody mediated cellular lysis. In addition, experiments using lysates or inactivated schizont material to immunise animals have not induced a protective response to *T.annulata*.

Immunity against sporozoites is not essential for a protective immune response since cattle infected with attenuated macroschizont-infected cell lines develop a protective immune response which does not recognise the sporozoite stage (Brown, 1990). However, a protective humoral response has been detected against the sporozoite stage of the parasite. Serum from immune cattle has been shown *in vitro* to be capable of neutralising sporozoite infectivity of lymphocytes (Gray and Brown, 1981) and, after successive sporozoite challenges, the neutralising activity has been shown to increase (Preston and Brown, 1985). Subsequently, monoclonal antibodies were raised against surface molecules from sporozoites and some of these were found to prevent sporozoite infection of lymphocytes *in vitro* (Williamson *et al.*, 1989). Two of the antigens recognised were, *SPAG-1* (sporozoite antigen1) and *SPAG-2* (sporozoite antigen-2). Western blot analysis of a monoclonal antibody raised against *SPAG-1* recognised a series of polypeptides in sporozoite extracts which are thought to be the result of the proteolytic processing of a single gene product. The gene encoding *SPAG-1* was found to contain regions with a high degree of homology to a bovine elastin repeat sequence (Hall *et al.*, 1992) and it has since been proposed that sporozoites may be using mimicry to bind to the bovine elastin receptors as a means of invading cells. However, this region has since been found to be highly polymorphic (Katzer *et al.*, 1994). Additionally, *in vitro* sporozoite infection, targets elastin receptor positive and negative cells populations with an equal frequency, thus making the elastin receptor an unlikely binding ligand for the parasite (Campbell *et al.*, 1994). This molecule, though, is a candidate for inclusion in a sub-unit vaccine and has been analysed to determine the

neutralising B-cell epitopes which lie towards the C-terminus (Boulter *et al.*, 1994; as described in section 1.4.8.3).

Similarly, immune responses against *T.parva* have also been found to be directed against the sporozoite stage. So far only humoral factors capable of neutralising sporozoite infectivity *in vitro* have been identified (Musoke *et al.*, 1982; Dobbelaere *et al.*, 1984; Musoke *et al.*, 1984). These immune responses in *T.parva* target *p67* (Nene *et al.*, 1992) and the 104kDa microneme-rhoptry protein (Iams *et al.*, 1990). The genes encoding these proteins have been cloned and sequenced. Immunisation trials using recombinant *p67* resulted in the protection of the majority of immunised cattle (Musoke *et al.*, 1992; Musoke *et al.*, 1993), indicating that a humoral immune response in cattle against the sporozoite could be protective *in vivo*, although the neutralisation titre does not correlate with protection (described in section 1.4.8.3).

A humoral response to merozoites and piroplasms has been found (Shiels *et al.*, 1989) and antigens recognised by the immune serum seem to originate from the merozoite stage (Glascodine *et al.*, 1990). An immune response to this stage may act primarily to block merozoite invasion of erythrocytes, reducing symptoms associated with the piroplasm stage of the Life-Cycle. This would possibly prevent transfer of the parasite to the tick, and disrupt the formation of a succeeding Life-Cycle generation. The recent ability to produce merozoites in culture from macroschizont-infected lymphocytes incubated at 41°C (Glascodine *et al.*, 1990) has provided a tool for future investigations into the role of anti-merozoite antibodies in protection. In addition, a highly abundant 30kDa antigen was identified on the surface of *T.annulata* (Ankara stock) merozoites (Glascodine, 1990). This molecule is a major polypeptide of merozoites and piroplasms and is strongly recognised by immune sera from cattle infected with *T.annulata*. It has a variable molecular mass in different parasite lines of clonal origin (Dickson and Shiels, 1993). The gene encoding the 30kDa molecule has been cloned and sequenced. Recent studies on the coding sequence have highlighted significant variability within the protein coding sequence between amino acids 40-60 which also corresponds to the sequence where the majority of N-linked putative glycosylation sites are predicted (Shiels *et al.*, 1995). The molecule has been shown to be glycosylated from periodic acid schiffs and the removal of glycosylation has also been

shown to correspond to a loss of species specific reactivity of bovine immune serum (Dickson and Shiels, 1993; Kawazu *et al.*, 1992). Therefore, secondary modifications play a role in the diversity of this molecule. The conservation of this molecule throughout the species of *Theileria* suggests that this molecule has an important functional role (Shiels *et al.*, 1995). The role played by this antigen with the bovine immune system is currently being investigated and work by Tanka, *et al.*, (1990) on the antigen in *T.sergenti* has shown that passive transfer of a monoclonal antibody directed against the molecule resulted in protection of calves against challenge by sporozoites.

1.4.5 Control Measures

Three main types of control for tropical theileriosis are available: chemotherapy, vector control, and vaccination. These measures in recent years have been reviewed by a number of authors; Brown (1990), Dolan (1989) and Tait and Hall (1990).

1.4.6 Treatment of Tropical Theileriosis

Although very effective, chemotherapy has been used sparingly for the treatment of *T.annulata* infection. This is mainly due to high costs of diagnosis and treatment. The most effective therapeutic drugs for the treatment of theileriosis are analogues of the naphthoquinones, menoctone (McHardy *et al.*, 1976), parvaquone (Gill *et al.*, 1981) and buparvaquone (McHardy *et al.*, 1985). These drugs specifically act against *T.annulata* by targeting the electron transport chain of the parasite, destroying macroschizonts within the leukocyte, piroplasms within erythrocytes (McHardy *et al.*, 1983; McHardy *et al.*, 1985). Halofuginone, parvaquone and buparvaquone were found to be active against both *T.annulata* and *T.parva* infection in cattle. The latter would seem to be the more effective treatment for *T.annulata* infection and like parvaquone, it has no significant side effects (Hashemi-fesharki, 1992). Halofuginone treatment of infected animals is toxic at levels close to the therapeutic dose (Schein and Voigt, 1979).

1.4.7 Vector Control

Tick infestations are currently controlled either by spraying or dipping cattle in acaricides such as butocarb or amitraz, once a week. The rationale for this treatment being that the acaricides residues are sufficient for about four days and sporozoite transmission

only occurs three days after tick attachment (Urquhart *et al.*, 1987). Also, acaricide treatment has the potential for controlling other tick borne diseases e.g. heartwater and babesiosis. In areas where indigenous animals reside natural immunity results in a state where challenge and protection are in equilibrium (endemic stability), the strategy should be limited to prevent excessive numbers of ticks coupled to full protection (by drugs or vaccination) of introduced susceptible animals. In regions of endemic instability or where susceptible cattle are reared, interval treatment with acaricides remains an option. Unfortunately this strategy is limited by both excessive cost and the selection of resistant ticks through long term repeated application of acaricides (Wharton, 1976). In addition to their high cost, most acaricides are toxic, cause environmental pollution and leave residues in meat and milk (Drummond, 1976).

1.4.8 Vaccination

1.4.8.1 Infection and Treatment

The infection and treatment method of immunising an animal relies on infecting cattle with either a stabilate of virulent sporozoites or infected ticks, and subsequently treating the resulting infection with chemotherapeutic agents during the latent period. Initially, tetracyclines were used for this application but have been substituted by buparvaquone more recently. Treated animals suffer from reduced symptoms of the disease and develop solid immunity to homologous challenge. At present this is the only method to provide immunity against *T.parva* infection, since an attenuated macroschizont vaccine is not available (Morzaria and Nene, 1990). The infection and treatment method also has been used with success to immunise cattle against *T.annulata* (Gills *et al.*, 1978, 1980) but, it is rarely used because it is more expensive than attenuated vaccines. In addition, infection and treatment may provide other potential problems such as the introduction of other pathogens, and for example, several parasite strains in a "cocktail" provided from the necessity of including cross protective immunity, could introduce parasite strains which were not already present in the region (Musisi, 1990).

1.4.8.2 Cell Culture Schizont Vaccine

Attenuated macroschizont infected cell lines are an efficient and widely used control measure against *T.annulata* infection. Cultivation *in vitro* of macroschizonts by Tsur (1945)

served as the basis for developing a schizont vaccine. Continuous culture of macroschizont-infected cell lines for prolonged periods results in attenuation of virulence (pathogenicity) and is achieved between 30 and several hundred passages depending on the isolate (Pipano, 1989). Inoculation of cattle with an attenuated cell line produces milder clinical symptoms within the animal and a lower parasitaemia. Usually, 10^6 - 10^7 infected cells are used per inoculation (Hall, 1988) and one immunisation is adequate to induce cross-protective immunity, which is reinforced by subsequent tick challenges within the field. However, Friesian cattle usually require a second immunisation from a heterologous schizont stock of a lower passage culture to provide full protection (Pipano, 1981). Vaccinated cattle show good protection against natural infections and heterologous challenge (Gill *et al.*, 1980; Hashemi-Fesharki, 1988). However, the immunity conferred by these vaccines often does not prevent the formation of piroplasms in the bloodstream of the animal, leading to the development of a carrier state by the host. Piroplasms arise either through new infections by feeding ticks or by a limited number of inoculated macroschizont cells differentiating to merozoites (Pipano, 1992).

Attenuated macroschizont vaccines have been developed in Israel (Pipano, 1981), Iran (Hashemi-Fesharki, 1988), India (Grewal, 1994) Kazakhstan (Sabanshiev, 1994) and China (Wenshun and Hong, 1994). So far, it has not been possible to develop an attenuated vaccine for *T.parva*, which has been postulated to be due to an inability to transfer schizonts into lymphocytes of the host recipient (Dolan, 1989). This could account for the suggestion by Musisi (1990) that the BoLA mis-match inhibits the development of immunity in *T.parva* but not in *T.annulata*. Unfortunately the application of an attenuated vaccine has a number of major drawbacks as postulated by Dolan (1989). These include a short shelf life, approximately 1 week at 20°C or 1 month at 4°C, the need for frozen transfer between laboratories and sites of immunisation, infection of immunised cattle with other pathogens, and the expense of testing and a long culture period for vaccine production. In addition, it is not clear whether attenuated cells can revert to virulent parasites following transmission through ticks.

1.4.8.3 Recombinant Vaccines

Antigens from all stages in the Life-Cycle of *T.parva* and *T.annulata* are being investigated with a view to developing sub-unit vaccines against tropical theileriosis and East Coast fever. The aim for such vaccines is to induce cross-immunity to animals at a low cost, with a vaccine that is easy to administer and has a long shelf life. A recombinant vaccine would have none of the potential hazards of the other vaccines previously described. A number of potential antigens have been characterised for inclusion in a molecular vaccine including the sporozoite specific antigens, *p67* from *T.parva* (Nene, *et al.*, 1992) and *SPAG-1* of *T.annulata* (Hall *et al.*, 1992), and the 30kDa major merozoite surface antigen, *TamSI* (Glascodine *et al.*, 1990).

The sporozoite surface antigen *SPAG-1* was identified by a monoclonal antibody which inhibited sporozoite penetration of bovine peripheral blood mononuclear cells *in vitro* (Williamson *et al.*, 1989, see section 1.4.4). The *SPAG-1* gene has since been isolated and sequenced (Hall *et al.*, 1992), and has recently been demonstrated to have cross reactive epitopes with the *p67* *T.parva* sporozoite antigen (Knight *et al.*, 1996, Nene *et al.*, 1992). The neutralising determinant on *SPAG-1*, recognised by the antibody which inhibited sporozoite penetration of host cells, was mapped to 16 amino acids which implicated the C-terminus of the protein as an immunologically relevant region that can be recognised by the bovine immune response (Boulter *et al.*, 1994). Vaccine trials using the C-terminal fragment of *SPAG-1*, expressed as a fusion protein in the e1 loop of the hepatitis B core antigen, generated high antibody titres of neutralising antibodies in the animals immunised. Some marginal protection of the disease was achieved in this small-scale trial, as assessed by the severity of the disease. The clinical symptoms of the disease were delayed, macroschizont parasitosis was reduced and there was significant reductions in the level of macroschizonts in lymph node smears. In addition, there was a lower piroplasm parasitaemia. Therefore, the trend overall would appear to suggest that immunisation with the C-terminal fragment may be beneficial (Boulter *et al.*, 1995).

Another sporozoite antigen recognised by a monoclonal antibody raised against *T.annulata*, which blocks invasion of sporozoites into host cells is *SPAG-2*. Western blot analysis of a monoclonal antibody against *SPAG-2* recognised a series of polypeptides of

molecular weights 150, 67, and 17-20kDa demonstrated that like *SPAG-1*, *SPAG-2* is also processed (Knight, 1993). The gene encoding *SPAG-2* was found to have no sequence homology with known genes when the gene banks were searched and, unlike *SPAG-1*, *SPAG-2* is not exclusively expressed during the sporozoite stage but also during the macroschizont stage as shown by Northern blot analysis (Knight, 1993). Unfortunately, vaccination trials using part of the *SPAG-2* antigen did not result in any detection of protection (Knight, personal communication).

The *T.parva* sporozoite antigen *p67*, a gene related to *SPAG-1*, has recently been used in vaccination trials. Immunisation with recombinant *p67* induced neutralising antibodies and protected six of the nine animals challenged with sporozoites (Musoke *et al.*, 1992; see section 1.4.4). Direct synthesis of *p67* in insect cells has recently been used to allow putative post-translation modification events to occur in an effort to increase the level of immunity induced by *p67*. Unfortunately, it was concluded that the production of recombinant *p67* from insect cells was not a significant advantage over recombinant *p67* expressed in *E.coli*. Thus it is unclear whether post-translational modification was occurring or whether it is relevant to immunity. Efforts are currently being made to improve the immunisation regime (Nene *et al.*, 1995). Some of the trials described above indicate that sporozoite antigens may induce protective immunity, however, that such a molecule should be capable of exhibiting protection alone is to some extent counter intuitive given that evidence suggests that sporozoites only exist free in the circulation for probably no more than 10 minutes (Fawcett *et al.*, 1982b). The subsequent severity of the disease is dependent on the dose of sporozoites (Radley *et al.*, 1974), indicating that achieving a high rate of sporozoite neutralisation should result in an effectively low dose and hence a less severe disease. To achieve a total elimination of sporozoites is probably unrealistic and thus it appears that a sub-unit vaccine containing antigens from all parasite life-cycle stages might be more effective.

Evidence from several studies have shown that immunity to theileriosis is mediated by cytotoxic cells (Eugui and Emery, 1981; Cox, 1982; Preston *et al.*, 1983) and it is thought that these recognise *Theileria*-associated antigens on the surface of the infected lymphocyte (Pearson *et al.*, 1979; Zinkernagel, 1979). Shiels *et al.*, (1986) provided

evidence that monoclonal antibodies, for example 4H5, recognise infection-associated epitopes present on the surface of a cell parasitized with *Theileria* and that such antibodies could mediate complement lysis of the infected cells (Preston *et al.*, 1986). Attempts to establish a direct correlation between parasite infection and the induction of the antigen recognised by 4H5 have as yet been inconclusive (Dando, thesis 1997). However, it is believed that the antigen which 4H5 recognises is host derived and expression of this antigen could involve subtle modifications of host self molecules. Only definitive identification of the antigen that is recognised by 4H5 will resolve this hypothesis. Therefore, it is possible that the antigens which are involved in the immunity to theileriosis are either closely associated with self molecules or involved in subtle modifications of the latter (Creemers, 1982). An infection-specific cell surface antigen in the purified form may have the ability to immunise cattle.

The merozoite stage is, like the sporozoite, invasive and also a potential target for a protective immune response. The most abundant and immunodominant antigen on the surface of the merozoite of *T.annulata* is a 30kDa surface antigen which is conserved throughout all species of *Theileria* analysed to date (Glascodine *et al.*, 1990; Kachani *et al.*, 1992; Dickson and Shiels, 1993). Recently, allelic forms of *TamS1* (*TamS1*-1 and *TamS1*-2) which encode the 30- and 32kDa major merozoite antigens of *T.annulata* respectively, were expressed in a *Salmonella typhimurium* aroA vaccine strain and *Escherichia coli* for inclusion in a sub-unit vaccine (d'Oliveira *et al.*, 1996). Preliminary immunisation trials with the recombinant antigens indicate that there may be some protection against challenge (d'Oliveira, *et al.*, 1997).

Polymorphism has been well documented in *Theileria* and needs to be addressed during the rational design of a multi-subunit recombinant vaccine against *Theileria* infection. When a molecule is considered for inclusion in a sub-unit vaccine it is essential to understand the degree and nature of variation in its structure. Ideally the aim is to identify an immunologically important region which is not polymorphic and will induce immune responses cross-protective against different strains of the parasite.

Initial evidence for strain variation of *T.annulata* was observed during the vaccination of calves by the infection and treatment method. It was demonstrated that calves vaccinated by this method were protected against homologous challenges but were not protected by heterologous challenge (Gill *et al.*, 1990), indicating the existence of different parasite strains from distinct geographical regions. Variation of geographically distinct stocks in *T.parva*, were characterised by raising monoclonal antibodies against different stocks. Monoclonal antibodies raised in this way were shown to be specific for certain *T.parva* strains demonstrating polymorphism between stocks (Minami *et al.*, 1983). Subsequently monoclonal antibodies were raised which demonstrated variation in *T.annulata* strains allowing strain typing (Shiels *et al.*, 1986). Variation between *Theileria* stocks has been characterised by monoclonal antibodies raised against macroschizonts and piroplasms (Minami *et al.*, 1983, Shiels *et al.*, 1986, Dickson and Shiels, 1993), demonstration of glucose phosphate isomerase polymorphisms (Melrose *et al.*, 1984; Wilkie *et al.*, 1986), DNA melting points between different parasite strains (Allsopp *et al.*, 1988), ELISA and western blot analysis (Kawazu *et al.*, 1992b), variations in restriction fragment length polymorphism (RFLP) patterns (Mozaria *et al.*, 1990; Katzer *et al.*, 1994, Ben Miled *et al.*, 1994; Shiels *et al.*, 1995), and characterisation of polymorphic strains using PCR (Ben Miled *et al.*, 1994).

The underlying basis for the diversity of the immunodominant 30-32kDa *T.annulata* merozoite/piroplasm antigen is probably differential glycosylation. It has been shown by Shiels *et al.*, (1995) that parasite stocks from different geographical regions can be characterised based on their *TamS1*, EcoRI RFLP profile. Sequence analysis of homologous genes in other species of *Theileria* have shown that there is a significant variability within a specific section of the coding region (amino acids 50-60) of the *TamS1* gene. This variable sequence was located in the region of the molecule where the majority of N-linked putative glycosylation sites were predicted and the molecule is known to be glycosylated from periodic acid shift staining (Dickson and Shiels, 1993). In addition, periodate treatment experiments resulted in the removal of species specific reactivity of bovine immune serum when tested against recombinant antigens suggesting that secondary modifications play a role in diversity. Detailed characterisation of the structure and immunogenicity of the secondary modification have not been achieved to date, and the possibility that the amino

acid sequence contributes to the antigenic diversity via primary or secondary structure has not been conclusively ruled out.

Further indications of polymorphism at the protein level were obtained by SDS-PAGE. The *T.parva* polymorphic immunodominant macroschizont antigen (PIM) was shown to have a size polymorphism (Toye *et al.*, 1991). Subsequently, similar size polymorphisms were observed in the *T.parva* immunodominant schizont surface antigen from different stocks using 2 dimensional SDS-PAGE followed by western blotting (Sugimoto *et al.*, 1992). This antigen is probably the same as PIM.

Polymorphism has also been detected in the *SPAG-1* molecule in *T.annulata*. Findings by Katzer *et al.*, (1994) have shown that the C-terminal half of *SPAG-1* is the most conserved region of the molecule and, interestingly, the N-terminus is also reasonably well conserved. It is possible that this conservation of sequence could be maintained for a functional purpose such as host cell invasion. Polymorphism is greatest between residues 213 and 502 in the second quarter of the molecule and is most extensive between amino acids 285 and 369. The variation is due to multiple gap/insertion and amino acid substitutions.

Within the context of vaccine development based on recombinant antigenic determinants, the definition of constant and variable regions will assume more relevance when more information about immunodominant molecules in *Theileria* is acquired. Ideally, a recombinant vaccine would be based on protective T-cell or B-cell epitopes which are identical for all *Theileria* species. However, polymorphism in *Theileria* has been well documented and needs to be considered during the rational design of an effective sub-unit vaccine. Therefore, when a molecule is considered for inclusion in a sub-unit vaccine it will be essential to understand the degree and nature of variation in its structure. Ultimately, it is likely that a sub-unit recombinant vaccine consisting of engineered antigens from sporozoites, schizonts as well as from merozoites/piroplasms will be necessary to induce protection against *Theileria* infection, initiated through the bite of a tick. Therefore, further work is necessary to identify schizont antigens which induce an immune response, and to

clarify immunologically important regions within known sporozoite and merozoite/piroplasm antigens.

1.5 Stage Differentiation in Protozoan Parasites

In general, differentiation from one life-cycle stage to another is fundamental to the biology of protozoan parasites and facilitates the expansion within and transmission between host and vector. Unlike somatic differentiation in multi-cellular eukaryotic systems, terminal differentiation does not take place. Instead the protozoan parasite cycles through a number of distinct life-cycle stages, one of which being the formation of sexual forms. One way to view stage differentiation in protozoa is to consider it as the net result of alterations to the control of gene expression. This in turn results in the production or shut down of polypeptides with metabolic or structural functions. Therefore, the signals which initiate a parasite to transform must ultimately exert their effect by altering the factors which determine the expression status of a particular set of genes, but how this occurs in response to changing extracellular environments encountered during the parasite Life-Cycle is unknown. Studies on developmental and cellular differentiation in other eukaryotic systems provide some clues to these events. Thus, the differential state of a cell during lineage development is probably determined by a hierarchy of regulatory factors, which control themselves by autoregulation, and control a number of target genes (Serfling, 1989). Precedents for several types of mechanisms exist, including homeotic regulation in *Drosophila* (McGinnis and Krumlauf, 1992) and sex determination in *Drosophila* and *C.elegans* (Hodgkin, 1990). Maintaining the level of gene expression via autoregulation can determine a set cellular state or in the case of the parasite a particular Life-Cycle stage. In order to move from this status, therefore, a differentiation signal must be transduced into altering the autoregulatory networks of the preceding Life-Cycle stage into those of the next programmed stage. A variety of signals have been shown to induce differentiation of protozoan parasites. Changes in temperature (Hulliger, 1965; Soete *et al.*, 1994; Van der Ploeg, 1985), pH (Soete *et al.*, 1994; Ziberstein *et al.*, 1991), growth to stationary phase (Vickerman, 1985; Sachs and Perkins, 1984, 1985), addition of drugs (De Gee *et al.*, 1994) and intermediates of the TCA cycle (Brun and Schonenberger, 1981) have all been implicated in triggering differentiation of one Life-Cycle stage to another. An alternative to autoregulation is that cells receive a constant stimulus which keeps turned on particular

regulators of gene expression. When the signal is removed or a new signal comes on then differentiation would occur.

It is unclear how these signals initiate differentiation or if there is any mechanism which is common to them directly. However, several models have been proposed for the modulation of differentiation by alterations to the extracellular environment. It has been suggested that in some instances the signal to differentiate is an exogenous factor produced by the parasite in order to prevent life-threatening parasitaemia. In *Trypanosoma brucei*, differentiation from long-slender to the short-stumpy form has been proposed, based upon *in vitro* studies on cultivation of the parasite, to be induced by the production of a factor (stumpy initiation factor) at a certain population density of long slender blood stream forms (Hesse *et al.*, 1995; Giffin and McCann, 1989; Seed and Sechelski, 1989; 1992; Pays *et al.*, 1993). It has been proposed that trypanosomes may constantly produce this factor, but only above a certain threshold does it induce the differentiation event. Therefore, the concentration of such a factor would be low as long as the population density of the parasite was low, allowing the establishment of parasitaemia before reaching a density which induces differentiation and prepares the parasite for transmission to the insect vector as the stumpy form is thought to ensure transmissibility (Hesse *et al.*, 1995).

Additional studies of *T.brucei* have implicated alterations in the parasite environment to be the source of stress factors which induce differentiation from the bloodstream form (via the short stumpy) to the procyclic form, found within the insect vector. Analysis of gene expression during the transition between bloodstream and procyclic forms detected increased expression and activity of a serine-threonine kinase, termed Nrk, due to translational control (Gale *et al.*, 1994), and a transient increase of adenylate cyclase activity was detected (Rolin *et al.*, 1993). The role of these two enzymatic activities in differentiation remains totally obscure. Expression of the variant surface glycoprotein (VSG) is terminated as soon as the bloodstream forms are placed in conditions which induce differentiation and in view of the rapidity of this phenomenon, it is thought that the very first event is cold-shocking-dependent blocking of RNA elongation (Alexandre *et al.*, 1988; Pays *et al.*, 1990). Secondly, within 2 hours of inducing differentiation, procyclin mRNA start to accumulate (Pays *et al.*, 1993). Inhibition of protein synthesis within bloodstream

forms has been shown to mimic these effects of differentiation (Dorn *et al.*, 1991) and a transient inhibition of protein synthesis has actually been detected *in vivo* during differentiation (Bass *et al.*, 1992). Taking all of these observations together, Vanhamme and Pays (1995) have proposed the following model of events occurs upon cold shock, a temperature-dependent inhibition of protein synthesis causes the rapid loss of labile protein involved in transcription elongation on the VSG expression site as well as in the repression of procyclin mRNA synthesis. None of the regulatory factors proposed in this model have been identified to date.

Two major environmental factors, temperature and pH, play an important role in *Leishmania* species, triggering differentiation of promastigotes (insect form) to the intracellular amastigotes (mammalian form). Changes to both of these factors take place following inoculation of the host and phagocytosis by host macrophages (Chang and Dwyer, 1976). Elevation of temperature appears to be a major signal in the induction of morphological changes into amastigotes in New World *Leishmania* species (Darling and Bloon, 1990; Eperson and McMahon-Pratt, 1989; Pan, 1984). However, changing temperature only, is not sufficient in most other *Leishmania* species for generation of a long-term amastigote-like culture. A combination of temperature elevation and a decrease in pH facilitates the establishment of stable amastigote-like cultures, which displays not only the typical amastigote morphology but also markers for differentiation, including specific cysteine proteinases, resistance to complement lysis, and increased infectivity (Bates *et al.*, 1992; Joshi *et al.*, 1993; Bates, 1993).

Temperature shifts are commonly found to induce parasite differentiation and the most striking example would seem to take place, not surprisingly, in parasites which transfer between a poikilothermic insect vector and a homeothermic mammalian host. Therefore, heat shock proteins (Hsps) must play important functions in parasitic and other organisms whose Life-Cycle imposes successive temperature shifts. It has been shown that these events can be associated with differential expression of heat shock proteins (Hsp) (Van der ploeg *et al.*, 1985). In *T.brucei*, mRNA transcripts of Hsp 70 and Hsp 83 were found to be 25 to 100 times more abundant in bloodstream forms compared to the procyclic insect stage (Van der Ploeg *et al.*, 1985). While a heat shock response in other eukaryotes

serves to shut down overall protein synthesis, elevated temperature, does not interfere with parasitic protozoa cell proliferation, indicating a different role for the heat shock response in these organisms. In *Leishmania* (Eperson and McMahon-Pratt, 1989) elevation of temperature appears to be significant stimulus for the induction of differentiation *in vitro*. However, heat shock proteins do not appear to play an active role in this response because as yet no direct correlation has been found between their expression and differentiation (Shapira, *et al.*, 1988; Zilbertstein *et al.*, 1991). It is conceivable that their expression is not strictly temperature-dependent, but rather stage-specific (Gerhards *et al.*, 1994).

An important event in the pathogenesis of toxoplasmosis is the interconversion between bradyzoite and the tachyzoite stage within the human host. Tachyzoites, a rapidly dividing Life-Cycle stage, are involved in an acute infection and the differentiation of the tachyzoites to bradyzoites, a dormant stage, correlates with the onset of protective immunity. Bradyzoites are located within cysts and are able to reconvert into tachyzoites (Gross *et al.*, 1996). However, the factors that influence either bradyzoite or reactivation of tachyzoites are unknown. Conversion from tachyzoite to bradyzoite is a gradual process with sequential expression of stage specific molecules and gradual morphological changes. Two possible models have been proposed concerning the correlation of host immunity and stage differentiation. Firstly, the host's immune system could stimulate differentiation either by a spontaneous conversion of the parasite with subsequent immune system-driven selection, or the immune system or other factors induce stage conversion. Recently conducted experiments on differentiation *in vitro* have shown the conversion model to be the most likely hypothesis (Bohne *et al.*, 1993; Soete *et al.*, 1994). Several reliable and efficient methods have been reported for inducing the transformation of tachyzoite to bradyzoite including the absence of immune factors, metabolic, chemical or physical stress factors such as alkaline pH, sodium arsenite or heat shock (Soete *et al.*, 1994). Similarly, Gazzinelli *et al.*, (1993) have shown that IFN gamma is able to inhibit the multiplication of tachyzoites *in vivo* and induces encystation, possibly by means of NO which is known as an inhibitor of iron-sulphur proteins involved in mitochondrial respiration (Stamler *et al.*, 1992). Together with this plus data from Tomavo and Boothroyd (1995) and Bohne *et al.*, (1994) suggests a connection between mitochondrial function and the differentiation of tachyzoites. Investigations using mitochondrial inhibitors demonstrated that inhibition of

mitochondrial function induces differentiation to the bradyzoite *in vitro*. In addition, it has been observed that after invasion of *T.gondii* into the host cell, mitochondria are usually located around the parasitophorous vacuole, indicating that they might be important for providing energy for the replicating parasite (Bohne *et al.*, 1994).

Differentiation of bradyzoites into tachyzoites occurs through an intermediate stage co-expressing bradyzoite and tachyzoite specific proteins, a stage which is also found during the transformation of tachyzoites into bradyzoites. The onset of differentiation has been defined by the expression of P30, the major surface protein of tachyzoite SAG-1 which is not expressed in bradyzoites, but occurs before parasite division, therefore, differentiation is not linked to division (Soete *et al.*, 1993). This appears to be the opposite of tachyzoite differentiation where multiplication of parasites starts before the expression of bradyzoite-specific proteins. Also bradyzoite-tachyzoite interconversion is not synchronous within a cell culture or even within a parasitophorous vacuole (Darde *et al.*, 1989; Soete *et al.*, 1993). This could suggest that internal factors can play a role in differentiation whereas only external factors had been supposed so far.

After the invasion of the red blood cell of the vertebrate host, *Plasmodium* parasites develop either into asexual multiplying schizont or into sexual forms, the gametocytes responsible for parasite transmission. Therefore, a mechanism has evolved which enables the parasite cell to activate one of two mutually exclusive developmental programmes (Bruce *et al.*, 1990). This mechanism must control expression of the appropriate stage-specific genes in the daughter cell, after the events of schizogony and merozoite invasion. The details of induction to differentiate into either form are still uncharacterised, however, commitment occurs in such away that each individual schizont produces progeny of merozoites which develop either into asexual or into sexual parasites. It has been postulated by Bruce *et al.*, (1990) that commitment in *P.falciparum* to differentiate into either form is predetermined prior to invasion of the red blood cells, following a developmental choice which has occurred in the parental schizont.

Molecular studies into the progression of differentiation on *P.falciparum* and the rodent *P.berghei* have established lines which after prolonged periods of propagation lose

the ability to differentiate into the sexual forms of the parasite. These include a block in the progression beyond morphological stage III of gametocyte maturation (Teklehaimnot *et al.*, 1987) and deficient production of male gametes (Vaidya *et al.*, 1993). In several cases the loss of gametocyte production has been attributed to a deletion of chromosome 9 (Forsyth *et al.*, 1990; Janse *et al.*, 1992) and it has been postulated that the terminal portion of chromosome 9 regulates an early step in the specialisation of the sexual cell. Other parasite sub-clones have been isolated which have lost the ability to produce gametocytes but maintain a full length chromosome 9 (Alano *et al.*, 1995). Therefore, these mutants are probably blocked in an early stage of gametocytogenesis different from that of the lines previously discussed. It would appear that in *Plasmodium*, that the mechanism for sexual development can be affected in several different genetic and functional ways.

1.5.1 Stage differentiation in *Theileria*

A clinical symptom of tropical theileriosis is the onset of fever of (41°C), which develops early in infection and remains until recovery or death (Barnett, 1977). This observation led to the hypothesis that elevated temperature occurring during the course of an infection could provide the stimulus for differentiation from macroschizont to merozoite stage. Work by Hulliger *et al.*, (1966) supported this hypothesis because it was shown that *in vitro*, differentiation could be induced by an elevation in culture temperature from 37°C to 41°C during a six day period. However, the exact nature of the part played by elevated temperature was unknown. In contrast *in vivo* studies carried out by Jarrett *et al.*, (1969), did not find a correlation between the appearance of merozoites and the onset of fever. Indicating that the process controlling the switch to merozoite formation functions on the basis of time rather than temperature. The timing of the fever was shown to be determined by the initial inoculation to the animal but this was independent of the appearance of merozoites. These findings led to the hypothesis that parasite differentiation may occur after a set number of mitotic divisions of the macroschizont infected cell. This so-called mitotic clock theory was also suggested to be involved in higher eukaryotic differentiation (Temple and Raff, 1986). However, the timing of differentiation from macroschizont to merozoite was shown to be susceptible to changes in the growth factors supplementing *in vitro* culture (Bogler *et al.*, 1990) and the correlation between division number and differentiation was lost. In a similar fashion as *Theileria* infected cell lines can undergo unlimited proliferation

without differentiation when cultured *in vitro* at 37°C, suggesting that like higher eukaryotic cells, the mitotic clock hypothesis is too simplistic to explain merozoite differentiation.

The study of temperature induction of macroschizont differentiation to merozoite in *Theileria* has been greatly aided by the isolation of macroschizont infected cell-lines with either enhanced or diminished abilities to differentiate *in vitro* at 41°C (Shiels *et al.*, 1992). Established *Theileria* infected cell lines cultured at 37°C *in vitro* undergo unlimited proliferation without differentiation. Exposure to an elevated temperature does not induce an immediate response, cells were cultured at 41°C for more than 24 hours before a proportion of the *Theileria* infected cells were induced to differentiate. Induction to differentiate in "enhanced" macroschizont infected cell-lines is characterised by an enlargement of the macroschizont and an increase in the rate of parasite nuclear division over the first two days at 41°C (Hulliger *et al.*, 1966; Glascodine *et al.*, 1990). This increase in parasite nuclear division is then coupled with a decrease in host cell division, and at a certain time point host cell division is completely inhibited. As the rate of host cell division is reduced, parasite cell division also becomes slower, owing to the association of the parasite with host cell spindle for separation of the macroschizont. Consequently, an increase in parasite nuclear division/growth and a decrease in parasite cell division led to the host cell's cytoplasm being completely filled with the enlarged macroschizont. No comparable size increase is seen in the differentiation diminished cell line incubated at 41°C (Shiels *et al.*, 1992). Thus, it was proposed from this work that the substantial enlargement of the schizont to a predetermined size or condition, or a reduction in the rate of parasite cell division is what triggers differentiation (Shiels *et al.*, 1992). For example it has also been found for other protozoan systems that alterations to the proliferative rate influences differentiation potential. Changing from logarithmic- to stationary-phase growth has also been shown to influence differentiation in *Trypanosoma brucei* (Vickerman, 1985), *Leishmania* (Sachs and Perkins, 1984, 1985) *Leishmania major* (Shapira *et al.*, 1988), *Plasmodium falciparum* and *T.gondi*.

Investigations by Shiels and co-workers demonstrated that macroschizont cell-lines did not immediately respond to temperature induction of differentiation as cells had to be cultured at 41°C for more than 24 hours before a proportion of the culture was induced to

differentiate. In addition a series of pulse experiments showed that an initial reversible phase exists before the macroschizonts are fully committed to differentiate, because incubation periods at 41°C correlated with increase levels of differentiation, returning the cultures to 37°C reduced differentiation compared to cultures continuously incubated at 41°C for 7 days. Thus, it was postulated that the longer the cells were incubated at 41°C the more likely they were able to reach a threshold which committed them to irreversible differentiation events.

Before the regulatory controls that operate during the differentiation of *Theileria* macroschizont infected cell-lines to merozoite can be defined, it is necessary to define the temporal order of gene expression reprogramming during this event. Using monoclonal antibodies and gene probes Shiels *et al.*, (1994) were able to define several molecular changes associated with differentiation. Expression of the 30 kDa merozoite surface polypeptide (*TamSI*) and the 117 kDa rhoptry antigen (*TamRI*) were shown to dramatically increase between days 4 and 6, while the level of major macroschizont polypeptides were found to decrease. The *TamSI* polypeptide and mRNA were detected early during the induction of differentiation and were detected unambiguously at day 2 of differentiation, but also occurred very faintly at day 0. Using pulse experiments similar to those described earlier, it was demonstrated that the expression of *TamSI* was reversible in the early stages of differentiation. Reversible elevation of the *TamRI* gene was not detected. The factor(s) which allow low level expression of *TamSI* in the preceding stages and during the reversible phase of differentiation were unknown. Low level expression in preceding stages appeared to be specific to *TamSI*, however, it may occur for other as yet undefined merozoite genes. It has since been proposed that the *TamSI* polypeptide is one of the first proteins to be expressed during the temporal order of induced differentiation of macroschizont to merozoite. Low level expression in preceding stages may therefore be due to an overlap of target regulation by macroschizont regulatory factors or due to a low level of factors which specifically regulate merozoite target gene expression. In addition there appears to be a secondary control of gene expression occurring between days 4 and 6 of differentiation where there is a dramatic increase of merozoite genes, and repression of macroschizont gene expression, possibly coupled to irreversible differentiation (Shiels *et al.*, 1994).

Recloning of the macroschizont infected cell-line with enhanced ability to differentiate resulted in the isolation of cell-lines which have a severely diminished ability to differentiate. RFLP analysis demonstrated that one of these attenuated lines appeared to be of an identical genotype to the enhanced parental cell-line but lacking an ability to fully differentiate. However, the loss of differentiation was not absolute and occasionally a low number of parasites would differentiate to merozoite when cultures were incubated at 41°C for prolonged periods (Shiels *et al.*, 1994). Therefore, the reduced ability to differentiate must have been caused by an alteration of the original cell line and a similar result has been observed during the propagation of *Leishmania* promastigotes *in vitro* (da Silva and Sacks, 1987). The observation that parasites were able to differentiate occasionally in the attenuated lines coupled to the finding that the *TamSI* level of up regulation during the reversible phase was higher in the differentiation competent cells lead to the hypothesis that the difference in the ability of these two cell lines to reach commitment was a quantitative one. Shiels *et al.*, (1994) proposed that a quantitative basis for differentiation could be explained by the postulation of a stoichiometric model for stage differentiation. In this model a mechanism is believed to operate, where the level or activity of factors controlling gene expression build up to a concentration threshold, which triggers commitment. Such a model could explain the reversibility of the differentiation process as in this phase the accumulation of the factors defining cell specificity would be below the threshold to commit the cell to irreversible differentiation. Removal of the stimuli necessary for inducing differentiation during this initial phase would result in the cell returning to the undifferentiated state, and increasing factor level, which was the result observed in the pulse field experiments.

The observation of asynchronous differentiation within a culture of differentiation enhanced cells is believed to be the inevitable consequence of a stochastic model, as the time taken to reach commitment to differentiation will be dependent upon how near or far each cell is from the commitment threshold. The regulatory factors involved in the concentration threshold have yet to be identified in *Theileria*. The stochastic model proposed by Shiels *et al.*, (1994) is reminiscent of differentiation events which occur both in other protozoan parasites and higher eukaryotic systems. A step process for committing to differentiation has been described for *T.brucei* differentiation from bloodstream form to

procyclic form (Czichos, *et al.*, 1986; Ehlers *et al.*, 1987; Pays *et al.*, 1993) and in higher eukaryotes, DMSO differentiation of either the human leukemic cell line HL-60 towards granulocytes (Tarella *et al.*, 1982) or erythroleukemia cells (Gusella *et al.*, 1976) has been shown to involve an initial reversible phase. Asynchrony of differentiation has been observed in *Trypanosomes* (Soete *et al.*, 1993, Heath *et al.*, 1990, Duszenko, 1990), in *Leishmania* (Sacks and Perkins, 1984) and in higher eukaryotes, such as DMSO induced differentiation of HL-60 cells (Tarella, *et al.*, 1982).

While commitment to differentiate is proposed to be determined by reaching a threshold mechanism of a factor(s), it is unclear whether this mechanism functions on the basis of time. From previous work it was postulated that disruption to the synchrony between parasite growth and division influence differentiation potential. Therefore to address whether differentiation occurs as a function of time, Shiels *et al.*, (1997) examined the effect of drugs which inhibited DNA synthesis (aphidicolin) or growth (oxytetracycline) on differentiation. Aphidicolin was shown to reduce the time taken for individual parasites to reach commitment while treatment with oxytetracycline had the opposite effect. Although the aphidicolin treated cultures at 41°C showed higher levels of differentiation, the stochastic nature of the differentiation process was retained. The drugs either shorten time taken to commitment or delay it. This differentiation does occur as a function of time which can be corrupted. The timing is associated with increased *TamSI* levels, supporting the stoichiometric model. However, aphidicolin does not induce differentiation at 37°C even though host cell division (and the parasite) cell cycle were inhibited. This suggested that blocking host or parasite cell division per se was not enough to initiate differentiation. Additional experiments demonstrated that at 41°C there was a disruption between protein synthesis and DNA levels and it was proposed that in order to reach commitment there has to be a stoichiometric increase in factors relative to DNA templates. This postulation was supported by data which indicated that the drugs either exacerbated this disruption and reduced timing, or inhibited it and delayed the time taken to reach commitment. In conclusion it appears that cell cycle and a reduction of host cell division are associated with parasite commitment but are not the cause. Instead by exacerbating a disruption between protein relative to DNA levels at 41°C using aphidicolin increased levels of differentiation can be induced. Therefore, Shiels *et al.*, (1997) propose that the differentiation clock is

controlled by the progressive build up of a regulatory factor(s) of merozoite gene expression relative to their DNA templates, until a quantitative commitment threshold is reached. This elaboration of Shiels *et al.*, (1994) stoichiometric mechanism, could account for the corruption of the timing mechanism to differentiation and could explain why a temperature shift does not appear to be necessary for initiation of the differentiation process *in vivo* (Jarret *et al.*, 1969). This could be explained by proposing that *in vivo* at the bovine body temperature (38-39°C) the parasite is held in an equilibrium between growth and division which favours a time dependent progression towards commitment. This equilibrium could be altered *in vitro* due to the selection of proliferation at 37°C which can be re-established by placing the cultures at 41°C allowing differentiation to be initiated. Another differentiation event in the Life-Cycle of *Theileria* which may have a corruptible timing mechanism controlling differentiation, could be the formation of sporozoites in the tick salivary gland. *In vivo*, an increase in temperature can induce sporozoite formation within the tick salivary gland, but at a reduced level and with a delayed timing, compared to the induction of a blood meal (Walker and McKellar, 1983). Therefore, *in vivo* differentiation to the sporozoite is initiated by temperature whereas for the merozoite this appears unnecessary.

1.6 Gene structure and Regulation in *Theileria*

DNA replication and gene expression in protozoan parasites is carried out by similar mechanisms to those found in other eukaryotic organisms. The study of gene expression in protozoa has uncovered several novel processes involved in transcription and gene arrangement, although in general, DNA replication, gene expression, RNA processing, translation etc. follow known eukaryotic processes.

1.6.1 Gene Structure

Extrachromosomal elements are important carriers of genetic information and have been found in all types of organisms, including protozoa. Examples of extrachromosomal elements in protozoa, include the mitochondrial DNA of trypanosomes, termed kinetoplast DNA. The kinetoplast consists of a huge network of intercatenated mini- and maxicircles (Benne, 1994), and circular elements found in *Leishmania* (Hightower *et al.*, 1987). A 6.5 kb and 7.1 kb linear double stranded extrachromosomal DNA element has been discovered

in *Theileria annulata* (Hall *et al.*, 1990) and *T.parva* respectively (Kairo *et al.*, 1994). Molecules of similar size have been detected in the DNA of *T.buffeli*, *T.mutans*, *T.sergenti* and *T.taurotagi* (Kairo *et al.*, 1994). The element in *T.parva* has the potential to encode three mitochondrial respiratory chain proteins, apocytochrome *B* and polypeptides I and III of cytochrome oxidase, and it has fragmented rDNA sequences. Functional genes for *cox1*, *cob* and organellar LSU and SSU rRNA are universal features of mitochondrial DNA (Gray, 1989). Hence, the structure, protein coding potential and scrambled rDNA sequences are reminiscent of those described in the mitochondrial DNA of *C.reinhardtii* (Boer and Gray, 1988; Michaelis *et al.*, 1990), suggesting that the extrachromosomal DNA element is part, if not all, of the parasite mitochondrial genome. The extrachromosomal element is a direct homologue of elements described in the malaria parasites *P.yoelii* (Vaidya *et al.*, 1989), *P.gallinaceum* (Aldritt *et al.*, 1989) and *P.falciparum* (Feagin *et al.*, 1992). The *P.yoelii* (Vaidya *et al.*, 1989; Suplick *et al.*, 1990) and *P.falciparum* (Feagin *et al.*, 1992) elements have the same protein and fragmented rDNA coding potential as the *T.parva* element, but there is a difference in the organization of the sequences. However, the most obvious difference between the extrachromosomal molecules is their structure: *Plasmodium* contains a tandemly linked element (Vaidya and Arasu, 1987; Joseph *et al.*, 1989), *Babesia bovis* contains a circular element (Jasmer *et al.*, 1990), while *T.annulata* (Hall *et al.*, 1990) and *T.parva* (Kairo *et al.*, 1994) contain a linear molecule. In addition, analysis between related apicomplexan parasites is complicated by the finding of a DNA fragment containing partial *cob* and *coxI* sequences dispersed within the nuclear DNA of *Toxoplasma gondii* (Ossorio *et al.*, 1991).

Evidence exists for extensive polycistronic transcription of the *Plasmodium* 6kb element, which implies that processing of large precursor RNAs is a necessary part of the expression of its genes (Ji *et al.*, 1996). Polycistronic transcription and processing have been reported for other mitochondrial genomes including the 16kb circular mammalian mitochondrial genomes (Clayton, 1984; Shadel and Clayton, 1993) and mitochondrial gene expression in *Trypanosoma brucei* (Feagin and Stuart, 1985; Feagin *et al.*, 1985; Jasmer *et al.*, 1985). Therefore, the proposed polycistronic transcription and processing of the mitochondrial genes on the 6kb element are consistent with known mitochondrial expression mechanisms. However, as yet there is no evidence for mRNA editing or

processing of any of the genes transcribed on the extrachromosomal element in *T.parva*. Northern blot analysis of RNA transcripts generated by the element by Northern blotting have corresponded in size to the full length cDNA for the open reading frame (Kairo *et al.*, 1994).

Novel structural arrangements of genes have been detected in *Theileria* and other member of the Apicomplexan family. Typically eukaryotic organisms have multiple copies of rRNA genes, which are tandemly arranged, but the organisation of *Theileria* rRNA genes appears to be unlike that of most eukaryotes. The ribosomal RNA genes are split between two separate chromosomes (Kibe *et al.*, 1994). This unusual rRNA gene organisation appears to be widespread among apicomplexan protozoa, though not universal, since it does not appear in *Toxoplasma gondii* (Johnson *et al.*, 1987). *Plasmodium* spp. have 4-8 rRNA transcription units (McCutchan, 1986) and *Babesia* have 3 (Dalrymple, 1990; Reddy *et al.*, 1991).

Another unusual cluster of *Theileria* genes is that of the *Tpr* family, a collection of repetitive sequences with protein coding potential found in *T.parva* (Baylis *et al.*, 1991). The *Tpr* family is transcribed in piroplasms, and apparently consists of a few potentially complete genes containing *Tpr3* and many partial copies of either *Tpr1* or *Tpr2* arranged in long arrays. Possible mechanisms by which the partial genes in the *Tpr* family could be expressed either in or as polypeptides include, DNA rearrangements, RNA splicing and gene conversion. This arrangement appears to be unique, although reminiscent of other systems including immunoglobulin genes (Alt *et al.*, 1987). A *BabR* multigene family in the protozoan parasite *Babesia bovis* has some resemblance to the *Tpr* family, however, the sequences are not significantly similar and there are relatively few individual genes (Cowman *et al.*, 1984).

1.6.2 Initiation

The study of gene regulation in *Theileria annulata* is in its infancy. One of the first indications of differentially controlled gene expression was demonstrated by Shiels *et al.*, (1986) using monoclonal antibodies raised against the macroschizont infected cell. These antibodies raised were shown to either react exclusively with one stage of the parasites Life-

Cycle e.g. with schizonts or with several stages e.g. schizonts and sporozoites. Therefore, it was possible to conclude that *Theileria* expressed antigens in a stage specific manner. Subsequently, additional monoclonal antibodies have been raised which are specific for macroschizonts of *T.parva* (Shapiro *et al.*, 1987). Therefore, immunological studies of antigen expression in *Theileria* demonstrated that a mechanism for regulatory control of stage specific gene expression must exist within the parasite, but as yet no component of this mechanism has been isolated. Further evidence was obtained by Northern blot analysis, indicating stage-specific patterns of RNA accumulation (Shiels *et al.*, 1994; Mason *et al.*, 1989; Swan *et al.*, 1996; Kinnaird *et al.*, 1996, see section 1.5.1).

Most of the genes so far isolated from *Theileria* show some form of stage regulated control throughout the Life-Cycle stages analysed. In addition, several genes have been shown to be constitutively expressed. Constitutively expressed genes investigated, include apocytochrome *B* (see section 1.6.1), a cysteine protease (Baylis *et al.*, 1992), a hsp 70.1 gene and 28S ribosomal RNA (Shiels *et al.*, 1994). The hsp 70.1 gene of *T.annulata* appears to be constitutively expressed in four stages of the parasites which occur in both the mammalian host and tick vector. The expression of the hsp 70.1 gene product is inducible by heat shock in both sporozoites and in macroschizont infected cell lines (Mason *et al.*, 1989). Induction of hsp 70.1 has been speculated to play a role in macroschizont differentiation to merozoites, but unlike other eukaryotes the expression of the heat shock genes does not interfere with cell proliferation, indicating that the heat shock response may play a different role in parasitic protozoa. Although these genes are constitutively expressed during the mammalian phase of Life-Cycle, very little is known about gene expression within the tick. It may become apparent that not all genes constitutively expressed in the bovine host are expressed during the stages in the tick host. This may be of particular interest with respect to the hsp genes within the poikilothermic host.

Stage specific regulation has been documented for a number of genes in *Theileria* by either Northern analysis or antigenic analysis. Transcription of the *SPAG-1* (Williamson *et al.*, 1989) and *p67* genes occurs only during the sporozoite stage of the parasite (Nene *et al.*, 1992). ORF-1, an open reading frame 5' to the *p67* of unknown function, is transcribed during the schizont stage and the open reading frame 3' to the *p67* gene, ORF-2, is

transcribed during the sporozoite, schizont and piroplasm stages. Expression of genes during differentiation of the parasite from one life-cycle stage to another does not appear to occur uniformly. For example, mRNA and polypeptide encoded by *TamSI* has been observed to occur several days before the expression of the *TamRI* 117 kDa rhoptry protein encoding gene during differentiation to merozoite, demonstrating a temporal order for the stage-specific expression of genes between one Life-Cycle stage to another (Shiels *et al.*, 1994). Also *TamSI* must be on the surface of the membrane before budding, so that the merozoite get its surface coat. The mRNA for the 117 kDa rhoptry protein occurs approximately 2 to 3 days prior to rhoptry formation, presumably indicating the need for nascent polypeptide prior to the formation of this organelle. These limited examples demonstrate the necessity for a sequential ordering of gene expression events during differentiation which allow a functional merozoite to be formed correctly. Unfortunately very little is known about the mechanisms allowing for these complex networks of control.

In higher eukaryotes, genes encoding mRNAs are transcribed by RNA polymerase II (pol II) the activity of which is initiated by transcription signals via promoters upstream to the start site of transcription (or RNA formation). Some promoter elements, such as TATA, GC and CCAAT boxes are highly conserved among many of the genes transcribed by pol II and interact with the basal transcription machinery (Pugh and Tjian, 1992). Other sequence elements are less common or unique and are implicated in specialised types of signal-dependent transcriptional regulation, such as controlling developmental expression, response to heat shock, hormones, and growth factors (Mitchell and Tjian, 1989). Within the group of protozoan parasites, analysis and characterisation of promoters has been limited. The lack of DNA transfection for *Theileria* has hampered the determination of the elements involved in gene transcription. However, recent work on the apicomplexan parasites *Toxoplasma* (Soldati *et al.*, 1995) and *Plasmodium* (Crabb *et al.*, 1996), where genetic transfection is available, indicates that this obstacle will soon be overcome. In Kinetoplastida (e.g. *Leishmania* and *Trypanosoma*) for which transfection is available, promoter analysis is complicated by the fact that transcription involves the synthesis of large polycistronic precursors from which mature RNAs are produced by transplicing (Cross, 1990). Only a few promoters have been characterised in trypanosomes, specifically for genes encoding surface antigens (VSG, PARP), ribosomal sequences and for some small

RNA genes (reviewed in Vahamme and Pays, 1995). Furthermore, gene transcription is unusual in trypanosomes in that a significant number of genes are apparently transcribed by RNA polymerase I (Pol I) e.g. VSG and PARP (Pays *et al.*, 1990; Clayton *et al.*, 1990; Zomerdijsk *et al.*, 1990). This is possibly due to the ability of the 5' cap, essential for mRNA functioning, to be added post-transcriptionally by trans-splicing to the pre-mRNA. Whereas for Pol II transcribed genes the cap is generated by guanylyl transferase (Lewin, 1990(a)) In most other systems Pol I transcribes only ribosomal coding genes.

There is no evidence of polycistronic transcription or trans-splicing in *Toxoplasma* or *Theileria*, to date. This should facilitate the identification of promoters and make likely that a polymerase similar in properties to the higher eukaryotic RNA polymerase II is responsible for the transcription of protein-coding genes in these organisms. Initial attempts to identify *Theileria* promoters has been restricted to cloning and sequencing analysis upstream of the initiation site of several stage specifically expressed genes. The upstream region of three *Theileria* genes have been cloned and the RNA initiation sites of these genes has been determined either by S1 mapping, primer extension or 5' RACE. Transcription initiation sites have been mapped for the hsp70.1 gene (Mason *et al.*, 1989), and *SPAG-1* gene (Katzner, thesis 1995). The beginning of the mRNA's for these genes has been mapped to 215 bp and 278 bp respectively, 5' of the ATG start codon. As yet only the hsp70.1 promoter shows sequence homology to DNA sequence motifs. Some homology was found to the heat-shock element binding sites consensus sequence but as the motifs are not a classical heat shock element it is unknown as to whether they do carry out this function, and a putative TATA box was found in the 5' region (Mason *et al.*, 1989). Putative TATA box sequences have been detected in the 5' regions of the 117 kDa rho-try and *SPAG-1* genes but due to their positions in relation to the transcription initiation site and the ATG start codon it seems unlikely that they are involved in the regulation of these genes, in a manner similar to the TATA motif of higher eukaryotes (Katzner, thesis 1995). Thus it is possible that a classical TATA motif in a position approximately 20-30bp upstream of the transcription initiation site is not necessary for the correct positioning of the RNA polymerase in certain *Theileria* genes. "TATA-less" promoters have also been identified in the apicomplexa parasite *Toxoplasma gondii* (Soldati *et al.*, 1994). In higher eukaryotes "TATA-less" promoters have generally been found associated with housekeeping genes

(Smale and Baltimore, 1989) and multiple sites of transcription initiation are often seen (Pave-Preux *et al.*, 1990).

To date, no details of novel promoter sequences or promoter binding proteins for *Theileria* have been published. However, several promoters have been characterised from the related apicomplexan parasites *Plasmodium* and *Toxoplasma*. Five different upstream sequences are known which are functionally active in *P.falciparum*. These are the 5' regions of the PfCAM, PfDHFR-TS, and PCDHFR-TS genes described by Crabb and Cowman (1996) and the hsp86 and hrp2 genes described by Wu *et al.*, (1995). In addition polypurine motifs have been reported in the promoter regions of several protozoan parasites. Mercier *et al.*, (1996) for example drew attention to the presence of an A/TGAGACG motif in *T.gondi* which is common to the four GRA genes and present in either the forward or reverse orientation. Interestingly, this motif is found within the promoter region of other *Toxoplasma* genes. This includes the core element demonstrated to function as selecting the transcription start site of the SAG1 gene, which contains a stretch of six 27bp repeats (Soldati and Boothroyd, 1995). Other related motifs have been reported in the promoter region of the tubulin genes of the ciliate protozoan *Tetrahymena pyriformis* (Barahona *et al.*, 1988) and a repeated motif (T/AGTGTAC) that resembles those found in both *Toxoplasma* and ciliate promoters was reported in the promoter regions of the *Plasmodium* GBP130 and KAHRP genes (Lanzer *et al.*, 1992a,b; 1993). It may therefore be possible that promoter motifs are conserved among apicomplexan parasites.

A highly divergent TATA-binding protein (TBP) has been characterised in the apicomplexan parasite *Plasmodium falciparum* (McAndrew *et al.*, 1993). Thus, it is very likely that TBP exists in *Theileria* and *Toxoplasma*. If it is involved in the promoter function of the *Theileria* promoters without a correctly positioned consensus eukaryotic TATA-box, it must recognise either a noncanonical TATA box or interact with another DNA-binding protein. Several cases have been reported where, in the absence of a TATA box, regulatory factors are believed to tether TFIID (of which TBP is part) or other components of the pre-initiation complex of RNA pol II to the promoter. For example, the Sp1 binding domain directs initiation of transcription in the hamster dihydrofolate reductase promoter (Blake *et al.*, 1990). The activation proteins recognising these sites are believed to

determine the specific initiation start site through protein-protein interactions (Kollmar and Farnham, 1993). Another transcription factor identified in *Plasmodium* is a highly conserved HMG-like protein (PF16) isolated from *Plasmodium falciparum* (Guntaka *et al.*, 1992).

Although factors are beginning to be analysed, there is as yet not much data on factors which control stage specific gene expression and what mechanisms bring about a change over in factor activity. At present the identification of transcriptional signals controlling *Theileria* gene expression during differentiation is limited to structural analysis, mainly because assays such as transfection, or *in vitro* transcription are not yet available. However, an important step towards the development of these techniques has been the isolation of macroschizont infected cell lines by Shiels *et al.*, (1994) which differentiate with enhanced capabilities allowing large amounts of parasites undergoing differentiation to be harvested. This system has the potential to allow studies to analyse regulation of gene expression during differentiation at an *in vivo* level.

1.6.3 RNA Processing and Termination

The majority of higher eukaryotic genes are composed of coding sequences, exons, interrupted by one or more, intervening sequences, introns. In *Theileria* only four genes have as yet been identified which have been shown to have introns; *SPAG-1* (Katzner, Thesis York 1995) and *ThaCRK2* (Kinnaird *et al.*, 1996) genes of *T.annulata*, and *P67* (Nene *et al.*, 1992) and the cysteine protease genes (Nene *et al.*, 1990) of *T.parva*. Analysis of the flanking regions of the introns identified that the splice donor and acceptor sites are in accordance with those for other eukaryotes, and revealed the conserved GU---AG nucleotides for splice sites found in introns isolated in *Plasmodium* (Scherf *et al.*, 1988; Weber, 1988).

Theileria mRNAs can be separated from ribosomal RNA species by their affinity to bind to oligo (dT) indicating that like most eukaryotic mRNAs *Theileria* mRNAs are polyadenylated at their 3' end. This has been confirmed by the generation of a *T.annulata* cDNA library (Shiels *et al.*, 1994) and the isolation from this library of several genes including the 30kDa major merozoite surface antigen and the 117 kDa rhoptry protein

(Shiels *et al.*, 1994). No detailed mapping of the precise 3' ends of *Theileria* genes has yet been performed. Sequence analysis of the hsp 70.1 sequence from *T.annulata* highlighted two possible motifs which have perfect homology to the AATAAA or ATTAAA consensus polyadenylation signals found in eukaryotic genes (Mason *et al.*, 1989).

Structural analysis, and mapping of the mRNA initiation site has defined regions where cis acting regulatory motifs for transcription possibly exist both for initiation and termination of *Theileria* genes. In the absence of assays, such as transfection, gel mobility shifts and *in vitro* transcription, it is impossible to address mechanisms of gene regulation at a functional level and define the structural elements responsible for the regulation of gene expression. An understanding of these processes will not only improve our knowledge about stage specific gene regulation during the parasite's Life-Cycle but, may also provide new targets for treating parasitic disease.

1.6.4 Post-Transcriptional Gene Regulation

Evidence for post-transcriptional control of gene expression in protozoa is becoming more and more evident, and is believed to account for the majority of the gene expression regulation in *Trypanosoma* procyclin units and a number of *Leishmania* genes (Ziberstein and Shapira, 1994). The 3'-untranslated region (UTR) of *Trypanosoma* procyclin mRNA contains elements which have been shown to play a role in post-transcriptional regulation which is conserved among different 3'-UTR's of procyclin mRNAs, as well as in the 3' UTR of an unrelated mRNA for a major surface antigen of the procyclic form of *T.congolense* (Bayne *et al.*, 1993). The procyclin 3' UTR might serve as a target for mRNA degradation in bloodstream forms (Jefferies *et al.*, 1991), however, the involvement of this element in the regulation of procyclin expression remains to be demonstrated. Evidence for post-transcriptional regulation in *Leishmania* was obtained from transfection experiments where the downstream intergenic regions of hsp 83 were replaced with increased stability of the mRNA normally observed in response to elevated temperature. This exchange eliminated the temperature-induced differential stability of the corresponding mRNA (Ziberstein and Shapira, 1994). Additionally in *Plasmodium*, post-transcriptional control has been observed in the stage specific expression of different ribosomal sub-units (Gunderson, *et al.*, 1987;

Waters *et al.*, 1989) and the gene encodes surface protein Pbs21 in *Plasmodium* (Paton *et al.*, 1993).

In general; gene expression, appears to be controlled via conventional mechanisms in *Theileria* parasites, but this does not mean that regulation will be limited to transcriptional control. Evidence for post-transcriptional control does indeed exist for the regulation of Hsp 90 gene in *T.parva* (Gerhards *et al.*, 1994) and the *Tacyp* gene in *T.annulata* (David Swan, personal communication). The hsp 90 gene is believed to be post-transcriptionally regulated in the piroplasm stage of the parasite (Gerhards *et al.*, 1994). Transcriptional and post-transcriptional control of the hsp family has been described for a number of other organisms, in particular for *Drosophila* (Morimoto *et al.*, 1990).

1.7 Objectives of this work

The precise cellular and molecular events which trigger differentiation of protozoan parasites have not yet been fully elucidated. Shiels *et al.*, (1994) have proposed a stoichiometric mechanism for *in vitro* differentiation of the macroschizont to merozoite stage in *T.annulata*. The underlying basis of this model is that a concentration threshold of regulatory factors must be reached before the parasite is committed to differentiate. Below this threshold differentiation is reversible as described in section 1.5. The *T.annulata* 30kDa merozoite surface polypeptide (*TamSI*) is initially expressed during the reversible phase of *in vitro* differentiation from macroschizont to merozoite stage and low level expression of this molecule can be detected in the preceding stage before differentiation (Shiels *et al.*, 1994). Low level expression of the *TamSI* molecule in the preceding stages is believed to be due either to a functional overlap of regulatory factors or due to a low level (or activity) of regulatory factors for merozoite stage specific expression (Shiels *et al.*, 1994). In addition, the expression of the 30kDa molecule is likely to be controlled by regulatory factors defining the early stages of the temporal order of molecular changes that occur during *in vitro* differentiation, and following commitment expression levels are very high. The 30kDa merozoite surface antigen is, therefore, a good candidate gene with which to examine the process of gene expression within the parasite and to characterise regulatory proteins which may be involved in the temporal control of differentiation. One of the primary objectives of this thesis was to isolate and characterise the regulatory elements controlling the expression of *TamSI*. To do this it was necessary to determine the level at which the expression of *TamSI* was controlled and to map the mRNA initiation site. In view of this limited ability to study promoter function and the lack of a DNA transfection system for *Theileria*, a further aim of this work was to develop a gel mobility shift assay suitable for studying nucleic acid-protein interactions for putative *TamSI* gene regulatory regions. Identification and characterisation of DNA-protein interactions with the *TamSI* 5' untranslated region is hoped to correspond to functional promoter elements. Additionally, this assay would allow the interactions of transcription factors with *TamSI* to be compared between diminished and enhanced cell line during differentiation from macroschizont to merozoite.

Stage differentiation is a fundamental event of protozoan biology allowing expansion and transmission of parasite populations. As changes to gene expression are central to stage differentiation it is possible to view these events as central to the process. Understanding the mechanisms which control gene expression changes will ultimately lead to an understanding of how differentiation is initiated. This study aimed to initiate work to investigate how changes to the control of *TamS1* gene expression is altered during differentiation from macroschizont to merozoite. The observation that the timing of differentiation can be altered extrinsically and correlates with changes in levels of *TamS1* gene expression. Therefore by identifying the molecular mechanisms of *TamS1* gene expression this could lead to the identification of a factor involved in the timing of differentiation events. Modulation of this factor could be relevant to control strategies based on the extrinsic modulation of apicomplexan differentiation.

1.8 Summary of work presented

Chapter 2: describes the cloning and sequencing of a ~8kb genomic fragment containing the *TamS1* sequence and investigations into the organisation of the open reading frames flanking the coding region of the *TamS1* gene.

Chapter 3: describes the characterisation of transcriptional control of the *TamS1* gene, mapping of the mRNA initiation site, and investigations into sequence comparisons to determine the promoter elements involved in controlling the expression of the *TamS1* gene.

Chapter 4: describes the strategies used to determine protein binding to the 5' non-transcribed region of the *TamS1* gene and the attempts made to determine the size of each polypeptide binding to the regulatory element.

Chapter 5: describes the characterisation of differences between differentiation competent and attenuated cell lines. These experiments involved investigating alterations in gene expression of a number of target genes, including *TamS1* and the binding of factors to the *TamS1* promoter element.

Chapter 6: is a general discussion of the findings of this thesis and their relevance to apicomplexan differentiation

2 The Cloning and Sequencing of the Regulatory Regions Controlling the Expression of the *TamS1* Gene

2.1 Introduction

Differential gene expression between Life-Cycle stages of protozoan parasites has been well documented but it is only recently that studies have begun to investigate the mechanisms which control this process. Stage-regulated genes encode molecules that are involved in biochemical pathways and structural functions as well as those which are involved in the evasion of the host defence mechanism (Parsons, 1990). The level at which gene expression can be controlled is at transcriptional and post-transcriptional levels. Transcriptional regulation can either be constitutive, suppressible or inducible; while post-transcriptional regulation can be controlled by mRNA stability, mRNA splicing or by regulating translation. Transcriptional regulation is predicted to be the most economical mode of regulation (Latchman, 1990) and the expression of many genes have been found to be controlled by this process.

Little is known about the process of gene regulation in *Theileria*. The first indication of controlled gene expression was demonstrated by Shiels *et al.*, (1986) by raising monoclonal antibodies against surface molecules of different parasite stages for vaccine development. Using a series of monoclonal antibodies, raised against *T.annulata* macroschizonts and piroplasms, Shiels *et al.*, (1990) were able to identify several stage specific antigens by indirect immunofluorescence assay (IFA). Three of the monoclonal antibodies raised against the piroplasm stage (5E1, 2D5 and 1E11) were positive against merozoites, but showed no reactivity when tested against macroschizonts. Monoclonal antibody 5E1 reacted intensely against the periphery of the parasite, and using protein A-gold to enable the monoclonal antibody to be visible by electron microscopy, 5E1 was demonstrated to bind to the outer surface of the merozoite suggesting that it may be a target for a protective immune response. 5E1 detected a 30kDa molecule in extracts derived from merozoite preparations of heat induced cultures, which has since been shown to be the most abundant and immunodominant antigen on the surface of merozoites and piroplasms of *T.annulata* (Dickson and Shiels, 1993).

Molecules with similar characteristics to the 30kDa polypeptide have been identified in other species of *Theileria*. In the *T.sergenti/buffeli/orientalis* group of parasites, stocks were characterised by the possession of a 33kDa or a 34kDa polypeptide (Sugimoto *et al.*, 1991; Kawazu *et al.*, 1992; Tanaka *et al.*, 1990). Experiments with a monoclonal antibody raised against the *T.sergenti* 32kDa merozoite surface molecule indicated that a humoral response may inhibit merozoite invasion of erythrocytes (Tanaka *et al.*, 1990). An immunodominant molecule of 32 kDa has also been described for the erythrocyte stages of *Theileria mutans*, and this antigen has been used in the development of a diagnostic test (Katende *et al.*, 1990). Thus, the 30kDa - 34 kDa polypeptide family is probably distributed throughout all the bovine species of *Theileria*, with antigenic diversity being present both between and within the different species.

Isolates and stocks of *T.annulata* usually consist of more than one parasite genotype (Shiels *et al.*, 1992; Ben Miled *et al.*, 1994). Shiels *et al.*, (1993), demonstrated that the major merozoite surface polypeptide is antigenically diverse and can vary in size. Genotypically distinct parasites were shown to possess either a 30 or a 32 kDa form of the antigen. The two molecules were shown to be closely related by peptide mapping and periodate oxidation experiments indicated that they are glycosylated (Shiels *et al.*, 1993). In order to characterise and compare the genes encoding the 30 (*TamSI-1*) and the 32 kDa (*TamSI-2*) merozoite antigens in more detail the nucleotide sequences representing the polypeptide coding region were elucidated. Both genes contained an open reading frame which was sufficient to encode a polypeptide of 281 amino acids with a predicted molecular mass of 32 kDa. Analysis of the predicted amino acid sequences highlighted a hydrophobic stretch at the C terminus of the molecule which probably functions to anchor the polypeptide in the merozoite membrane. Also, the erythrocyte-binding sequences Lys-Glu-Leu (KEL) and Lys-Glu (KE) described for *Plasmodium* (Molano *et al.*, 1992) were found at a number of positions within the *TamSI-1* and *TamSI-2* sequences. Comparisons with the published amino acid sequences of the major merozoite polypeptide predicted for *T.sergenti* and *T.buffeli*, and the partial amino-acid sequence predicted for the *T.parva* gene, showed that all five polypeptides are likely to be structurally related (Shiels *et al.*, 1995).

Studies aimed at defining the regulatory control of the 30kDa molecule at the nucleic acid level confirmed that the polypeptide is highly expressed in the merozoite and piroplasm stages of the parasite. Observations were made that low level expression of the merozoite gene, *TamS1* could be detected in the preceding macroschizont life-cycle phase (Shiels *et al.*, 1994) which was attributed to either merozoite regulatory factors that operate at a low level in the macroschizont or alternatively due to a functional overlap of factors that regulate gene expression in the different stages. Investigations in to whether low level expression is found in all stages of the parasite was not determined due to an inability to study the stages within the tick. However, basal expression of all parasite genes throughout the Life-Cycle does not appear to occur, as *SPAG-1* gave no signal against macroschizont or piroplasm RNA (Williamson *et al.*, 1989 Shiels *et al.*, 1994).

Expression of the 30kDa merozoite surface polypeptide was assessed during the differentiation from macroschizont to merozoite *in vitro*. Levels of *TamS1* RNA increased dramatically between days 4 and 6 of differentiation coinciding with the loss of macroschizont antigens (Shiels *et al.*, 1994). Thus, there is a temporal link between the events which positively and negatively control stage specific gene expression. In addition pulse experiments showed that differentiation and the expression of *TamS1* was reversible during the first 2 days of induced differentiation. If the inducement to differentiate was removed during this period the merozoite genes were down regulated and the parasite returned to the macroschizont cellular state. It was also found that following the initial reversible phase the parasite becomes committed to differentiation and *TamS1* was expressed at very high levels. Therefore, the expression of the *TamS1* appears to be linked to the temporal order of differentiation within the macroschizont and alteration to its control is one of the earliest molecular alterations associated with differentiation. This would imply that *TamS1* is controlled by a regulatory factor(s) which may also be involved in the commitment step of differentiation. Therefore understanding the mechanisms which differentially control *TamS1* during merozoite production could shed light on the molecular events which initiate and regulate the transformation of one parasite stage to the next

Aims

Given the high level of mRNA production and the association between the differential control of *TamS1* expression with the temporal ordering of the major phases of the differentiation process, it would appear that *TamS1* is a good candidate for studying the mechanisms which control gene expression during stage differentiation of *T.annulata*. The aim of the work presented in this chapter was to initiate these studies by isolating and sequencing regions of genomic DNA which were likely to contain the regulatory elements controlling the expression of *TamS1*. In addition, as little is known about the genomic arrangement of genes which are stage regulated in *T.annulata*, the second aim of this chapter was to determine the position and expression of any open reading frames flanking the *TamS1* protein coding sequence.

2.2 Materials and Methods

2.2.1 Phage Stocks and Liquid Lysates

A genomic DNA library was constructed (by Dr Kinnaird, WUMP, University of Glasgow) in the vector λ DASH using DNA isolated from merozoites that had been partially purified by differential centrifugation from a 500ml culture of the enhanced cloned cell line D7, incubated at 41°C for 7 days. Dr Kinnaird screened the genomic library and isolated 3 λ DASH clones (using a *TamSI* cDNA probe representing the 30kDa merozoite surface antigen). These clones were the starting point for the work described in this chapter. *Eschericia coli* strain XL-1Blue MRF (Stratagene) was the host bacterium used to plate out one of the λ DASH clones. The bacteria were streaked out onto 2XYT plates (1.6% bactotryptone, 1% yeast extract, 5% sodium chloride, 1.5% bactoagar) containing 100 μ g/ml⁻¹ tetracycline and incubated overnight at 37°C and the following day, an overnight culture was prepared by picking a single colony from this plate into 5ml of 2XYT medium (as 2XYT plates minus bactoagar) containing 0.2% maltose and 10mM magnesium sulphate and incubated overnight at 37°C in an orbital shaker. 50ml of 2XYT medium was inoculated with 0.5ml of overnight culture, and incubated in an orbital shaker at 37°C until an absorbance of 0.5 OD at 600nm was obtained (corresponding to a cell density of 2.5×10^8 cells ml⁻¹). The cells were harvested by centrifugation at 800g for 5 minutes at 4°C, and resuspended in 5ml of 10mM magnesium sulphate. Plating bacteria were stored at 4°C and were viable for up to 2 days.

The titre of the λ DASH clone, was determined by serial 10 fold dilutions in SM buffer (50mM Tris.HCl (pH7.5), 10mM magnesium sulphate, 100mM sodium chloride, 0.01% w/v gelatine). 100 μ l aliquots of each dilution were incubated with 100 μ l of XL-1Blue MRF cells in 10mM magnesium sulphate by incubation at 37°C for 15 minutes. 3ml of top agarose BBL (1% Bacto-trypticase, 0.5% sodium chloride, 0.7% low EEO agarose (Sigma)) at approximately 50°C was added to the cells and the samples poured on to a BBL plates (1% trypticase, 0.5% sodium chloride, 1.5% bactoagar). Once set the plates were inverted and incubated at 37°C overnight. The plaques were counted and the titre determined from appropriate dilutions.

Phage stocks were prepared using a dilution of bacteriophage, estimated from the titrations described above, that would almost generate confluent lysis of the plated bacterial lawn. 100µl aliquots of this dilution was mixed with 100µl of plating bacteria, incubated at 37°C for 15 minutes and plated out as previously described. Following incubation overnight at 37°C, or until lysis was confluent, 3-4ml of SM buffer was added and the plates were placed at room temperature on a shaker for 5 hours. With a sterile pasteur pipette, as much as possible of the SM buffer was harvested. The top layer of agar was then scraped off and added to the pooled phage and SM buffer. 0.1 ml of chloroform was added and following an overnight incubation at 4°C the top agarose SM buffer suspension was centrifuged at 4000g for 10 minutes at 4°C. High titre phage supernatant was then removed and stored at 4°C, after the addition of chloroform to individual aliquots.

To prepare phage liquid lysates XL-1blue MRF cells, first grown overnight in LB media (1% bactotryptone, 0.5% yeast extract, 1% sodium chloride) supplemented with 0.2% maltose and 10mM magnesium sulphate defined as above were inoculated into 100ml of LB media to a turbidity of 0.1 OD at 600nm (OD_{600nm}). The culture was then incubated at 37°C shaking vigorously on a orbital shaker until the OD_{600nm} reached 0.45-0.6, 1ml of 1M magnesium sulphate was then added to each flask plus $2-3 \times 10^{10}$ plaque forming units (pfu) of the λ DASH clone and the culture was incubated for a further 6 - 8 hours. During this period the OD_{600nm} was followed and was seen to rise to approximately 1.2, before dropping to 0.4-0.6. 200µl of $CHCl_3$ was then added, the flask shaken for a further 15 minutes and stored at 4°C.

Phage DNA was isolated and purified using the Promega Magic Lambda Prep kit (Promega Corporation) according to the manufacturer's protocol. 10ml aliquots of liquid lysate culture were mixed with 40µl of nuclease mixture (0.25mg/ml RNase A, 0.25mg/ml DNase I, 150mM NaCl, 50% glycerol), followed by incubation at 37°C for 30 minutes. On completion of the reaction, 4 ml of phage precipitant (33% polyethylene glycol (PEG-8000), 3.3M NaCl) was mixed gently with the lysate and placed on ice for 30 minutes. Samples were centrifuged at 10,000g for 10 minutes and the supernatant removed. Pelleted phage was resuspended in 500µl of phage buffer (150mM NaCl, 40mM Tris.HCl (pH7.4), 10mM $MgSO_4$) and transferred into fresh microfuge tubes, followed by the addition of 1 ml of

purification resin (Promega) which was mixed with the phage suspension by inverting. The resin/lysate mix was pushed through a Wizard Minicolumn (Promega) using a syringe attached to the luer-lock extension of each Minicolumn, and the column was washed by gently pushing through 2ml of 80% isopropanol. A spin at 12,000g for 20 seconds dried the resin, and the DNA was eluted with 100µl of dH₂O preheated to 80°C, the eluate being collected by centrifugation at 12,000g for 20 seconds.

2.2.2 Restriction digests of DNA and agarose gel electrophoresis

All restriction endonuclease digests were performed according to the procedures recommended by the manufacturer (Gibco-BRL) using the buffers supplied and the temperature for optimal activity of the enzyme, 37°C. In general, 0.2 to 1µg of DNA was made up to a total volume of 20µl with dH₂O, mixed with 2µl of the appropriate 10x REact buffer (Gibco-BRL) and 1 to 2 units of restriction enzyme (Gibco-BRL). Following completion of digestion the samples were incubated at 65°C for 5 minutes, chilled on ice and centrifuged for 5 minutes at 12,000g. Plasmid DNA was digested for an average of two hours, while genomic DNA was left to digest overnight.

Restriction digested DNA samples were separated by agarose gel electrophoresis as described by Sambrook *et al.* (1989). 0.7 to 1% agarose gels were prepared by dissolving agarose in an appropriate volume of TAE buffer (40mM Tris-acetate, 1mM EDTA) with 0.5µg/ml ethidium bromide. When DNA fragments were to be excised from an agarose gel, low melting point agarose (Sigma) was used and the gels were cast at 4°C. DNA samples were diluted in 6 x agarose gel loading buffer (0.05% bromophenol blue, 100mM EDTA, pH 7.5, 22% Ficoll) and electrophoresed at 100V until it was possible to estimate the size of the DNA fragments of interest, or at 20V overnight in TAE buffer. The size of DNA fragments were estimated by comparison to the 1kb maker ladder (Gibco-BRL), which ranges from 12kb to 75bp. DNA fragments were then visualised using a UV transilluminator (366nm).

2.2.3 Mapping of λ DASH clones by restriction digestion and Southern blotting

The method of capillary transfer of DNA from agarose gel to nylon membrane (Amersham, Hybond N) was adapted from the standard Southern blotting method

(Sambrook *et al.*, 1989, Southern 1975). DNA was purified from the λ DASH clone (see section 2.2.1), digested with a variety of restriction enzymes and run on a 0.7% agarose gel (see section 2.2.2). After electrophoresis the agarose gels were washed twice in denaturing solution (1.5M NaCl, 0.5M NaOH) for 15 minutes per wash, then once in neutralising solution (2M NaCl, 0.5M Tris-HCl, pH 7.4) for 20 minutes. The gel was then equilibrated in phosphate transfer buffer (50mM phosphate buffer, 25mM NaH_2PO_4 pH 5.5 with 25mM Na_2HPO_4) for 30 minutes, before being placed on a platform of 3MM paper which had its ends immersed in phosphate transfer buffer. Hybond N membrane cut to the same size as the gel was laid on top followed by 3 sheets of 3MM paper, a stack of paper towels and a light weight. DNA was transferred overnight and fixed to the nylon membrane by exposure to 150 Joules short wave Ultra-violet radiation using a GS Gene linker (Biorad).

Following UV cross-linking the membranes were prehybridised in 10ml of Church and Gilbert hybridisation solution (0.5M Na_2HPO_4 , 3.4% Orthophosphoric acid (pH 7.2) 1.75% SDS, 1mM EDTA, Church & Gilbert, 1984) at 65°C for 30 minutes, using the Hybaid cylinder and hybridisation oven system (Hybaid). Probes were prepared using the Random Priming DNA Labelling Kit (Boehringer-Mannheim Pharmaceuticals) according to the manufacturer's instructions: selected restriction digested DNA fragments were excised from low melting point agarose gels, melted in 1.5 ml of dH_2O per gram of agarose slice and boiled for 5 minutes. Approximately 25ng of denatured DNA was randomly labelled during the synthesis of a complementary DNA strand using 50 μCi $\alpha^{32}\text{P}$ dCTP, (3000 ci/mmol), 3 μl of 0.166mM, dATP, dTTP, dGTP mixture, 2 μl of 10x hexanucleotide mixture (Boehringer-Mannheim Pharmaceuticals) and 1 μl (2 units) of Klenow enzyme. The reaction was incubated at 37°C for 30 minutes, before it was boiled for 3 minutes to denature the double stranded probe and added to the 10ml of Church and Gilbert hybridisation solution. Hybridisation was allowed to proceed overnight at 65°C. The nylon membranes were then washed three times with 1x SSC (0.15M NaCl, 0.15M trisodium citrate, pH 7.0), 5% SDS at 65°C, before being wrapped in Saran wrap and exposed to Kodak X-OMAT film overnight in a X-ray film cassette at -70°C.

2.2.4 Cloning and transforming λ DASH genomic clones

DNA fragments identified in 2.2.3 as containing genomic *Theileria annulata* sequence flanking the *TamSI* open reading frame were inserted into the pGem (Promega) and pBluescript (Stratagene) vectors to facilitate further sub-cloning and sequencing. DNA was first digested with the appropriate enzymes, see section 2.2.2. and restriction digests subjected to electrophoresis through a 0.7% low melting point TAE agarose gel. Bands of interest were excised and purified either by salt phenol extraction or using the QIAquick gel extraction kit, according to manufacturers instructions (Qiagen).

DNA fragments purified by the salt phenol extraction method were initially electrophoresed on low melting point TAE agarose gels. The required fragment was identified by UV illumination, excised from the gel and the gel slice was weighed. The agarose was melted at 65°C for 15 minutes before adding a volume of 5M NaCl which was equivalent to 1/9th of the weight of the agarose slice (i.e. 100 μ l/0.9g). The DNA was extracted by the addition of an equal volume of phenol saturated with 0.5M NaCl and vigorous shaking for 10 minutes. The aqueous and phenol phases were then separated by centrifugation at 13,000g for 10 minutes, before the aqueous phase was transferred to a fresh eppendorf and before further extraction with an equal volume of chloroform isoamyl alcohol. After shaking for 1 minute, the samples were centrifuged at 13,000g for 1 minute and the aqueous phase removed and transferred to a fresh eppendorf. 1 μ l of glycogen was then added and the DNA fragment precipitated by the addition of 2 volumes of ethanol followed by incubation at -20°C for 30 minutes. DNA was pelleted by centrifugation at 13,000g for 10 minutes and washed with 75% ethanol, before drying for 5 minutes under vacuum. DNA pellets were resuspended in sterile dH₂O and stored at -20°C.

DNA fragments were also purified from agarose gels by the QIAquick Gel Extraction Kit (Qiagen) on the basis of ion exchange chromatography. The band was excised from an agarose gel, 3 volumes of Buffer QX1 (Qiagen) was added and the agarose was dissolved at 50°C for 10 minutes. The sample was placed in a QIAquick spin column, centrifuged at 12,000g for 1 minute before disposal of the drain through fraction from the collection tube. To wash the extracted DNA fragment, 0.75ml of PE buffer (Qiagen) was added, and the column centrifuged for a further minute at 13,000g. An extra spin at 13,000g

removed residual buffer, to prevent the presence of ethanol from interfering with subsequent reactions. DNA was eluted with 50µl of dH₂O, the eluate being collected by centrifugation.

Vector DNA was prepared for cloning by digestion with an appropriate restriction enzyme(s) in order to generate cohesive ends for ligation (see 2.2.2). To prevent vector-vector ligations, vector DNA cut with a single restriction enzyme was dephosphorylated with calf intestinal alkaline phosphatase (CIAP). 20µg of digested vector was extracted once with Tris saturated phenol (pH 7.8) and once with Tris saturated phenol:chloroform (1:1), ethanol precipitated (as previously described above) and resuspended in 22.5µl of dH₂O. 2.5 µl of 10x CIAP buffer (500 mM Tris.HCl, pH 8.5, 1mM EDTA) and 30 units of CIAP (Gibco-BRL) was then added, and the mixture incubated at 37°C for 60 minutes. To inhibit the CIAP, 2.5µl 500mM EDTA was added to the mixture followed by heating to 65°C for 45 minutes, phenol extraction and ethanol precipitation.

Ligation reactions consisted of an insert:vector ratio of 3:1, dissolved in a final volume of 20µl. DNA was made up to a total volume of 17µl with dH₂O and 2µl of 10 x T4 DNA ligation buffer (500mM Tris. HCl, pH 7.8, 100mM dithiothreitol, 1mM ATP, 25µg/ml⁻¹ bovine serum albumin) and 1-2 units of T4 DNA ligase (Gibco-BRL) was added. The ligation mixture was then incubated overnight at 13°C.

Competent cells were prepared from a 1ml aliquot of an overnight culture of *E.coli* XL-1Blue cells mixed with 50ml of 2XYT broth and incubated at 37°C with shaking, until a OD_{600nm} of 0.3 was reached (see 2.2.1). The cells were harvested by centrifugation at 4°C for 10 minutes at 4,200g. The cell pellet was resuspended in 25ml of ice cold 50mM CaCl₂, placed on ice for 45 minutes, harvested as before and resuspended in 5ml of ice cold 50mM CaCl₂. Cells were transformed into 200µl of competent cells by adding 1 to 20µl of DNA (of approx. 1mg/ml) and incubating on ice for 45 minutes. The cells were heat shocked at 42°C for 2 minutes and returned to ice for 5 minutes. 400µl of 2XYT broth was then added to the cells which were incubated at 37°C for a further 45 minutes to allow the expression of antibiotic resistance markers encoded by the plasmid. The cells were spread out onto 2XYT plates containing the appropriate selective antibiotic, and 50µl of X-gal (20mg ml⁻¹ 5-

bromo-4 chloro 3 indoly β -D galactosidase, in dimethylformamide, Gibco-BRL) and 2 μ l of IPTG (10mg ml⁻¹ isopropylthio- β -D galactosidase, Gibco-BRL) which had previously been spread onto the plate surface and allowed to dry. Inoculated plates were incubated overnight at 37°C and the next day white putative recombinant colonies were picked and streaked onto fresh 2XYT plates containing the appropriate selective antibiotic. Plasmid DNA was then made from overnight cultures of selected colonies by picking one colony into 5ml of 2XYT media containing the appropriate selective antibiotic and incubating overnight on an orbital shaker at 37°C.

2.2.5 Preparation of plasmid DNA

(A) Alkaline Lysis Minipreparation Method

1.5ml aliquots of overnight recombinant bacteria culture were pelleted by centrifugation at 13,000g for 5 minutes. The supernatant was removed and the pelleted cells were resuspended in 100 μ l of resuspension solution (50mM Tris.HCl, pH 7.5, 10mM EDTA, 100 mg ml⁻¹ RNase A). The cells were lysed by the addition of 200 μ l of freshly prepared lysis buffer (200mM NaOH, 1% SDS) and then 150 μ l of neutralisation solution (2.55M KAc, pH 4.8) was added. The cells were vortexed and, centrifuged at 13,000g for 5 minutes, before the supernatant was transferred to a new eppendorf. DNA was extracted by an equal volume of phenol/chloroform see 2.2.4, the aqueous layer was recovered after centrifugation, precipitated with two volumes of ethanol. The DNA was then pelleted by centrifugation, washed in 70% ethanol, centrifuged, vacuum dried, and resuspended in 30 μ l of sterile dH₂O and stored at -20°C.

(B) Magic Minipreps purification kit (Promega)

Alternatively, plasmid purification was carried out using the magic miniprep purification kit (Promega). According to the protocol given by the manufacturer, 1 to 3 ml of cells from an overnight bacterial cell culture were pelleted by centrifugation at 13,000g for 5 minutes. The cell pellet was taken up in 200 μ l of cell resuspension solution, lysed by the addition of 200 μ l of lysis solution and neutralised by the addition of 200 μ l of neutralisation solution. After inverting the eppendorf several times, the samples were centrifuged at 13000g for 5 minutes and the clear supernatant was decanted into a new microfuge tube. 1ml of wizard miniprep DNA purification resin, a silica based resin, was

mixed with the supernatant by inverting and the samples were pushed through a Wizard Minicolumn (Promega) using a syringe attached to the luer-lock extension of each Minicolumn. The DNA was washed by gently pushing 2ml of wash solution (200mM NaCl, 20mM Tris.HCl, pH 7.5, 5mM EDTA, diluted with 95% ethanol to a final ethanol concentration of 55%) through the Minicolumn. A spin at 13,000g dried the resin and the DNA was eluted with 50µl of dH₂O, the eluate being collected by centrifugation in a new microfuge tube.

2.2.6 Generation of Exo III deletions via Erase-a-Base System (Promega)

The Erase-a-Base System (Promega), was used, following the manufacturer's instructions, to generate unidirectional deletions of DNA, inserted within a plasmid vector. 5µg of closed circular DNA was linearized with two different restriction enzymes within the multiple cloning site, fanning one side of the insert. The direction with which the deletions were to proceed was determined by the restriction enzymes sites selected. The first enzyme digest generated a 3' overhanging end close to the binding site of the oligonucleotide primer used for sequencing, resistant to Exonuclease III (Exo III) digestion, while the second digest generated a blunt end or a 5' overhang proximal to the insert, susceptible to Exonuclease III digestion (see 2.2.2). Restriction digested DNA was run on a 0.7% agarose gel (see 2.2.2) and the band of interest was excised and purified using the salt phenol extraction method, see 2.2.4. Pelleted DNA was dissolved in 1x Exo III buffer (66mM Tris.HCl, 0.66mM MgCl₂) and warmed at 37°C. 450 units of Exo III were mixed with the DNA and at 30 second intervals, 2.5µl samples were removed into eppendorfs placed on ice, which contained 7.5µl of S1 nuclease mix (40mM potassium acetate, pH 4.6, 338mM NaCl, 1.35mM ZnSO₄, 6.8% glycerol, 60 units of S1 nuclease). The tubes were then incubated at room temperature for 30 minutes, before adding 1µl of S1 stop buffer (0.3M Tris base, 0.05M EDTA) to each sample and heating the tubes to 70°C for 10 minutes. 2µl of each time point sample was removed, diluted 1:1 in agarose sample buffer and analysed (by agarose gel electrophoresis (see 2.2.2)) to determine the extent of the Exo III digestion. The remainder of the samples were transferred to 37°C and 1µl of klenow mix (20mM Tris.HCl, pH 8.0, 100mM MgCl₂, 4 units of Klenow DNA polymerase) was added. Incubation of the samples was carried out for 3 minutes at 37°C before adding 1µl of dNTP mix, and incubating for a further 5 minutes at 37°C. Ligation was carried out in a total volume of

15µl in 1x T4 ligase buffer and (1 unit) of T4 Ligase at 13°C overnight (see 2.2.4). Half of the total volume of the ligation products were transformed in to XL-1blue cells (see 2.2.4), plated out onto 2XYT plates containing the appropriate selective antibiotic and incubated overnight at 37°C. Recombinant colonies were randomly picked the following day, placed in 5ml of 2XYT media and grown overnight on an orbital shaker. Plasmid DNA was prepared from each overnight using the alkaline lysis minipreparation method (see 2.2.5) and analysed by restriction enzyme digestion and agarose electrophoresis. Stocks of bacteria containing the deleted plasmid/insert DNA were prepared by pelleting cell which were resuspended in 1ml of 2XYT medium containing 10% glycerol and stored at -70°C. DNA was prepared from samples selected for sequencing using the Magic Minipreps purification kit (Promega),

2.2.7 DNA Sequencing

(A) Manual DNA Polymerase Sequencing of Double Stranded DNA

Sequencing was carried out using the dideoxy chain termination method of Sanger and co-workers (1977). A Sequenase version 2.0 DNA sequencing kit (United States Biochemical) was used for all manual sequencing reactions. Double stranded DNA was prepared using the Promega Wizard Miniprep method and denatured in 200mM NaOH (freshly made) for 5 minutes. 0.4 volume of 5M ammonium acetate pH 7.5 was added, immediately followed by 4 total volumes of ethanol and incubated at -20°C for 2 hours. The DNA was pellet by centrifugation at 13000g, washed in 100% ethanol, dried and resuspended in 7µl of dH₂O to give a concentration of between 300-500ng ml⁻¹, and 2µl of sequencing buffer (200mM Tris.HCl, pH7.5, 100mM MgCl₂, 250mM NaCl). The T7, T3 or SP6 sequencing primers were annealed to the DNA at 65°C for 2 minutes and allowed to cool slowly. Sequence reactions were generated by combining the DNA/primer mix with 2µl of 1x labelling mix (1.5µM dGTP, 1.5µM dCTP, 1.5µM dTTP), diluted sequenase enzyme version 2.0 (8.75mM Tris.HCl, 4.375 mM DTT, 0.4275 mg/ml BSA, sequenase version 2.0 enzyme), 1µl of 0.1M DTT and 0.5µl of ³⁵S dATP, (1000 Ci/mmol) and incubating the reaction at room temperature for 5 minutes. 2.5µl of dideoxy termination mix (80µM dGTP, 80µM dATP, 80µM dCTP, 80µM dTTP, 50mM NaCl, 8µM ddGTP or ddATP or ddTTP, or ddCTP) was then added into each of 4 reaction mixtures together with 3.5µl of the template/primer/enzyme mix. The reactions mix were then incubated for a

further 5 minutes at 37°C, before termination with the addition of 4µl of stop solution (95% formamide, 20mM EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol FF). The samples were denatured at 80°C for 2 minutes prior to loading 2.5µl of each reaction onto 6% acrylamide, 0.5xTBE non-gradient gels. The 6% acrylamide gels were prepared by dissolving 210g of Urea in 72.5ml of a 40% acrylamide stock solution (Scotlab), 50ml of 10x TBE (0.9M Tris base, 0.9M boric acid, 20mM EDTA and the final volume made up to 500ml with dH₂O. Before polymerisation was initiated, 75ml of acrylamide mix was filtered and degassed, followed by the addition of 75µl tetra-methyl-1,2-diaminoethane (TEMED) and 150µl freshly made 25% ammonium persulphate (APS). The gel was cast and left to set for approximately 60 minutes, followed by the assembly of the sequencing apparatus. The upper and lower tanks were filled with 0.5xTBE buffer and the gel was pre-run at 50-55 watts to preheat the gel and sequencing plates. After the gel was loaded, electrophoresis was carried out for the desired time at 50-55 watts, approximately 1500V. After completion of electrophoresis the gels were fixed in 20% ethanol, 20% acetic acid for 15 minutes dried onto 3MM paper at 80°C for 2 hours using a Biorad gel drier. Finally the gels were exposed to Kodak X-OMAT S film in a x-ray cassette at room temperature for 24 hours.

(B) Automated DNA Polymerase Sequencing of Double Stranded DNA

The LiCor sequencer (LiCor Corp) was used for all automated sequencing reactions. All samples were prepared using the Fluoro Sequenase kit (CamBio) which uses priming oligonucleotides covalently coupled to an infrared fluorophore (laser dye IRD41; commercial primers used included T3, T7 and Sp6). The Licor automated sequencer protocol uses a PCR amplification step of the template DNA. As the DNA migrates through the sequencing gels the sequence image is collected by a microscope/detector, following excitation of labelled fragments by a laser diode emitting at 785nm. Once the gel has finished the sequences were manually edited for any necessary corrections.

For each sample, 500ng of template DNA was combined with 2.0pmol of IRD41 labelled primer, 2.5µl of 10x sequencing buffer (0.5M Tris.HCl, pH 9.3, 25mMMgCl) 1µl of BioPro thermostable DNA polymerase (Bioline) and dH₂O to a total of 17µl. In 4 labelled thermocycler tubes, 2.0µl of the appropriate SequiTherm™ Long-Read Termination mix (50mM NaCl; 180 µM 7-deaza-dGTP or 180 µM dATP or 180 µM dCTP or 180 µM

dTTP) was combined with 4 µl of the template/primer/enzyme/mix. 20µl of mineral oil was placed on top of each reaction mixture and the tubes were placed in a thermocycler programmed with the following cycles: initial 95°C for 5 minutes; followed by 95°C for 30 seconds (denaturing step), 60°C for 30 seconds (annealing step) and 70°C for 1 minute (elongation step) for a total of 25 cycles. After the program was complete, 4µl of stop solution was added to each reaction and the samples denatured prior to loading onto a 6% Long Ranger™ gel, prepared by dissolving 21g of ULTRA pure urea (Gibco-BRL) in 6ml of 10x TBE and 6ml of Long Ranger™ gel mix (50% FMC), and the final volume was made up to 50ml with dH₂O. The acrylamide solution was filtered and degassed before polymerisation was initiated by the addition of 25µl TEMED and 250µl 10% APS freshly made. The gel was pre-run using 1xTBE buffer, until the running temperature reach 50°C, 1µl of each sample was loaded and electrophoresis was continued overnight.

2.2.8 Culture of *T.annulata* Cell lines and a Bovine B Cell Lymphosarcoma Cell line

The cloned macroschizont infected cell lines D7 and D7B12 isolated by Shiels *et al* (1992, 1994) were maintained in culture at 37°C in a 5% CO₂ atmosphere. Routine cell culture was carried out using RPMI 1640 media (Gibco) supplemented with 15% heat-inactivated foetal calf serum (Sigma), 8µg/ml streptomycin, 8U/ml penicillin, 600ng/ml amphotericin B and 0.05% NaHCO₃. A bovine B cell lymphosarcoma cell line, (BL20 Morzaria *et al.*, 1982) was maintained as described for the *Theileria* infected cells, except myoclone calf serum (Gibco) was substituted for heat inactivated foetal calf serum.

The differentiation of competent (D7) and attenuated (D7B12) macroschizont infected cloned cell lines was carried out by incubation at 41°C. The cultures were passaged, using fresh supplemented medium, every two to three days to give a cell density of approximately 1.4x10⁵ cells/ml. Morphological examination of cultures during differentiation time course was carried out by light microscopy of Giemsa stained cytopsin preparations. Slides were prepared by spinning 60µl of culture at 1500 rev min⁻¹ (240g) for 5 minutes in a Shandon cytopsin 2. The slides were then air dried for 10 minutes and fixed in methanol for 15 minutes, before staining in Gurr's improved R66 Giemsa stain (BDH) (4% solution in water) for 40 minutes, followed by a brief wash with dH₂O and air drying.

Slides were examined using a Leitz Wetzlar SM-Lux light microscope fitted with an oil immersion x100 objective lens.

2.2.9 Isolation of RNA, gel electrophoresis and Northern blotting

TRI Reagent (Molecular Research Centre) was used to purify RNA from piroplasms and *T.annulata* infected cell lines. Macroschizont infected cells pelleted from 200ml of culture by centrifugation at 1000g for 5 minutes or 500µl of pellet piroplasms provided by F.McDonald, were resuspended in 5ml of TRI reagent. The samples were incubated at room temperature for 5 minutes, before the addition of 1ml of Chloroform. The samples were then covered tightly, shaken vigorously for 15 seconds and placed at room temperature for further 15 minutes. The resulting mixture was centrifuged at 12,000g for 15 minutes at 4°C. Following centrifugation, 60% of the aqueous phase was transferred to a fresh tube. RNA was precipitated by the addition of 2.5ml of isopropanol to the aqueous phase and incubated at room temperature for 10 minutes, before centrifuging at 13,000g for 10 minutes. The resulting RNA pellet was washed in 75% ethanol, air dried and resuspended in diethyl pyrocarbonate (DEPC) treated dH₂O. The RNA concentration of each sample was determined by taking an optical density (OD) reading, at wavelengths 260 and 280nm, and calculated as 1.0 OD unit at 260nm equals 40mg ml⁻¹. The purity of the RNA sample was determined by the 260/280nm ratio (Sambrook *et al.*, 1989).

RNA was separated by formaldehyde/agarose gel electrophoresis. Formaldehyde/agarose (1.2%) denaturing gels were prepared by melting 3.0g of agarose (Sigma, low EEO) in 195ml of dH₂O. When cooled to hand temperature, 50ml of 5x MOPS buffer (0.2M 3-[N-morpholino]-2-hydroxypropanesulfonic acid (MOPS) (pH 7.0) 50mM NaOAC, 5mM EDTA) and 5ml of formaldehyde (37%) were added and the agarose was poured into a casting frame. The gels were allowed to set at room temperature before being removed to the cold room for 15 minutes. 20µg of total RNA, for each sample, was denatured in 7.9 µl formaldehyde, 22.5µl formamide and x1 MOPS buffer, by heating to 55°C for 15 minutes before adding 5µl of formamide/EDTA/dye mix (0.3% bromophenol blue, 0.3% xylene cyanol FF, 10mM EDTA, pH 7.5, 97.5% deionised formamide). Electrophoresis was carried out in a fume hood overnight at 30 Volts, in recirculating (negative to positive electrode) 1x MOPS buffer, until the xylene cyanol dye front was

approximately one inch from the bottom of the gel. RNA standards (Gibco BRL) were run in conjunction with the samples of interest. The standard ladder, 0.24kb to 9.5kb, was visualised by removing the marker lane from the gel and staining in ethidium bromide (1mg ml⁻¹) for 20 minutes, followed by destaining with dH₂O for 2 hours after electrophoresis. Bands were visualised by exposure to short wave ultraviolet light (366nm). The remainder of the gel was equilibrated in phosphate transfer buffer (see section 2.2.3) for 15 minutes. RNA was transferred onto nylon membrane as described for Southern blotting (see section 2.2.3) then transferred before blotting overnight onto a Hybond N membrane (Amersham), see 2.2.3. The RNA was fixed to the nylon membrane by exposure to Ultra-violet radiation (see 2.2.3) and the blot was prehybridised in Church and Gilbert buffer. A probe was prepared using the Random Priming DNA Labelling Kit (see 2.2.3) prior to hybridisation.

2.3 Results

2.3.1 Characterisation of the λ DASH Clones

To initiate studies aimed at defining the regulatory elements that control the differential expression of *TamS1*, a genomic library was constructed using DNA isolated from purified *T.annulata* Ankara (D7) merozoites. Screening of the genomic library with a *TamS1* cDNA probe encoding the 30kDa piroplasm surface protein of *T.annulata*, resulted in a number of positive plaques. Three representative plaques were purified through subsequent rounds of screening and were donated by Dr Kinnaird. The λ DASH clone 1.1 was randomly selected for further characterisation with the proviso that if the clone did not represent the intergenic regions overlapping both 5' and 3' regions of the *TamS1* protein coding sequence, characterisation of other clones would be carried out. DNA was isolated from the phage following the preparation of a liquid lysate (see 2.2.1). To estimate the size of the cloned λ DASH insert, DNA was restriction digested with enzymes BamHI, HindIII, EcoRI and Xba I, and the DNA fragments separated by agarose gel electrophoresis. Analysis of the digested DNA revealed that the λ insert size was approximately 20kb, if not greater (Figure 2.1). Therefore, the λ DASH insert was too large to either sequence or subclone without further characterisation. Firstly it was necessary to identify fragments flanking the 5' and 3' ends of the protein coding region of *TamS1* which were considered large enough to contain putative regulatory domains. If it was found that these regions did not contain the control domains, for example, if evidence for polycistronic transcription and processing was found, then subsequently mapping and characterisation of overlapping 5' and 3' genomic fragments was to be performed.

The 5' and 3' ends of *TamS1* were identified using restriction enzymes known to cut in the polypeptide coding region, coupled with hybridisation using 5' and 3' specific cDNA probes. DNA prepared from the λ DASH clone 1.1 was digested with restriction enzymes Hind III x (Ava I, Hae I, Sal I), EcoR I x (Ava I, Hae I, Sal I), and Xba I x (Ava I, Hae I, Sal I), Southern blotted and hybridised with the TA37d probe (Figure 2.2). This probe represents a deletion of the *TamS1* cDNA insert generated for sequencing purposes and contained a 261bp segment located to the 5' end of the known cDNA sequence. The TA37d probe hybridised to DNA fragments over 6kb in size for the restriction digests Hind III, Hind III + Ava I, EcoRI, EcoRI + Ava I, Xba I and Xba I + Ava I (Figure 2.2; Lanes 1, 2, 5,

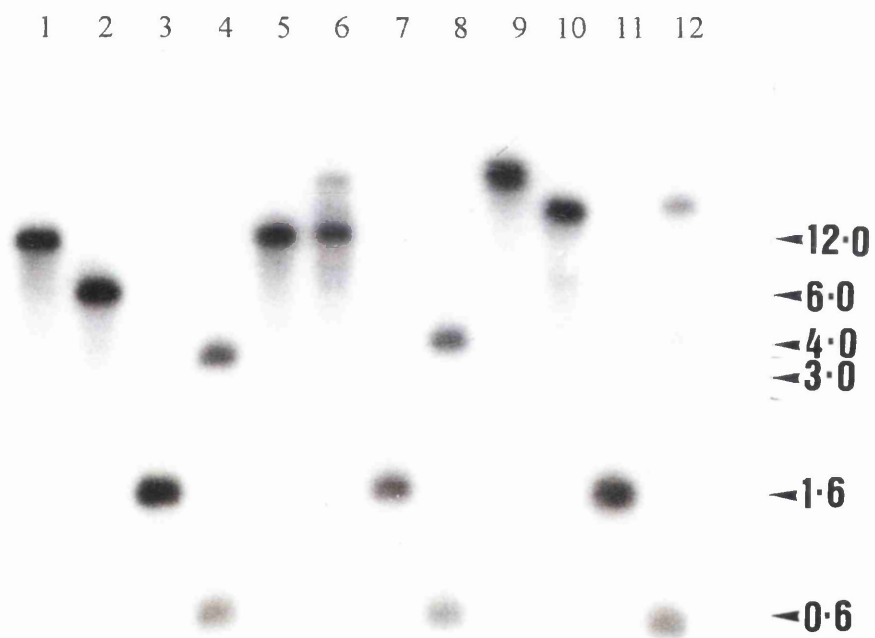
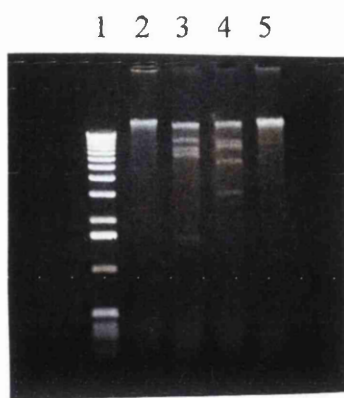
6, 9 and 10) which were considered to be too large for initial sub-cloning and sequencing. Common to the Hind III + Sal I, EcoR I + Sal I and Xba I + Sal I digests (Figure 2.2; Lanes 4, 8, and 11) was the detection of a 600bp fragment, in addition to larger fragments. Analysis of the *TamSI* cDNA sequence revealed two Sal I sites internal to the protein coding sequence of *TamSI* which are separated by 615bp. It was therefore concluded that the hybridisation of the probe to the ~600bp fragment corresponded to the Sal I restriction enzymes sites internal to the protein coding sequence. In addition to the Sal I fragment the TA37d probe hybridised to 4.0kb and 3.0kb bands for the restriction digests EcoR I + Sal I and Hind III + Sal I, (Figure 2.2; Lanes 8 and 4) respectively. From the limited information on the size of apicomplexan intergenic regions these fragments were postulated to be large enough to incorporate the 5' end of the *TamSI* coding region plus the majority of the intergenic sequence regulatory elements. In addition both fragments which contained putative elements were of a suitable size to sub-cloned into standard plasmid vectors.

Identification of the two Sal I restriction enzymes sites within the protein coding sequence of *TamSI* also provided an opportunity to identify the 3' intergenic region. It was therefore decided to determine whether a Hind III (which does not cut in the coding region) + Sal I digest of the λ DASH clone would generate a genomic fragment corresponding to the 3' downstream sequence which was of suitable size to allow sub-cloning. λ DASH 1.1 DNA clone was cut with Hind III, Sal I or Hind III + Sal I to completion. Digests were run in triplicate on an agarose gel and Southern blotted. The filter was divided into three and each section hybridised with probes representing either the 5', middle or 3' sections of the *TamSI* cDNA generated using the internal Sal I restriction enzyme sites (see Figure 2.4). The 3' probe hybridised to a Hind III + Sal I (Figure 2.4C; Lane 3) insert of 4.0 kb which was of suitable size for sub-cloning and considered large enough to contain a significant section of the 3' intergenic sequence region. Additionally, the 3' probe hybridised to a Sal I ~4.0kb fragment, identifying a Sal I site just upstream of the Hind III site. 3.0kb (Figure 2.4A; Lane 3) and 600bp (Figure 2.4B, Lanes 2 and 3) fragments were detected by the 5' and centre probes respectively, confirming the previous digest obtained with Sal I and the 5' region TA37d probe. The resulting restriction map is shown in Figure 2.3.

Figure 2.1. Agarose electrophoresis analysis of genomic clone isolated from λ DASH genomic library. The genomic clone was digested with the restriction enzymes; Lane 1 BamHI; Lane 3, Hind III; Lane 4, EcoRI; and Lane 5 Xba I. Lane 1 shows the 1kb Lambda DNA Marker (Gibco-BRL) ranging from 12kb to 75bp

Figure 2.2 Autoradiograph of Southern blot of Lambda Dash clone TamS1 1.1 probed with TA37d. Arrowheads at the right hand-side indicate the position in kilobases (kb) of DNA markers. The tracks contained the following digests of the Lambda Dash clone *TamS1*.

- Lane 1) Hind III,
- Lane 2) Hind III + Ava I,
- Lane 3) Hind III + Hae I
- Lane 4) Hind III + Sal I,
- Lane 5) EcoR I,
- Lane 6) EcoR I + Ava I
- Lane 7) EcoR I + Hae I
- Lane 8) EcoR I + Sal I
- Lane 9) Xba I
- Lane 10) Xba I + Ava I
- Lane 11) Xba I + Hae I
- Lane 12) Xba I + Sal I



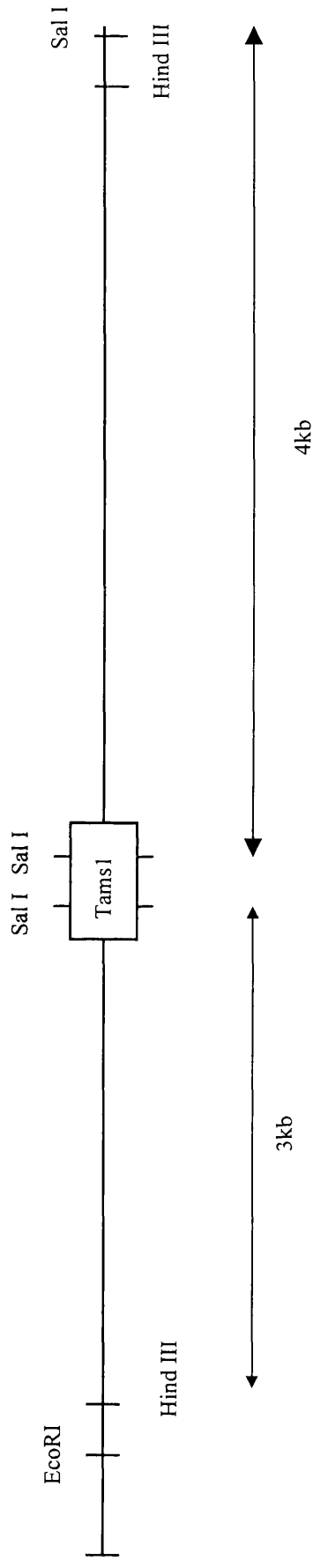


Figure 2.3 Restriction enzyme map of TamS1 1.1 bacteriophage clone. The restriction pattern was derived from the autoradiograph of a Southern blot of Lambda Dash clone TamS1 1.1 probed with the 5' region of *TamS1* cDNA, central 600bp Sal I fragment of *TamS1* cDNA and the 3' region of *TamS1* cDNA. The arrow indicate the distance between the Hind III restriction digest site and the Sal I site internal to the *TamS1* gene.

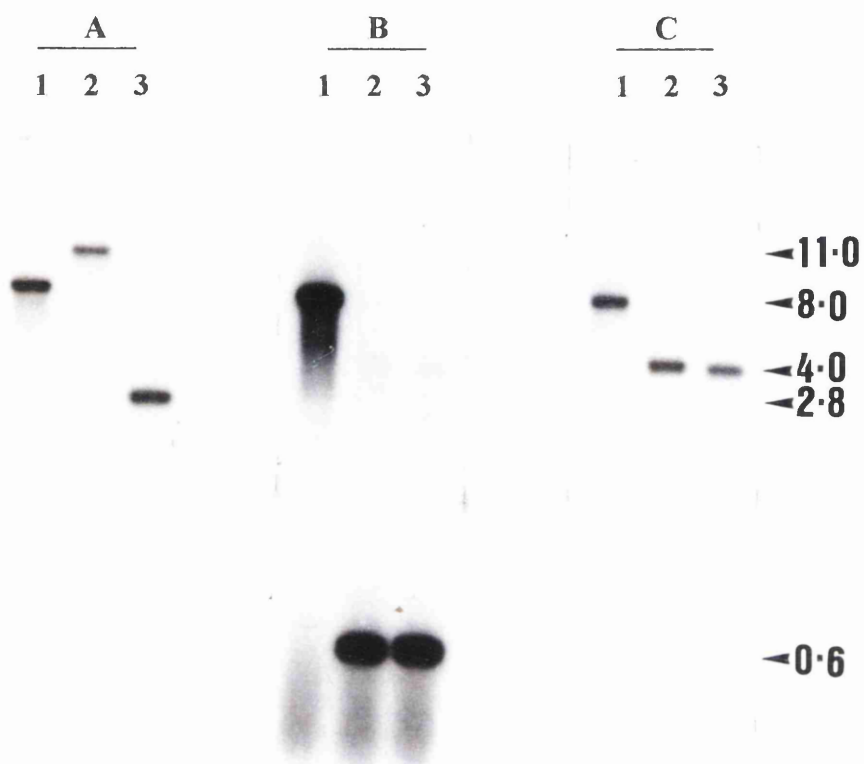


Figure 2.4 Autoradiographs of bacteriophage DNA Tams1 1.1 cut with (Lane 1) HindIII, (Lane 2) Sal I, and (Lane 3) Hind III x Sal I. Hybridized with (A) 5' region of *Tams1* cDNA, (B) central 600bp Sal I fragment of *Tams1* cDNA and (C) 3' region of *Tams1* cDNA. The position of each band is indicated by an arrowhead and in kilo-bases.

2.3.2 Subcloning the 5' and 3' regions of the *TamSI* gene

Restriction mapping and Southern blotting identified two fragments which were considered, from the known 3' end of the cDNA and the prediction that the 5' untranslated mRNA would be < 500bp, to represent a significant proportion of the intergenic regions 5' and 3' to the *TamSI* protein coding sequence. It was decided, therefore, to sub-clone the 5' 3.0kb Hind III x Sal I fragment and 3' 4.0kb Hind III + Sal I, respectively, into the pGem3zf vector for further characterisation. The sub-cloning procedure is summarised in Figure 2.5.

The 3 kb and 4 kb genomic fragments, separated from the bacteriophage vector, were excised, purified from the gel (see 2.2.4) and ligated into pGem3zf DNA which was also digested with Hind III and Sal I. Seven white transformants were picked for each ligation (see 2.2.4). Isolated DNA from the 5' genomic insert transformants was digested with Hind III x Sal I and 3' genomic transformants were linearized with the restriction enzymes EcoRI. Positive sub-clones were identified by Southern blotting and hybridisation with the appropriate 5' or 3' cDNA probe (Figure 2.6). 5' genomic sub-clones 3, 5, 6, and 7 (Figure 2.6A; Lanes 3, 5, 6, and 7) hybridised to the 5' cDNA probe and contained DNA fragments of the expected size (3.0 kb), sub-clones 6 and 7 were selected for subsequent sequence analysis (Figure 2.6 A; Lanes 6 and 7). Sub-clones 1 and 2 of the 3' genomic transformations contained DNA fragments of the expected size to which the 3' cDNA probe hybridised to and both were subsequently sequenced (Figure 2.6 B; Lanes 1 and 2). The selected sub-clones were manually sequenced in both directions (see 2.2.7). Comparison of the nucleotide sequence of each clone with the *TamSI* cDNA sequence to which they overlapped, either 5' or 3', gave an identity of 99.8%. It was, therefore, concluded that the 5' and 3' ends of the coding region plus intergenic sequence had been identified and sub-cloned.

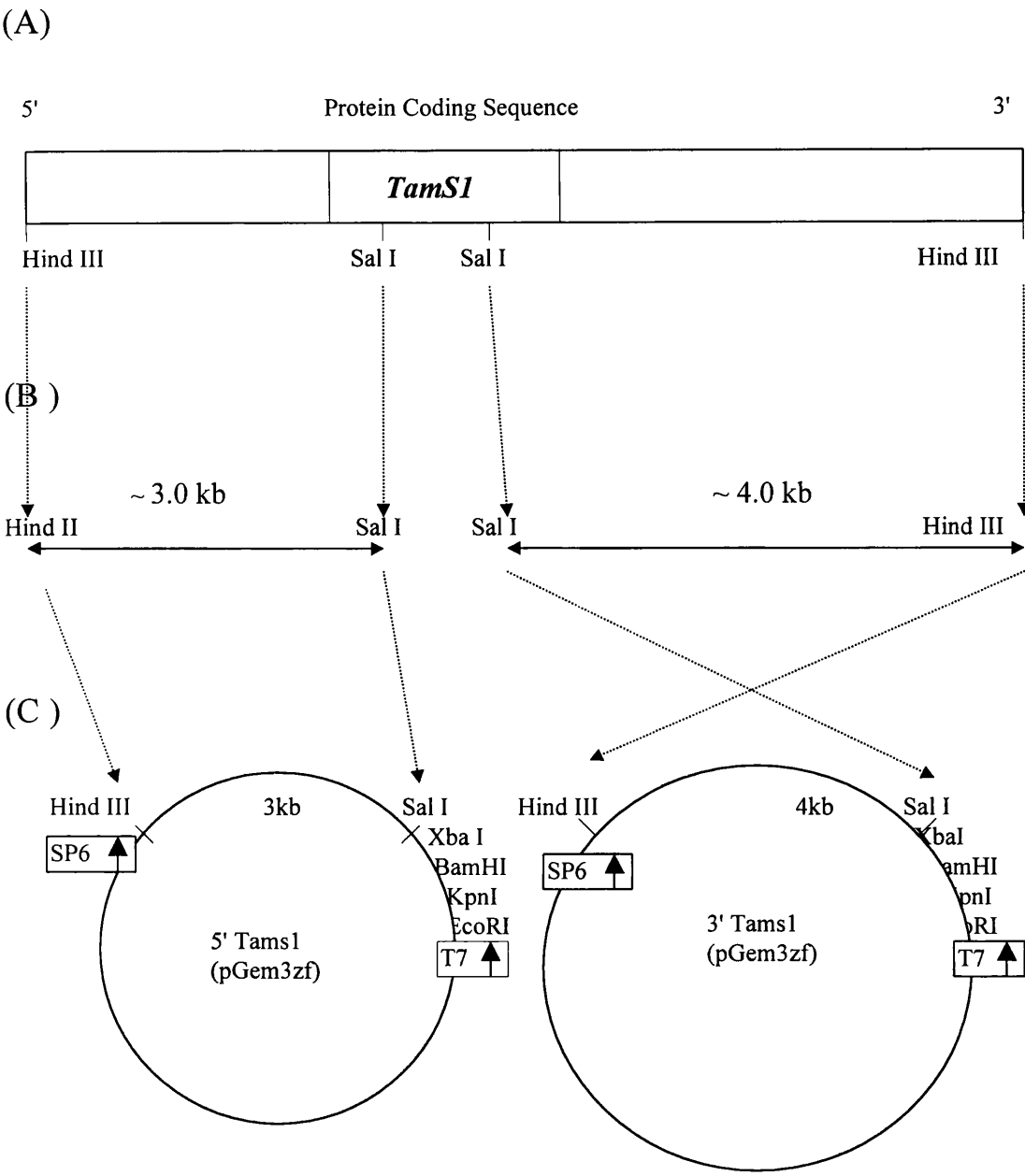


Figure 2.5 A diagrammatic representation of the sub-cloning of the 5' and 3' *TamS1* regions in to the vector pGem 3zf. A) shows a representation of the genomic *TamS1* fragment from the original Lambda DASH clone 1.1; (B) shows how the intergenic regions of *TamS1* were sub-cloned in two fragments, one was 3 kb and the other 4 kb; (C) depicts the final pGem 3zf plasmids.

Figure 2.6. Southern blot analysis of genomic sub-clones following digestion with Hind III + Sal I or EcoRI restriction enzymes and probing with cDNA probes as indicated. A * indicates the sub-clones subsequently selected for sequencing.

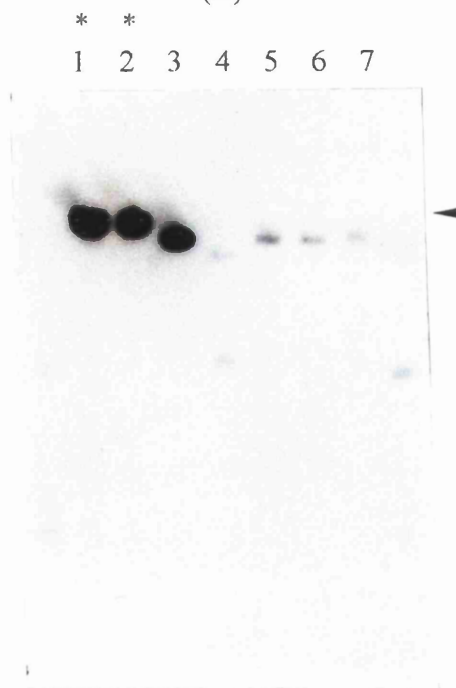
A: 5' genomic sub-clones digested with restriction enzymes Hind III + Sal I and probed with the TA37d probe. The position of the insert is indicated by an arrowhead and in kilobases.

B: 3' genomic sub-clones linearized with the restriction enzymes EcoRI and probed with the 3' cDNA probe. The position of the linearized ~7kb vector:insert band is indicated by an arrowhead.

(A)



(B)



2.3.3 Initial Sequencing of the 3kb Hind III + Sal I 5' *TamSI* intergenic fragment

To sequence the 3.0kb 5' intergenic genomic Hind III x Sal I fragment a set of Exo III deletions of this insert were generated using the methodology outlined in section 2.2.6. DNA for each deletion clone was analyzed by agarose electrophoresis and 11 out of the 30 randomly selected clones contained a range of deletions, which were subsequently sequenced. Preliminary analysis of the sequence data demonstrated that the 3.0kb clone had not been sequenced in its entirety and that 3 gaps existed within the nucleotide data. It was decided that due to the lack of success in generating a complete collection of overlapping deletions spanning the 3.0kb insert, that attempts would be made to sub-clone the genomic insert into smaller fragments for subsequent sequence generation. 5' *TamSI* plasmid DNA was restriction digested with enzymes Hpa I, Hinc II, Sac I, Ava I, BamHI + Ava I and BamHI and the DNA fragments analyzed by agarose gel electrophoresis (see Figure 2.7).

The Hinc II restriction enzyme cut within the genomic DNA to give three distinct bands of approximate sizes 800bp, 1.2 kb and 4.0kb (Figure 2.7; Lane3). Further analysis of the sequence data from the 3.0kb 5' intergenic fragment confirmed that a Hinc II enzyme site was present approximately 1.2kb from the 5' internal Sal I enzyme site. It was therefore concluded that the 1.2kb fragment contained both coding and intergenic sequence for *TamSI* and that the 4.0kb fragment consisted of 1.0kb of intergenic sequence and vector. It was decided to sub-clone the 1.2kb and 0.8kb fragments and religate the 4.0kb vector/insert band.

Plasmid DNA was cut to completion with the restriction enzyme Hinc II, the 1.2kb and 0.8kb fragments were isolated and blunt ended, ligated into the pGem7zf and pGem3zf vector DNA digested respectively with Sma I or HincII (see sections 2.2.2. and 2.2.4). The 4 kb fragment containing the pGem3zf vector plus 1 kb of 5' intergenic insert was religated. Three white transformants were picked for each ligation and screened for the presence of inserts either by Hind III + Sal I digestion of the pGem3zf DNA or BamHI + EcoR I digestion of the pGem7zf DNA. All three colonies screened for each ligation contained insert of the appropriate size and a representative clone of all three inserts was manually sequenced from either end using the T7 and Sp6 primers. Comparison of the nucleotide sequence of each clone with the sequence data from the deletions previously obtained

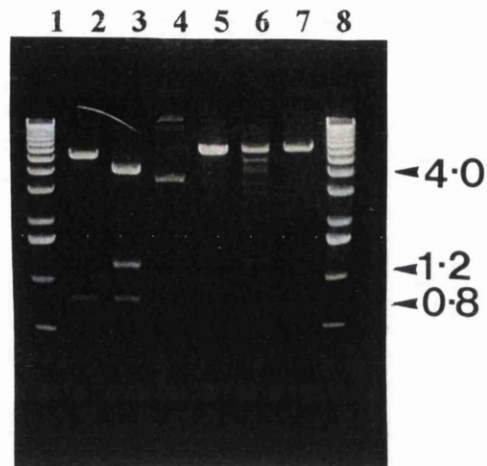


Figure 2.7. Agarose electrophoresis analysis of the 5' genomic clone of the 30kDa molecule from the λ DASH genomic library. The 5' genomic clone was digested with restriction enzymes; Lane 2) Hpa I, Lane 3) Hinc II, Lane 4) Sac I, Lane 5) Ava I, Lane 6) Ava I + BamHI, Lane 7) BamHI. Lanes 1 and 8 show the 1kb Lambda DNA Marker (Gibco-BRL) ranging from 12kb to 75bp. The position of the Hinc II fragments are indicated by arrowheads and in kilobases

deduced the order and orientation of each clone in relation to one another. Therefore, as previously concluded the 1.2kb fragment contained the 5' intergenic sequence proximal to the protein coding region of *TamSI*. The 0.8kb fragments splits the 1.2kb and 1.0kb fragments and the 1.0kb fragment is the most distal region relative to the protein coding region of the gene. Out of the three 5' *TamSI* 0.8kb clones, 1 had the fragment inserted in the opposite orientation to the other two sub-clones which aided in the generation of deletions to complete the sequence of this insert on both DNA strands. The sub-cloning procedure and the position of each clone in relation to each other is summarised in Figure 2.8.

2.3.4 Sequencing of the sub-cloned 5' *TamSI* intergenic fragments

Having obtained the 0.8kb, 1.0kb and 1.2kb sub-clones described above it was feasible to complete the sequence of the 5' intergenic fragment spanning from the upstream HindIII site to the Sal I site within the 5' protein coding region of the gene. From the nested deletions of the original 5' *TamSI* sub-clone, sequence data covering 1.2kb of the coding strand proximal to the polypeptide start site had been obtained. In addition sequence covering the Hinc II restriction sites had been derived from these deletion clones. Therefore, only the 5' strand of the 1.2kb sub-clone had still to be sequenced while for completion both strands of the 0.8kb and 1kb fragments had to be sequenced. All of the sub-cloned 5' *TamSI* intergenic inserts were greater than 500 base pairs in length and were too long to be sequenced conveniently from a single primer binding site on the vector. In order to obtain the complete sequence data from each sub-clone, unidirectional deletions were generated for both strands.

The orientation of the 1.2kb HincII 5' intergenic fragment within the pGem7zf vector had previously been determined by sequencing, see Figure 2.8. To generate unidirectional deletions allowing sequencing of the 5' strand of the 1.2kb insert, Exo III resistant and susceptible restriction sites were chosen within the vector, Sac I and BamHI respectively, and deletions generated as described previously. A total of 50 transformants were randomly selected from 12, 30 second deletion time points and screened to confirm deletion size by restriction digestion to release the insert and agarose electrophoresis. 14

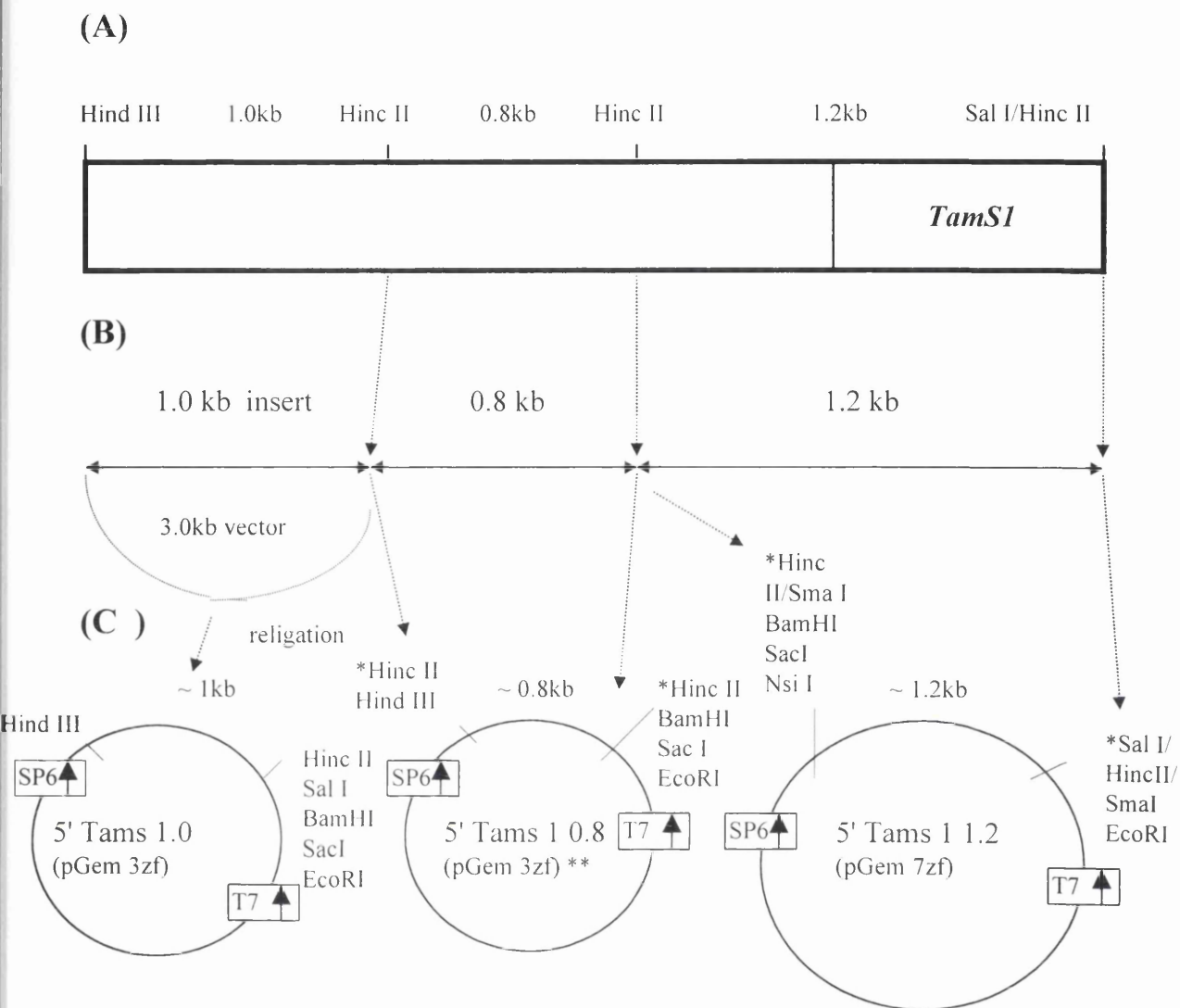


Figure 2.8 A diagrammatic representation of the sub-cloning 5' *TamS1* into plasmids 5' Tams 1.0, 5' Tams1 0.8, and 5' Tams1 1.2. A) shows a representation of the genomic clone 5' *TamS1* cloned into the vector pGem 3zf; B) represent the 5' region of *TamS1* divided into three Hinc II restriction fragments of 1 kb, 0.8 and 1.2 kb C) depicts the sub-cloned inserts in the pGem plasmids, the 1.2 kb fragment was sub-cloned into the pGem7zf vector, the 0.8 kb insert as sub-cloned into the pGem3zf vector, while the 1.0kb construct was generated by religaion of the original 5' *TamS1*.1 sub-clone following Hinc II digestion

* Sites removed by blunt end cloning procedure

** This fragment was also sub-cloned in the opposite orientation

sub-clones were identified that represented a set of deletions spanning the 1.2 kb fragment. All 14 clones were subsequently sequenced.

The isolation of two clones containing the 0.8 kb fragment in alternate orientations provided an opportunity to construct unidirectional deletions for both DNA strands using the same restriction enzyme sites in the polylinker of the vector. Plasmid DNA was cut with BamHI to generate a 5' overhang proximal to the Hinc II cloning site of the insert and Sac I to generate a 3' overhang to the insert/vector Hinc II site (see Figure 2.8) and Exo III digestion was as described in section 2.2.6. 10 recombinant clones, randomly selected from 7 time points were screened for deletions of the desired size. 6 deletions for the 5' sense strand and 5 for the anti-sense strand were of appropriate size for further analysis and were subsequently sequenced using the T7 primer binding site.

The sub-cloned Tams1.0 was obtained by religation of 5' *TamSI* following digestion with restriction enzyme Hinc II. Unfortunately, this strategy only left restriction sites suitable for generating deletions which would allow the anti-sense to be sequenced. Therefore, it was necessary to reclone the 1kb insert so that deletions for sequencing the coding strand could be generated. 5' Tams1.0 plasmid DNA was digested with restriction enzymes Hind III + Hinc II, and sub-cloned into the pBluescript SK (+/-) vector. DNA was prepared from 10 white colonies and screened for the presence of insert by digestion with Kpn I + EcoRI and gel electrophoresis. 9 out of the 10 clones contained insert of the appropriate size and two of these clones were confirmed by sequencing.

Deletions of the 1kb 5' intergenic fragment in either the pGem3zf vector or pBluescript Sk (+/-) vector were generated following the digestion of plasmid DNA with restriction enzymes, Sac I + BamHI or Sac I + EcoR I respectively. 4 Exo III deletion time points were taken and 30 recombinant clones were screened for deletions of appropriate sizes to allow the anti-sense strand to be sequenced. In addition 40 recombinants were screened for the deletions for the sense strand. DNA inserts were analysed by digestion with restriction enzymes Hind III + Kpn I and analysed by agarose gel electrophoresis. Selected deletions were then sequenced as described using either T7 or T3 primers.

2.3.5 Initial Sequencing of the 4kb Hind III + Sal I 3' intergenic fragment

The 3' Hind III + Sal I genomic fragment cloned into the pGem3zf vector is approximately 4kb in length. This sub-cloning strategy only allowed the generation of (see Figure 2.5) unidirectional deletions which were suitable for determination of the sense strand sequence. Following Exo II digestion and analysis of transformants, 23 deletions were selected and subsequently sequenced. Preliminary analysis of the sequence data generated showed that the 4kb 3' intergenic clone overlapped with the published cDNA sequence and that the 4kb insert had been completely sequenced on the sense strand. Further analysis of the sequence data revealed that the KpnI and Xho I restriction sites were located in the 3' half of *TamSI* intergenic region, ~ 2.1kb and ~2.4kb respectively, from the internal Sal I site to the protein coding sequence. Using this information it was decided to clone a 2.1kb XbaI x XhoI and a 2.4kb Kpn I x Hind III fragment to determine the sequence of the anti-sense strand of the 3' intergenic region. An advantage of this strategy was that it generated an overlap between the sub-clones allowing sequence determination of the KpnI and XhoI restriction site junctions.

The sub-cloning procedure is summarised in Figure 2.9. DNA was prepared from 10 white colonies and analysed by selected restriction enzyme digestion and gel electrophoresis. All of the selected clones contained an insert of the appropriate size and 2 clones for each ligation reaction were confirmed by sequencing the insert terminal regions. Comparison of the nucleotide sequence of each clone to the 3' sequence of the complementary strand previously determined gave a 98.7% identity.

The cloning sites within the polylinker of the pGem7zf vector utilised for sub-cloning the ~2.1kb and ~2.4kb 3' intergenic fragments, allowed the sequence of the anti sense strand to be determined from Exo III deletions generated following digestion of both plasmids with restriction enzymes Sac I and Hind III. DNA prepared from each of the 58 randomly selected recombinants for each 3' deleted plasmid were screened for inserts of the appropriate size by digestion with restriction enzymes Nsi I + Apa I and analysed by gel electrophoresis. Selected deletions were subsequently sequenced.

The sequence data of the genomic DNA flanking the 5' and 3' ends of the coding region of the *TamS1* gene (cDNA) was assembled into two contigs using the gel assemble programme. As shown in Figure 2.5 both contigs overlapped with the 5' and 3' respective termini of the published cDNA sequence and the restriction analysis performed during the sub-cloning. This confirmed that the correct regions had been sub-cloned and sequenced to allow the generation of 1 large contig of 8kb. The DNA sequence covering this contig is given in Appendix I of this thesis.

2.3.6 General Analysis of Sequence

The complete sequence of *TamS1* plus the 5' and 3' contigs is 8068 base pairs in length and has a base composition of 32.8% G+C, 67.2% A+T. This is comparable with *T.parva* which is estimated to be 31% G+C (Nene *et al.*, 1992), but higher than *Plasmodium falciparum* which has an unusually low G+C content of 18% (Weber, 1988; Pollack *et al.*, 1982). The overall nucleotide compositions of protein coding sequences, total intergenic sequence, and 5' and 3' flanking sequences across the ~8kb contig are shown in Table 2.1. In addition, the % G+C content has been plotted across the entire 8kb sequence, see Figure 2.15. Analysis of this data reveals, as expected, that the G+C% is highest within open reading frames and that G+C rich regions, relative to the remaining intergenic (or total) sequence, exists immediately 5' to the these open reading frames. These G+C rich islands may be indicative of control elements for each of the protein coding sequences. Analysis of the composition of the first 400 bases 3' to each open reading revealed an A+T rich sequence in the immediate 3' flanking region of each gene. The role of this A+T rich sequence is unknown, however, work on other eukaryotic genes suggest that the termination process for transcription may require a T-rich or A-rich sequence immediately 3' to the gene.

The overall base compositions of the open reading frames within the 8kb contig is 40% G+C which is comparable with *T.parva* at 42.5%. All of the coding sequences within the 8kb contig have a small excess of T and A over G and C. The nucleotide composition of *TamS1* had been previously determined as 42.7% G/C (Shiels *et al.*, 1994) and the Ta-ORF-1 has a similar G+C content of 40%.

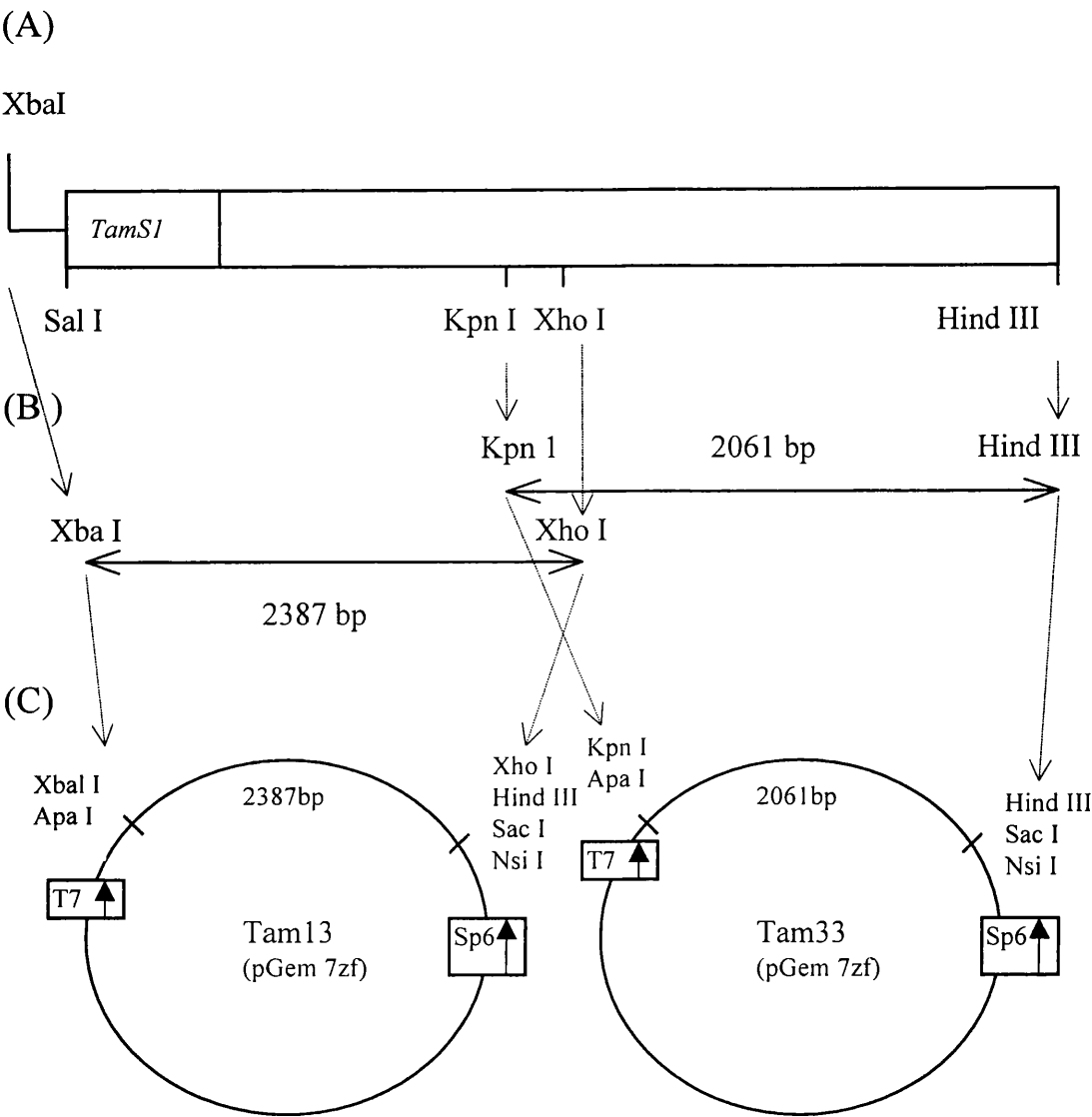


Figure 2.9 A diagrammatic representation of the sub-cloning of the 3' *TamSI* into plasmids Tam13 and Tam33. A) shows a representation of the genomic clone 3' *TamSI* cloned into the vector pGem 3zf; B) shows how the 3' region of *TamSI* was sub-cloned into two fragments of 2061bp and 2387bp inserts; C) depicts the final pGem7zf plasmids.

2.3.7 Open Reading Frames

Sequence analysis of the compiled 8kb contig identified four open reading frames, including *TamS1*, see Figure 2.10-2.13. A complete open reading frame was identified downstream from *TamS1*, Ta-ORF-1 and the two other partial open reading frames were detected at either end of the contig sequence, Ta-ORF-2 and Ta-ORF-3, see Figure 2.14

The potential product of Ta-ORF-1 is 303 amino acids in length, as shown in Figure 2.11. The nucleotide composition of the Ta-ORF-1 gene (see Figure 2.11 or Appendix 1 nucleotides 5816-4906) has been determined as 60.2% A/T and 39.8% G/C. In order to determine whether Ta-ORF-1 was expressed by either of the macroschizont, merozoite or piroplasm stages a Northern blot was hybridised with a 1.2kb Ta-ORF-1 probe generated from the clone Tam13. This analysis showed that Ta-ORF-1 detected an abundant mRNA of 1.8kb, in RNA isolated from the infected cloned cell line D7 cultured at 37°C and during differentiation towards the merozoite (Figure 2.16). This mRNA was also detected in total RNA isolated from the piroplasm stage, but at a reduced level. Reprobing the blot with the *TamS1*, cDNA probe confirmed previous finding of Shiels *et al.*, (1994) in that an abundant 1.1kb mRNA was detected which is expressed at low levels by the macroschizont and is significantly up regulated during differentiation to the merozoite, and is expressed at a high level in the piroplasm (Figure 2.16). It was concluded that while Ta-ORF-1 and *TamS1* are expressed in the same stages the temporal pattern of expression of the genes in the stages analysed is different. Unfortunately, no RNA from the sporozoite or tick stages of the parasite was available in order to determine the complete expression of the gene. The data from the Northern blot suggests that Ta-ORF-1 is expressed and has a structural/functional role in *Theileria* parasites.

Analysis of the amino acid sequence for Ta-ORF-1 with the GCG program MOTIFS or FASTA did not reveal any similarities with other polypeptide sequences in the databases. The deduced 303 amino acid residue sequence of Ta-ORF-1 has a calculated molecular mass of 34.2kDa. It is a highly hydrophilic molecule which has 4 predicted N-linked glycosylation sites but no other recognisable sequence domains. The predicted amino acid sequence from Ta-ORF-1 was compared with that of the equivalent predicted open reading frame from the *T.parva* contig (Chapter 3). An overall degree of identity of 66.1% was

2641 AACTTATAATAATGTTATATTAAAAATATGACAAATTAATGTGTGAAATAAATAAGTAAA
 2701 GTGGAGGGGAAATTCTGTCTTAATATTTAACTGGGTCTAGTGACTTAAAAAAGAAAGCAGG
 2761 TTATTGGTTTGAAAAAATAATAAATTCATCTTTTTATAGTCATCAAAAAATGAAAAATGG
 2821 ATAAATATACATGTTATTTTCATTAAGATGGACTTCAATGGAGGATAAGGCATTAGGCGTG
 2881 ATGAGTGCATAGATACAGATAAACATGCACACAATTTGTAGGGCGACATTGTTTTGTATG
 2941 GTGAATGCATATCCATCTCACTTCACTTATAGTTATTGGTCTTTTTTTCCATAACATCCA
 3001 CCCAATTAGTTAATTTTTTAATATTTAAATCGCTCACTAGTCTGCCCTTTCTTATCTTTTT
 3061 ATAATATAATTATTTGAGATGTTGTCCAGGACCACCCTCAAGTTCTTATATTTGAGCTTC
 M L S R T T L K F L Y L S F
 3121 TTCGTTATCTCATCCGTTAATGCTGCAAATGAGGATGAAAAGAAAAAGGAGGAAAAAAA
 F V I S S V N A A N E D E K K K E E K K
 3181 GATGTTGTTCTTGATGTTACTCTCACTTCATGTGAGAATGTAACCTTTAAAAACGTCGAC
 D V V L D V T L T S C E N V T F K N V D
 3241 TCTAACACCACTGAGTTAACTGTGCGGATGGCTACCGTTTCAAGACCTTAAAGGTCGGA
 S N T T E L T V A D G Y R F K T L K V G
 3301 GACAAGACCTTGTTCAATGTTGACACCTCAAAACATACCCCAGTACAGGCATTCAAACCTT
 D K T L F N V D T S K H T P V Q A F K L
 3361 AAGCATGAATCCGATGAGTGGTTCAGACTTAATCTTCAACCCTGCCAGCCAAAGATGTTT
 K H E S D E W F R L N L H P A Q P K M F
 3421 AAGAAGAAGGGAGACAAGGAATATTCTGAGGTCAAATTCGAGACCTACTACGATGATGTC
 K K K G D K E Y S E V K F E T Y Y D D V
 3481 TTGTTCAAGGGGAAATCCGCCAAGGAAGTAGATGCTTCCAAGTTCGAAGATACATCTTTG
 L F K G K S A K E L D A S K F E D T S L
 3541 TTCACCTCCTCCGCTTCGGCACTGGAAAAGATGTACACCTTTAAAAAGGAATTTAAACCT
 F T S S A F G T G K M Y T F K K E F K P
 3601 TCCAAAGTCACCTTCGACAAAGAAAGTTCGGAAAACCAAACAATGCCAAGTATCTTGAA
 S K V T F D K K E V G K P N N A K Y L E
 3661 GTTGTTGTTTTTGTGTTGTTCTGATTCCAAGAAGTTCGTCAAACCTCTACTACTTCTATACC
 V V V F V G S D S K K F V K L Y Y F Y T
 3721 GGTGAAGTCAAGGTTGAAGGAGACCTACTTCGAGCTTAAGGACGATAAGTGGGTTCAAATG
 G D S R L K E T Y F E L K D D K W V Q M
 3781 ACACAGGCAGATGCAACAAGGCCTTGAATGCCATGAACTCATCTGGTCAACCGATTAC
 T Q A D A N K A L N A M N S S W S T D Y
 3841 AAACCAGTTGTCGACAAGTTCTCCCCCTTGCAGTCTTCGCCTCAGTACTCATCGTCTTC
 K P V V D K F S P L A V F A S V L I V F
 3901 TCATCAGTCCTTTACTTCCTTTAAACCCATGTTTCGTAACAACCTTATCAACTTTTAAAC
 S S V L Y F L *
 3961 AATTTTGATAATTTGTATACAATTCGAGAACTAAATAACTAGCTTAAGTCATTATATGC
 4021 CACTTAATTTTATACTTTAAACTTATATGTTTAGATATAACTTCAACAGCTATAGGTCAT
 4081 CAATATCCAAGTCACGGTAAAAATCGCGGCCCTCACGCTAACCAGCTAAAAGCAGAACAGC
 4141 AAACCTCTGGTTATCGCAGCTTCTCTTCACTTTCGTCATCTCTTTCCACAGTGTGCAATG
 4201 AACCTGTAAAAATTAAGTGAATTGTAAAAATGAAATTATTAATCTTACTGAAAAGACGGT

Figure 2.10. The nucleotide and derived amino acid sequence of the *TamSI* gene. Regions of sequence overlap between published cDNA sequence and the genomic contigs are underlined. The numbers depicted the position of the sequence on the 8kb contig (see Appendix 1).
 * denotes termination codon TAA

7321 ATTTCTTTAAGTATTTTTCTTCTCTACTGCGATACCCCTCCAGTTGACACATTATACTTA
7381 TTTTCCAGGCTTTTTATTCTCTTAGATAGATTTTCCAGGTGTTTCTCGTATTCTTTACAT
7441 AACATTTCTTCCGCTAACTTTTGTTCCTTGCAGTGTTAGCTTTTCACATGGCATATATCTC
7501 TTTTGGAGGTGAGAATACTTAAAGTTGCTCGTTTTTCTATTAAATGTGTTGATTTTGGTA
7561 GAGAACGTACCATTCTAGCAACAAATTTGAATCTTCTAGGCCTTTTGACAAGTTTCCGC
7621 TTCGGAATCGATTAGTTTCCTTGATTCCCTACAGATAATTTCAATTGGAGTTAGTTACCTC
7681 TATTAGTTTCAAATCAAAATACTCCTTAGGTTTGTACTGGTCCGCAGTCTTCCGTGAGGA
 * I L I S R L N T S T R L R G H P
7741 ATAGGATGTTTCACACTGTTCGTATGCCAGCAGGTTTACGAACTCTTCATTTATTAAGTGG
 I P H K V S D Y A L L N V F E E N I L Q
7801 TTATACTTGGACTGCGATTCATTAAATGTATTTTCTAGGTCATTTACAAAACTATTGAA
 N Y K S Q S E N F T N E L D N V F V I S
7861 TTGAAGACTACTGGTCTTGGCAGCTTTCTCTGGATGACTTGTGTCTCCAGGAGTTCTCTT
 N F V V P R P L K R Q I V Q T E L L E R
7921 TCTTCTTGTTTTTTCTTTTCCAATTCCTGTTTACGTCTTTCTATCTCTTCCATATCCAAA
 E E Q K K K E L E Q K R R E I E E M D L
7981 TCTTGTTCAACTTCTTCCTCTTCCATGGTTTCAATATCTGGAATTGTAAGTTCAACATTC
 D Q E V E E E E M T E I D P I T V E V N
8041 CCTTCTGGCTCTGGGAGGTTTGACAAGC
 G E P E P L N S L

Figure 2.12 The nucleotide and predicted amino acid sequence of open reading frame 2. The numbers depicted the position of the sequence on the 8kb contig (see Appendix 1).

1 AAGCTTGCGAACAGTCTCCAAGGGCTTCTCTGAGACGAAGTTCGCTCAATCAATGTTGTC 60
A C E Q S P R A S L R R S S L N Q C C P
61 CAGGTCTGAAGTTTTACGACTGGATGGTTAAGATTTACCGGGATCAGAAGGGTAGTAATC 120
G L K F Y D W M V K I Y R D Q K G S N P
121 CTCAATTATCTACCTTCGATTATGATAACTACGTCTACACCTATAACCCATATCGCAAAGA 180
Q L S T F D Y D N Y V Y T Y N P I A K K
181 AGAGCTAGTTCAGCTTATTCTAAGAATCCGTTCTCCTATTTCGTATAAAATCTTTTTTCTT 240
S *
241 TTTACTTTTATAGGTTTTCTTCTCATATTTGTTCCCCAGGTTCTAAGAAGGACGTTATT 300
301 TGTGATGATTTCTTTTTCTTCTGGGACACTGAAATGCAGTTTAAATATTTGAGAACAAAT 360
361 TGGCTTTAAATTAATGAAGTATTGTTATTTATTGAGTTTAAGATTTTGTTATTTTAAAT 420
421 TATTTGGCAAATTTGGTTTAAATTGAGTCATTTTAAATTTAATTATTCACCTTACACATTTA 480
481 TTATTCTCACCTCTATAAATAAATTACTTCTGGTTTTTTTAAACCCCATATTTATTTGAA 540
541 ATTTTAAATGTTTATACGCAAATTATATTATCTAATTTTCAGTTTCACGAAAAACAAGAAA 600

Figure 2.13. DNA sequence of open reading frame 3 plus 400bp of 3' intergenic sequence. The predicted amino acid sequence is shown below. The numbers depicted the position of the sequence on the 8kb contig (see Appendix 1).

determined between the two predicted polypeptides, with the most conserved region lying in the C terminal half (78% identity). The amino acid sequences diverge in the N-terminal half of the molecule. Comparisons of the Ta-ORF-1 sequence with *TamR1*, representing the merozoite 117kDa rhoptry gene, or *TamS1* and *Tams2*, encoding the polypeptide of >30kDa located on the surface of the merozoite, showed no significant homology with either the nucleic acid or protein sequence.

The partial open reading frames, Ta-ORF-2 and Ta-ORF-3 showed no sequence homology with any known gene in the databases searched or to any known conserved polypeptide motifs. Additional, preliminary studies into the expression of Ta-ORF-2 and -3 suggest that Ta-ORF-2 may be expressed during the differentiation of merozoites to macroschizonts. The Ta-ORF-3 probes did not hybridise with RNA made from either macroschizont, merozoite or piroplasm stages of the parasite (results not shown).

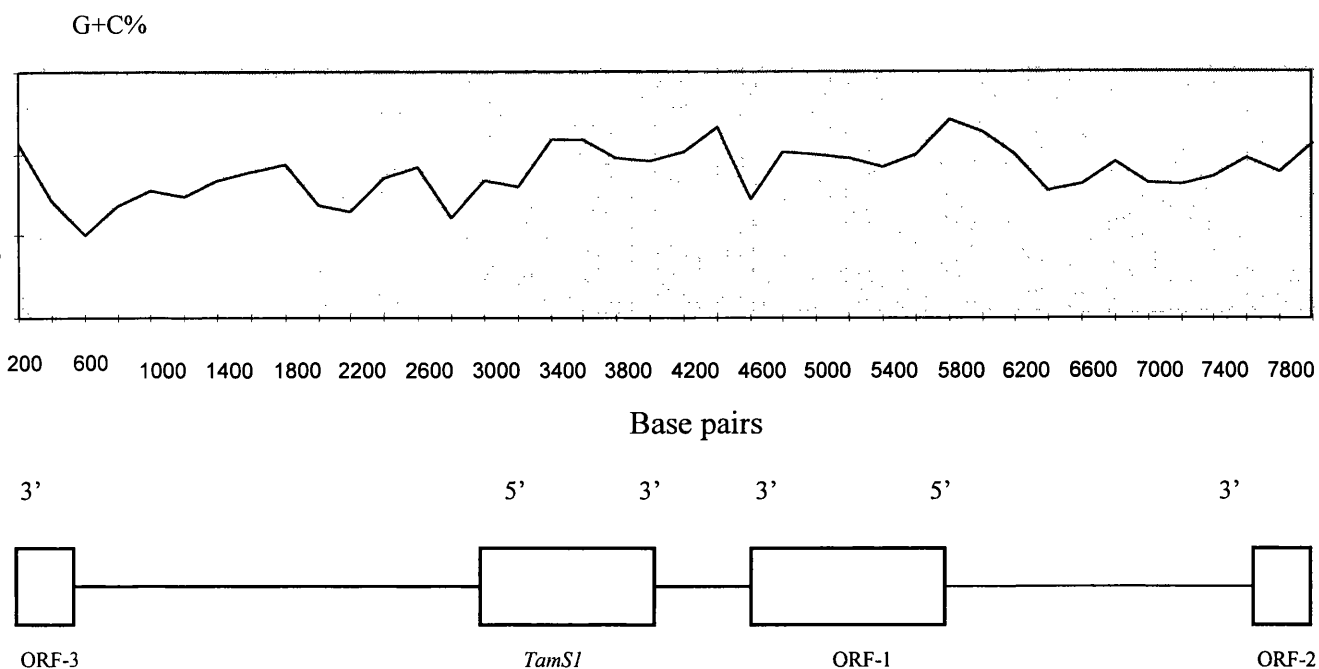


Figure 2.15 Distribution of nucleotide composition across the ~8kb genomic clone. The diagram represents the position of each open reading frame in comparison to the G+C content.

Table 2.1 Nucleotide composition of different regions of the 8.0kb genomic fragment.

	G +C %	Number of nucleotides
Protein coding sequence	40	2418
Total intergenic sequence	34	5726
5' flanking sequences	38	780
3' flanking sequences	31	1460

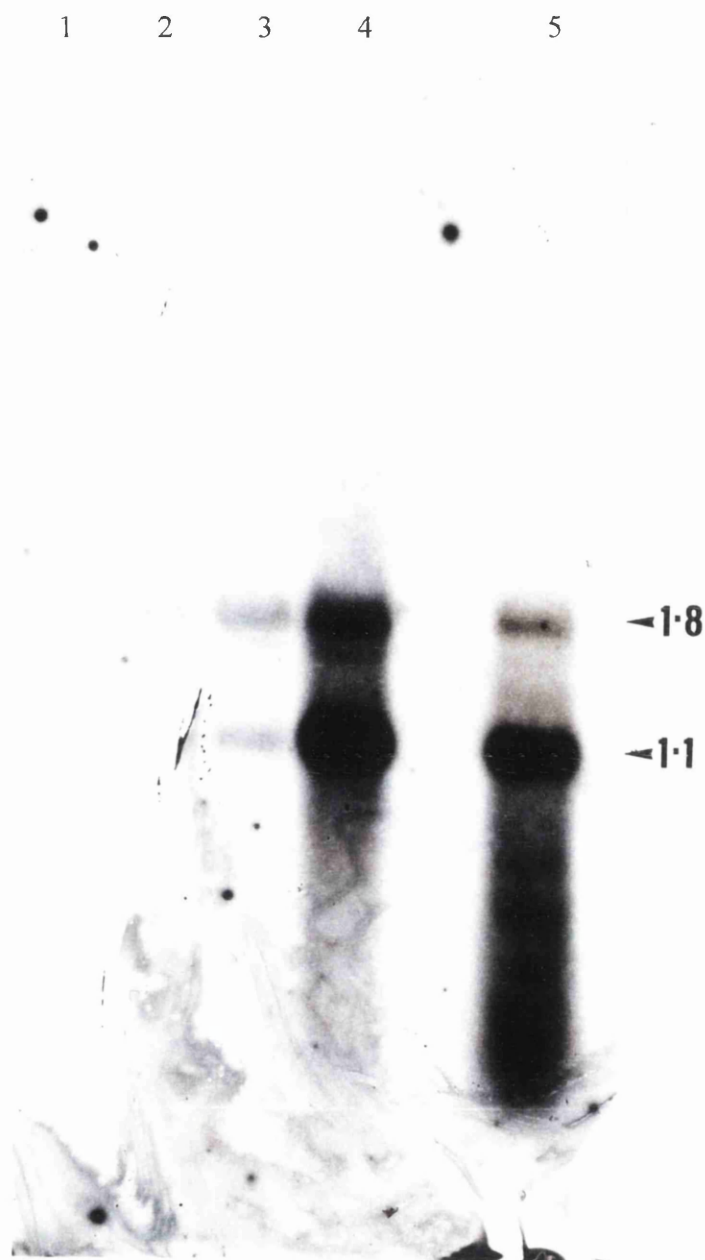


Figure 2.16 Analysis of RNA levels during a differentiation time course of macroschizont infected cell line D7 and Piroplasm RNA by Northern blotting. RNA was isolated from piroplasms and from D7 parasite infected cells at 37°C and every 2 days following incubation of the culture at 41°C. The blot was hybridized with the *Tams1* cDNA and Ta-ORF-1 probes. The size of the bands estimated in kilobases relative to RNA markers are indicated by the arrowheads. Macroschizont D7 culture incubated at 37°C, Lane 1; and at 41°C, Lane 2, Day 2; Lane 3, Day 4; Lane 4, day 6 and piroplasm RNA, Lane 5.

2.4 Discussion

To clone the regulatory elements controlling expression of the *TamS1* gene, a genomic library, constructed from DNA isolated from *T.annulata* (Ankara) merozoites, was screened with a cDNA probe encoding the 30kDa merozoite surface antigen. Three representative plaques were purified and DNA from one of the clones was restriction mapped with respect to *TamS1*. To maximise the likelihood of cloning the regulatory elements, a 3kb 5' intergenic genomic and a 4kb 3' intergenic genomic fragment were cloned and sequenced (Figure 2.3). The sequence data from the genomic fragments was assembled and allowed the generation of 1 large genomic DNA contig of 8kb (see Appendix 1).

Analysis of the genomic 8kb contig, containing *TamS1*, identified 4 open reading frames in various orientations (Figure 2.14). The genes flanking *TamS1* are of unknown identity and only open reading frame Ta-ORF-1 3' to *TamS1* was sequenced in its entirety. The nucleotide and predicted amino acid sequence of Ta-ORF-1 gene was compared to other polypeptides in the Genbank and Swiss Prot databases by FASTA search, however, no significant homology to other sequences was detected. Comparisons with known *Theileria* genes was also unsuccessful. Investigations into the expression of the Ta-ORF-1 product detected an abundant message of 1.8kb during day 6 of macroschizont differentiation to merozoites which was also observed at a much lower level at day 4. Interestingly, the RNA was also found in macroschizont infected cells at 37°C and showed an increase at day 2 of the time course. This temporal pattern of mRNA levels was similar to that detected by the *TamS1* cDNA and *TamR1* probes (Shiels *et al.*, 1994), although the level of the probe representing *TamR1* mRNA appears lower during the early phase of the time course. Low level expression of the Ta-ORF-1 gene was detected in RNA isolated from macroschizont infected cell lines incubated at 37°C and basal expression of the *TamS1* and *TamR1* in macroschizont cells has been previously demonstrated by Shiels *et al.*, (1994). However, basal expression did not seem to operate for all parasite genes as there was no detectable mRNA for the *SPAG-1* gene in the macroschizont stage. Only a limited number of examples of highly expressed merozoite genes have been isolated to date, and it is possible that basal expression of merozoite genes in the preceding stage of the parasite Life-Cycle stage may turn out to be a common occurrence. The role of the low level expression is unknown,

however, Shiels *et al.*, (1994) have postulated that it might relate to priming of the parasite for the next differentiation event and two possible mechanisms which explain how it occurs were suggested (see 2.1).

Northern blot analysis of Ta-ORF-1 revealed low level expression in the piroplasm stage of the Life-Cycle. It is however, unclear whether the Ta-ORF-1 product is expressed in piroplasm stage or if expression may be attributed to low level merozoite contamination or production within erythrocytes. Further investigations into RNA and polypeptide expression of Ta-ORF-1 are needed. Sporozoite and other tick stages of the parasite have, as yet, not been examined for the expression of the Ta-ORF-1 gene, but may be informative if expression of the gene was found to be limited to the Life-Cycle stages within the leukocyte or bovine host cells. The polypeptide and nucleic acid sequence of the Ta-ORF-1 gene did not show any similarity to other polypeptides sequences in any of the gene banks searched. However, the absence of a group of hydrophobic amino acids would suggest that the polypeptide is present in the cytoplasm of the parasite but without an antibody raised against the molecule, this cannot be proven.

Open reading frames Ta-ORF-2 and TaORF-3 are positioned at either end of the 8kb contig (Figure 2.14) and neither sequence showed any similarity to other polypeptide sequences in either the Swissplot or Genbank databases. Only the 3' end of each open reading frame was present within the 8kb. Comparisons of the full nucleotide and predicted amino acid sequence for each gene within the above databases may resolve their identity and function. Preliminary Northern analysis suggests that Ta-ORF-3 is not expressed in either macroschizont, merozoite, or piroplasm stages of the parasite and it is unclear whether this open reading frame is functional while Ta-ORF-2 was shown to be expressed at low levels in differentiating macroschizont cultures. Therefore, even though similarities in expression exist between 3 of the 4 open reading frames within the 8kb contig, no two genes are expressed to the same degree and with the same temporal order during the parasite's Life-Cycle. Different gene expression patterns of flanking genes have also been demonstrated by Nene *et al.*, (1992), suggesting that it is not necessary for genes expressed at the same time within the life-cycle to be clustered together.

Genes are often expressed in a temporal sequence, the activation of one gene triggering the expression of another (downstream) gene, ultimately leading to a cascade of expression which accompanies the transition of one cellular state to the next for the parasite, possibly differentiation. Alternatively, genes or related groups of genes are expressed co-ordinately, that is they respond simultaneously and usually to the same degree, to a regulatory signal. Genes that are co-ordinately regulated are often linked together in the genome and can be transcribed from a promoter at the 5' end of the gene cluster into a single RNA molecule, a polycistronic transcript. Most prokaryotic and some parasite transcription units are polycistronic, however, clustering of genes in apicomplexa parasites is not well documented, to date, and appears to be limited to rRNA sequences and the mitochondria (extrachromosomal element). Regardless of their relative order in the mitochondrial genome (extrachromosomal element) there is no firm evidence suggesting that the genes encoding polypeptides of the electron transport chain are expressed as large precursor RNA's within *Theileria*. However, polycistronic transcription and processing have been reported for the *Plasmodium* mitochondrial genome (Ji *et al.*, 1996) and is predicted to be involved in the expression of the *Trypanosoma brucei* mitochondrial genome (Feagin and Stuart, 1985; Jasmer *et al.*, 1985).

In higher eukaryotic organisms, individual genes are transcribed into individual monocistronic RNAs encoding single proteins. Therefore, genes which are co-ordinately expressed are transcribed as individual genes and tend not to be closely associated in the genome. As the genes (ORF's) defined in the 8kb *TamS1* contig are not co-ordinately expressed it seems likely that transcription operates monocistronically. Also genes which are polycistronically expressed are normally positioned 5'-3'- 5'-3', therefore, since the *TamS1* and Ta-ORF-1 reading frames are ordered 5'-3'-3'-5', this argues against this mechanism because the coding strands of mRNA do not match. However, as differential expression can occur, following precursor processing, at the level of mRNA stability in protozoan parasites a polycistronic mechanism, cannot be discounted. Indeed while many of the genes so far isolated in *Plasmodium*, *Toxoplasma*, *Babesia* and *Theileria* are believed to be transcribed monocistronically this hypothesis has only been proven for a relatively low number of apicomplexan parasite genes by nuclear run on analysis (Lanzer *et al.*, 1992a). Further studies analysing expression of *TamS1* by nuclear run on methodology are

necessary before a conclusion on the mechanism of transcriptional control can be made and on what role mRNA stability plays in controlling the differential control of this gene.

Nucleotide sequence comparisons between the 30-kDa merozoite surface molecule in *T.parva* and *T.annulata* has demonstrated a 75.5% identity (Shiels *et al.*, 1995) and similarities between *T.parva* p67 and *SPAG-1* of *T.annulata* have shown a similarity of 54% (Nene *et al.*, 1992). It seems likely, therefore, that ORF-1 cloned and sequence for both *T.parva* and *T.annulata* with an identity of 66.1% is the same gene or from the same gene family. Unfortunately, the length of the genomic *T.parva* clone (Chapter 3) did not include the two open reading frames found at either end of the 8kb contig. It is therefore, impossible to determine whether the gene organisation between *T.parva* and *T.annulata* is conserved.

While the expression patterns of *TamS1* and Ta-ORF-1 are similar during differentiation of macroschizonts towards the merozoite, sequence comparisons of the intergenic regions of both genes did not reveal distinct similarities which could be involved to co-ordinate the expression of these genes. This may mean that if expression of these genes is monocistronic then recognition of the important regulatory elements is by distinct factors. This possibility is supported by the differential regulation observed in the piroplasm stage, that there must be multiple factors involved in the control of expression of these two genes, at least within this stage. The lack of definition of shared motifs between intergenic regions of *TamS1* and Ta-ORF-1, however, necessitated that a different strategy was needed to define DNA sequence motifs involved in the control of *TamS1* gene expression. Interestingly, the size of the 3' intergenic region of both genes, where the 3'-3' regions are together, is much shorter than the 5' intergenic sequences. The 5' intergenic sequences may be longer due to the presence of important regulatory control elements, a hypothesis supported by G+C rich islands flanking the protein coding start site of each open reading frame. Additionally the 3'-3' intergenic region flanking both genes is very A+T rich suggesting that it may have a role to play in the termination process of transcription. Termination sites for eukaryotic RNA polymerase II genes have been shown to be either within a few hundred bases beyond the polyadenylation site, for example the α -globin gene (Sheffery *et al.*, 1984), or a few thousand bases beyond the polyadenylation site, for example the α -amylase gene (Hagenbuchle *et al.*, 1984). Comparisons between the

terminator signals for several genes, have implicated a long stretch of A or T residues in the process of transcriptional termination signals. Alternatively, the A or T rich 3' intergenic sequence may play a role in mRNA stability.

3 Structural Analysis of the 5' Upstream Region of the *TamSI* Gene

3.1 Introduction

During the differentiation of the macroschizont to merozoite *in vitro*, the 30kDa polypeptide is expressed differentially, as indicated by the analysis of steady state RNA and protein accumulation (Shiels *et al.*, 1994). The molecular mechanisms regulating this differential expression are unknown but, as indicated in section 1.5.1, are of importance because up regulation of *TamSI* gene expression is one of the earliest detectable events of the differentiation process. Furthermore, a direct correlation between the observed elevation with the ability to form merozoites (or reach commitment) was reported. Therefore, identification and analysis of structural motifs which direct the expression of *Theileria* genes during differentiation may lead to a greater understanding of the molecular mechanisms involved in determining merozoite formation.

The current picture of transcriptional control in eukaryotes is a complex one, where each of three RNA polymerases depends upon a particular set of transcription factors for correct initiation. The transcriptional control regions of eukaryotic protein coding genes transcribed by RNA polymerase II (pol II) can be separated into at least two categories: a core promoter and upstream (or downstream) regulatory elements. Promoters have been defined as modulatory DNA structures containing a complex array of cis-acting regulatory elements required for accurate and efficient transcription and for controlling differential expression of a gene (Polyanovsky *et al.*, 1990). Each gene can also carry a unique array of proximal and distal enhancer elements that are important for elevating, activating or repressing transcriptional activity (Tjian and Maniatis, 1994). Eukaryotic promoters and enhancers are regulated by a combination of sequence specific DNA binding proteins in combination with general transcription initiation factors plus accessory factors. Many eukaryotic protein coding gene promoters contain a TATA element, which is located approximately 25 to 30bp upstream of the transcription start site (Benoist and Chambon, 1981; Ghosh *et al.*, 1981). Interaction of a TATA binding protein with this sequence specifies the site of initiation in many of these promoters (Jones *et al.*, 1987; Schmidt *et al.*, 1989). In addition, it is generally accepted that the simplest promoter includes a TATA box and a transcription start site, or only a TATA box (reviewed in Breathnach and Chambon,

1981; Buratowski *et al.*, 1988). Other common important control elements highly conserved among many genes transcribed by pol II and which interact with the basal transcriptional machinery are GC and CCAAT boxes (Pugh and Tjian, 1992).

Protozoan parasites progress through a number of distinct life-cycle stages. Therefore, expression of genes in more than one developmental stage may require, in some cases, distinct combinations of transcription factors, especially if the same gene is expressed in response to different extracellular signals in different Life-Cycle stages. One single promoter region, therefore, may not always be sufficient to accommodate all the required information. Several genes have been described in eukaryotes that use alternative promoters during development. Both insulinlike growth factor I (IGF-I) and insulinlike growth factor II (IGF-II) genes are regulated by multiple promoters that are active in a variety of embryonic and adult tissues and are subjected to developmental and tissue specific regulation (van Dijk *et al.*, 1992; Gilmour *et al.*, 1994). Alternative promoter usage can influence gene expression in very diverse ways. The level of transcription initiation can vary between alternative promoters, and translational efficiency of mRNA isoforms with different leader exons can differ: they can also direct tissue specificity, differential responsiveness to extracellular signals and the generation of protein isoforms with variable amino termini (for review see Ayoubi and Van de ven, 1996).

To date, attempts to identify promoter elements in *Theileria* have been restricted to cloning DNA regions upstream of the transcriptional initiation site of several stage specifically expressed genes for sequence analysis. As yet only the promoter for the constitutively expressed gene hsp70.1 has shown sequence homology to DNA sequence motifs. Homology was found to the consensus sequence of the heat-shock element binding site and a putative TATA box was reported in the 5' region (Mason *et al.*, 1989). Putative TATA box sequences have been detected in the 5' region of several stage specifically expressed genes. These motifs are not believed to be functional, however, due to their positions in relation to the transcriptional initiation site and the ATG start codon (Katzer, thesis 1995; Fraser McDonald, personal communication).

Several promoters have been characterised for the related apicomplexan parasites, *Plasmodium* and *Toxoplasma*. Perhaps the most surprising revelation of all the studies conducted on *Toxoplasma*, is that to date no canonical TATA box has been found. In *Plasmodium*, with a non-coding region consisting of ~90% A/T content (Weber, 1988; Pollack *et al.*, 1982), a number of TATA boxes have been defined, although they are not a universal feature. In addition, McAndrew *et al.*, (1993) have reported the isolation of a gene encoding a divergent TATA-binding protein from *P.falciparum*. Thus it is very likely that a similar TATA-binding protein exists in *Toxoplasma* and *Theileria* and, if it is involved in transcriptional initiation of genes which lack a TATA box in their promoter, must work by either recognising a noncanonical TATA box or via indirect interaction with another DNA binding protein (Soldati *et al.*, 1995).

Polypurine motifs have been reported to be conserved between protozoan parasites but, the functional role of these motifs is, as yet, unknown. For example, Mercier *et al.*, (1996) drew attention to the presence of an A/TGAGACG motif which is common to four GRA genes and the SAG1 genes in *Toxoplasma* (Soldati and Boothroyd, 1995). In addition, a related motif, T/AGTGTAC, was reported in the putative promoter regions of the *Plasmodium* GBP130 and KAHRP genes (Lanzer *et al.*, 1992a, b; 1993). Interestingly, the purine rich sequence elements of both the SAG1 and GRA genes are present in tandem arrays, like the GC-rich motif recognised by the SP1 transcription factor of higher eukaryotes (Kadonaga *et al.*, 1986). Moreover, consistent with a SP1 like motif, it has been demonstrated that the A/TGAGACG motif acts in an orientation independent fashion (Mercier *et al.*, 1996).

The direct involvement of the purine motifs in the regulation of stage specific gene expression seems unlikely. In *Toxoplasma* the A/TGAGACG motif is found in the promoter regions of genes that are expressed in both tachyzoite and bradyzoite stages of the parasite (Mercier *et al.*, 1996). In addition, a purine rich sequence element in the *Plasmodium* GBP130 promoter region, expressed at the erythrocytic stage, was found to be homologous to the core region of the SV40 enhancer (Weiher *et al.*, 1983) and to a similar sequence motif in the upstream region of the *P.knowlesi* circumsporozoite gene (Altaba *et al.*, 1987). Therefore, it is expected that these purine motifs are involved in more general

transcriptional processes, since they are present in genes transcribed at different states of a parasite's Life-Cycle.

As described above, the primary control of gene expression lies at the level of transcription, however, in some cases the rate of synthesis of a particular protein can be altered without a change in the transcription rate of the corresponding gene. In principle, such post-transcriptional regulation could operate at any of the many stages between transcription and translation of the corresponding mRNA in the cytoplasm. One possible mechanism for regulating expression post-transcriptionally is by differential removal (splicing) of sequences which intervene between polypeptide coding regions of a gene. However, the inability to detect introns within multiple genomes sequences of *TamSI* (Frank Katzer, personal communication), removes the possibility that differential splicing operates in the expression of this gene.

An important factor controlling the amount of proteins produced is the stability of mRNAs. Degradation of a mRNA transcript proceeds from the 3'- end and deadenylation may lead to the removal of the methyl-guanyl cap at the 5'-end of the mRNA so that degradation also proceeds from that end. The more rapidly degraded an RNA is, the less protein it is likely to produce. Hence an effective means of gene regulation is by changing the stability of an RNA species in response to a regulatory signal. Stability of RNA can be conferred by motifs generally repeated several times in the 3' untranslated region of the transcript. These motifs have either the ability to increase stability or to greatly increase the rate of degradation. For example the stability of the H3 mRNA is controlled by a 30 nucleotide sequence at the extreme 3' end of the molecule (Latchman, 1990) that confers the potential to form stem-loop structures by intra-molecular base pairing (Mullner and Kuhn, 1988). It has been suggested that changes in stability might be brought about by alterations in the folding of this region of the RNA in response to a specific signal. Similarly the destabilisation of the mRNA encoding a number of cytokine and proto-oncogenes is conferred by a motif AUUUA which is repeated several times in their 3' untranslated region. Transposition of these sequences to the corresponding position in more stable mRNAs greatly increases their degradation (Hawkins, 1996).

In addition to specific sequence motifs, Poly (A) containing mRNAs are stabilised by Poly(A)-binding protein (PABP), a cytoplasmic protein that binds to a site containing at least 27 A residues in their Poly(A) tails. mRNAs with a smaller number of A residues are degraded so rapidly that they are generally undetectable (Hawkins, 1996).

The final stage in the expression of a gene is the translation of its messenger RNA into protein. In theory the regulation of gene expression could be achieved by production of all possible mRNA species by a cell followed by selection of particular mRNAs for translation into protein. Therefore, regulation of translation could operate via modifications in translational apparatus affecting the efficiency of translation of RNAs or by modifications in the RNA which affect the way in which it is translated by the ribosome. Modification of RNA can be mediated either by sequences in various parts of the RNA, involve secondary structure similar to that observed for mRNA stability and the use of different translation initiation codons. Furthermore, once the primary translation products are produced they may need further processing to acquire full functional properties. Thus, theoretically regulation of *TamS1* gene expression during differentiation could occur by any of the mechanisms outlined above. However, the temporal and quantitative correlation between mRNA and polypeptide levels observed during differentiation time course suggest that primary control is likely to be at the level of transcription and/or mRNA stability.

3.1.2 Aims.

The primary aim of work presented in this chapter was to identify the structural motifs which could be associated with the developmental regulation of the *TamS1* gene. Initially experiments were to be carried out to define whether the regulation was mainly controlled at the transcriptional or post-transcriptional level. It was also necessary to map the transcriptional start site of the gene and to identify potential poly A addition signals. A final aim of the work carried out in this chapter was to clone and sequence 5' and 3' intergenic regions from other *Theileria* species in order to pinpoint conserved domains which could be of functional significance. Designated domains may be shared across gene types and this was to be tested for by comparison with the available intergenic regions of the *TamR1*, *SPAG1*, and known control motifs of eukaryotic genes, in general.

3.2 Materials and Methods

3.2.1 Nuclear run on analysis.

Parasite nuclei were prepared using the method of Lanzer *et al.*, (1992). All steps were carried out on ice or at 4°C. 200ml of differentiating D7 macroschizont infected cells (Shiels *et al.*, 1994) were spun down at 1500g for 5-10 minutes. The cells pellet were resuspended in 12ml of ice cold nuclear wash buffer (20mM PIPES, pH 7.5, 15mM NaCl, 60mM KCl, 14mM β -mercaptoethanol, 0.5mM EGTA, 4mM EDTA, 0.5mM spermidine, 0.15mM spermine, 0.125 mM PMSF) and transferred to a dounce homogenizer. 270 μ l of a 10% NP40 solution was added and seven strokes with a pestle were applied. The cell homogenate was analysed under phase contrast microscopy to estimate cell breakage, transferred to a sterile Sorvall tube and spun at 1500g for 5 minutes to separate the host and parasite nuclei. Supernatant was transferred to eppendorfs and the 1500g pellet discarded. The supernatant was then spun at 5000g before being washed in 1ml of nuclear wash buffer and respun at 5000g. The 5000g pellet was resuspended in 50 μ l of 2x Elongation buffer (100mM Tris, pH 8.0, 50mM NaCl, 100mM KCl, 2mM MgCl₂, 4mM MnCl₂, 2mM DTT, 0.15mM spermine, 0.5mM spermidine, 10mM creatine phosphate, 2mM GTP, 2mM CTP, 2mM ATP, 25% glycerol) on ice, followed by the addition of 20 μ l of sterile DEPC treated dH₂O, 125 units/ml of RNasin (Promega) and 30 μ l of ³²P labelled UTP (300 Ci/mmol). The nuclei were incubated at 37°C for 5 minutes before stopping transcript production by heating at 70°C for 5 minutes. Following cooling of the reaction mix to 37°C, 10 μ g/ml of RNase-free DNase was added and incubated at 37°C for 5 minutes. 100 μ g/ml of *E.coli* tRNA was then added and unincorporated [α -³²P] UTP was removed by size exclusion chromatography through a Nuc-Trap Probe purification column (Stratagene). The column was pre-run with 70 μ l of STE buffer (100mM NaCl, 20mM Tris.HCl, pH 7.5, 10mM EDTA) and the reaction was passed down the equilibrated column using a syringe. Radiolabelled RNA was eluted by passing 150 μ l of STE buffer down the column. The eluate was collected into a fresh sterile eppendorf tube containing 100 μ l of Church and Gilbert buffer. The probe was either immediately frozen at -20°C or used directly in hybridisation experiments.

DNA fragments cloned into the pGem vectors, to be hybridised with the labelled RNA, were slot blotted onto nylon membrane filters. 1 μ g of sample DNA was diluted in

25mM phosphate buffer (pH 5.5) to a total volume of 50µl. The solution was boiled for 3 minutes, chilled and the DNA transferred onto nylon membrane pre-soaked in phosphate buffer, using a hybri-slot manifold (Gibco-BRL 24 well filtration manifold) under vacuum. The DNA was fixed to the nylon membrane, (as described previously, section 2.2.3), by exposure to 150 Joules short wave Ultra-violet radiation using a GS Gene linker (Biorad). Following UV cross-linking the membranes were prehybridised in 10ml of Church and Gilbert hybridisation solution (section 2.2.3) at 45°C for 30 minutes, with rotation, using the Hybaid cylinder and hybridisation oven system (Hybaid). The labelled RNA was added and hybridisation carried out for 48 hours. The filter was then washed three times with 1x SSC buffer and 5% SDS for 30 minutes at 45°C and exposed to X-ray film for 6 days.

3.2.2 5' Rapid Amplification of cDNA Ends (5' RACE)

The 5' RACE system (Gibco-BRL) and was used to determine the transcriptional start of *TamSI*. To carry out this method it was necessary to isolate piroplasm Poly A+ RNA, because it was known that high levels of *TamSI* RNA was present in this stage. Isolation of *T.annulata* piroplasm total RNA was carried with the Tri reagent (section 2.2.9) and used for isolation of mRNA using the Poly A Tract mRNA Isolation System (Promega). 1mg of total piroplasm RNA was combined with RNase-free sterile dH₂O to a final volume of 2.43ml and heated to 65°C for 10 minutes. The RNA was allowed to cool following the addition 10µl of Biotinylated-Oligo (dT) probe and 60µl of 20x SSC. Streptavidin-paramagnetic particles were then added to the annealing reaction and the tubes incubated at room temperature for 10 minutes. Using a magnetic stand the streptavidin-paramagnetic particles were collected, then washed in 1.5ml of 0.1x SSC; this was repeated four times. Finally the mRNA was eluted by resuspending the streptavidin-paramagnetic particles in 1.0 ml of RNase free dH₂O. The streptavidin-paramagnetic particles were collected and the eluted mRNA transferred to a sterile RNase-free eppendorf.

First strand cDNA synthesis was generated by combining 1µg of piroplasm mRNA with 2.5 pmole of the *TamSI* gene specific primer, GSP1 (see Table 3.1). The RNA was denatured by incubating the mixture at 70°C for 5 minutes, before cooling on ice for 1 minute. The contents of the tube were combined with 2.5µl of 10x reaction buffer (200mM Tris-HCl, pH 8.4, 500mM KCl), 3µl of MgCl₂, 1µl of 10mM dNTP mix (10mM each,

dATP, dCTP, dGTP, dTTP) and 2.5µl of 0.1M DTT. The sample was incubated for 2 minutes at 42°C, before addition of 1µl of Superscript Reverse Transcriptase, followed by a further incubation at 42°C for 30 minutes. The reaction was stopped by heating at 70°C for 15 minutes, before digesting the mRNA template with 1µl of RNase H, at 55°C for 10 minutes. The resulting cDNA was purified using a Glassmax DNA isolation cartridge (Stratagene). 120µl of binding solution (6M NaI) was added to the sample and transferred to the Glassmax cartridge. Following a 13,000g spin for 20 seconds, the column was washed once with 400µl of 1x wash buffer (Stratagene) and twice with 400µl of cold (4°C) 70% ethanol. The cDNA was eluted with 50µl of sterile dH₂O preheated to 65°C and centrifugation at 13,000g for 20 seconds. The single stranded cDNA fragment was now tailed with Poly dC by combining 10µl of cDNA with 2.5µl of 2mM dCTP, 1.5µl of 25mM MgCl₂, 2.5µl of 10x reaction buffer and 7.5µl DEPC-treated dH₂O; incubating at 94°C for 3 minutes; chilling on ice and adding 1µl of terminal deoxynucleotidyl transferase (10units/µl); followed by a incubation at 37°C for 10 minutes. The enzyme was then heat inactivated at 65°C for 10 minutes.

To PCR amplify the dC-tailed cDNA product, 5µl of tailed cDNA was combined with 2µl of anchor primer (Stratagene), 2µl of GSP2 (see Table 3.1) (prepared as a 10µM solution) 1µl of 10mM dNTP mix, 3µl of 25mM MgCl₂, 4µl of 10x reaction buffer, 5µl of diluted Taq DNA polymerase, and dH₂O to a final volume of 50µl. 30µl of mineral oil was placed on top of the reaction mixture and the tube incubated in a thermocycler for the following cycles: 94°C for 60 seconds (denaturing step) 57°C for 30 seconds (annealing step) and 72°C for 2 minutes (elongation step), followed by a final step of 10 minutes at 72°C. A total of 30 cycles was performed and the amplified sample analysed using agarose gel electrophoresis. The 5' Race product was excised from the gel, purified using the QIAquick gel extraction method and cloned in the TA Cloning Vector (Invitrogen), as described in section 2.2.4.

3.2.3 DNA amplification by Polymerase chain reaction (PCR)

The primers used for PCR are listed in Table 3.1. In all cases the PCR reaction mix was made up to 100µl and contained 1µg of each primer, 100ng of template, 9µl of 11.1x PCR stock buffer (45mM Tris-HCl, pH8.8, 11mM NH₄SO₄, 4.5mM MgCl₂, 6.7mM 2-β-

mercaptoethanol, 4.4 μ M EDTA, pH 8.0, 1mM dATP, 1mM dCTP, 1mM dGTP, 1mM dTTP, 113 μ g/ml BSA) and 1 μ l of Taq polymerase (amplitaq). 30 μ l of mineral oil was placed on top of each reaction mix and the tubes were incubated in a thermocycler for the following cycles: initially the mixture was incubated at 95°C for 4 minutes, followed by 25 cycles of 1 minute at 95°C (denaturing step), 1 minute at 45°C (annealing step) and 1 minute at 72°C (elongation step). The PCR products were separated by agarose gel electrophoresis, from which fragments were purified using the QIAquick gel extraction method following manufacturer guidelines as described in section 2.2.4.

3.2.4 Generation of genomic DNA mini-library and cloning of the *TamSI* gene from *T.parva*.

50 μ g of genomic DNA was digested with restriction enzymes, using standard conditions (section 2.2.2), to completion overnight at 37°C. The samples were run on a 0.7% agarose TBE gel until maximum resolution of the DNA fragment size to be isolated was obtained. Viewing under UV radiation, a scalpel cut was placed across the track, 0.5cm below the region of DNA to be eluted. Dialysis tubing was inserted into the cut, folding under the gel. The DNA was run onto the membrane for 3 hours at 100 volts. The length of the run was monitored under UV radiation by visualising how far the 1kb DNA marker bands migrated. Turning the current down to 20 volts, the tubing was removed quickly from the gel using plastic handled forceps. The tubing was placed in 1ml of low salt elution buffer (0.2M NaCl, 20mM Tris, pH 7.3, 1mM EDTA), and rinsed several times in the buffer before the tubing was discarded. The DNA was then purified using an Elutip column. The column was pre-washed with 2ml of high salt buffer (1M NaCl, 20mM Tris, pH 7.3, 1mM EDTA) and then primed by pushing 2ml of low salt elution buffer through the column. 1ml of gel eluted DNA was then passed over the column and this was repeated to increase efficiency of DNA binding. To elute the DNA, 200 μ l of high salt buffer was passed down the column, and DNA precipitated from the eluate by the addition of 2 volumes of ethanol and incubation at -70°C for 30 minutes.

<u>Name of Primer</u>	<u>Sequence of primer 5' to 3'</u>	<u>Position on <i>TamS1</i> genomic Sequence (Appendix I)</u>
GSP 1	cgaacttggaagcatctagttccttggcgg	3526-3497
GSP 2	caucaucaucaugggtcttgaaacggtagccatccgcgacag	3289-3260
1945	gcgaattcgggtgtagtattgctgaag	2416-2436
1944	gcaagcttttgcgttatgagctgtcc	2544-2524
1963	gtcgacgggtaaggacagctcataacg	2517-2540
1964	tctagagaatttcctccactttactt	2715-2694
1947	gcggatccaaataaagtaaagtggagggaaat	2684-2712
1946	gcggtaccccatcttaatgaataacatg	2850-2830
Tams1DB	gcgctcgagcttgtctccgaccttaagggcttt	3306-3282

Table 3.1: The oligonucleotide primers used for PCR and 5' RACE system. The name of the primers, their sequence, the origin of the sequence and the position they map to on the sequence given in Appendix I is listed.

2.5µg of Lambda ZAP Express vector DNA was cut with appropriate restriction enzymes (see sections 2.2.2) for one hour before dephosphorylating the 5' terminal nucleotides with calf intestinal alkaline phosphatase (see 2.2.4). The digested DNA was phenol:chloroform extracted before ethanol precipitation and the pelleted DNA was resuspended in 5µl of sterile dH₂O. Gel eluted *T.parva* DNA fragments were then ligated into the Lambda ZAP Express vector by combining 1µl of digested Lambda ZAP Express DNA with, 0.5µl 10x ligation buffer (see 2.2.4), 0.5µl 10mM rATP, 1µl of T4 DNA ligase and 2µl of *T.parva* DNA. Incubation of the ligation mixture was carried out for 2 days at 13°C and then packaged with the Stratagene Gigapack II packaging extract. A sample of freeze/thaw extract (Stratagene) was thawed quickly and 1µl of the ligation mix immediately added. The packaging reaction was placed on ice before adding 15µl of Sonic extract (Stratagene) and the mixture incubated at 22°C for 2 hours. 500µl of SM buffer and 20µl of chloroform were then added and the supernatant was stored at 4°C, before titration of plaque forming units as described previously (see 2.2.1).

To screen the λ Zap express mini-library, 50,000 plaque forming units were plated out onto BBL plates (section 2.2.1) and incubated overnight at 37°C. The following day the plates were chilled at 4°C for 2 hours, before plaque lifts were taken by placing nitrocellulose filters on the agarose for 2 minutes. During this period the position of the filters were marked with a needle. Removed filters were immersed in denaturing solution (1.5M NaCl, 0.5M NaOH) for 2 minutes before neutralising for 5 minutes by immersion in 1.5M NaCl, 0.5M Tris-HCl, pH 8.0. The filters were then rinsed in 0.2M Tris-HCl, pH 7.5, 2x SSC and the DNA was fixed by baking the filters at 80°C for 2 hours, under vacuum. Following fixation, the filters were hybridised with a radiolabelled probe (section 2.2.3) for 24 hours, washed 3x 30 minutes in 1x SSC, 5% SDS at 65°C and exposed to X-ray film overnight at -70°C. On development of the autoradiograph, positive plaques were identified and picked into SM buffer containing 20µl of chloroform and stored at 4°C overnight. In addition, one negative plaque was also picked to serve as a control in subsequent rounds of screening. Secondary and tertiary screens were performed as above until phage were derived from a single plaque.

To excise the plasmid containing the *T.parva* genomic insert from the λ ZAP, DNA, a 250 μ l aliquot of a positive pure phage stock (containing approximately 1×10^5 phage particles) and 1 μ l of R408 helper phage (approximately 1×10^6 pfu ml⁻¹) were absorbed onto 200 μ l of XL-1blue MRF' cells (Stratagene, see section 2.2.1). 5ml of 2XYT media was added and the mixture incubated for 2.5 hours at 37°C in an orbital shaker. The sample was heated to 70°C for 20 minutes before centrifuging at 4000g for 15 minutes. The supernatant was then decanted into a sterile tube and stored at 4°C. 200 μ l of XL-1Blue MRF' cells was then mixed with 200 μ l of phagemid stock and incubated for 15 minutes at 37°C before the addition of 200 μ l of 2XYT media and a further incubation at 37°C for 45 minutes. 200 μ l of the sample mix was then plated out onto LB/kanamycin plates and the plates incubated overnight at 37°C.

3.3 Results

3.3.1 Isolation of gene clones encoding the major merozoite antigen in *T.parva*

To try to define conserved sequences which may function as regulatory elements controlling the expression of the *TamS1* gene, the 5' and 3' intergenic regions of the related gene from *T.parva* (*TpmS1*, Shiels *et al.*, 1995) were cloned. As attempts to clone this gene from a *T.parva* genomic library with the *TamS1* cDNA probe were unsuccessful it was decided to construct a genomic DNA mini-library using DNA isolated from *T.parva* piroplasms (Muguga stock).

Restriction sites suitable for generating the mini-library were identified by digesting genomic *T.parva* DNA with, BamHI, Hind III, Kpn I, EcoR I and Sal I. The DNA was separated on an agarose gel, Southern blotted and probed with the 5' cDNA probe of *TamS1* (see 2.3.1). The resulting autoradiograph is represented by Figure 3.1(A). The blot was then stripped, reprobed with the 3' cDNA probe of *TamS1* (see section 2.3.1) and the result obtained shown by Figure 3.1(B). These results showed that both probes detected a single Hind III band at 6kb. This could be due to recognition of the same fragment by both probes or the presence of two Hind III fragments which were very close in size. If a Hind III site was close to 5' or 3' end of the probes used then only one intergenic region would be obtained, necessitating further cloning. From the limited information in chapter 2, and the size of the *TamS1* coding region, it was thought that there was a reasonable chance that a single 6kb Hind III fragment may contain sufficient additional sequence both 5' and 3' to the respective probes to allow a meaningful analysis. Thus, it was decided to construct the mini-library from genomic *T.parva* DNA cut with the restriction enzyme Hind III. The DNA was separated on an agarose gel before being electroeluted, ligated and packaged as described in section 3.2.4. A number of positive plaques were identified and purified to homogeneity by successive rounds of screening, using the cDNA probe of the *TamS1* gene (see 3.2.4).

Figure 3.1 Southern blot analysis of *T.parva* genomic DNA following restriction digestion and hybridisation with (A) 5' *TamSI* cDNA probe and (B) 3' *TamSI* cDNA probe (see section 2.3.1). Arrowheads at the right hand-side indicate the position in kilobases of the DNA markers.

Lane 1) BamHI

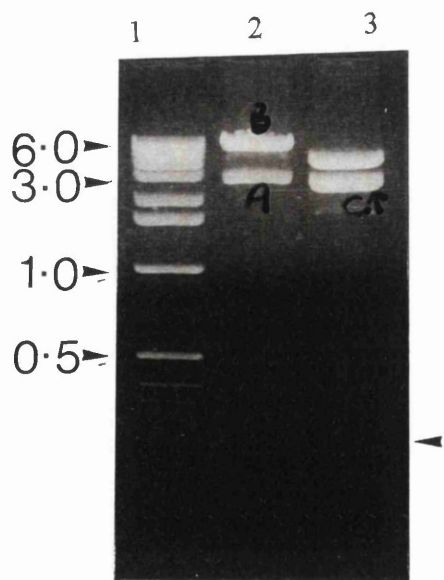
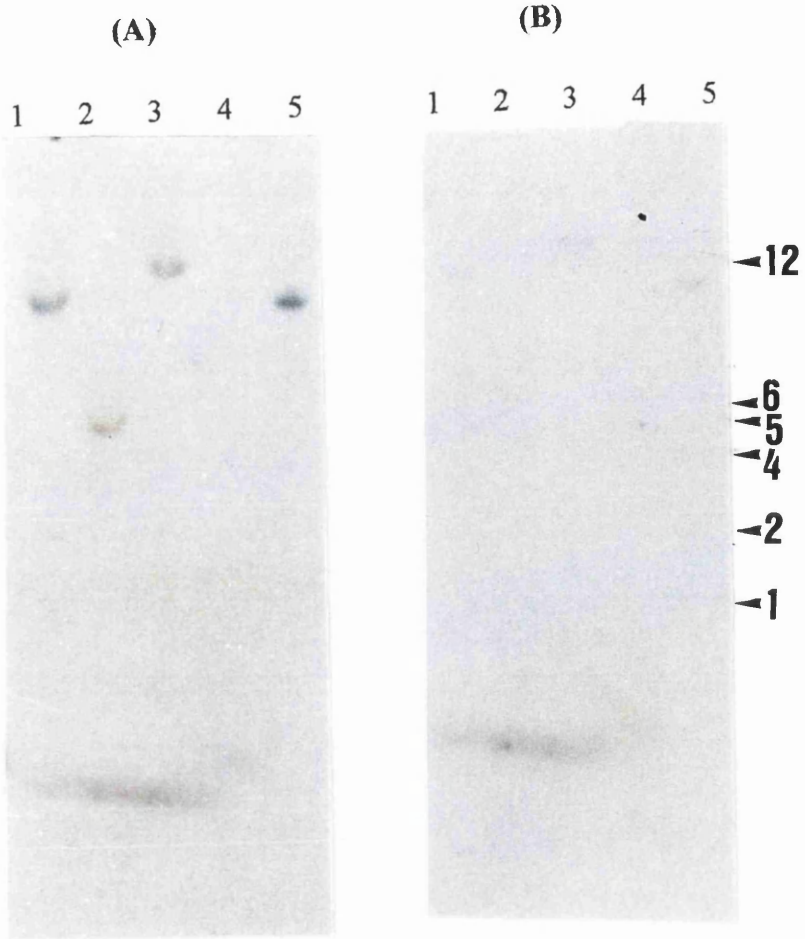
Lane 2) Hind III

Lane 3) Kpn I

Lane 4) EcoRI

Lane 5) Sal I

Figure 3.2. Agarose electrophoresis analysis of the TpmS13.2 genomic clone isolated from a *T.parva* genomic mini-library. The cloned DNA was digested with restriction enzymes: Lane 1) λ kb DNA ladder, Lane 2) Acc I and Lane 3) Acc I + Hind III, and Lane 1) shows the λ kb DNA ladder track. Arrowheads at the left hand-side indicate the position in kilobases (kb) of the DNA markers. The arrowhead on the right hand side of the diagram highlights a AccI fragment of ~250bp.



To excise the genomic DNA from the λ ZAP clone, XL-1Blue MRF' cells were co-infected with purified λ ZAP phage and filamentous helper phage R480 (as described in section 3.2.4). The size of the genomic insert in the resulting plasmid was estimated by digestion with Hind III followed by agarose gel electrophoresis and found to be ~5.5kb. Sequence data previously obtained by PCR (Shiels *et al.*, 1995) revealed that there was a Acc I site internal to the protein coding sequence of *TpmS1*. Restriction mapping demonstrated that the genomic insert plus vector could be divided into two fragments of approximately 2.7kb and 6.0kb when digested with Acc I (see Figure 3.2, Lane 2). The 2.7kb fragment consisted of the *TpmS1* genomic insert (band A) and the 6.0kb fragment of vector plus genomic insert (band B). Digestion of the excised clone with Hind III +Acc I resulted in the generation of three fragments, two of which were of approximately 2.7kb (Figure 3.3; Lane 3, band C). As digestion with Hind III releases an insert of 5.5kb and the 4.5kb band corresponds in size to linearized pBK-CMV vector DNA, it was concluded that the 2.7kb bands resulted from an AccI site in the vector polylinker for the 2.7kb band obtained with the AccI digest.

Following the result of the Southern blot, it was therefore decided, to sub-clone 3' and 5' regions of the *TpmS1* genomic clone using the Acc I sites within in the polylinker of the pBK-CMV vector and within the protein coding sequence of *TpmS1*. Plasmid DNA was digested with restriction enzyme Acc I and the products were analysed by gel electrophoresis on a low melting point agarose gel. Two fragments were excised from the gel, a 2.7kb fragment containing the 5' region of *TpmS1* and a 6kb fragment which contained the pBk-CMV vector plus 2.7kb of the 3' region. The 5' fragment was sub-cloned into the AccI site within the Bluescript vector and the 6kb fragment was religated. The 5' clone was given to Ms McKellar for sequencing.

3.3.2 Sequencing of the 3' untranslated of the *TamS1* related gene in *T.parva*

The 3' genomic insert of the *TpmS1* gene religated into pBK-CMV vector, was approximately 2.0kb in length. Unfortunately, this strategy only left restriction sites suitable for generating unidirectional deletions which would allow determination of the sense strand sequence. Plasmid DNA was digested with restriction enzymes Hind III + Kpn I, to generate a 5' overhang susceptible, and a 3' overhang resistant to Exo III, respectively. Exonuclease

digestion of the insert was performed as described previously in section 2.2.6. 64 randomly selected transformants were screened for deletions of appropriate size by digestion of plasmid DNA with restriction enzyme Pst I and gel electrophoresis. 18 deletions were selected for further analysis and were sequenced manually using the T7 sequencing primer. Preliminary analysis of the data indicated that the fragment had been completely sequenced on one strand.

To facilitate the sequencing of the anti sense strand, a Hind III x Pst I fragment was sub-cloned into the pGem3zf vector (see 2.2.4). DNA was prepared from 10 white colonies and screened for the presence of insert by digestion with Kpn I and agarose gel electrophoresis. 7 out of the 10 clones screened contained insert of the appropriate size and two positive clones were confirmed by sequencing. Deletions of the ~2.0kb 3' intergenic fragment were constructed following digestion with Kpn I x BamHI. The resultant deletions were transformed into *E.coli* XL-1Blue competent cells and the screened for the presence of appropriate size inserts by digestion with EcoRI and Hind III, followed by sequencing (see section 2.2.7).

The sequence data of the genomic DNA flanking the 3' end of the coding region of the *TpmS1* gene was assembled into a contig using the GCG gelaassemble program. Unfortunately, analysis of the 3' end of the 5' intergenic sequence contig (sequenced by Ms McKellar) did not show an overlap with the 5' end of the 3' intergenic sequence contig. To confirm that the correct fragment for the 5' intergenic region of *TpmS1* had been sub-cloned and sequenced the excised genomic clone, TpmS13.2, was digested with Hind III + Acc I and Acc I and hybridised with the 5' *TamS1* cDNA probe (see Figure 3.3(A)). The probe detected a band of 2.7kb for both digests (highlighted with an arrowhead) corresponding in size to the band observed in Figure 3.2. The Southern blot was stripped and reprobed with the 3' *TamS1* cDNA probe which hybridised to a 0.25kb band (Figure 3.3(B)). Re-analysis of the restriction digest of the original excised *T.parva* genomic clone with enzymes Hind III + Acc I and Acc I, confirmed that a ~250bp Acc I fragment had been overlooked during the original sub-cloning of the original clone (see Figure 3.2, band highlight with an arrowhead). To obtain overlapping sequence across the AccI sites, including the 250bp sequence, of the

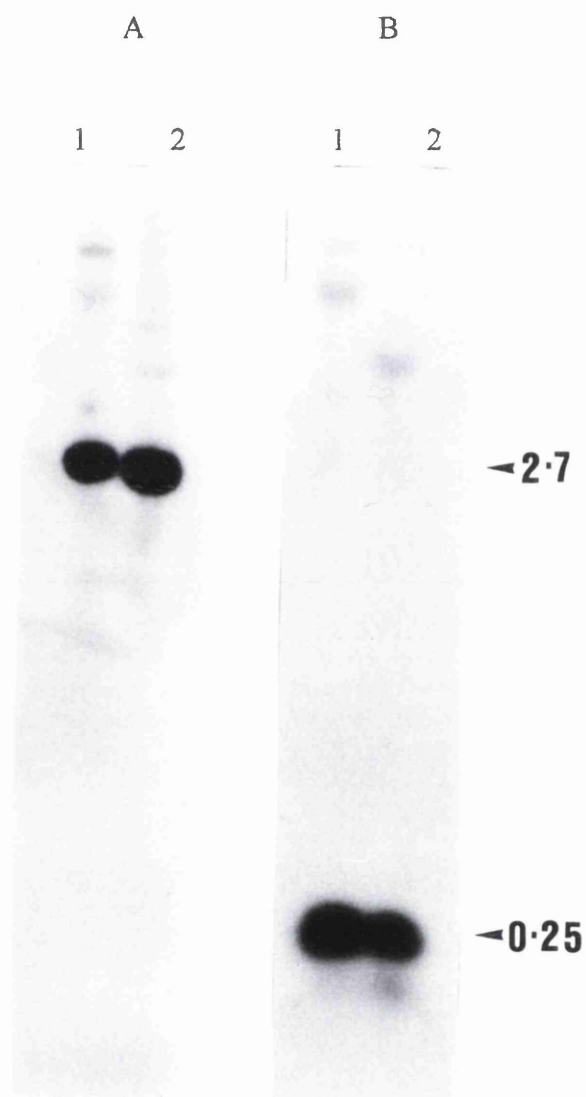


Figure 3.3. Southern blot analysis of genomic clone isolated from the *T.parva* genomic library. Following digestion with Acc I (Lane 1) and Acc I +Hind III (Lane 2) the blotted DNA was hybridized with the 5' (A) and 3' (B) *TamS1* cDNA probes (see section 2.3.1). The position of each band is indicated by an arrowhead and in kilobases.

5' and 3' sequence contigs of *TpmS1* fragment, primers were made using known sequence from the protein coding region of *TpmS1*, upstream of the known *AccI* site and downstream of the previously undefined *AccI* site within the 3' intergenic fragment. The ~250bp gap was sequenced in either direction by automated sequencing using DNA from the excised *TpmS13.2* genomic clone. The accumulation of this data allowed for the generation of 1 genomic contig of ~5kb. The DNA sequence covering this contig is given in the Appendix II of this thesis.

3.3.3 Sequence comparison of the *T.annulata* and *T.parva* contigs

The *T.parva* and *T.annulata* contigs were compared over 5kb of sequence, which included the polypeptide reading frames and the immediate flanking intergenic sequences of the *TamS1* and *TpmS1* genes (Appendix III). The sequences were found to be conserved across the two species giving 80% identity between the open reading frames and 73.3% identity between the compared intergenic regions. The most significant region of sequence diversity was observed in the region between the polypeptide coding region of *TamS1* and *TpmS1* and the unidentified ORF-1 sequence. As this region appears to consist of the 3' intergenic regions of the *TamS1/TpmS1* and the ORF-1 genes, the lack of conservation may indicate that little of this sequence has a functional role. In contrast the sequences of both sets of genes is highly conserved within 40-60 base pairs 3' to the polypeptide stop codons, the noted divergence occurring downstream of this position (see Appendix III).

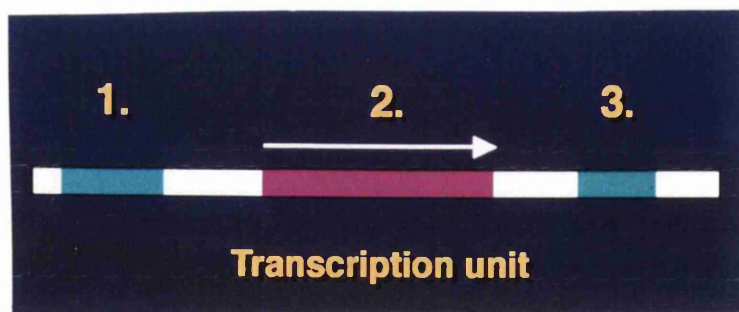
Enhancers can act from a distance, independent of orientation and either from the 5' or 3' end of a gene (Hames and Glover, 1988). Therefore, to determine if sequences which could function as enhancers for either *TamS1* or ORF-1 were present, the 3'intergenic region was examined for homology to defined enhancers of known eukaryotic transcription factors using the GCG program patternfinds. Unfortunately, no known enhancers were found to be conserved between both the *T.annulata* and *T.parva* sequences which could either be due to the region of conservation of enhancer being quite short and missed on general match up, or possibly, though unlikely, it was present in 3' untranslated RNA sequence. In addition to analysing the 3' intergenic sequences of the *TamS1* and ORF-1 genes for enhancers, analysis for signals indicative of RNA processing and polyadenylation was carried out. In most higher eukaryotes RNA processing of a mRNA transcribed by pol II requires a AAUAAA

sequence plus GU and U-rich elements downstream. Unfortunately, no such conserved elements were detected in either the *T.parva* or *T.annulata* sequences of the *TamS1* or ORF-1 genes. However, there are examples of natural Poly(A) sites which are not dependent on these sequence elements, for example β -globin gene expression in *Xenopus* oocytes (Mason *et al.*, 1986). Therefore, known Poly (A) signals may not be relevant to the polyadenylation of transcribed mRNA in *Theileria*.

3.3.4 Transcriptional regulation of the *TamS1* gene

Investigations into the expression of the *TamS1* gene by Shiels *et al.*, (1994) demonstrated that there is an up regulation of *TamS1* RNA production during differentiation to the merozoite which is maintained by the piroplasm stage of the parasite. Clues to where control motifs are likely to be located, can be made by knowing whether primary control of gene expression is at the transcriptional or post-transcriptional level. It was therefore decided to perform nuclear run on analysis to determine if altered transcriptional control contributes to *TamS1* differential expression. In addition, this assay could be used to confirm whether transcription of the *TamS1* gene was mono-cistronic or polycistronic. The *TamR1* gene encoding the 117kDa rhoptry protein, was analysed in parallel.

Parasite nuclei were isolated from 200ml of cell culture which had either been incubated at 37°C or incubated at 41°C for 6 days to induce macroschizont differentiation to merozoites (see 3.2.1). Preformed transcriptional complexes were allowed to elongate in the presence of labelled UTP. The labelled nascent RNA was purified from unincorporated [$\alpha^{32}\text{P}$] UTP by size exclusion chromatography through a NucTrap Probe purification column (Stratagene, see 3.2.1) and donated by Dr Swan. The radiolabelled nascent RNA was used as a probe for DNA fragments specific for the *TamS1* and *TamR1* gene coding regions and their respective 5' and 3' intergenic sequences (Figure 3.4). Low level hybridisation was observed to the coding regions of both the *TamS1* and *TamR1* genes by radiolabelled nascent RNA prepared from macroschizont cells (Figure 3.4A). The intensity of this signal was greater with DNA hybridised with nascent RNA prepared from day 6 differentiating macroschizont cultures and no hybridisation was observed to the intergenic regions of either gene (Figure 3.4B). This would indicate that transcription



(A)

2

TamR1



TamS1



(B)

2 1 3

TamR1



TamS1

Figure 3.4 *TamS1* and *TamR1* genes define independent transcription units. The schematic drawing reveals the organization of each gene. DNA fragments spanning each region of the genes are presented and numbered. The coding region is labeled 2, the 5' intergenic sequence, labeled 1 and the 3' intergenic region, labeled 3. The DNA fragments were hybridized with radiolabelled, nascent RNA generated by nuclear run on analysis from A) macrophage cells and B) differentiating macrophage cells incubated at 41°C for 6 days.

for both genes is monocistronic and that the 5' intergenic region of each gene must contain at least the minimal signals for transcription initiation and termination. Furthermore, the observation of an increase in nascent transcript production by differentiating cell between the promoter activity of macroschizont and differentiating macroschizont cells, plus RNA indicates that the previously detected accumulation of *TamS1* mRNA (Shiels *et al.*, 1994) must, at least in part be due to an elevation of transcription rate.

3.3.5 Mapping the Transcriptional Initiation Site of the *TamS1* Gene

Monocistronic transcription units usually produce a single transcript encoding a single type of polypeptide. In addition, each unit is governed by initiation and termination signal which determine the length of the individual transcript. Transcription by eukaryotic RNA polymerases I and II are dependent, to a major extent, on nucleotide sequences located within about 100bp surrounding the 5' end of their respective transcription RNA initiation sites and often on additional sequences which can be located up to several kilobase pairs upstream from this position. Therefore, mapping of the RNA initiation site of the *TamS1* gene could aid in a screen for motifs which may bind transcription factors and control production of the 30kDa polypeptide.

The RNA initiation site of the *TamS1* gene was mapped using the 5' Race assay (Gibco-BRL). In this method the first strand cDNA is synthesised from mRNA and tailed with Poly dC residues. The cDNA is then amplified by PCR using a gene specific primer and a Poly dG (anchor) primer and the amplified product sub-cloned and sequenced. Two primers were designed which were located downstream of the ATG (methionine) start codon. The first primer, GSP 1, maps to a position 3526 to 3497 (see Appendix I), 418 nucleotides downstream of the ATG while the second primer, GSP 2 is located at 3289 to 3260 (see Appendix I), 237 upstream of GSP1. mRNA was isolated from piroplasms and reversed transcribed using the GSP1 oligonucleotide as a primer. The product was treated with RNase H to remove the mRNA template and the single stranded cDNA was purified before being tailed using terminal deoxynucleotide transferase. The product was subsequently amplified by PCR using the anchor primer (provided by the kit) and the GSP 2 primer. The resulting PCR reaction was analysed on an agarose gel, but unfortunately no visible bands

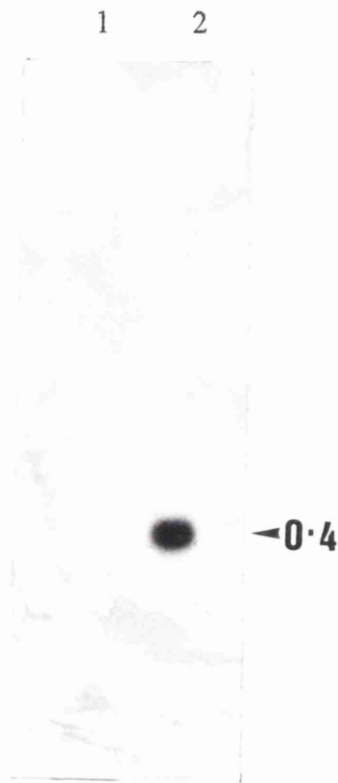


Figure 3.5 Southern blot analysis of 5' RACE products derived from *TamSI* mRNA. The position of the band is indicated by an arrowhead and in kilobases. Lane 1) cDNA template, Lane 2) cDNA template amplified by one round of PCR.

were detected (data not shown). The agarose gel was Southern blotted and probed with the *TamS1* cDNA, and the resulting autoradiograph is shown in Figure 3.5. The *TamS1* cDNA insert hybridised to a band of approximately 400bp.

To amplify enough product to allow sub-cloning, 10µl of the PCR reaction was used in a secondary round of PCR. The 400bp band, now visible on an agarose gel, was purified by the QIAquick gel extraction method and cloned into the pGem T vector. DNA minipreps were prepared using the alkali lysis method and clones containing insert were identified by agarose gel electrophoresis. Two clones containing the PCR amplified insert were identified and sequenced. The beginning of the PCR sequence for both *TamS1* clones was mapped to an adenosine residue 117bp 5' of the ATG start codon which is highlighted by an arrow in Figure 3.6. The sequence of the nucleotides flanking this position, CTCACCTTTC, is in accordance with the transcription initiation site consensus sequence, Py Py C A Py Py Py Py Py (Corden *et al*, 1980) and it was concluded that this was likely to be the major transcription initiation position of the *TamS1* gene.

3.3.6 Comparative Sequence Analysis of the 5' Sequence of Major Merozoite Surface Antigen Genes

As functional DNA protein binding domains can be conserved across a wide range of species, it was decided to clone and sequence the 5' intergenic region of the *T.hirci* and *T.sergenti* and compare these sequences with those already obtained for *TamS1* and *TpmS1*. An initial analysis between the *TpmS1* and *TamS1* upstream sequences revealed that the most extensive homology occurred over the first 500-600bp upstream of the transcriptional start site. Therefore, it was decided, if possible, to PCR amplify this region using genomic DNA from *T.sergenti* and *T.hirci*. Primers 2735 and Tams 1DB were used to amplify a 771bp fragment from *T.hirci*, and primers CAT4a and Tams 1DB were used to amplify a ~320bp fragment from *T.sergenti* DNA. The PCR products were cloned into pGem-T and sequenced. The sequence comparisons between *TamS1*, *TpmS1*, *ThmS1*, over the first 608bp upstream of the ATG codon are shown in Figure 3.6.

The overall level of identity between *TamS1* and the *TpmS1* and *ThmS1* fragments was high, making it difficult to identify small conserved DNA motifs. However, as

suggested for *Plasmodium*, it is possible that control elements will contain a higher content of G+C nucleotides when compared to the overall intergenic region (Lanzer *et al.*, 1992a, b). Interestingly, several regions which show conservation across the species had a higher G+C content (see boxed regions in Figure 3.6) and this was most evident in the conserved upstream region (Figure 3.6; Box A) proximal to the transcription start site. More specifically a number of palindromic motifs containing G or C residues were found within the conserved boxes of sequence. For example, a heptamer TCACACT and the related pentamer ACACA were located within boxes A and J (Figure 3.6), while other related palindromes AGTGA and TGTGT were located to boxes F and D (Figure 3.6). All of these sequences are G rich and conserved between the three sequences making them candidate sequences for motifs involved in the transcriptional regulation of the major merozoite antigen.

During the course of this study sequence information was published showing that *T.parva* and *T.annulata* were more closely related to one another than to the *T.sergenti/buffeli* group. This may explain why a PCR product was only obtained with one of the primer combinations used and this only generated 178bp of sequence data for comparative analysis. In general this sequence showed extensive divergence from the other sequences. However, a 10mer motif (Figure 3.7, Box 2; TTTTAATATTT) 54 bases upstream of the ATG polypeptide initiation codon showed an exact match over all 4 sequences. The mapping of the transcriptional start site of *TamSI* places this motif within the untranslated region of mRNA, confirmed by comparison with available cDNA sequence from *T.sergenti* (Kawazu *et al.*, 1992). A second region of homology is also represented in Figure 3.7. However, this sequence represents the primer used in the PCR amplification and the homology is based on the presumed annealing of the primer at this position.

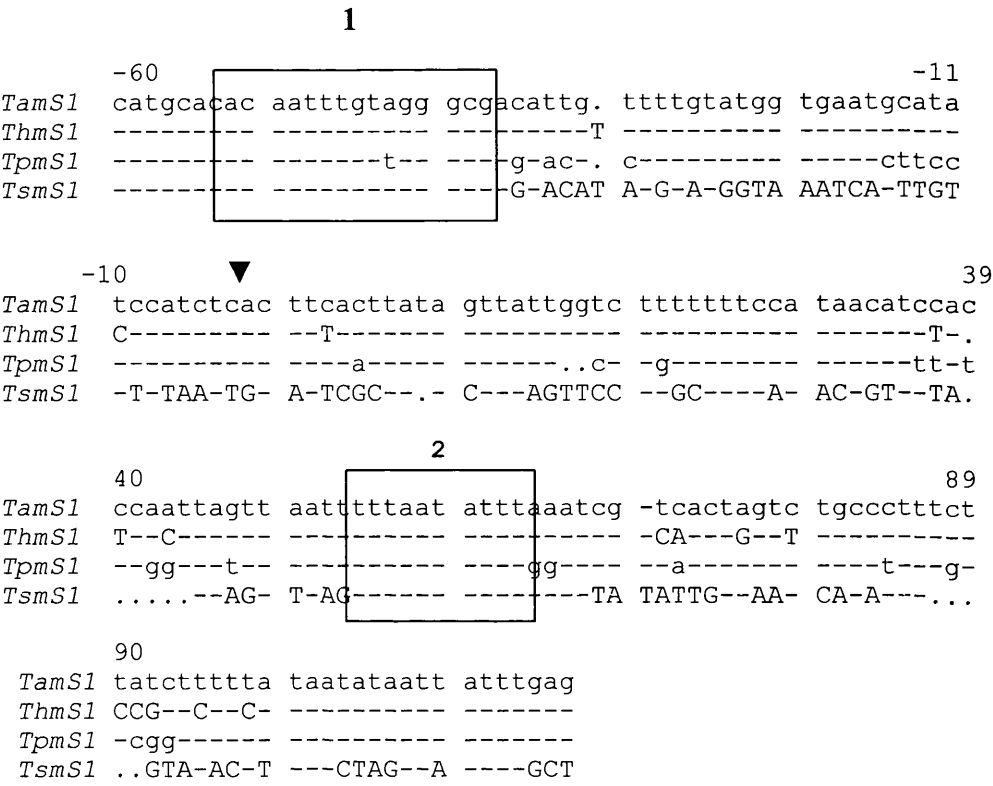


Figure 3.7 Comparison of *TamS1* like genes isolated from different *Theileria* species over the first 178bp of 5' region. Numbers represent the sequence position with reference to the transcription initiation site (marked by an arrowhead and is highlighted in bold). All alignments are made to the *TamS1* sequence and dash represents identity, dot represents a gap, and any substitutions are shown by the relevant letter. The number boxes surround regions of high homology.

3.3.7 Putative Promoter binding Sites in the 5' Untranslated Region of *TamS1*

For eukaryotic RNA polymerase II to transcribe a gene, an array of proteins must be assembled at the promoter region (Buratowski *et al.*, 1994), and a number of these transcription factors must specifically recognise a promoter sequence motif. With the hope of identifying putative DNA binding sites of proteins which may be involved in the regulation of *TamS1* expression, the sequence 5' to the RNA initiation site was examined for homology to defined binding sites of known eukaryotic transcription factors. In this analysis only motifs conserved across the sequences from *T.parva*, *T.annulata*, and *T.hirci* were considered and those sequences identified are shown in Figure 3.8

Seven binding sites of known eukaryotic transcription factors had homology to the 5' region of *TamS1*. Of these, 3 binding sites were known to act as putative TATA boxes and are capable of binding the transcription factor TFIID (the TATA-binding protein (TBP)) directly, which is the first step in the assembly of the initiation complex in higher eukaryotes (Buratowski, 1994). Two of the binding sites consisted of the classical TATA box motif, TAT t/a A t/a and the third motif, GATAAA (betaP-F1), has been shown to be a putative TATA box for human Factor VIII, a glycoprotein essential for blood coagulation (Figuiredo and Brownlee, 1995). Interestingly, only the GATAAA binding site was positioned upstream of the transcription start site. The 2 traditionally recognised TATA boxes were situated between the transcription initiation site and the ATG start codon, see Figure 3.8.

The c-mos_DS1 binding site, TGGTTTG, Figure 3.8, has been shown to control the cell type-specific expression of the c-mos proto-oncogene. Analysis of this motif has demonstrated that it acts as a repressor in an orientation-independent manner over large distances (van der Hoorn, 1987). Therefore, the role of this motif in controlling the expression of the *TamS1* gene could, if functional, relate to stage specificity, preventing expression in those stages of the Life-Cycle where the *TamS1* polypeptide is not required.

Interestingly 3 different elements, MRE, GR-MT-IIA and IBP-1_CS2 are inducible enhancers. The MRE element is induced by heavy metals (Labbe *et al.*, 1991), GR-MT-IIA is induced by the many factors which affect the expression of the glucocorticoid hormone

(Ram *et al.*, 1995) and the IBP-1_CS2 elements can be induced by phorbol ester (Cohen *et al.*, 1991). It would seem unlikely that the *TamS1* gene is controlled by a classical inducible enhancer. The *TamS1* gene product is a surface antigen and appears to be constitutively expressed by merozoites and piroplasms. However, the possibility that gene expression is upregulated during differentiation by the production/activation of factors which bind to enhancer type motifs is possible.

Any of the 7 known transcription factor binding elements in Figure 3.8 could play a role in *TamS1* gene expression. However, without either a functional assay or a method for identifying DNA-protein interactions it was impossible to determine whether these motifs were identified by chance or because they play a role in regulating gene expression, and have been conserved across *Theileria* species.

3.3.8 Comparative Sequence Analysis of the upstream sequences of the *TamS1*, *TamR1*, Ta-ORF-1 and *SPAG-1* genes.

To test whether any of the promoter like motifs identified in the upstream region of the *TamS1* gene could be involved in the general transcriptional process, upstream regions of the *TamR1* (McDonald, unpublished results), Ta-ORF-1 and *SPAG-1* (Katzer, 1995) genes were compared with the 5' intergenic region of the *TamS1* gene. Using the GCG program pileup, the sequences were compared over the first 1kb of intergenic sequence (Figure 3.9) the postulation being that general motifs would be shared between the different genes and located to a similar position relative to the transcriptional start site.

The overall identity between the 4 sequences was low, however, it was possible to identify small regions of conserved sequence. These could range from a few bases in length, to pentamers and even a 16mer. More specifically a number of palindromic motifs, previously noted in section 3.3.6, were found to be present in the upstream sequences of all four *T.annulata* genes, providing further evidence for a possible role in transcription regulation (see Figure 3.9, sequences underlined and labelled, I, II and III).

```

(3)
-482 ATTAAACTCAAAGACATCACACTCTGATGAAAGTGAAGGCGGGTAAGGAC
-432 AGCTCATAACGACAAATTCAGTATAAAAACATAATGAATGATTGTTTTAA
-382 GGAAATGATGTGTGGAATCACGGTTATTTGTGCGGGAAATTATGCATGGA
-332 TATTATCTATTTAACTTATAATAATGTTATATTTAAATATGACAAATTAA
(2)
-282 TGTGTGAAATAAATAAAGTAAAGTGGAGGGAAATTCTGTCCTAATATTTA
(4)
-232 ACTGGGTCTAGTGACTTAAAAAAGAAAGCAGGTTATTGGTTTGAAAAAAT
(5)
-182 AATAAATTCATCTTTTTTATAGTCATCAAAAAATGAAAAATGGATAAATAT
-132 ACATGTTATTTTCATTAAGATGGACTTCAATGGAGGATAAGGCATTAGGCG
(5) (1)
-82 TGATGAGTGCATAGATACAGATAAACATGCACACAATTTGTAGGGCGACA
-32 TTGTTTTGTATGGTGAATGCATATCCATCTCACTTCACTTATAGTTATTG
18 GTCTTTTTTTCCATAACATCCACCCAATTAGTTAATTTTAATATTTAAA
68 TCGCTCACTAGTCTGCCCTTTCTTATCTTTTATAAATAATAATTATTTGAG
118 ATGTTGTCCAGGACCACCCTCAAGTTCTTATATTTGAGCTTCTTCGTTAT

```

Figure 3.8 Factor binding sequence motifs identified in the 5' untranslated region of *TamS1* by GCG patternfind program. Numbers represent the sequence position with reference to transcription initiation site, marked by an arrowhead and highlighted in bold. 5 transcription factor consensus sequences have been underlined: (1) MREe (Labbe *et al.*, 1991); (2) GR-MT-IIA (Ram *et al.*, 1995); (3) IBP-1_CS2 (Cohen *et al.*, 1991); (4) c-mos_DS1 (van der Hoorn, 1987); and (5) betaP-F1 (Figueirido and Brownlee, 1995), the two classical TATA box motifs are boxed, as is the ATG start codon.

Larger regions of sequence homology were detected between the intergenic sequences. For example, two regions of homology which are A+T rich were located within boxes A, and highlighted in bold and italics. The homology between the *TamR1* and ORF-1 sequences only differ by one base (see Figure 3.9; boxes A) while the *TamS1* and *TamR1* sequences (Figure 3.9; highlighted in bold and italics) are the reverse of each other and a significant distance apart. Both of these sequences are located over 100 bases from the RNA initiation site of each gene but neither of motifs are conserved between all of the intergenic regions analysed. Therefore, although A+T rich, it is doubtful that these motifs are acting as an alternative TATA box, involving the recognition of the DNA by a TFIID or equivalent factor prior to binding by the RNA polymerase. An alternative possibility is that the conservation of the A type sequences within the intergenic regions of genes expressed in the merozoite stage, is an indication that they maybe required for regulating expression during this part of the parasite Life-Cycle.

Three of the seven known promoter elements shown to have homology with the 5' region of *TamS1* were detected in the intergenic regions of the *TamR1*, *SPAG* and ORF-1 genes. However, the betaP-F1 element was only detected in the *TamR1* 5' region and as its position is over 700 bases from the RNA initiation site it is unlikely that this element is used as a TFIID binding site. The inducible elements GR-MT-IIA and IBP-1_CS2 are located in the *TamR1* and *SPAG-1* sequences respectively. The GR-MT-IIA element was found within the *TamR1* sequence while the IBP-1_CS2 motif was detected in the upstream region of *SPAG-1* while it is possible that, as mentioned previously these elements could function in regulation of stage specific gene expression, the presence of the IBP-1_CS2 in *SPAG-1* makes this unlikely for this element.

For many promoters, control elements have been identified to be within the first 100bp upstream of the transcriptional start site. Therefore, it was decided to compare the first 100 bases upstream of the initiation site for the *TamS1*, *TamR1* and *SPAG-1* genes in isolation from the rest of the intergenic sequence (See Figure 3.10). As for the *TamS1* intergenic sequence (see section 3.3.7) no putative TATA box was detected immediately upstream of the transcription start sites for either the *TamR1* or *SPAG1* genes. Interestingly several regions showed conservation across the intergenic regions of the genes, specifically

to a CTTTTT motif. This motif was repeated several times within the *TamR1* sequence and a further search of the 1kb intergenic region, in addition to the *SPAG-1* sequence, was also shown to be present in the upstream region of the *TamS1* gene. Indeed the only motif that could be considered to be conserved was the CTTTTT sequence. This sequence was repeated in *TamR1* and was present in *SPAG-1*. Although the sequence was not detected within the first 100bp of *TamS1* 5' sequence, it was shown to be present further upstream. Similar, but incomplete motifs were detected in the ORF-1 sequence.

Figure 3.9 Comparison of the intergenic regions of the *TamS1*, *TamR1*, *SPAG-1* and ORF over the first 1kb 5' to the ATG start codon. The numbers represent the sequence position with reference to the ATG start codon of each gene. The transcriptional start site are marked with labelled arrowheads: *TamR1*, mR; *TamS1*, mS; and *SPAG-1*, SP. Sequence which show a match with higher eukaryotic motifs (see 3.3.7) are labelled (1), IBP-1_CS2 (2) and GR-MT-IIA (3). Conserved palindromes and pentamers are labelled I) CACA, II) ACACA, III) AGTGA, and IV) CTTTTT. The sequences which are boxed or in bold italic were identified as conserved sequences and their significance is discussed in the text.

F-1

ion

ked

ich

and

A.

ere

-1022

-973

TamR1	GGCCATGAAG	GTAAATTAAA	AGGTTATAAC	TGAGCATAGG	TCCGCCGGCG
SPAG1	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
TamS1	~~~~~	~~CCAAATTA	AGCAAAGACG	TCAGAGTAAT	GTTCGAAACA
ORF-1	~~~~~	~GCCTAAATT	GGTAAACTGT	TACCCTTAAA	TCCAAAAACT

-972

(3)

-923

TamR1	TTTGTGGACT	TTAATCCCAG	AAAAGCAATG	ATATCTTCAG	CTTGTATAAA
SPAG1	~~~~~	~~~~~	~~AATCAATG	TTATTTTTTAC	TAAAATTTGT
TamS1	ACAAAGCATG	TGTCCTGTTT	AACTCAAATT	AGTC ATTTTA	AATAAATTAA
ORF-1	TTGTTGTTTG	GGTATCTTCT	GGTCTAAGGC	TAATTCCCTA	CCAAATCCAA

-922

II

-873

TamR1	AGTGGCATGG	TGGATACCGA	ACATGAA...	AACTGAAACT	CAAGAATAAA
SPAG1	AACATTTAAT	CTCATACCTA	ATATGAACTT	CATGTAAACA	TAGGATTCAC
TamS1	GGGGTAAAAT	CAAATACACA	TCAAAAGTAT	AAATATATAA	AATACCTTAG
ORF-1	CTCATAGCAT	TTTTAAACCA	AACTTCCAA	CTTTTCGTAT	TTAACATAT

-872

-823

TamR1	ATGTTTTTTG	ATTTTTATTG	TATTTAGAAG	CTCTCGGAAT	..AATAATAA
SPAG1	TTCAATAATT	ACAGTTATAG	AATGAATAAT	TAACCAAAAT	TGATTTCTAT
TamS1	AATTAAAAAG	ATCTCTCTAA	AAGAATCAAT	GCAAGGATTC	TCTTGGGTAG
ORF-1	ACGATTCATA	ATTTTGCCTT	AACCTGGGAT	TTTATTTAAG	GGTATATTGG

-822

I

-773

TamR1	CAGTTAATTG	GCATTAAAAT	TTTGTTTTAT	TTCCACAATG	ATAATATAAT
SPAG1	AAAATAAATA	TTTAACTATT	TAGATTAGAT	TTTCGAATTT	ATGGTTGTGT
TamS1	GGATGAGAGG	CACCGTTGCC	TATACGACAG	ATATAGAGTG	TTAGGATCGA
ORF-1	TTGTTTTGTA	TATTTTCACC	AAGTTTGCCT	GATCAAACCA	CTCATTTATA

-772 I (1)

-723

TamR1	CTGATAAACA	TATTATATTC	TACTTTTAAT	ATCTTCTAGA	AATATA.AAA
SPAG1	TCACAAAAGG	CATATCATAA	CATTTTTTCT	AACGGCGAAA	AAATTA.TTA
TamS1	CAAGCACAGT	AGTTTAATTA	CATATTACAA	TCTGGATAGA	TGTATAGAAA
ORF-1	AAAATGGATA	AAAATTGGAA	TAATGAAACC	AATTCTGAGG	AAAATTTGTC

-722

-673

TamR1	TACAAGTTCA	ACATGGATTA	ATGGCATTAA	ATAGTGTACT	ATGATTAAA.
SPAG1	AAAAATTATA	AAACTAAAGT	TTTTAATGAA	ATATATGAAT	AATATGTAAT
TamS1	TAAAAAGAGC	ATACACACCT	CTATCTGTAA	CTCTCTGGAG	AATAAATAGG
ORF-1	TAATATGAGC	.CTCGAATCC	CTATCACTGG	GACAAAGGCT	ATTATCTAAG

(A)

-672

623

TamR1	.ATGGAA ATAT	TTTAAAAATAA	TTAGACAAGA	TA ATAATAAA	AAAATTAATT
SPAG1	TAAAGTATAA	TTATGGTTAA	TTAGACTTTT	AGATTCAAAT	TTATTAAATT
TamS1	GTGTACTATT	TGCTGAAGAA	AATTTTCAAA	ATCCGGGGCCG	GTAATATCTG
ORF-1	TGCTTAAAAA	AGACCGGTAG	GCAAATAAAC	AT ATAAATAA	AAATTAAT AT

(2)

-622

II+III-573

TamR1	ATTTCAAATA	AACAGTATAA	ATTTTTAATA	TTTGTCAATG	GAAGATAGTT
SPAG1	GTGTGTAATC	CTATACTTAA	AAATCTCAT.	TTTTTGAATA	ATTCAAATTC
TamS1	AAGGAGACTG	ATTAAACTCA	AAGACATCAC	ACTCTGATGA	AAGTGAAGGC
ORF-1	TTTATTCAGG	CATAAGTGTA	AGCGTACGAG	GTATAGCAGT	TAACACAACC

-572

III

-523

TamR1	AAAAAGGGAA	TTTATATTTT	GTAGTGAAAT	GTGTAAATCA	ACAATGTCCA
SPAG1	TACATTTAAA	TTCATATTAT	ATATGTTAAT	TTAGATAT.A	TAAATGATGA
TamS1	GGGTAAAGGAC	AGCTCATAAC	GACAAATTCA	GTATAAAAC	ATAATGAATG
ORF-1	AGCCAAATTA	AGACAGCGGG	AAGCGCTTGG	AAATCAGAAG	TAAAGTACGA

	-522	(2)				-473
TamR1	AATATTGTCT	AAACGGCCAT	TATTTACTTA	TTAAATTATT	TCAACAATTT	
SPAG1	GTTTAAGTGA	AAATTGAGGG	GAATTTAATA	TGCGATAGAT	TAATTGTTTG	
TamS1	ATTGTTTTAA	GGAAATGATG	TGTGGAATCA	CGGTTATTTG	TGCGGGAAAT	
ORF-1	ATCAAACGTC	GAAAGCAAAA	CGTAAGCTAA	CTAATTAAAT	AAAAATATGT	
	-472					-423
TamR1	CAAGATTTTT	TATTCATTCA	ATTTAACGGC	AACATTCCTT	ACACATCTCA	
SPAG1	GCTAGCAGAA	TGAGCTCAAA	TAAAGAGCTA	GCCAATGTCA	TTAAAGCTCC	
TamS1	TATGCATGGA	TATTATCTAT	TTAACTTATA	ATAATGTTAT	ATTAAAATAT	
ORF-1	AGCGGAACAC	TGGACGATGT	TTCATTCCCA	GATTTACTGG	GCTCGAGTAA	
	-422					-373
TamR1	GATTTAGACA	ACCTTCTATT	CGTGATTTAC	AT.TAGATGT	GTAGTAGGGT	
SPAG1	AATAAACCCA	AGATTAGACT	CGTTCACTAG	ATATACTAAT	AAAATCCATC	
TamS1	GACAAATTAA	TGTGTGAAAT	AAATAAAGTA	AAGTGGAGGG	AAATTCTGTC	
ORF-1	CCGAGATGTT	AGTGTAATA	ATCGGATCTC	GGAAACCTCA	AGCGTCACCA	
	-372		II	I		-323
TamR1	GCATTGTATT	ATAGGAAATC	CTTTACACAT	TCAAATCACG	CGTGAGCCCT	
SPAG1	ATTATTTTTG	TGCATTTATC	ATCGAGTTAT	TTTCACAAAA	AATTATACTA	
TamS1	CTAATATTTA	ACTGGGTCTA	GTGACTTAAA	AAAGAAAGCA	GGTTATTGGT	
ORF-1	ACGACCAAGA	TCGCCTCTCC	AAGGAACCTC	GAAGTACCAG	ATACAAGTTC	
	-322	II	IV	(3)		-273
TamR1	AGTTAGACAT	TGTTCATAAT	TTGTGCACCA	GGGTGTCCTC	AATTGGATAT	
SPAG1	ACACACA ACT	GTATAAGACC	GTGTTTAGTC	TTTTTCCTTA	AATTTGCTAC	
TamS1	TTGAAAAAAT	AATAAATTCA	TCTTTTTTATA	GTCATCAAAA	AATGAAAAA.	
ORF-1	AAACATAGAA	TTTCAAGCGT	TAAATCAGTT	GCCAGCCCAG	GATTCCATAG	
	-272	II			(SP)↓	-223
TamR1	GTGT.TAGAC	AAAAACCAAG	AGTGTGCAGG	AATACTTCCC	CATACCTTAA	
SPAG1	CTATATCTGT	AAAACACAAG	CTTATGTTTT	ACAAATTACT	ATAAAAAACA	
TamS1TGGAT	AAATATACAT	GTTATTTTCAT	TAAGATGGAC	TTCAATGGAG	
ORF-1	TAGTCGTGAC	AGCAGCTACT	CCA ACT CGTT	ATCTTTCAGA	AATTCCAACA	
	-222			(1)	I-173	
TamR1	GTTATTTAGA	CAATCATTTT	ATCTTTATTT	TATTTATATA	TTGTTAGCCA	
SPAG1	AATAAAACAA	AAACACACTC	GTTTTGAGAT	AATTTTCTTA	ATAACAAGTG	
TamS1	GATAAGGCAT	TAGGCGTGAT	GAGTGCATAG	ATACAGATAA	ACATGCACAC	
ORF-1	GCGATGACAC	TGATAATATC	CTGGGGTACC	GAGTACCGAC	TTCCTGATGC	
	-172		IV	IV	(mS)↓	123
TamR1	TATTGAAATG	AGGCTAATTT	ACTTTTTTGTT	CTTTTTATTT	GGTCTTTTTT	
SPAG1	TTTTAAAAAG	GTAGGACTTT	ATCACCTAAA	AGCAGAATAG	TCTCAAATGC	
TamS1	AATTTGTAGG	GCGACATTGT	TTTGTATGGT	GAATGCATAT	CCATCTCACT	
ORF-1	CCTGAAATGA	AGGGCAACTA	CCCCTAGGG	TTTACCAACT	TCATCGGAAG	
	-122					-73
TamR1	CCAATTCATT	TGTTGGAGCA	GACTCTGTTT	TAGACTTAGA	CTTTTACGAC	
SPAG1	ATTGAGATTA	GAGGCTCCCT	GATAATTGAC	TAAAAATGTT	ATATTACACA	
TamS1	TCACTTATAG	TTATTGGTCT	TTTTTTCCAT	AACATCCACC	CAATTAGTTA	
ORF-1	AGGTGTAGAG	TCACCACCTC	CACCTCCACC	TTCAGTGCCG	TACCCAGCCA	
	-72	(mR)↓				-23
TamR1	TCTGAAAAGG	TGTATCATTC	TTCATCACTA	CAAAATAATT	ACTCAGTTGA	
SPAG1	GCTTTTTTATT	CGATTAATTT	ATAACAATTA	TTTATAAATA	TTACTTATCG	
TamS1	ATTTTTTAATA	TTTAAATCGC	TCACTAGTCT	GCCCTTTCTT	ATCTTTTTTAT	
ORF-1	ACCTAACGGG	ATCTACCAAG	GTTTTTCAGCA	GTATGCAACA	AGGTTGAAGA	

	-22		1
TamR1	AGTTTTCCA~	~~~~~	~~
SPAG1	AGATAGTTTC	TTTAAA~~~~	~~
TamS1	AATATAATTA	TTTGAGAT~~	~~
CRF-1	GTAATTCGAG	ACCTCAAATG	TT

Figure 3.18 Comparison of the intergenic regions of the *TamS1*, *TamR1* and *SPAG1* genes over the first 100bp 5' to the transcription initiation site of each gene. The nucleotide sequence position with reference to the transcription initiation site of each gene. The conserved CTTCT motif is underlined.

	-100				-51
<i>TamR1</i>	A <u>CTTTTT</u> TGTT	<u>CTTTTT</u> TATTT	GGT <u>CTTTTT</u>	CCAATTCATT	TGTTGGAGCA
<i>TamS1</i>	AGGATAAGGC	ATTAGGCGTG	ATGAGTGCAT	AGATACAGAT	AAACATGCAC
<i>SPAG1</i>	AACACACAAC	TGTATAAGAC	CGTGTTTAGT	<u>CTTTTTC</u> CTT	AAATTTGCTA
	-50				-1
<i>TamR1</i>	GACTCTGTTT	TAGACTTAGA	CTTTTACGAC	TCTGAAAAGG	TGTATCATTC
<i>TamS1</i>	ACAATTTGTA	GGGCGACATT	GTTTTGTATG	GTGAATGCAT	ATCCATCTCA
<i>SPAG1</i>	CCTATATCTG	TAAAACACAA	GCTTATGTTT	TACAAATTAC	TATAAAAAAC

Figure 3.10 Comparison of the intergenic regions of the *TamS1*, *TamR1* and *SPAG-1* genes over the first 100bp 5' to the transcription initiation site of each gene. The numbers represent the sequence position with reference to the transcription initiation site of each gene. The conserved CTTTTT motif is underlined.

3.4 Discussion

The protozoan parasite *T.annulata* has a complex Life-Cycle alternating between a vertebrate and an invertebrate host. During the Life-Cycle changes in gene expression control are manifest as alterations to the pattern of RNA and protein accumulation. Understanding how these alterations come about could provide insight into what determines how and when the parasite switches from one stage to another. It was the aim of this chapter to initiate investigation into how the *TamS1* gene is regulated during differentiation and delineate possible nucleic acid motifs that may be associated with this process.

Transcription across the *TamS1* and *TamR1* intergenic regions is discontinuous as demonstrated by nuclear run on assay. Using radiolabelled nascent RNA as a probe for DNA fragments representing coding, 5' and 3' intergenic regions of the *TamS1* and *TamR1* genes, no hybridisation was observed to the intergenic regions, whereas the coding sequences hybridised to nascent RNA. This indicates that transcription is monocistronic for both genes and it was concluded that the intergenic regions contain the minimal necessary signals for transcription initiation and mRNA termination/poly A addition. In addition, comparison of transcription activity from nuclei isolated from infected cells cultured at 37°C with nuclei isolated from D7 infected cells differentiating at 41°C indicates that the *TamS1* and *TamR1* genes are transcriptionally regulated during the merozoite stage (Figure 3.4). Eukaryotic genes transcribed monocistronically by RNA polymerase II frequently contain transcription signals upstream of the RNA initiation site and a termination signal 3' to the coding sequence. As it appears that certain *T.annulata* genes are transcribed in a discontinuous manner it is possible that a polymerase similar in properties to the higher eukaryotic RNA polymerase II carries out this function in *T.annulata*. This would be confirmed if transcription were found to be sensitive to α -amanitin, as this enzyme is known to inhibit RNA polymerase II in a wide range of organisms (Sentenac, 1985).

A single RNA initiation site was observed for the *TamS1* gene. This was also shown for the Hsp 70.1 (Mason *et al.*, 1989), *SPAG-1* (Katzer, 1995) and *TamR1* genes (McDonald, unpublished data). The transcription start site was mapped to 118 bases from the ATG start codon of the 30kDa polypeptide coding sequence, and is in accordance with the transcription initiation site consensus sequence Py Py CA Py Py Py Py Py (Corden *et*

al., 1980) for the *T.annulata*, *parva* and *hirsi* sequences (Figure 3.6). The *T.sergenti* sequence was divergent over the transcription initiation site from that of the three related sequences described above. This could be an indication that RNA initiation of the *TsmS1* molecule differs in position from the other related genes, a possibility which is supported by the larger size of the coding region of the *TsmS1* gene.

At present the identification of transcriptional signals of *Theileria* is limited to a structural analysis of the sequence upstream of the RNA initiation site. Sequence analysis of the *TamS1* intergenic region has revealed several features characteristic of eukaryotic promoter regions. For example, 7 different known eukaryotic transcription factor binding DNA sequence motifs were identified, conserved across the species analysed. Two TATA boxes were identified but were found to be located before the RNA initiation site. It is, therefore, unlikely that they regulate the expression of the *TamS1* gene in a similar manner to the TATA motifs of other eukaryotes. Alternatively, the classical description of a higher eukaryotic TATA box may not be functional in *Theileria*. A possible non-classical TATA box was found 57 base pairs upstream of the transcriptional start site. The GATAAA motif has been shown to be a putative TATA box for the human Factor VIII (Figuiredo and Brownlee, 1995) and could therefore be involved in the expression of the *TamS1* gene. It is also possible that, the TATA-binding protein (TBP) in *Theileria* may not recognise any known TATA box sequences. The recent characterisation of the TBP protein in a related parasite, *Plasmodium falciparum* (McAndrew *et al.*, 1993), suggested that this molecule was likely to perform the same general function as the TBPs from higher eukaryotes but would require alterations in its interaction with DNA. Therefore, if the TBP protein of *Theileria* was altered with respect to its DNA recognition sequence it was impossible to determine this without functional binding studies with recombinant *Theileria* TBP.

The high A+T content of the *Plasmodium* and *Theileria* parasites may make it necessary for an alternative sequence to have the same function as the TATA box of other eukaryotes. Soldati *et al.*, (1994) proposed for *Toxoplasma*, however, that certain genes have "TATA-less" promoters. Thus, in addition to the recognition of an alternative motif by a *Theileria* TBP, it is also possible that for some genes, including *TamS1*, positioning of the RNA polymerase is determined by a unique motif and/or polypeptide factor. These motifs

could be in a different position or even be relatively independent of the transcriptional initiation site.

It was hoped that the search for potential elements within the *TamS1* 5' intergenic region would be simplified by cloning and sequencing the homologous intergenic regions from *T.parva*, *T.hirci*, and *T.sergenti*. A sequence comparison between the first ~600 bases of the 5' untranslated region of the *TamS1*, *ThmS1*, *TpmS1* genes revealed that the 5' intergenic sequence shows high conservation between these species. Thus, motifs of the expected size (10-20bp) were difficult to identify because of the extent of homology. Given the high A+T content of the *Theileria* promoter it could be predicted that the chance of G+C rich sequences being conserved, even across these quite closely related species, is more remote and may point to a functional role for such motifs. Interestingly the G+C content of conserved domains (Figure 3.6; Boxes A, D, F and J) was higher than the intergenic region overall, and was highest in the conserved region proximal to the transcription initiation site. Sequence palindromes, pentamers, and heptamers were also located in the conserved boxes. In the absence of functional data it can only be speculated that the conserved GC rich sequences play a role in the regulation of the *TamS1* gene. However, the AGTGA pentamer is part of the IBP-1_CS2 motif indicating that the motif could have functionality. The IBP-1_CS2 motif is an inducible enhancer which can be induced by phorbol ester (Cohen *et al.*, 1991). Expression of the *TamS1* gene is specific to certain Life-Cycle stages therefore, it is unlikely that its expression is controlled by an inducible enhancer. Rather expression is more likely to be controlled by a cell type specific enhancer. Therefore, if this motif is involved in *TamS1* expression it is likely that it interacts with regulatory factors in a different manner from that described previously.

The MREe motif TGCACACA is conserved in the *TamS1* upstream region, and indicates that the conserved ACACA motifs could have a functional role. The MREe motif functions as an enhancer which is induced by heavy metals (Labbe *et al.*, 1991). It is, however, unlikely that the ACACA motif acts as an enhancer stimulated by heavy metals but does provides evidence that the described pentamer like motifs can be involved in binding of regulatory factors. Also the complementary pentamer TGTGT has been found in a number of stage regulated genes in *Toxoplasma gondii* (Mercier *et al.*, 1996) and

resembles an element (T/A GTGTAC) reported in the upstream region of the *Plasmodium* GBP130 gene (Lanzer *et al.*, 1992a). Therefore, it may be that these types of sequences play a general role in the control of apicomplexan gene expression, which awaits further definition. In addition to the MREe and IBP-1_CS2 motifs, two other enhancers were identified c-mos_DS1, and GR-MT_IIA. The c-mos_DS1 functions as an enhancer controlling cell type-specific expression of the c-mos proto-oncogene (van der Hoorn, 1987) and GR-MT_IIA functions as a binding element for the glucocorticoid hormone receptor (Ram *et al.*, 1995).

The comparison of the 5' untranslated region of the *TamS1* gene with the related region in *T.sergenti* demonstrated a greater diversity than that observed with the *T.parva* and *T.hirci* data. Phylogenetic analysis of the major merozoite surface antigens has shown that the *T.parva* sequence has a greater similarity to the *T.annulata* sequence than *T.sergenti* (Shiels *et al.*, 1995). Therefore, as expected the greatest amount of diversity between 5' intergenic sequences is observed with species furthest apart on the phylogenetic tree. It is unfortunate, therefore, that more sequence data from *T.sergenti* could not be obtained. Interestingly within the *T.sergenti* sequence there was one region of homology consisting of a 10mer motif 54 bases upstream from the ATG start codon of the *TamS1* gene. The 10mer is located prior to the transcriptional start site, is highly conserved amongst all four species of *Theileria* analysed, but shows no homology to known eukaryotic transcription factor binding elements. As this motif is located before the RNA initiation site it is less likely that it is involved in the transcriptional regulation of the *TamS1* gene. However, regulatory motifs downstream of the transcriptional start site have been reported for a number of genes (Lehn and Bustin, 1993); and it is possible that this motif functions in a post-transcriptional manner. The motif was not absolutely conserved or present in a similar position within the untranslated regions of *TamR1*, *SPAG-1* or Ta-ORF-1. Therefore, if it does have a regulatory role it is probably specific to *TamS1* homologs or genes expressed in a similar temporal manner with respect to merozoite differentiation.

The *T.sergenti* sequence was isolated by PCR using primers designed from the 5' intergenic DNA fragment and coding regions of *TamS1*. It is interesting that although a number of upstream primers used in the attempts to PCR amplify the 5' intergenic region of

T.sergenti, only the CAT4 primer was successful. Therefore, this primer must have a reasonable degree of homology across *Theileria* species. Taken into consideration with the limited homology present over the rest of the upstream sequence this may indicate conservation of a domain involved in transcriptional control of *TamS1*.

It was not possible to obtain additional 5' sequence from the *T.sergenti* genomic DNA using PCR with the primers that were available. Furthermore, as the primer used represented the most highly conserved regions from the other species analysed, it is unlikely that more sequence could have been obtained by this strategy. To do this it may be necessary to either screen genomic libraries or mini-libraries for the required sequence. Alternatively, it maybe possible to clone the complete intergenic region between the *TpmS1* gene and the gene immediately 5' to the *TsmS1*. This would require the gene 5' to the *TamS1* present in the same position in *T.sergenti* and that primers which recognised conserved amino acid sequence of this gene were available. However, because the *T.annulata*, *T.sergenti* comparison does have the potential to pinpoint putative control motifs, further studies could aim to isolate more 5' sequence of the *TsmS1* gene. Alternatively the technique of inverse PCR (Williams, 1989; Pang and Knecht, 1997) could be used. This method has been employed recently to isolate the upstream region of the *T.annulata* HSP90 gene (D.Smith, York personal communication).

When the *TamS1* 5' intergenic region was compared with upstream regions of other *T.annulata* genes, the overall identity was low. However, it was possible to identify small conserved regions between the 5' intergenic sequences of the *TamR1*, Ta-ORF-1, *SPAG-1* and *TamS1* genes. Palindromes and pentamers were identified which were found to be conserved between all of the sequences. These included the AGTGA and ACACA pentamers described above. The *SPAG-1*, *TamR1*, ORF-1 and *TamS1* are all expressed either at different stages of the parasite's Life-Cycle or during different points of differentiation to the merozoite. Therefore, if functional, these pentamers must be involved in the general transcriptional control of *Theileria* genes, and if they are involved in the regulation of gene expression during differentiation it is likely to be in an accessory rather than a direct role. A similar conclusion has been made for pentamers believed to be

involved in the expression of certain *Plasmodium* genes (Lanzer *et al.*, 1992a, b), but their specific function awaits further definition.

In conclusion a number of motifs have been found with homology to higher eukaryotic domains which function in transcriptional control, and GC rich palindromic sequences are repeated in both *TamSI* and *TamRI* upstream regions now identified. However without a functional assay or further structural/biochemical analysis it is not possible to identify which of these putative domains are involved in the regulation of *TamSI* gene expression. Ideally, transfection studies with a range of deletions of the 5' untranslated region of *TamSI* linked to a reporter gene would be used to carry this out. Unfortunately, no transfection system is available for *Theileria*. However, another strategy to identify regions involved in transcription control is to use nuclear extracts from either differentiating macroschizont cell lines, merozoites or piroplasms in assays developed to identify DNA sequences which specifically bind nuclear polypeptides.

4 Identification and Characterisation of a Nuclear Factor Binding Element in the Upstream Region of the *TamS1* Gene

4.1 Introduction

Many studies carried out on a wide range of eukaryotic cell types have led to the identification of cis-regulatory DNA sequences which bind proteins that control the level of gene transcription (see section 1.6.2; Maniatis *et al.*, 1987; Mitchell and Tjian, 1989). The activity of a promoter is determined by firstly its architecture, consisting of the type of control elements contained in the promoter and the context in which they appear, and secondly by the regulatory state of the proteins that recognise these control elements. It is well known that in terms of differential gene expression during development that control can be exerted by modification of access to the regulatory domains, by for example methylation of histone packaging, or altering the milieu of transactivating factors both qualitatively and quantitatively.

Data from preceding chapters describe the cloning, sequencing and analysis of the 5' intergenic region of the *TamS1* gene, which encodes a 30kDa merozoite antigen. Furthermore, it was found that regulation of *TamS1* expression during differentiation is through, at least in part, elevation of transcription levels, and a number of putative motifs were defined which could function in transcription control. However, identification of motifs on the basis of homology across related species of *Theileria* and comparison with known control sequences defined for a wide range of eukaryotic systems was of limited use in the absence of further data indicating functional importance. If these motifs are functional it is likely that they interact with polypeptide factors located to the nucleus. Moreover, a correlation between motif-factor interaction with gene transcription at different developmental stages could provide further information on the elements (protein and DNA sequences) and nature (activation or suppression) of the mechanisms governing *TamS1* gene expression during differentiation of *T. annulata*.

In vitro techniques developed to study protein-DNA interactions can provide a large amount of information pertinent to the regulation of gene expression. This information is directly relevant to, and can be complemented by functional assays such as *in vitro*

transcription and gene transfection. One of the relatively simple procedures most useful in identifying relevant cis-regulatory DNA sequences and putative transcription factors, is the electrophoretic mobility shift assay (EMSA) (Fried and Crothers, 1981; Strauss and Varshavsky, 1984).

This assay typically involves the addition of purified, nuclear extracts or total cell extracts to linear radiolabelled double-stranded DNA fragments, which under incubation conditions allow complex(es) to be formed. Samples are then electrophoresed under non-denaturing conditions to separate the protein-DNA complexes from unbound DNA, and retarded bands due to factor bound DNA can be visualised by autoradiography (Lane *et al.*, 1992). With a little modification this assay can also be used to measure kinetic parameters and to assess the sequence specificity of a particular protein-DNA interaction under investigation (Fried, 1989, Revzin, 1987).

Another widely used *in vitro* technique in which the precise location of a bound protein is detected by the protection it affords the DNA against chemical or enzymatic attack, is known as DNA footprinting (Galas and Schmitz, 1978). In general this assay involves the limited degradation of an end labelled DNA fragment by an agent which cleaves the DNA, preferably using conditions in which each DNA fragment is cut only once generating a set of radioactive DNA sub-fragments of different lengths. The number and length of fragments produced will depend on the sequence specificity of the agent used for cleavage. If the DNA is complexed with a protein(s) before being treated with the cleavage agent, the nucleotides to which the protein is bound will be protected against cleavage. Thus, when paired samples, are prepared for incubation in the presence and absence of cell nuclear extracts, compared by denaturing electrophoresis, a number of bands will be missing in the sample where nuclear protein has bound to a specific DNA sequence. The resulting gap in the DNA ladder is termed the footprint. A number of DNA-footprinting techniques have been developed which use different agents to cleave the DNA (Larkin, 1993).

Once a protein binding element has been identified, *in vitro* techniques can be used to determine the relative molecular masses of polypeptides interacting with DNA sequences.

An advantage of these techniques is that they allow characterisation of the factor from crude material to provide vital information about the factor before complete purification of the polypeptide(s) is undertaken. Furthermore, although screening an expression library using DNA-binding site probes can be carried out with minimal characterisation of the factor(s) of interest, pre-characterisation allows verification of the identity of putative positive gene clones and provides some indication of the probability of the success of using this screening protocol.

Determination of the molecular weight of the individual transcription factors involved in a complex can be characterised by (UV) crosslinking the factor to a radiolabelled DNA motif followed by SDS-PAGE analysis. The UV cross-linking assay relies on irradiation of DNA-protein complexes with UV light causing the formation of covalent bonds between pyrimidines and certain amino acid residues in the DNA-binding domain of the polypeptide in close proximity to the DNA. Thus, transcription factors can be selectively labelled as a consequence of specific binding to a DNA sequence before rapid determination of its molecular weight under denaturing conditions in a SDS-polyacrylamide gel (Chodosh *et al.*, 1986).

South-Western blotting can also be used to estimate a DNA binding factor's mass. In this method DNA binding polypeptides are identified by incubating Western blots of crude nuclear extracts separated by SDS-PAGE with radiolabelled double stranded DNA probes containing the binding motif. South-Western blotting is particularly useful for the detection of previously undiscovered DNA-binding proteins which specifically bind to particular nucleotide motifs (Knepel *et al.*, 1990; Tully and Cidlowski, 1993). A modification of the South-Western technique is to use labelled DNA to screen gene expression libraries for phage plaques producing recombinant protein which bind to the DNA probe (Sambrook *et al.*, 1989; Kwok *et al.*, 1990; Hoffmann *et al.*, 1990).

The advantages of several of the techniques described above is that not only are they relatively simple procedures but that they can be used with relatively crude cell extracts. Whole cell or nuclear extracts are easy to prepare enabling the entire DNA-binding content of the cell to be examined and allow access of multi-protein complexes to the DNA binding

motif. Moreover, a correlation of factor activity with (or between factor activity and) the physiological state of the cell can be assessed directly by growing cells under different conditions before extract preparation (Arcangioli and Lescure, 1985).

The generation of extracts from apicomplexan parasites is problematical. Problems can include the limited supply of parasites, the small size of the parasite nuclei, and difficulties encountered in isolation of parasites or parasite nuclei free from host cell contamination. These problems have been overcome in *Plasmodium* by growing erythrocytic cultures of the parasite. Erythrocytic cultures have the advantage that they do not contain host nuclei and that after lysis, the parasite can be significantly purified from host material following centrifugation (Lanzer *et al.*, 1992a, b). However, the large-scale production of parasites using this methodology is a considerable task.

The ability of *T.annulata* schizonts to immortalise the host cell allows unlimited production of parasite infected cells. This coupled to the ability to induce a high level of differentiation in individual cloned cell lines provides a system where, potentially, *in vitro* derived extracts from differentiating and non-differentiating cells could be compared for detection of factors which bind to 5' upstream regions of stage regulated genes. Unfortunately, it was considered likely that the generation of parasite nuclear extracts from macroschizont cell lines would require the removal of the host cell nuclei from parasite material. This was because due to the size and complexity of the host nucleus relative to the parasite (particularly under non-differentiation conditions) it was thought that host non-specific interactions could mask parasite representation and that parasite factors would be relatively low. In the event that these drawbacks were unable to be removed one possibility was to use piroplasms isolated from erythrocytes, as it was previously shown that the *TamS1* gene is highly expressed by this stage.

Aims 4.1.2

As outlined above the EMSA has been used in many systems to define the interactions of cis-acting elements and sequence-specific trans-acting factors, and is ideally suited to quickly screening for such interactions with crude nuclear extracts. Even with the potential problem of host contamination it was decided to use this technology as a first step

towards identification of factors which interact with the *TamS1* 5' region. The primary aim of this chapter, therefore, was to screen for DNA protein interactions by preparing parasite enriched nuclear extracts and performing EMSA analysis across the conserved sequence of *TamS1* 5' intergenic region. If this proved successful a second aim was to define the specific binding site of any motifs and if possible, the number and size of the polypeptide which bound to them. In addition, analysis of extracts derived from macroschizont compared to cells cultured at 41°C could provide information on the molecular mechanism(s) operating during differentiation and link this to the quantitative and qualitative changes of *TamS1* gene expression previously observed (Shiels *et al.*, 1994); and this comparison was the final aim of work presented in this chapter.

4.2 Material and Methods

4.2.1 Isolation of Macroschizont or Differentiating Macroschizont Nuclei and Extract Preparation

Isolation of *Theileria* nuclei and preparation of nuclear extracts were based on the method of Lanzer *et al.*, (1992) with the addition of a differential centrifugation step for enrichment of parasite nuclei relative to host. 100ml of culture of parasite infected cells were spun down at 400g for 5-10 minutes. The cell pellet was resuspended in 12ml of ice cold nuclear wash buffer (20mM PIPES, pH 7.5, 15mM NaCl, 60mM KCl, 14mM β -mercaptoethanol, 0.5mM EGTA, 4mM EDTA, 0.5mM Spermidine, 0.15mM Spermine, 0.125 PMSF) and transferred to a dounce homogenizer. 780 μ l of a 10% NP40 solution was added and seven strokes with a pestle applied. The cell homogenate was then transferred to a sterile Sorvall tube before being spun at 400g for 5 minutes. The resulting supernatant was transferred to eppendorf tubes and pellet resuspended in 600 μ l of solution C (i) (20mM HEPES pH 7.9, 1mM EDTA, 1mM DTT, 1mM PMSF). The supernatant was re-centrifuged at 2000g and the pellet resuspended in 6ml of nuclear wash buffer before a final spin at 2000g. The resulting pellet was then resuspended in 600 μ l of solution C(i) and the volume of the 400g and 2000g pellets plus buffer measured. NaCl from a 3M stock solution was added to a final concentration of 300mM and the samples left on ice for 30 minutes, with gentle vortexing every 5 minutes. After centrifugation at 2000g for 5 minutes, the supernatant was dispensed and snap frozen in liquid nitrogen. All extracts were stored in liquid nitrogen until use. Nuclear extracts for the BL20 cell line were prepared in an identical manner to the 400g pellet.

4.2.2 Indirect Immunofluorescence Assay

An indirect immunofluorescence assay (IFAs) based on the method of Minami *et al.*, (1983) was performed using material fixed onto cytospin slides. Slides were prepared by spinning 100 μ l of culture or material from different points of the nuclear isolation procedure in a Shandon Cytospin at 240g for 5 minutes. When necessary, the cell/nuclear density was adjusted by dilution in phosphate-buffered saline (PBS) (150mM NaCl, 16mM $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 4mM $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, 4mM KCl, pH 7.2). The slides were then air dried, and fixed by immersion in acetone at -20°C for 15 minutes, followed by air drying. 20 μ l of the first antibody was spotted onto the fixed cells and the slide placed in a humidified box

for 30 minutes at room temperature. Unbound antibody was removed with 2x 5 minute washes in PBS. The slide was then air dried before the addition of 20 μ l fluorescein isothiocyanate (FITC) conjugated anti-mouse IgG (Sigma), at a dilution of 1:100 in TBL culture media. The slide was again incubated in a humidified box for 30 minutes at room temperature, washed in PBS as before and counter-stained with Evans Blue (0.1% in PBS) for 5 minutes. The slide was mounted with a few drops of 50% glycerol/H₂O containing 2.5% w/v 1,4 diazabicyclo(2.2.2.)octane (DABCO), pH 8.0. Immunofluorescence was analysed with a leitz Ortholux II fluorescent microscope and an Orthomat-W camera attachment.

4.2.3 SDS-polyacrylamide gel electrophoresis

The method for SDS-polyacrylamide gel electrophoresis (SDS-PAGE) was adapted from Laemmli (1970). The Biorad Protean II and mini Protean II gel electrophoresis systems were used. 30ml of 10% resolving gel mix, composed of 10ml 30% acrylamide-bis mix (30% acrylamide, 0.85% N-N Bis-methylene acrylamide, Scotlab), 7.5ml 1.5M Tris-HCl pH 8.8, 0.3ml 10% (w/v) SDS, 11.9ml dH₂O was poured after the addition of 0.3ml 10% (w/v) ammonium persulphate (freshly made), and 20 μ l of TEMED (Sigma) to initiate polymerisation. The gel was then overlaid with water saturated 2-butanol and allowed to set. Following polymerisation, the overlay of water saturated 2-butanol was removed and the top of the gel rinsed in dH₂O. 12ml of 4% stacking mix consisting of 2ml 30% acrylamide-bis, 1.875ml 0.5M Tris-HCl pH 6.8, 0.15ml 10% (w/v) SDS, 10.9ml water, was then poured on top of the resolving gel after the addition of 0.15ml 10% (w/v) ammonium persulphate (freshly made), and 15 μ l of TEMED to initiate polymerisation. The volume of the resolving and stacking gel mixes were reduced to 10ml and 3ml respectively when the Protean II mini gel apparatus was used.

Protein samples were boiled in 1x sample buffer (diluted from 4x stock, 0.25M Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 10% β -mercaptoethanol, 0.001% bromophenol blue) for 5 minutes and cooled prior to loading. High and low molecular weight markers were generally loaded into the first and last tracks of the gels, and electrophoresis was carried out overnight at 50-60V. Following electrophoresis, the gels were either processed for South-Western blotting or fixed and stained in a solution of 20% methanol, 10% acetic acid and

0.5mg ml⁻¹ Coomassie Brilliant Blue R-250, for approximately 60 minutes before, destaining in the same solution lacking the Coomassie dye.

4.2.4 Preparation of labelled DNA probes for retardation assay and South-Western assays

Larger fragments were generated by PCR using the 5' intergenic region of the *TamS1* gene as a template. The primers used for each PCR reaction are listed in Table 4.1. In all cases the PCR reaction mix was made up to 100µl containing 1µg of each primer, 100ng of template DNA, 9µl of 11.1x PCR buffer, 1µl of amplitaq and 60µl of dH₂O. The reaction mix was incubated at 95°C for 4 minutes and this was followed by 25 cycles of 1 minute at 95°C, 1 minute at 45°C and 1 minute at 72°C, with a final incubation at 72°C for 10 minutes. The PCR products were separated from unannealed primers by agarose gel electrophoresis, and purified using the QIAquick gel extraction method. The PCR fragments were then sub-cloned in to the TA cloning vector prior to labelling. Each fragment was released from the vector using at least one restriction enzyme which generated a 5' overhanging that would allowed, the incorporation of cytosine residues upon end fill reaction. After digestion the DNA was analysed on a low melting point agarose gel and the band of interest excised and purified using the QIAquick gel extraction kit. The isolated restriction fragment was then radiolabelled by end filling using klenow fraction of DNA polymerase. Standard labelling reactions, were carried out for 40 minutes at 37°C and consisted of 1µg of template DNA, 1µl of 10x Klenow buffer (500mM Tris-HCl, pH 7.2, 100mM MgSO₄, 1mM DTT, Promega), 0.5mM of each, dATP, dTTP, dGTP, 10µCi of [α^{32} P]-dCTP, 0.5 µl of Exo minus-Klenow polymerase (Promega). Unincorporated nucleotides were then removed using a Nucrap Column (Stratagene), see 3.2.1.

The primers used for generating double stranded oligonucleotides are listed in Table 4.1 and the complementary pairs are designated for each with the lower case (a) the sense strand or (b) representing the anti-sense strand of the *TamS1* sequence. In all cases 20µl of each oligonucleotide (1mg/ml) was added to 5µl 10x Polynucleotide Kinase buffer, and made up to a final volume of 50µl with dH₂O. The reaction mix was incubated for 5 minutes at 80°C before being allowed to cool slowly to room temperature. Once annealed, 4pM of double stranded oligonucleotide was labelled with 10 units of T4 polynucleotide kinase,

160 μ Ci γ -³²P ATP and 1x Polynucleotide Kinase buffer (700mM Tris-HCl, pH 7.8, 100mM magnesium chloride, 50mM dithiothreitol, Promega), in a final volume of 10 μ l. The reaction mix was incubated for 40 minutes at 37°C, and the unincorporated radioactive nucleotides separated from the labelled primer using Stratagene Nucrap Columns (see 3.2.1).

4.2.5 Gel Mobility Shift Assay

Gel retardation experiments were carried out according to the method of Dent and Latchman (1993), using the Biorad Protean II gel electrophoresis system. A 4.5% non-denaturing gel mix composed of 6.75ml 40% acrylamide-bis, 4ml 10xTBE (0.9M Tris, 0.9M boric acid, 20mM EDTA) and 50.25 ml dH₂O was poured and polymerisation initiated by the addition of 57.5 μ l TEMED and 120 μ l of freshly made 25% APS. Following casting the gel was left for approximately 60 minutes to set and the assembled rig placed at 4°C. Many protein-DNA interactions are unstable and by performing the gel mobility shift assay at low temperature it is possible to increase the sensitivity of the assay. The upper and lower tanks were filled with 0.5 xTBE and the gel was pre-run for between 1-2 hours.

Standard protein binding reactions comprised of 4 μ l nuclear extract, 1 μ l of radiolabelled probe, 2 μ g of poly dI-dC (200ng/ml) and 14 μ l of binding buffer (10mM HEPES pH 7.9, 5% Ficoll), and were performed on ice for 40 minutes. For competition experiments, cold competitor oligonucleotides were incubated with extract, at variable fold excess relative to the labelled probe, 20 minutes prior to the addition of the labelled probe, and the reactions were incubated for a further 40 minutes. Samples were loaded following pre-electrophoresis and was performed while the gel was still connected to the electrical supply. Bromophenol blue can rapidly disrupt some protein-DNA complexes and so was omitted from all samples. A dye lane consisting of Bromophenol blue in a 10% glycerol solution was run in parallel with the samples, and electrophoresis was performed at 230V in 0.5x TBE buffer. The duration of each run was dependent upon the size of the DNA fragment used in each binding experiment and was monitored by the mobility of the bromophenol blue marker dye lane. On completion the gel apparatus was disassembled and the gel was transferred onto filter paper, dried and exposed to X-ray film.

<u>Name of Primer</u>	<u>Sequence of primers 5' to 3' used for PCR</u>	<u>Position on <i>TamS1</i> genomic Sequence (Appendix I)</u>
1945	gcgaattcgggtgtagtattgctgaag	2416-2436
1944	gcaagctttttgtcgttatgagctgtcc	2544-2524
1963	gtcgacgggtaaggacagctcataacg	2517-2540
1964	tctagagaatttcctccactttactt	2715-2694
1947	gcggatccaaataaagtaaagtggagggaaat	2684-2712
1946	gcggtacccccatcttaatgaataacatg	2850-2830
TA3051	gggtctaggacttcaatggaggataagg	2845-2869
TA3052	gggagctcgaagtgagatggatgcatcacc	2939-2964
	<u>Complementary oligonucleotides for double stranded probes</u>	
GATA35a	atgctagagtgcatagatacagataaacat	2882-2905
GATA35b	gcatatgtttatctgtatctatgcactcta	2902-2882
TAT1a	atgcattgtttctatggtgaatg	2927-2947
TAT1b	gcatcattcaccatagaaaacaat	2947-2927
CAT1a	atgcgcacacaatttgtagggcgac	2906-2927
CAT1b	gcatgtcgcctacaaattgtgtgc	2927-2906
CAT2a	atgcagataaacatgcacacaatttgta	2895-2920
CAT2b	gcattacaaattgtgtgcatgtttatct	2920-2895
CAT3a	atgcagggcgac	2920-2927
CAT3b	gcatgtcgcct	2927-2920
CAT4a	gatcacaatttgtagggcga	2910-2926
CAT4b	gatctcgcctacaaattgt	2926-2910
CAT1M1a	atgcgcactggatttgtagggcgac	2906-2927
CAT1M1b	gcatgtcgcctacaaatccagtgc	2927-2906
CAT1M2a	atgcgcacacaattacgagggcgac	2906-2927
CAT1M2b	gcatgtcgcctcgtaattgtgtgc	2927-2906
CAT1M3a	atgcgcacacaattttaggcgtac	2906-2927
CAT1M3b	gcatgtacgcctacaaattgtgtgc	2927-2906

Table 4.1 Oligonucleotide primers used to generate gel shift assay probes by annealing or PCR amplification of *TamS1* genomic DNA.

4.2.6 Preparation of Single Stranded Plasmid DNA

XL-1Blue cells transformed with pGem7zf plasmid containing the concatenated CAT1 double-stranded sequence element were donated by Ms McKellar. 2mls of an overnight culture was used to inoculate 100ml of 2XYT broth. The culture was shaken at 37°C for 30 minutes, before the addition of 0.8ml of helper phage R408 (Stratagene), and then continually shaken vigorously for a further 6 hours. The supernatant harvested by pelleting the cells at 13,000g for 15 minutes, was poured off into a fresh tube and spun again for 15 minutes. Phage was precipitated by adding 0.25 volumes of a solution containing 20% polyethylene glycol, 3.75M ammonium acetate and incubated on ice for 30 minutes. The phage pellet was collected by centrifugation at 13,000g for 15 minutes, resuspended in 0.4ml of TE and extracted twice by phenol/chloroform. The resulting aqueous phase was transferred to a fresh tube and 0.5 volume of 7.5M ammonium acetate and 2 volumes of ethanol were added and the tube left at -20°C for 30 minutes. The precipitated DNA was then pelleted by centrifugation at 13,000g for 15 minutes. The pellet was resuspended in 15µl of dH₂O

4.2.7 Preparation of Radiolabelled Probe for UV Crosslinking

To 5µg of single stranded (+)pGem7zf vector containing the concatenated CAT1 insert, 15ng of T7 sequencing primer and 10µl of 10x medium salt restriction enzyme buffer (0.5M NaCl, 100mM MgCl₂, 10mM DTT, 100mM Tris-HCl pH 7.5) were added, and the final volume was made up to 74µl with sterile dH₂O. The mixture was heated to 90°C for 5 minutes before being allowed to cool slowly to room temperature. Once cooled, 10µl of crosslinking nucleotide mixture (0.5mM dGTP, 0.5mM dATP, 0.5mM 5-bromo-2'-deoxyuridine triphosphate (BrdU), and 50µM dCTP), 1µl of 0.1M DTT and 10ml of [γ -32P] dCTP (3000ci/mmol) were added and the contents mixed. 25U of klenow was then added and the solution incubated for 1.5 hours at 16°C. Klenow was inactivated by incubation at 65°C for 10 minutes and 25 units of the restriction enzymes Xba I and Hind III were added and the tube incubated for 2 hours at 37°C. The labelled/modified DNA was ethanol precipitated and resuspended in 20µl of dH₂O before being run on a 0.7% agarose gel. The labelled modified probe was then purified from the gel matrix using the QIAquick gel extraction kit before use.

4.2.8 UV Crosslinking Procedure

Binding reactions with extract were performed in a 96 well plate and comprised of 8µl of extract, 2µl of radiolabelled probe, 2µg of poly dI-dC, and 28µl of Binding Buffer. Reactions were incubated for at least 10 minutes on ice before being exposed to UV radiation (305nm). After exposure to the UV source, 5µl of DNase I buffer (0.1M CaCl₂, 0.1M MgCl₂), 4µg DNase I (Promega) and 1 international unit (IU) micrococcal nuclease (Sigma) were added to each sample. The samples were incubated at 37°C for 30 minutes before the addition of 40µl of 2x protein sample buffer and boiling for 2 minutes. The boiled samples were loaded onto an SDS-PAGE gel and electrophoresed at 50-60V overnight until the bromophenol blue dye front reached the bottom of the gel. Following electrophoresis, the gels were stained with Coomassie Blue (see 4.2.3), dried on a piece of Whatman 3MM paper and exposed to X-ray film at -70°C.

4.2.9 South-Western Blotting

Samples were prepared by mixing nuclear extracts with SDS-page sample buffer and boiling for 3 minutes. The samples were loaded onto a 10% SDS-page gel to resolve protein with molecular masses in the range of 10-200 kDa. High and low molecular weight markers (Sigma) were run in the outer tracks of the gel. Electrophoresis was terminated when the marker dye reached the bottom of the gel. The Biorad mini Protean II gel kit was dismantled and proteins transferred from the polyacrylamide gel on to a nitro-cellulose filter using the method of Sambrook *et al.*, (1989). The nitro-cellulose membrane was cut to the dimensions of the gel and placed on top, followed by sheets of 3MM Whatman paper soaked in transfer buffer (25mM Tris, 192 mM glycine, 20% methanol) on either side of the gel and nitro-cellulose filter. The resulting sandwich was placed in the miniblott cell (Biorad) so that the nitro-cellulose membrane faced the anode, and electrophoresis was carried out in transfer buffer with a current of 300mA for 1 hour at 4°C. Efficiency of transfer was determined by staining the filter for 5 minutes with 0.2% Ponceau-S (Sigma) in 3% TCA, followed by destaining in distilled water, Marker lanes were cut off and stored. The filter was then washed in Buffer 2 (50mM Tris-HCl, pH 8.0, 2mM β-mercaptoethanol, 25mM NaCl, 1mM EDTA) before being completely immersed in buffer 1 (10mM Tris-HCl, pH 8.0, 15mM Mg Acetate, 7mM KCl, 10mM β-mercaptoethanol, 1x Denharts solution diluted from a 50x

stock: 5g Ficoll, 5g polyvinylpyrrolidone, 5g BSA and made up to a total volume of 500ml with dH₂O) and stored at 4°C overnight. The filter was removed from buffer 1 and washed briefly in buffer 2. Each filter was hybridised with 200pmol of P³² labelled oligonucleotide in 3ml of buffer 2 and 10mg/ml of the non-specific competitor, Poly dI-dC. Hybridisation was carried out at room temperature for 3 hours before the filters were washed several times in buffer 2 and exposed to X-ray film overnight at -70°C.

4.3 Results

4.3.1 Analysis of Nuclei from Macroschizont cell lines

The literature describes many methods for the preparation of nuclear extracts, several of which were developed for eukaryotic cells in cultured *in vitro*. However, preparation of *T.annulata* nuclear extracts from macroschizont cell lines is complicated by the presence of the host nucleus. Due to the large difference in size between host and parasite nuclei and the problems this could generate in detecting and identifying parasite specific shifts, it was considered essential to perform enrichment of parasite material away from host.

Several methods for isolating *T.annulata* nuclei from macroschizont infected cells have been used, including the french press combined with ficoll gradients (David Swan, personal communication). This procedure was successful in isolating parasite nuclei, however, it suffered from significant host nuclear contamination and loss of parasite material. Attempts were also made to isolate the parasite from the host using a bionebuliser. This procedure relied on compressed gas to propel macroschizont culture against a barrier generating an aerosol which depending on droplet size can disrupt eukaryotic cell membranes. The preliminary experiments were partially successful, in that extracts generated from the isolated parasites were sufficiently active to demonstrate specific binding to the *TamR1* promoter (C. Robertson, unpublished data). Unfortunately, in this study no binding activity to the promoter region of the *TamS1* gene was found using these extracts. In addition to these contradictory results, the bionebulisation procedure could result in significant destruction of host nuclei, if not carefully calibrated, and was not particularly amenable to rapid processing of large cell/culture volumes. With this last disadvantage in mind a crude nuclear purification and extraction method was selected, which had been successfully employed to produce active extracts from both higher eukaryotic (Dent and Latchman, 1993) and *Plasmodium* infected cell cultures (Lanzer *et al.*, 1992a, b). For *T.annulata* infected cells it was necessary to introduce an additional centrifugal step into the protocol in an attempt to separate parasite nuclei from the host.

The aim of the final protocol was to generate as pure nuclei as possible but in as short a time as possible to increase the probability of preserving the activity of any nuclear

factors. To achieve both aims was likely to result in a compromise of either purity or speed. In the first instance, therefore, preparations were kept as simple as possible, and the purity of the nuclei assessed by IFA. To assess the purity of the resulting preparations this analysis was undertaken using monoclonal antibody 5E1, raised against the 30kDa merozoite membrane polypeptide, and anti-TamR1, an anti-rabbit sera raised against the 117kDa rhoptry protein. DAPI staining of nuclei was used to assess host contamination.

DAPI staining of intact macroschizont and differentiating macroschizont infected cells, in Figure 4.1, clearly highlights the size difference between the parasite and host nuclei (host nuclei are marked with the arrow heads; Figure 4.1; Plates 1A and 2A). When isolated, parasite nuclei appeared in clumps (Figure 4.1; Plates 1B and 2B), with some large clumps containing hundreds and perhaps thousands of nuclei. Visual analysis of parasite nuclei preparations would suggest that host nuclei contamination was low relative to the initial culture and this was confirmed by comparison of the 400g pellet to the 2000g pellet for host nuclei (Figure 4.1; comparing plates 1B and 1C, 2B and 2C). This analysis indicated that a reasonable separation between host and parasite had occurred and it was concluded that the ratio of host to parasite nuclei within the 400g pellet would not interfere with the extraction of and investigations into parasite DNA binding proteins.

Figure 4.2 represents IFA analysis using mAb5E1 on slide preparations of intact cells (Figure 4.2; Plates 1A and 2A), isolated parasite nuclei (Figure 4.2; Plates 1B and 2B) and isolated host nuclei (Figure 4.2; Plates 1C and 2C) prepared from day 0 (37°C) (Figure 4.2; Column 1) and differentiating (41°C) macroschizont infected cell cultures (Figure 4.2; Column 2). Very faint reactivity was exhibited by intact cells (Figure 4.2; Plate 1A) and isolated parasite nuclei (Figure 4.2; Plate 1B) from non-differentiating macroschizont infected cells at 37°C. No reactivity was observed with isolated host nuclei (Figure 4.2; Plate 1C) prepared from this culture. Faint reactivity with the macroschizont infected cells at 37°C is probably due to the low level expression of this antigen in non differentiating cells described previously, while reactivity with this nuclear extract is either artifactual or recognition of merozoites generated from occasional differentiating cells produced by sub optimal culture conditions. Strong reactivity was detected with the intact differentiating macroschizont infected cells incubated at 41°C for 6 days (Figure 4.2; Plate 2A), and some

positive reactivity was detected in the parasite nuclei prepared from this culture (Figure 4.2; Plate 2B). However, as for the culture incubated at 37°C, no reactivity was detected with isolated host nuclei (Figure 4.2; Plate 2C). Thus this analysis suggested that the parasite membrane content had been significantly reduced in the generation of the parasite nuclear fraction but that this was not absolute.

Reactivity of the anti-TamR1 serum demonstrated similar reactivity to that of the monoclonal antibody 5E1 (Figure 4.3). No reactivity was detected with whole macroschizont-infected cells incubated at 37°C (Figure 4.3; Plate 1A), nor the isolated host nuclei (Figure 4.3; Plate 1C), but some reactivity was found with the parasite nuclei preparations (Figure 4.3; Plate 1B) which, as for monoclonal antibody 5E1, could be due to a low level of merozoite production induced by sub-optimal culturing or from the unusual pattern of reactivity compared to differentiating cells (see punctuated reactivity, Figure 4.3, plate 2B) due to artifactual reactivity. Analysis of cells cultured at 41°C for 6 days with the anti-TamR1 serum showed positive reactivity specific to the differentiating parasite for the intact cells (Figure 4.3; Plate 2A) and parasite nuclei preparations (Figure 4.3; Plate 2B), but no reactivity was observed with the host nuclei (Figure 4.3; Plate 2C). This pattern of reactivity confirmed that the parasite is being partially purified from host material but also showed that the rhoptry organelle was not being separated from the parasite nuclei. In addition comparisons between the results with mAb5E1 and anti-*TamR1* indicated that the parasite cytoplasmic membrane was removed to a greater extent than the rhoptry proteins. Given the relative sizes of the rhoptry organelle and the parasite nucleus combined with the centrifuged forces employed this contamination of the nuclei preparation was not expected.

4.3.2 Preparation of nuclear extracts

Extracts of both parasite and host nuclear fractions, by salt extraction, were obtained (see section 4.2.1). To confirm that partial purification of parasite nuclear proteins had occurred, samples for each part of the purification step including intact cells were analysed by SDS page gel followed by staining with Coomassie blue. This analysis was performed for infected cells cultured at 37°C.

Figure 4.1 DAPI staining of fixed slide preparations of the cloned cell line D7 after cu
were incubated at 37°C (Column 1) or at 41°C for 6 days (Column 2).

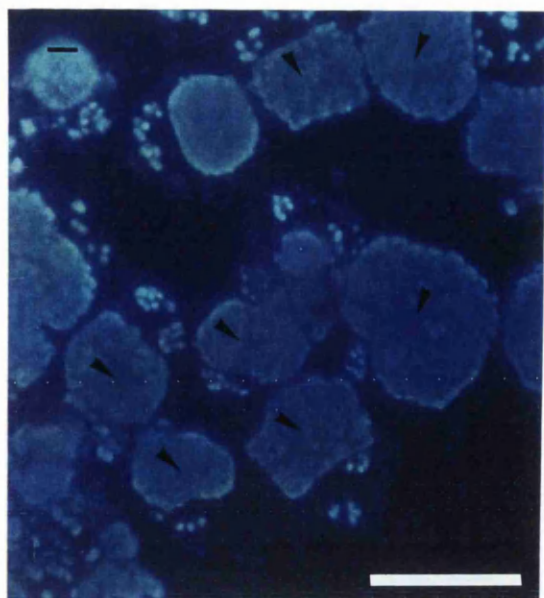
Row A :-Whole cell preparations

Row B:- Parasite nuclei

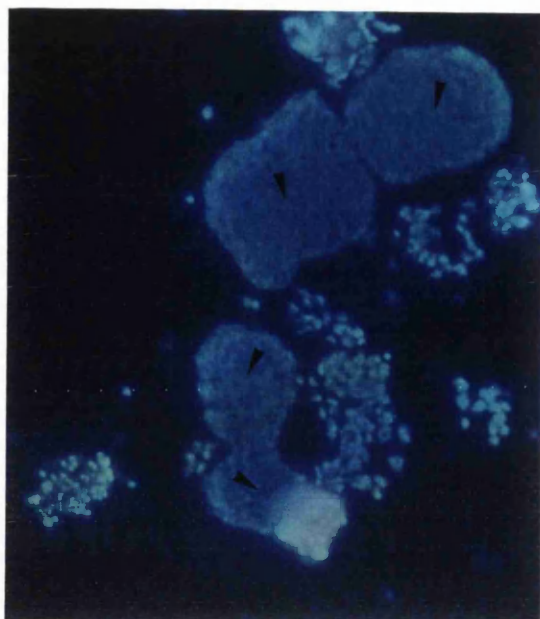
Row C:- Host nuclei

The magnification bar shown in plate 1A applies to all of the plates (Bar =36µm).

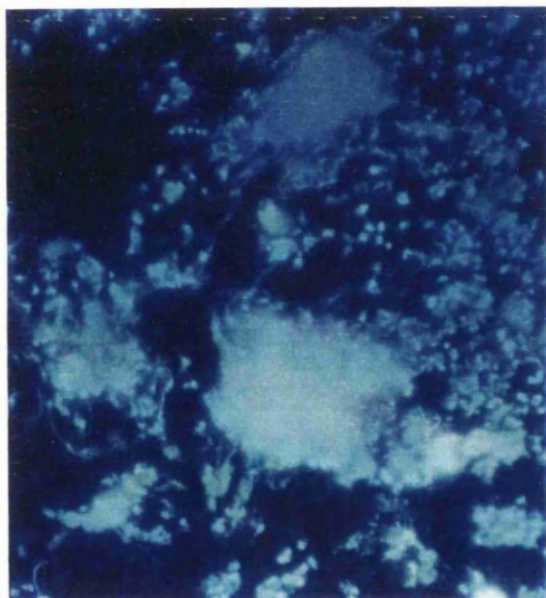
1



2



A



B



C

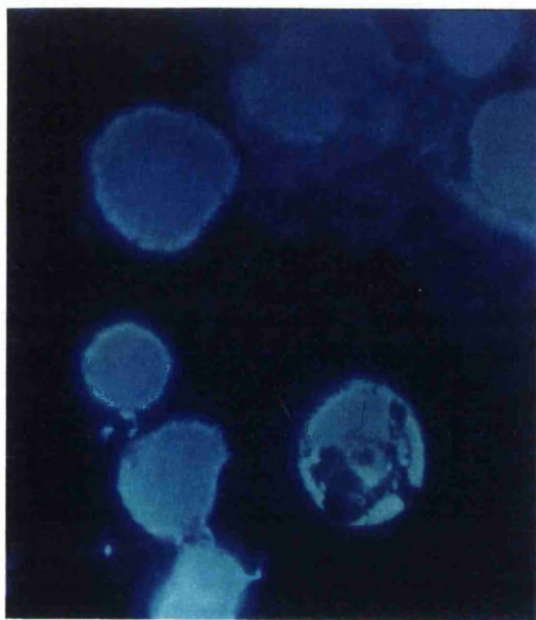
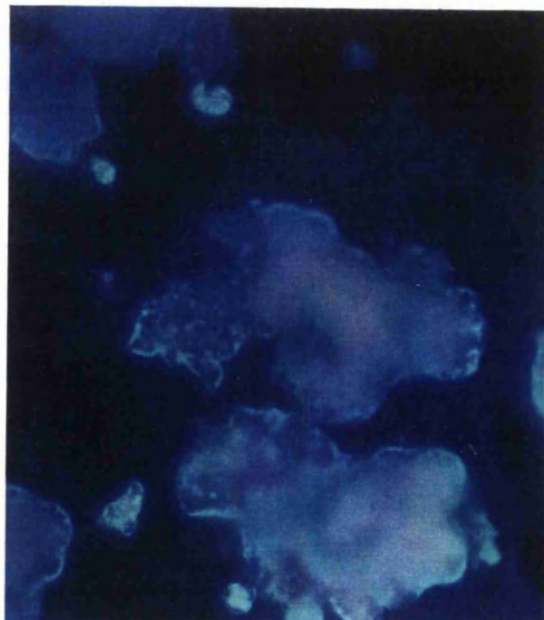


Figure 4.2 Immunofluorescence reactivity of monoclonal antibody 5E1 on fixed preparations of cloned cell line D7 after cultures were incubated at 37°C (Column 1) 41°C for 6 days (Column 2).

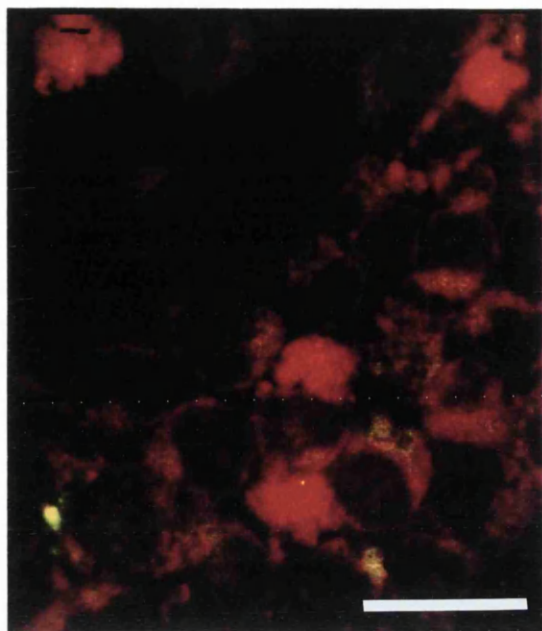
Row A:- Whole cell preparations

Row B:- Parasite nuclei fraction

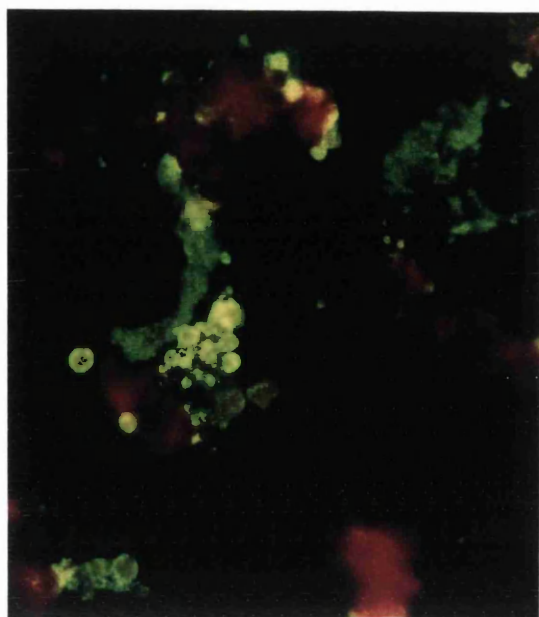
Row C:- Host nuclei fraction

The magnification bar shown in plate 1A applies to all of the plates (Bar =50µm).

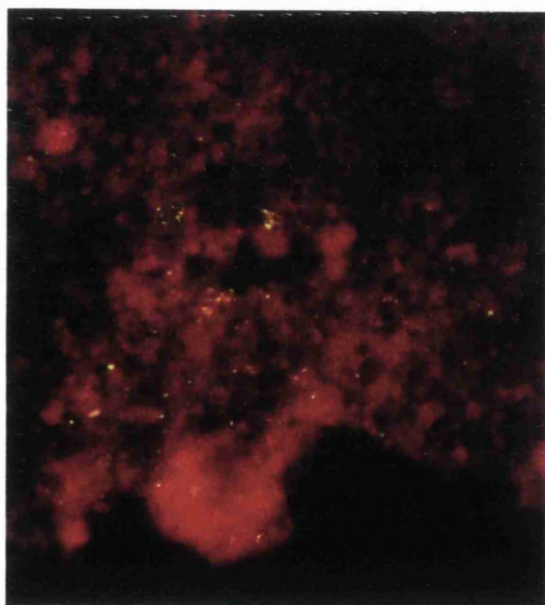
1



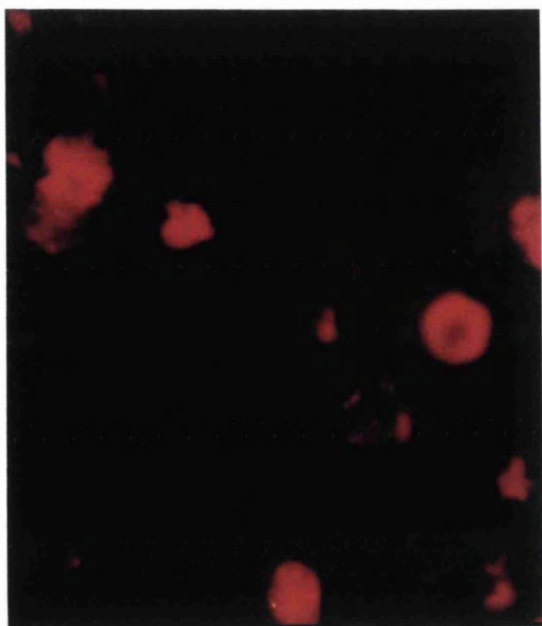
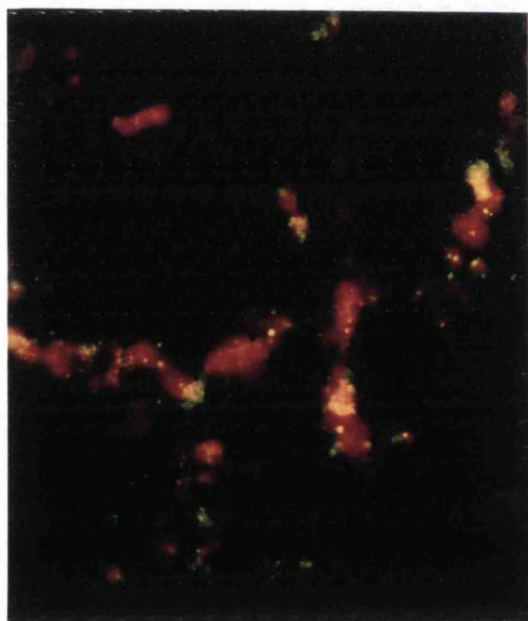
2



A



B



C

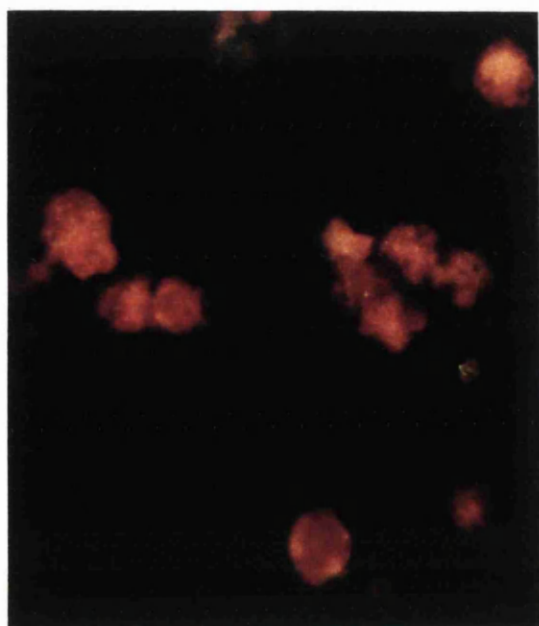


Figure 4.3 Immunofluorescence reactivity of anti-TamR1 serum on fixed slide preparations of cloned cell line D7 after cultures were incubated at 37°C (Column 1) or at 41°C for 6 days (Column 2).

Row A:- Whole cell preparations

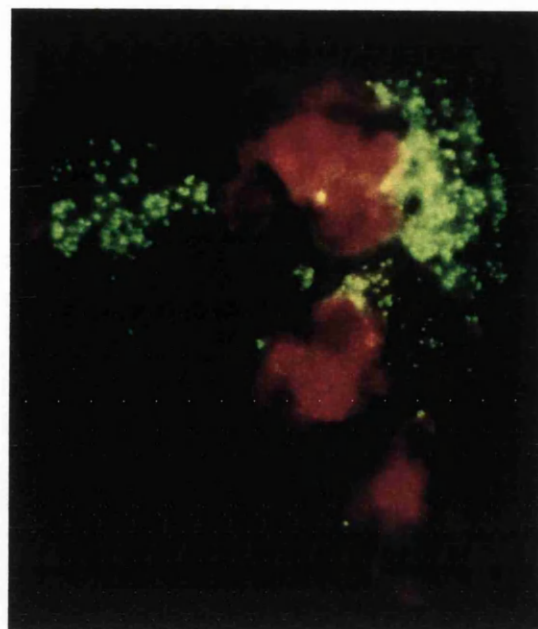
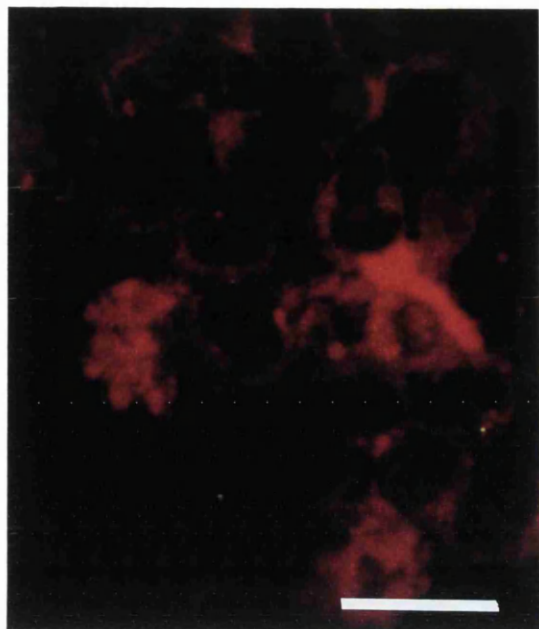
Row B:- Parasite nuclei fraction

Row C:- Host nuclei fraction

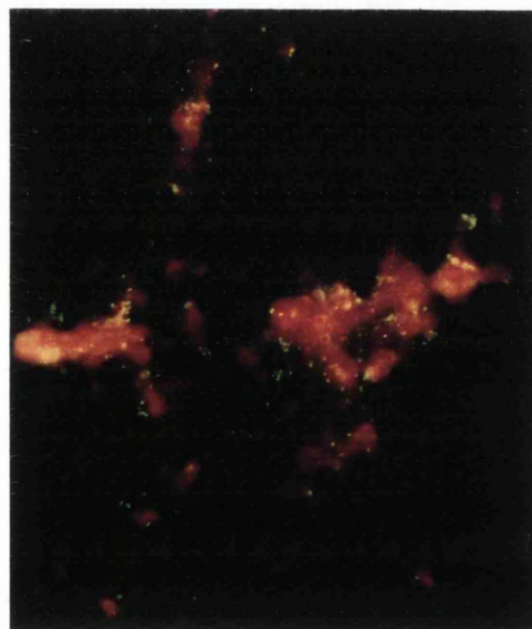
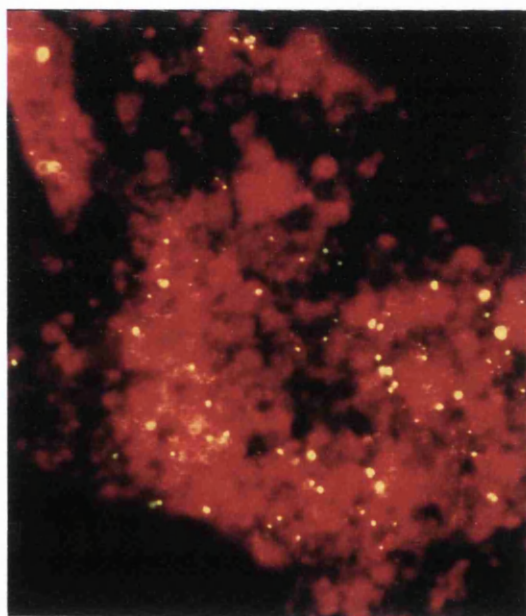
The magnification bar shown in plate 1A applies to all of the plates (Bar =50µm).

1

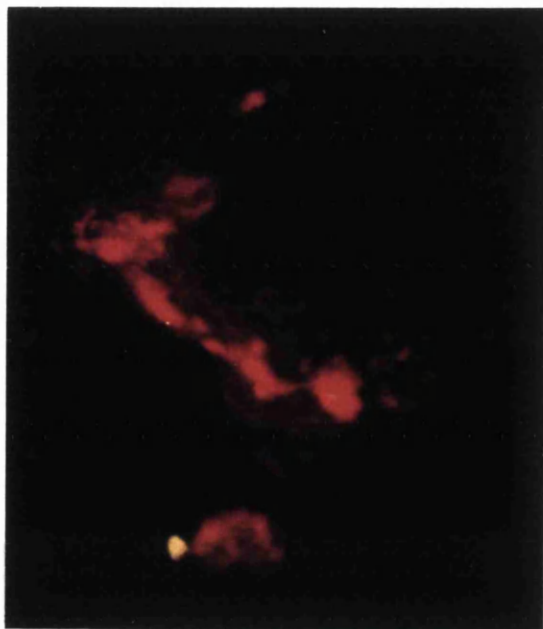
2



A



B



C

As shown in Figure 4.4 a visibly different profile of proteins was observed between total cell, host nuclei and parasite enriched nuclear fractions. For example, a pronounced band of 205kDa observed in the total cell sample (Figure 4.4, Track 1 marked with an arrowhead) was significantly reduced in host nuclei fractions and absent in parasite samples. In addition, two bands of approximately 50kDa present in host nuclei fractions, were absent in both total cell and parasite samples (Figure 4.4; Track 2 marked with arrowheads). Many other such examples were clearly visible. Thus it appeared that partial purification of nuclear proteins was occurring.

Salt extraction of nuclear fractions from host and parasite nuclei resulted in a significant loss of polypeptide levels (Figure 4.4; comparing Tracks 2 and 3, 4 and 5). The most notable decrease in protein yield was observed for extracts prepared from parasite nuclei (Figure 4.4; comparing Tracks 4 and 5), however, no clear qualitative changes were observed between these fractions.

DNA Mobility Shift Assay

4.3.3 Subcloning the 5' untranslated region of the *TamS1* gene

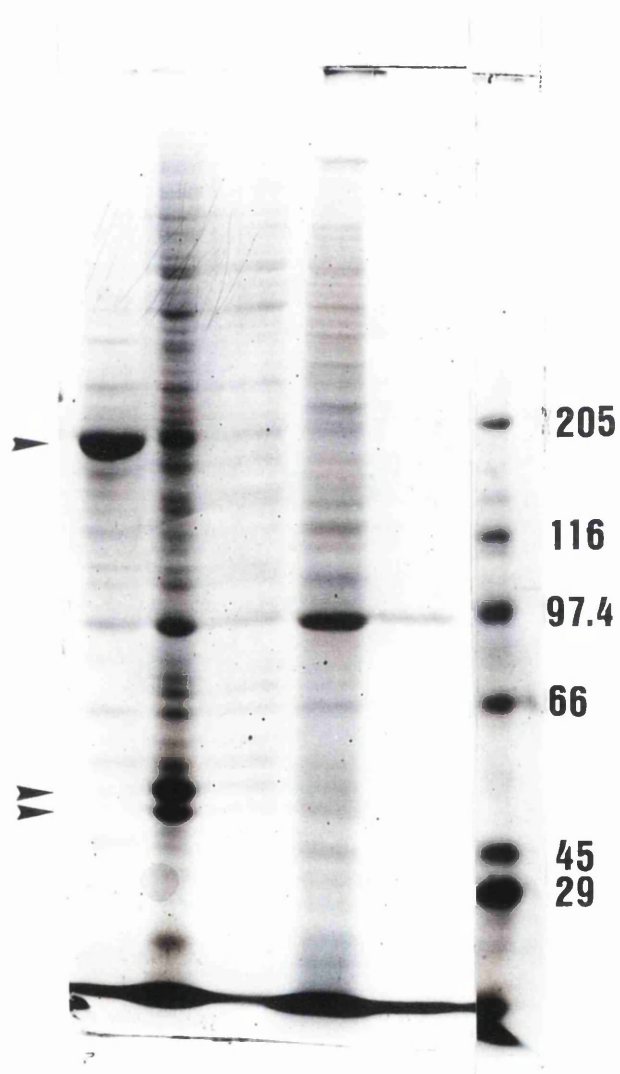
As a functional assay was not available to characterise the 5' intergenic region of the *TamS1* gene, it was decided to screen this region for sites which bound nuclear polypeptides using EMAS. In addition to nuclear extracts, the application of the mobility shift assay requires the selection of a DNA probe. Either restriction fragments or synthetic oligonucleotide probes may be used, but the size of the fragment is normally kept below 250 base pairs to enable the clear distinction of the probe from any complexes.

Approximately 2.0 kb of the 5' intergenic region for the 30kDa merozoite molecule had been cloned, see chapter 2. Based on the observation that significant homology was found between the *T.parva* and *T.annulata* sequences over the first 600bp of intergenic sequence proximal to the RNA initiation site of the *TamS1* gene, it was decided to use this region to screen for nuclear factor binding motifs. 4 fragments (A-D) spanning this region were amplified by PCR from the 5' proximal boundary of the 600bp homologous sequence towards the transcription start site. The primers used were 1945 x 1944, generating a 120bp fragment (Tams1pA); 1947 x 1946, generating a 170bp fragment (Tams1pB); 1964 x 1963,

Figure 4.4 Coomassie blue stained SDS-polyacrylamide gel of total cell extract material from different points of the nuclear isolation procedure. The positions of molecular weight makers are indicated in kDa. Tracks were loaded with protein samples based on equal volume, prepared from the following stages of making nuclear extracts from D7 macroschizont cells incubated at 37°C.

- Track 1 – Total cell extract
- Track 2 - Host nuclei fraction
- Track 3 - Host enriched nuclear extract
- Track 4 - Parasite nuclei fraction
- Track 5 - Parasite enriched nuclear fraction
- Track 6 - Markers

1 2 3 4 5 6



generating a 130bp fragment (Tams1pC); and TA3051 x TA3052, generating a 108bp fragment (Tams1pD). The primers were designed so that an overlapping 3' to 5' junction was present, so that a motif at the boundary of each fragment would not be split. The PCR products were cloned into pGem-T vector and sequenced. Sequence comparisons with the genomic sequence of the 5' region of the *TamS1* gene confirmed that the regions of interest had been cloned. Figure 4.5 shows the position of each fragment cloned in relation to the gene and the transcriptional start site.

4.3.4 Gel Shift Analysis of the 5' Untranslated Fragments of the *TamS1* Gene

Differentiation to the merozoite is characterised by an enlargement of the macroschizont and an increase in parasite nuclear division which can consequently lead to the host cell cytoplasm being completely filled by parasite (Shiels *et al.*, 1992). Thus differentiating parasites have a significant parasite load per cell, providing a large amount of parasite nuclear matter for investigation. In addition from the nuclear run on data it was considered possible that an elevation in factors which bind to the *TamS1* 5' upstream region and increase *TamS1* transcription rate may occur during differentiation. Therefore, initial investigations using EMSA were made using nuclear extracts prepared from the macroschizont cell line D7 incubated at 41°C for 6 days.

The concentration of probe DNA and binding proteins must be optimised for the binding reaction to ensure optimal and quantitative results. Determination of the concentration of nuclear extract for optimal shift resolution was undertaken by decreasing the concentrations of parasite extract which were reacted under standard binding conditions (see section 4.2.5). Thus probes, Tams1pA, Tams1pB, Tams1pC and Tams1pD, were generated (see section 4.2.4) and radiolabelled as described and incubated with 0.48, 0.24, 0.12, 0.06, 0.03, and 0.015µg of protein extract. The products were separated on a 0.5 x TBE, 4% native polyacrylamide gel and visualised by autoradiography.

As shown in Figure 4.6, all of these fragments resulted in a multiple banding pattern when incubated with extract. In general more distinct bands appeared following dilution of the extract to 0.24µg (Figure 4.6; Lane 3). The smear at the higher concentrations was most likely due to non-specific binding of multiple factors (Figure 4.6; Lane 2). A number of

bands were also present in the probe alone tracks; which may be due to secondary structure of the probe (Yamada *et al.*, 1990). The bands specific to the extract are marked with an arrowhead in Figure 4.6. Thus, this experiment gave an indication of the level of extract relative to probe necessary for visualisation of shifts due to polypeptide/probe interaction. However, in the absence of additional experimental data, these results cannot be interpreted as being due to specific complexes of DNA binding protein(s) interacting with the probe. Furthermore, it was unclear from this analysis whether the shifts obtained were specific or enriched in parasite nuclear extracts.

4.3.5 Evidence for Parasite Associated Mobility Shifts

During the preparation of nuclear extracts, parasite nuclei were separated from the host nuclei by differential centrifugation and IFA showed that contamination by host nuclei was significantly reduced, see section 4.3.1. However, due to the relative sizes of host and parasite nuclei, the possibility of mobility shifts of the 4 PCR fragments, illustrated in Figure 4.6, being due to host nuclear factor could not be ruled out conclusively. Additionally, DNA-protein complexes binding to any of the fragments could be constitutive protein-complexes, specifically involved in the basic machinery of the transcriptional initiation process, or parasite stage specific factors, involved in the regulation of *TamS1* during differentiation. It was therefore, necessary to determine whether any of the shifts were (a) due to parasite derived factor and (b) were associated with differentiation events.

In order to characterise the nature of each electrophoretic band, probes, were reacted with nuclear extracts of D7 cells incubated at 37°C, D7 cells incubated at 41°C for 6 days and uninfected BL20 cells line. Extracts from host and parasite enriched nuclear fractions were reacted with each probe under standard binding conditions (see section 4.2.5) at a protein concentration of 0.24µg, and the results are shown in Figure 4.7.

By extending the running time of the gel, the single electrophoretic band previously observed for the *Tams1pA* probe (Figure 4.6A; Track 6) with differentiating parasite nuclear extracts, was shown to consist of two electrophoretic forms, marked with arrow heads (Figure 4.7A; Track6). In contrast, the *Tams1pA* probe formed only one distinct complex with parasite enriched nuclear extracts derived from D7 cells cultured at 37°C

Schematic of the 5' intergenic region of the 30kDa molecule

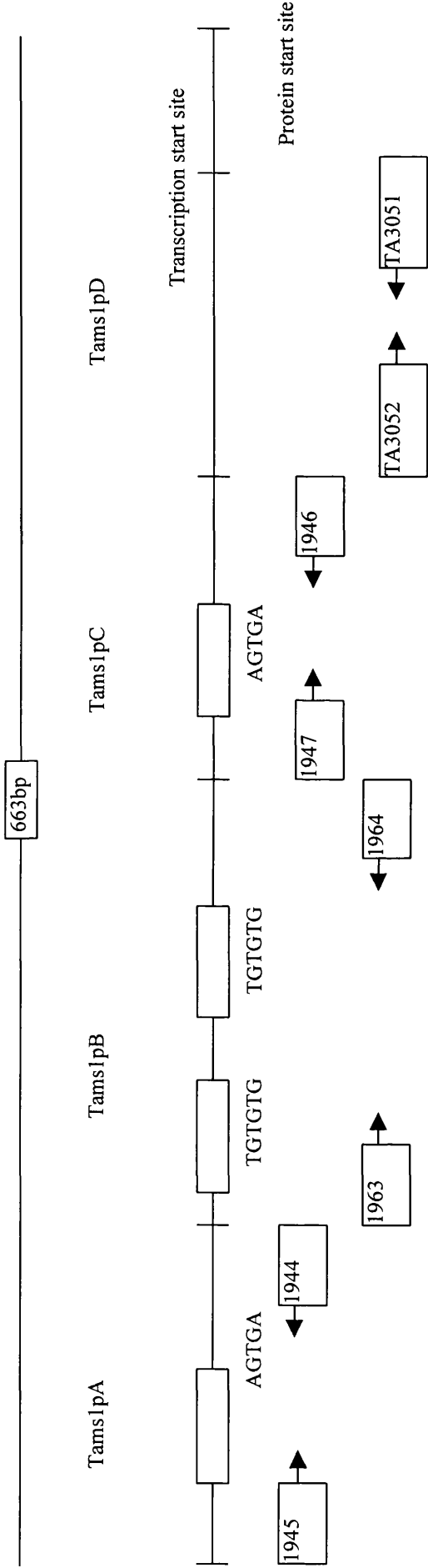


Figure 4.5 The promoter region of the 30kDa molecule. The schematic drawing reveals the organisation of PCR products investigated for DNA binding elements. The rectangles denote regions highly conserved between *T.hirci*, *T.parva* and *T.annulata*. The sequences denotes conserved palindromes found within each of the conserved block. The boxes with attached arrows denote the position and organisation of oligonucleotides used in each PCR reaction and were designed so that an overlapping 3' to 5' junction was present. The initiation site of the *TamSI* gene is indicated.

Figure 4.6 Analysis of protein concentration upon DNA binding complexes. Decreased protein concentrations of parasite nuclear extracts from D7 macroschizont-infected cells incubated at 41°C for 6 days were reacted with labelled fragments (A) Tams1pA, (B) Tams1pB, (C) Tams1pC and (D) Tams1pD. The concentration of protein indicated in the reaction under standard binding conditions is the total protein concentration. An excess of non-radiolabelled fragment is indicated by the unbound probe.

Lane 1) Probe only

Lane 2) 0.48 µg

Lane 3) 0.24 µg

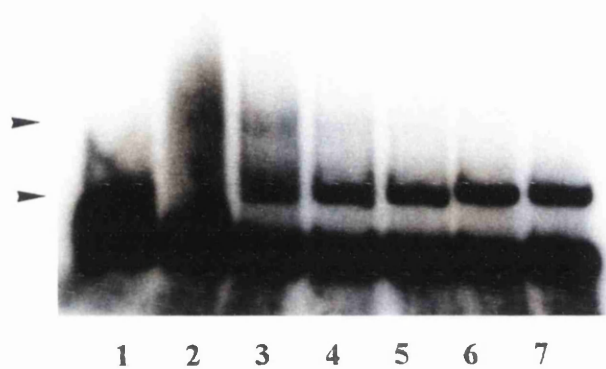
Lane 4) 0.12 µg

Lane 5) 0.06 µg

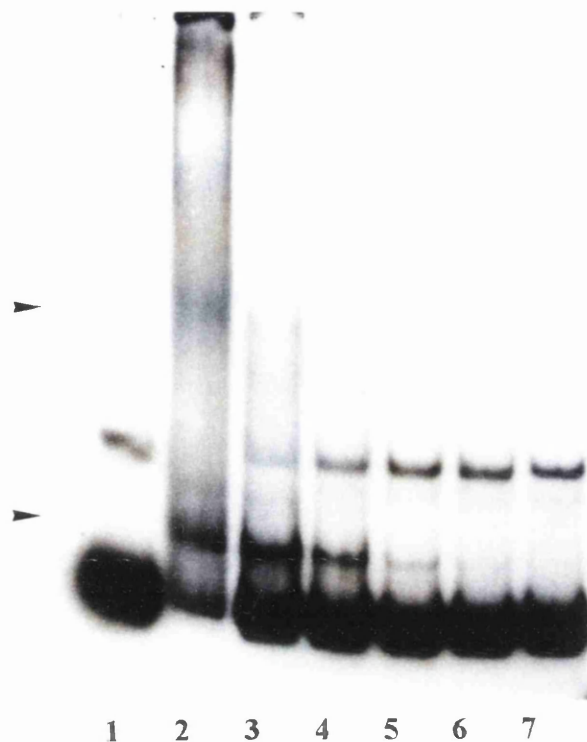
Lane 6) 0.03 µg

Lane 7) 0.015 µg

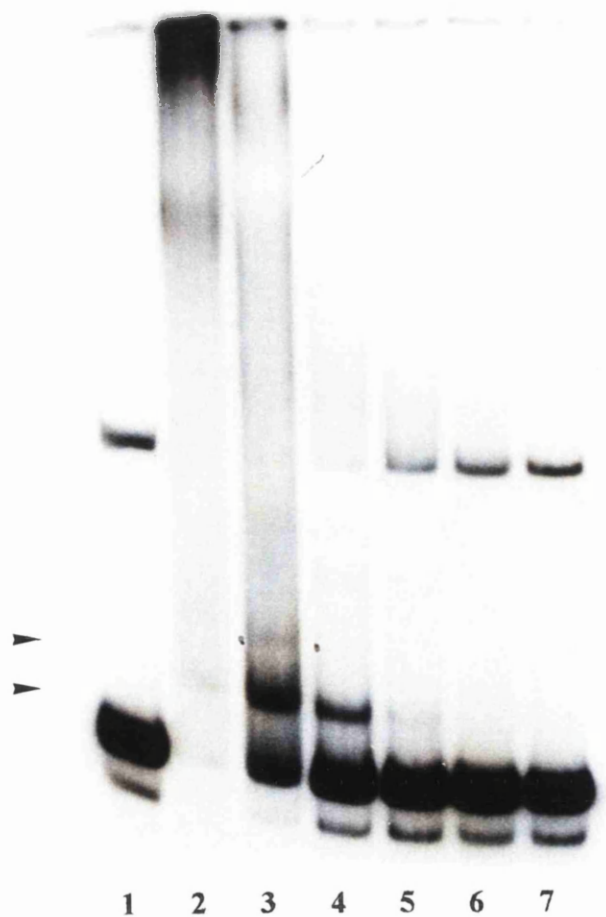
A



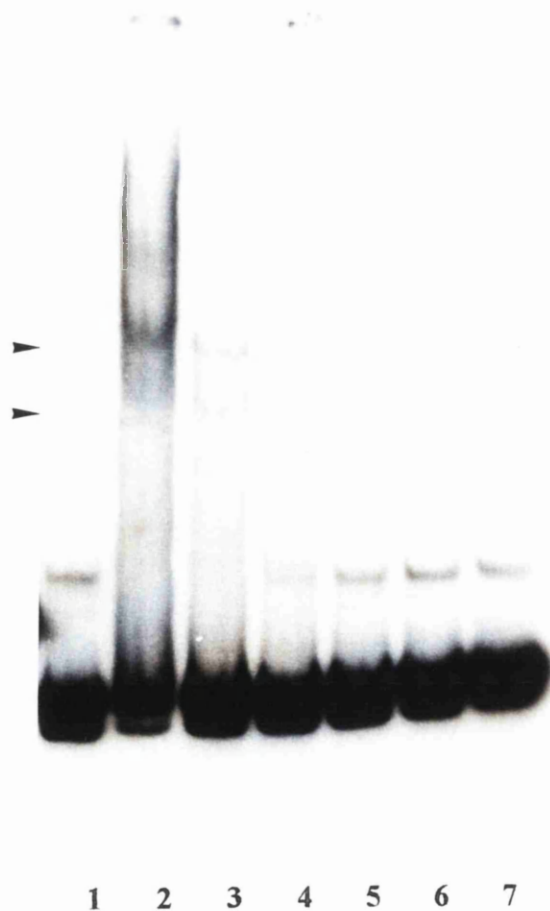
B



C



D



(Figure 4.7A; Track 4). This band was also obtained with the parasite enriched extract from differentiating D7 cells, the greater intensity of the shift possibly being due to a greater parasite load in these cells. Multiple very faint complexes were observed with nuclear extracts prepared from host nuclei from both differentiating D7 (Figure 4.7A; Track 5) and D7 cultures incubated at 37°C (Figure 4.7A; Track 3). No complexes were observed to form when the probe was incubated with extracts prepared from the BL20 cell line (Figure 4.7A; Track 2). Comparisons between extracts derived from host and parasite nuclear fractions indicate that mobility alterations to the Tams1pA probe are due to parasite specific factors. In addition, the upper band observed with the differentiating parasite enriched extract maybe a result of this process (Figure 4.7A; Track6).

The multiple electrophoretic forms of probe Tams1pB first observed in Figure 4.6 (Figure 4.7B; Track 3) appear specific to neither the parasite or host (Figure 4.7B,). Extracts prepared from macroschizont cultures and BL20 cells interact with the Tams1pB probe forming one distinct band, shown by an arrowhead, and several faint bands, some of which are blurred. Several of the very faint bands, present in Tracks 4 and 6, appear to be associated with parasite enriched extracts, however, the lack of intensity of these bands made it difficult to study them in detail and to confirm that they were not present in host nuclei. Similar conclusions were made from the results obtained with the Tams1pC probe (Figure 4.7C). Therefore, although parasite specific protein binding may occur for both Tams1pB and C fragments, the standard binding conditions used in this experiment could not detect them at a significant level and the possibility that weak shifts in parasite enriched extracts were simply due to a greater level of non-specific binding could not be discarded.

The probe Tams1pD contains the transcription initiation site for the *TamS1* gene and is the fragment upstream closest to the protein start site. Based on the considerable evidence that important regulatory motifs have been found within this region for multiple promoters from other systems, it was predicted that control sequences of the *TamS1* gene could be located to this probe. Two distinct complexes were found to interact with the Tams1pD probe (Figure 4.7D) when incubated with extracts prepared from parasite nuclei from either differentiating (Figure 4.6D; Track 6) or 37°C D7 cultures (Figure 4.7D; Track 4), shown by arrowheads. A single band, separate to those seen in the parasite extracts, was observed

Figure 4.7 Gel mobility shift analysis of parasite specific DNA complexes. DNA fragment containing the 30kDa polypeptide promoter; (A) Tams1pA, (B) Tams1pB, (C) Tams1pC and (D) Tams1pD were incubated with crude nuclear extracts derived from parasite enriched and host enriched D7 cell nuclear fractions and a nuclear derived form from uninfected BL20 cells (a bovine B cell lymphosarcoma cell line (Morzaria *et al.*, 1982)).

Lane 1) Probe only

Lane 2) BL20 crude nuclear extract

Lane 3) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 4) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 5) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 6) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

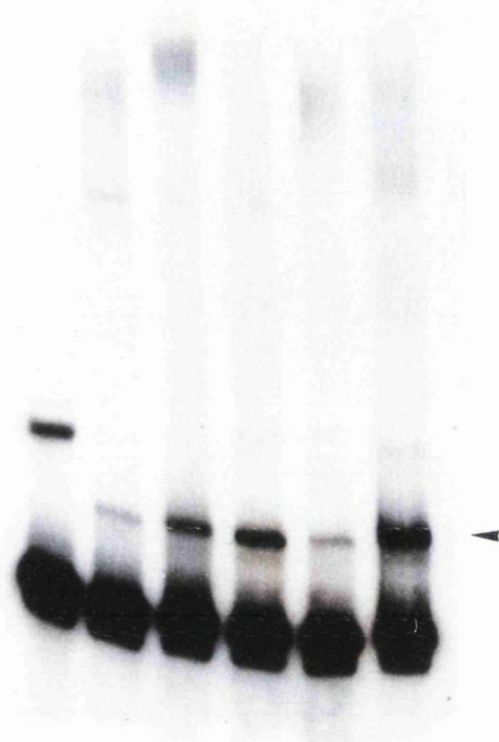
(A)

1 2 3 4 5 6



(B)

1 2 3 4 5 6



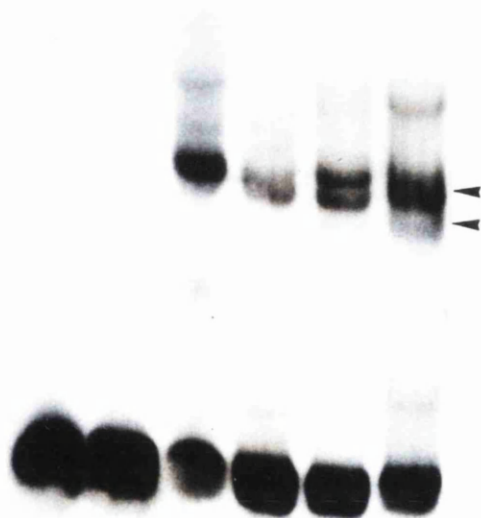
(C)

1 2 3 4 5 6



(D)

1 2 3 4 5 6



when the Tams1pD probe was incubated with BL20 extracts (Figure 4.7D; Track 2). In contrast, extracts prepared from the host enriched nuclear fractions of D7 cells either showed no discernible shifts interactions (Figure 4.7D; Track 5), or a smear which indicated non specific binding (Figure 4.7D; Track 3).

4.3.6 Specificity of Binding to Tams1pD Probe

The data presented in the above section identified DNA fragments Tams1pA and Tams1pD as containing potential motifs which bind factors in extracts of parasite enriched nuclear fractions. It was decided that work in this study should centre on the Tams1pD fragment due to its close proximity to the transcriptional start site and because parallel experiments carried out by Ms McKellar indicated that the observed shifts were most intense with this probe. The specificity of binding to the Tams1pD fragment was evaluated further by competition experiments in parasite enriched nuclear extracts from both 37°C and 41°C D7 cultures. Unlabeled Tams1pD probe was added to each reaction at an excess of x100.

The results shown in Figure 4.8 demonstrate that the binding of parasite specific factors from macroschizont (Figure 4.8; Track 4) and differentiating macroschizont parasite enriched extracts (Figure 4.8; Track 6) to the Tams1pD probe are reproducible and strong, supporting the data presented in 4.7D and generated by Ms McKellar. Addition of the unlabelled competitor (Figure 4.8; Tracks 7-10), prior to incubation with the radiolabelled probe resulted in a decreased intensity for all of the detected complexes, including the single major host complex first observed in the BL20 extracts (compare Tracks 7 and 9 with the single major BL-20 complex (shown with arrowhead) in Track 2 and also see host cell enriched nuclear extracts; Tracks 3 and 5). These findings indicate a specific interaction between the parasite enriched nuclear extracts and Tams1pD probe. However, competition with a non-specific competitor was necessary to confirm this conclusion.

4.3.7 Defining the Position of Binding Sites Within the Tams1pD Probe

Tams1pD is a 126bp fragment containing the transcriptional start site at its 3' end and 120bp of upstream sequence. To enable a more precise positional determination of the binding sites within the Tams1pD probe it was possible to either sequentially delete the

Figure 4.8 Competition EMSA with probe Tams1pD. The Tams1pD sequence element was incubated with crude nuclear extracts derived from parasite infected cultures D7 and BL20 cell line prior to incubation with radiolabelled probe (lanes 7-10). For competition experiments unlabeled Tams1pD fragment was added.

Lane 1) Probe only

Lane 2) BL20 crude nuclear extract

Lane 3) Host enriched crude nuclear extract prepared from D7 cells incubated at 37°C

Lane 4) Parasite enriched crude nuclear extract prepared from D7 cells incubated at 37°C

Lane 5) Host cell enriched crude nuclear extract prepared from D7 cells incubated at 37°C for 6 days

Lane 6) Parasite enriched crude nuclear extract prepared from D7 cells incubated at 37°C for 6 days

Lane 7) Host cell enriched crude nuclear extract prepared from D7 cells incubated at 37°C

Lane 8) Parasite enriched crude nuclear extract prepared from D7 cells incubated at 37°C

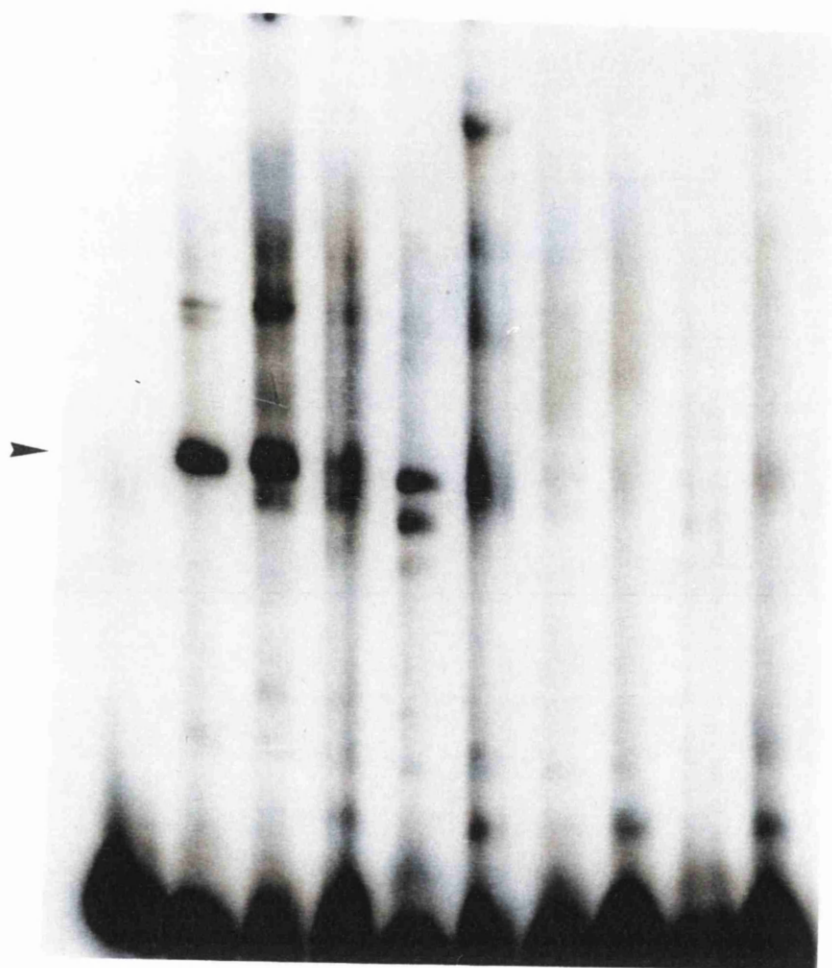
Lane 9) Host cell enriched crude nuclear extract prepared from D7 cells incubated at 37°C for 6 days

Lane 10) Parasite crude enriched nuclear extract prepared from D7 cells incubated at 37°C for 6 days

COMPETITION →

NONE

Tams1pD



1 2 3 4 5 6 7 8 9 10

probe or compete for binding of factors using double stranded oligonucleotides representing sub-regions of the probe. Deleting the probe in a controlled manner would have been problematical due to the relatively small size of the fragment. It was decided, therefore, that competition analysis would be more practical, and could be supported by subsequent experiments testing the ability of the competing oligonucleotides to give a specific shift in EMSA. Three double stranded oligonucleotides, GATA, TAT1 and CAT1 were generated which, when combined, cover the regions of the Tams1pD probe which are conserved across the *Theileria* species analysed (Figure 4.17).

Competition experiments were performed using unlabeled double stranded oligonucleotides added to each reaction at an excess of x100 and the results obtained shown in Figure 4.9. The oligonucleotide GATA, containing the GATA-1-MC-CPA binding sequence (Zon *et al.*, 1991), did not compete the shift with labelled Tams1pD when present at 100 fold excess with either parasite (Figure 4.9; Lanes 12 and 14) or host (Figure 4.9; Lanes 11 and 13) enriched extracts. Complex formation with extracts prepared from parasite nuclei from either 37°C D7 cells or differentiating D7 cultures were competed by both unlabelled CAT1 and TAT1. Competition with these oligonucleotides appeared to be directed against parasite enriched extract/probe interactions because competition of the shift obtained with host enriched nuclear extracts from D7 cells cultured at 37°C was minimal (Figure 4.9; Lanes 7, 11 and 21). No complexes were observed when extracts prepared from parasite nuclei from D7 cells incubated at 41°C were incubated with the Tams1pD probe.

In contrast, the mobility shifts associated with parasite enriched nuclear fractions were significantly competed with cold TAT1 (Figure 4.9; Lanes 22 and 24) and almost completely removed with cold CAT1 (Figure 4.9; Lanes 8 and 10). This result showed that the extent of competition with TAT1 and CAT1 varied in the degree of effectiveness. Competition of binding to the Tams1pD fragment by both the CAT1 and TAT1 competitors suggested that either nuclear protein(s) recognise sequences spanning the two oligonucleotides or complete binding sites are present on one or both oligonucleotides. In addition, complete competition by the CAT1 sequence relative to the partial competition of TAT1 may relate to the possession of a higher affinity binding site. Sequence comparison

showed that both oligonucleotides share the sequence TTTGTA which could, in combination with differences in flanking sequence, account for the similar qualitative, but different quantitative, effect of these oligonucleotides in the competition experiments.

4.3.8 Confirmation of DNA-Protein interactions with Oligonucleotides GATA, CAT1 and TAT1

Sequence-specific DNA-binding proteins are often capable of binding to a series of variations on a basic consensus sequence. A binding site that is further removed from the consensus may bind the factor(s) less strongly, and therefore, be a weaker site for transcription activation (Dent and Latchman, 1993). Sequence flanking the core binding site may also affect binding. Differences in affinity of different sites for the same factor(s) can contribute to the differential effects of a single transcription factor on different genes. In order to study the comparative affinities of the CAT1 and TAT1 sequences further, direct binding reactions were set up using radiolabelled oligonucleotides CAT1, TAT1, GATA and nuclear extracts. In addition, cross-competition reactions were set up, using a 100-fold excess of unlabelled oligonucleotide. The introduction of a competitor would not only allow the specificity of any binding activity to be determined but could also assess the affinity of binding of CAT1 relative to TAT1. More importantly these direct binding experiments would determine if binding of fractions to the Tams1pD probe was due to sequences spanning the oligonucleotides or motifs located within CAT1 and/or TAT1.

The results represented in Figure 4.10 demonstrated the ability of BL20, 37°C D7, and differentiating D7 extracts to form complexes with the CAT1 oligonucleotide (Figure 4.10; Lanes 2 to 6). A single complex mobility shift was observed when the probe was with BL20 extracts (Figure 4.10; Lane 2) which was also observed in the parasite (Figure 4.10; Lane 4) and host enriched nuclear extracts (Figure 4.10; Lane 3) from D7 cells. Additional mobility shifts were also observed both with host and parasite enriched fractions of D7 cells incubated at 37°C and parasite enriched fractions incubated at 41°C; but not in uninfected BL-20 cells or host nuclear fractions of D7 cell cultured at 41°C. From the reduction in levels of the mobility shift associated with the uninfected BL20 extracts in both extracts of parasite enriched nuclei, it was concluded that this shift represented binding by a host factor which was depleted by preparation of parasite nuclei and differentiation of infected cells. In

Figure 4.9 Analysis of the position of protein binding within the Tams1pD probe. A labelled PCR fragment Tams1pD was incubated with parasite or host enriched nuclear extract derived from 37°C or 41°C cultures. For competition experiments unlabeled oligonucleotides CAT1, GATA and TAT1 containing sequence elements of the labelled Tams1pD fragment were incubated with extract prior to the probe. Lanes 6 to 9 were competed with oligonucleotide CAT1, lanes 10 to 13 were competed with oligonucleotide GATA and lanes 14 to 17 were competed with oligonucleotide TAT1.

Lane 1) probe

Lane 2) BL20 crude nuclear extract

Lane 3) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 4) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 5) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 6) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 7) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 8) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 9) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 10) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 11) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 12) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 13) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 14) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 15) Probe

Lane 16) BL20 crude nuclear extract

Lane 17) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 18) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 19) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 20) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 21) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 22) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 23) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 24) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

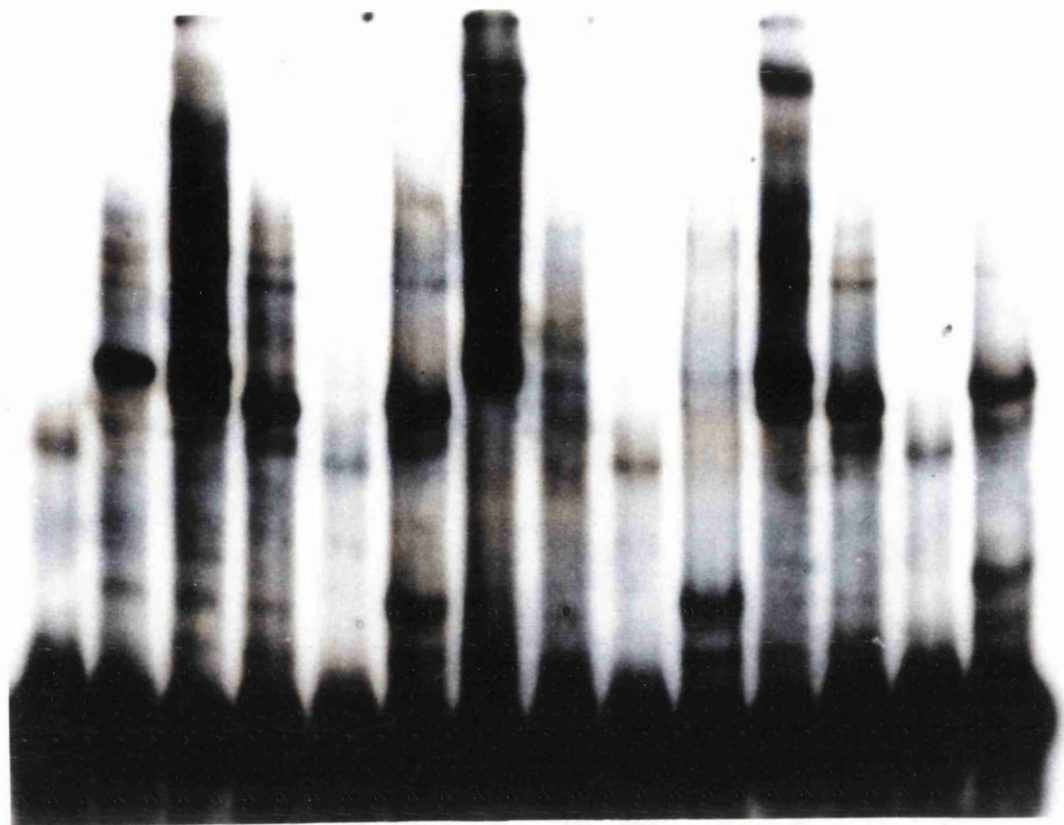
Arrowheads on Figure 4.9 denote shifts detected in host (H) or enriched parasite (P) nuclear extracts. The case letter on bottom of tracks denote origin of extract: hs, host specific; h, host enriched; p, parasite enriched.

COMPETITION →

NONE

CAT1

GATA

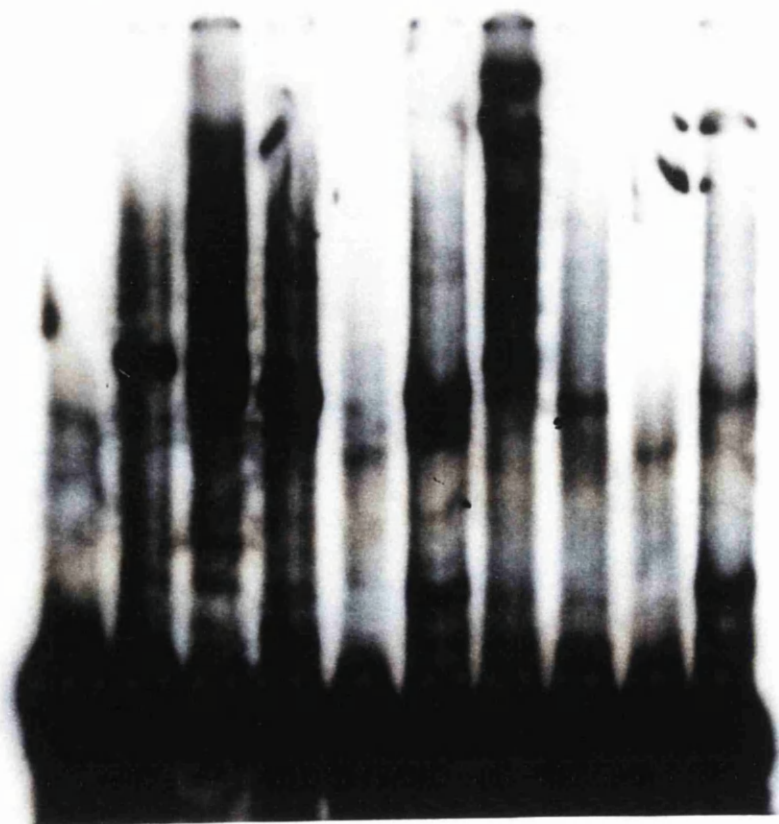


1 2 3 4 5 6 7 8 9 10 11 12 13 14

COMPETITION →

NONE

TAT1



15 16 17 18 19 20 21 22 23 24

contrast, the association of the additional shifts (marked A, B, and C on Figure 4.10) with parasite enriched fractions and differentiating cultures indicated that these were due to parasite factors; the presence of the B mobility shift in the host fraction of D7 37°C cells demonstrating contamination of host with parasite material.

As previously described (section 4.2.7) the GATA oligonucleotide contains the GATA-1 binding element. It is, therefore, not surprising that extracts prepared from BL20 cells, and D7 37°C cultures where the parasite load is low, should form complexes with this oligonucleotide (Figure 4.11). Four mobility shifts, were observed with the BL20 extract (Lane 2) and both parasite and host enriched nuclear extracts (although at low level) of D7 37°C cells (Lanes 3 and 4). In contrast, host extracts prepared from differentiating cultures did not give a mobility shift with the GATA oligonucleotide (Lane 5), while a very faint shift was observed for the parasite enriched extract (Figure 4.11; Lane 6). The complexes, formed by the interaction of the GATA probe and extracts, were competed by both unlabelled GATA (Figure 4.11; Lanes 7-10) and TAT1 (Figure 4.11; Lanes 21-24) but were found to be stable in the presence of unlabelled CAT1 (Figure 4.11; Lanes 11-14). In conclusion it appears that binding to the GATA oligonucleotide was biased towards host enriched extracts.

The experiment represented by Figure 4.12 demonstrated direct mobility shifts of the TAT1 oligonucleotide, most notably with the uninfected BL-20 nuclear extract and host enriched nuclear extract from D7 cells incubated at 37°C. These mobility shifts were competed to some extent with all three cold oligonucleotides, and although competition was greatest with cold TAT1 it was not absolute. Therefore, it was difficult to conclude that a sequence specific mobility shift was occurring with the TAT1 probe. However, more pertinent to the present study was the absence of a shift associated with parasite enriched or differentiation. It was concluded, that under the conditions tested, a self-contained binding domain was not present in the TAT1 sequence.

In summary, the data from these direct binding experiments indicated that the CAT1 sequence specifically bound factors from parasite enriched nuclear extracts. In contrast no evidence was produced which indicated specific binding of the GATA or TAT1 probes to

Figure 4.10 EMSA of the CAT1 sequence element with nuclear extracts. End L
oligonucleotide CAT1 was incubated with parasite or host enriched nuclear extract derived from 37°C
cultures. For cross competition experiments unlabeled oligonucleotides CAT1, GATA and TAT1 inc
with extract prior to addition of labelled probe. Lanes 7 to 10 were competed with oligonucleotide
lanes 11 to 14 were competed with oligonucleotide GATA and lanes 21 to 24 were compete
oligonucleotide TAT1.

Lane 1) probe

Lane 2) BL20 crude nuclear extract

Lane 3) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 4) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 5) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 6) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 7) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 8) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 9) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 10) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 11) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 12) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 13) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 14) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 15) Probe

Lane 16) BL20 crude nuclear extract

Lane 17) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 18) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 19) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 20) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 21) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 22) Parasite enriched nuclear extract prepared from cells incubated at 37°C

Lane 23) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

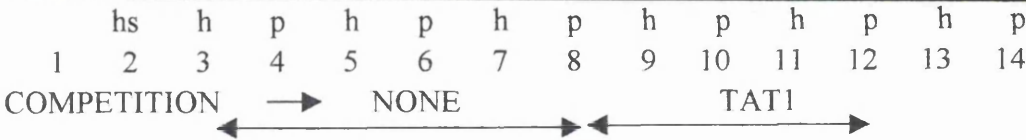
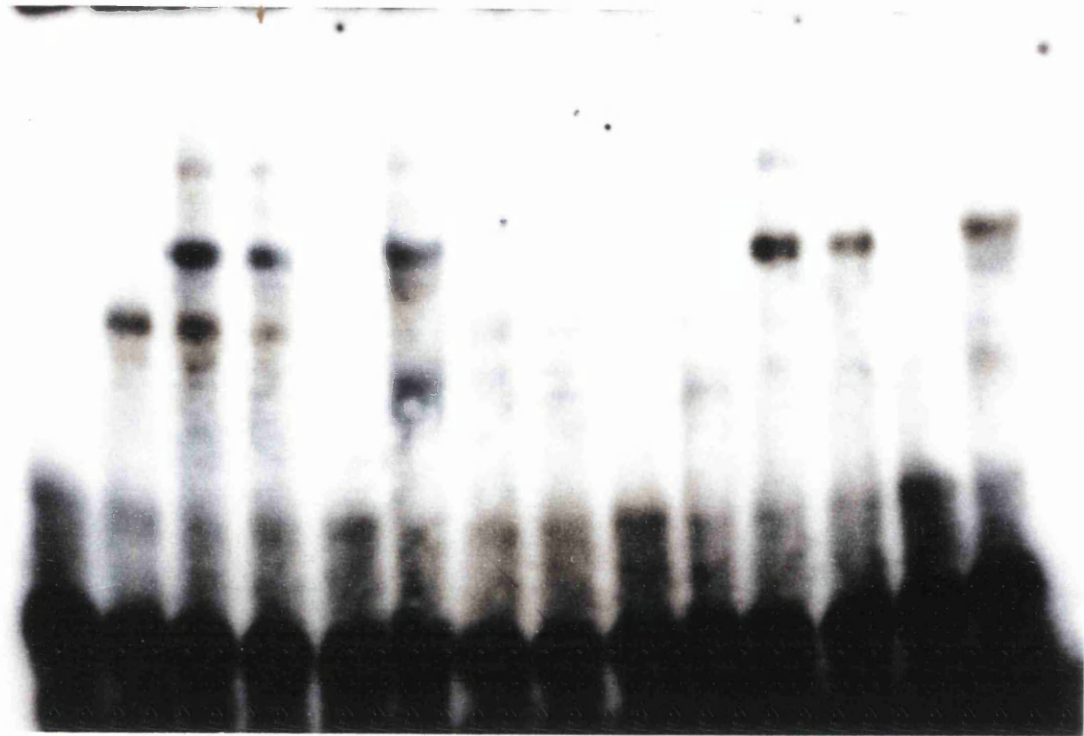
Lane 24) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B
Lower case letter on bottom of tracks denote origin of extract: hs, host specific; h, host enriched; p,
enriched.



COMPLEX

← A
← B
← C



COMPLEX

← A
← B
← C

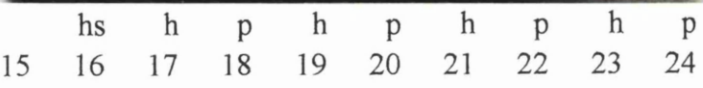
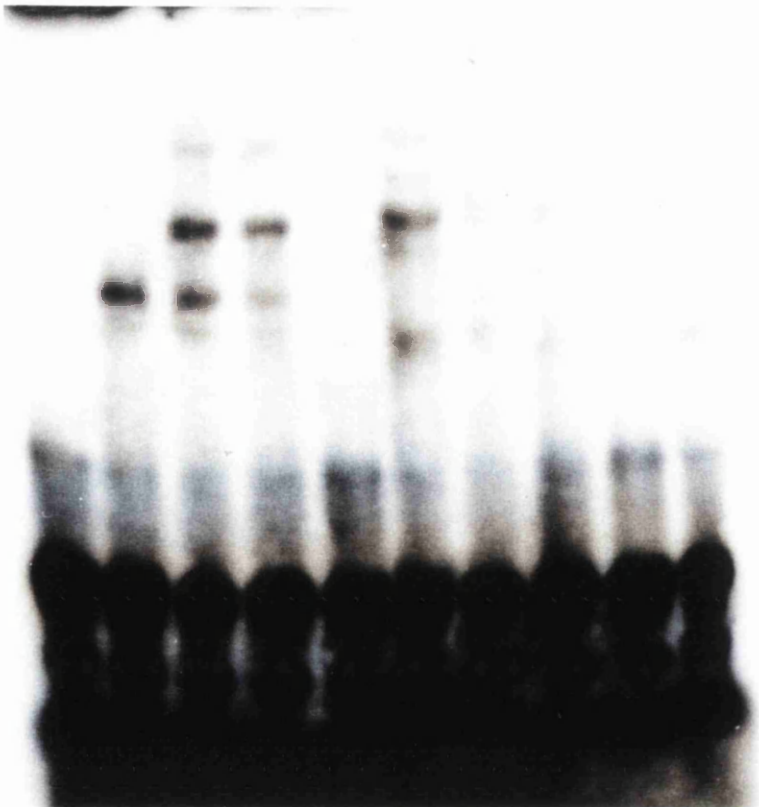


Figure 4.11 EMSA of the GATA sequence element with nuclear extracts. End 1 oligonucleotide GATA was incubated with crude nuclear extract derived from parasite or host enriched or 41°C D7 cultures. For cross competition experiments unlabeled oligonucleotides GATA, CAT1 and incubated with extract prior to labelled probe. Lanes 7 to 10 were competed with oligonucleotide GATA, lanes 11 to 14 were competed with oligonucleotide CAT1 and lanes 21 to 24 were competed with oligonucleotide TAT1. Lane 1) probe

Lanes 2) BL20 crude nuclear extract

Lane 3) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 4) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 5) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 6) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 7) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 8) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 9) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 10) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 11) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 12) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 13) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 14) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 15) Probe

Lane 16) BL20 crude nuclear extract

Lane 17) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 18) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 19) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 20) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 21) Host enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 22) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C

Lane 23) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

Lane 24) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

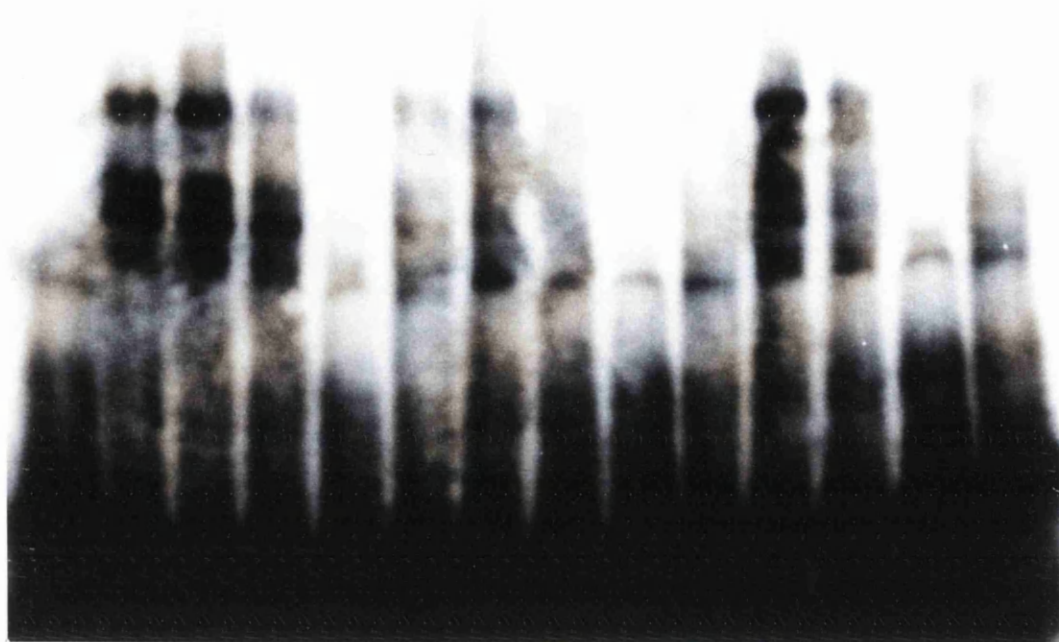
Lower case letter on bottom of tracks denote origin of extract: hs, host specific; h, host enriched; p, parasite enriched.

COMPETITION →

NONE

GATA

CAT1



hs

h

p

h

p

h

p

h

p

h

p

h

p

h

p

1

2

3

4

5

6

7

8

9

10

11

12

13

14

COMPETITION →

NONE

TAT1



hs

h

p

h

p

h

p

h

p

15

16

17

18

19

20

21

22

23

24

Figure 4.12 EMSA of the TAT1 sequence element with nuclear extracts. End 1 oligonucleotide TAT1 was incubated with crude nuclear extract derived from parasite or host enriched or 41°C D7 cultures. For cross competition experiments unlabeled oligonucleotides TAT1, CAT1 and were incubated with extract prior to addition of labelled probe. Lanes 7 to 10 were compete oligonucleotide TAT1, lanes 11 to 14 were competed with oligonucleotide CAT1 and lanes 21 to 24 competed with oligonucleotide GATA.

- Lane 1) probe
- Lanes 2) BL20 crude nuclear extract
- Lane 3) Host enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 4) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 5) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 6) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 7) Host enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 8) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 9) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 10) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 11) Host enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 12) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 13) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 14) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 15) Probe
- Lane 16) BL20 crude nuclear extract
- Lane 17) Host enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 18) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 19) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 20) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 21) Host enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 22) Parasite enriched nuclear extract prepared from D7 cells incubated at 37°C
- Lane 23) Host enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days
- Lane 24) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days

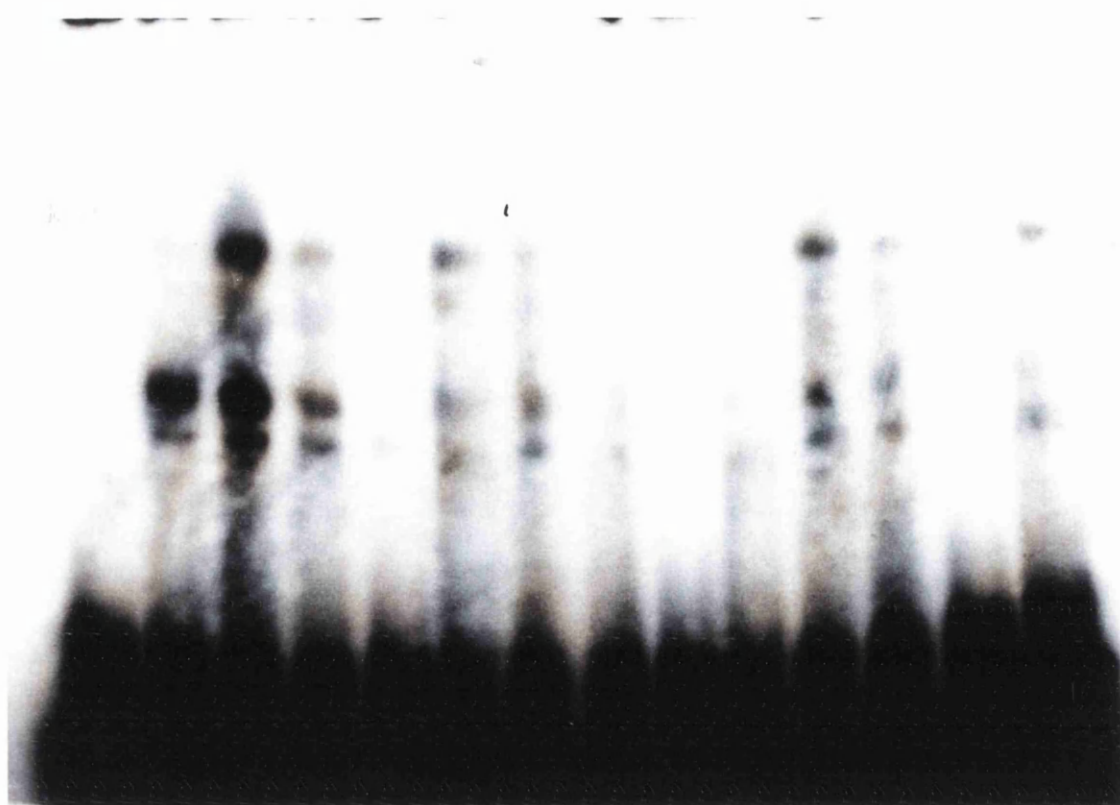
Lower case letter on bottom of tracks denote origin of extract: hs, host specific; h, host enriched; p, parasite enriched.

COMPETITION →

NONE

TAT1

CAT1



hs h p h p h p h p h p h p
1 2 3 4 5 6 7 8 9 10 11 12 13 14

COMPETITION →

NONE

GATA



hs h p h p h p h p
15 16 17 18 19 20 21 22 23 24

parasite associated nuclear factors although, in a similar fashion to the Tams1pD probe, a degree of competition of binding to CAT1 was obtained with cold TAT1. Based on these results it was decided to study binding to CAT1 in more detail as it appeared to possess a core sequence element responsible for the binding of parasite associated nuclear factors to the 120bp Tams1pD probe.

4.3.9 Effect of Poly d(I-C) concentration on Protein Binding to CAT1

Poly d(I-C) is used in binding reactions to reduce low affinity, non-specific binding of probes with proteins in nuclear extracts. However, complexes can be intolerant of variations in poly d(I-C) concentrations and this can be an indication of how stable a complex is when bound. Thus, poly d(I-C) can be used to assess complex stability by investigating the effect of different concentrations on the level of mobility shift formation. Probe CAT1 was incubated with increasing concentrations poly d(I-C), in the presence of parasite enriched nuclear extracts from D7 cells incubated at 41°C for 6 days shown in Figure 4.13.

As shown in Figure 4.13, between CAT1 and nuclear factors all concentrations of poly d(I-C) allow efficient complexes to form, and in the absence of poly d(I-C) the complex becomes more non-defined (Figure 4.13, Lane 2). The reduced clarity of this mobility shift may be due to a higher level of low affinity non-specific protein binding to the CAT1 probe which competes with specific complex formation. It was concluded from this experiment that the CAT1 complex was stable over a 10-fold increase of poly d(I-C) relative to the level used in the standard EMSA conditions.

4.3.10 Analysis of Protein Binding to CAT1

From the EMSA experiments performed with CAT1, three mobility shifts were detected which are likely to represent different complex formation of nuclear factors with the DNA probe. Of these complexes two A and B were detected when parasite enriched extracts were made from both 37°C cultures and differentiating cultures of D7, although the level of complex formation frequently appeared greater with extracts representing differentiating cells. The third complex (C) appeared only to be detected when extracts from differentiating cells were used in the analysis. However, it was unclear whether the

Figure 4.13 Binding of radiolabelled CAT1 in the presence of Poly d(I-C). End labelled oligonucleotide CAT1 was incubated with parasite enriched nuclear extract derived from D7 infected cell cultures incubated at 41°C for 6 days and increasing concentrations of specific competitor poly d(I-C) as indicated.

Lane 1) Probe only

Lane 2) no poly d(I-C)

Lane 3) 100ng/ml poly d(I-C) stock

Lane 4) 200ng/ml poly d(I-C) stock

Lane 5) 300ng/ml poly d(I-C) stock

Lane 6) 400ng/ml poly d(I-C) stock

Lane 7) 500ng/ml poly d(I-C) stock

Lane 8) 600ng/ml poly d(I-C) stock

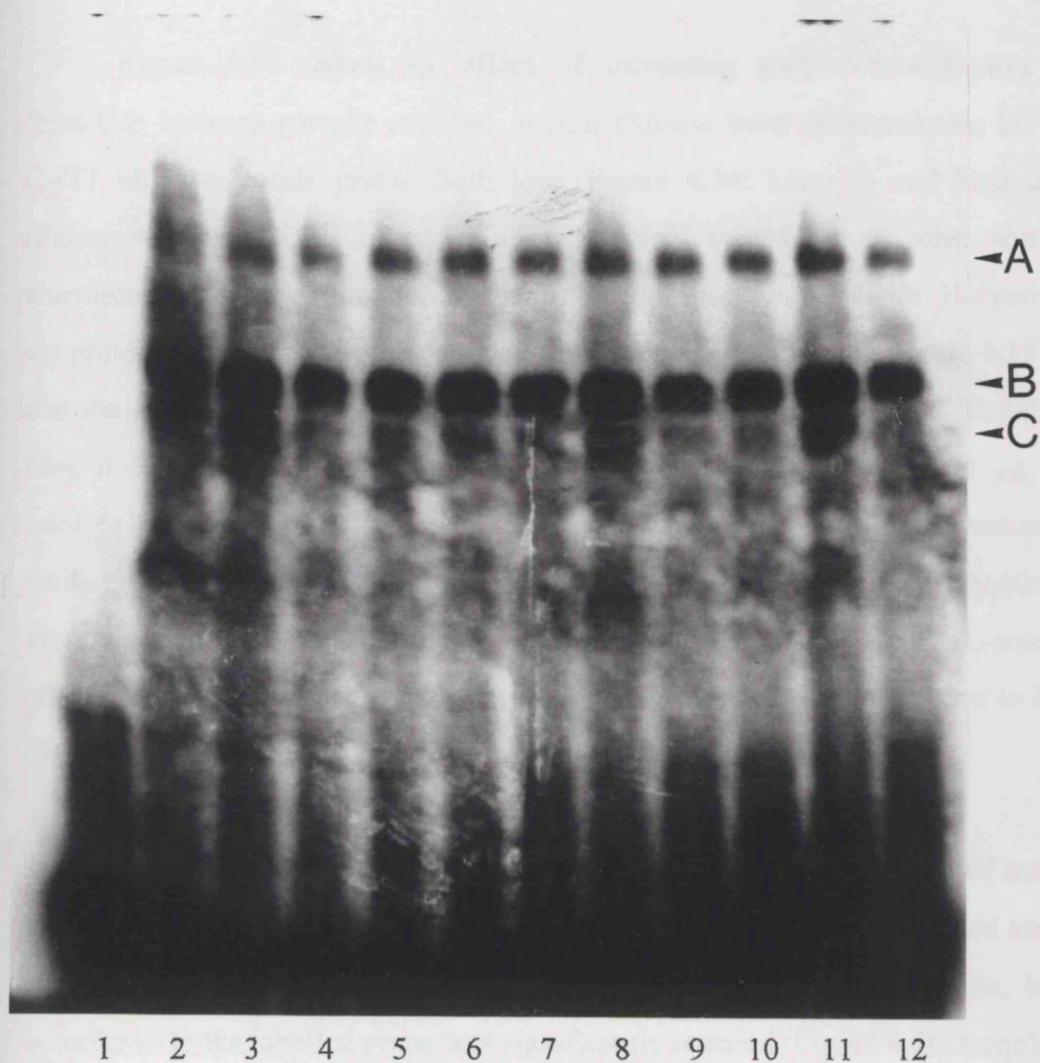
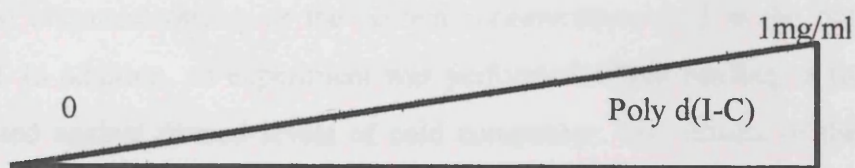
Lane 9) 700ng/ml poly d(I-C) stock

Lane 10) 800ng/ml poly d(I-C) stock

Lane 11) 900ng/ml poly d(I-C) stock

Lane 12) 1mg/ml poly d(I-C) stock

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B, C).



formation of the different complexes were related to each other and whether they could have different properties relative to each other. For example, different complexes could display differences in the affinity of the nuclear factor(s) to the DNA probe. To investigate possible differences in affinity, titration experiments were conducted where either the level of labelled oligonucleotides, or the protein concentration used in the binding reactions was altered. In addition, an experiment was performed where binding to the CAT1 probe was competed against titrated levels of cold competitor. The results of these experiments are shown in Figures 4.14, 4.15, and 4.16.

Figure 4.14 shows the effect of increasing probe concentration on complex formation between parasite enriched nuclear extracts from differentiating D7 cells and the CAT1 oligonucleotide probe. Both low (Figure 4.14; Lane 2) and high concentrations (Figure 4.14; Lane 12) of probe allow efficient complexes to form with no obvious alteration in complex formation as the probe concentration increases. However, titration of the protein concentration of the parasite enriched extract shown in Figure 4.15 demonstrates that the intensities of the three detected mobility shifts decreased at different rates. As the protein concentration drops from 0.48 μ g (Figure 4.15; Lane 1) to 0.032 μ g (Figure 4.15; Lane 6) the intensity of shift C was seen to significantly decrease, and eventually disappear, while the intensities of shifts A and B dropped but remained visible. Interestingly when the protein concentration was reduced to below 0.12 μ g the complex represented by shift A altered as it appeared that the intensities of shift A became greater relative to B at the lower extract concentrations.

The effect of altering the level of cold CAT1 on the formation of complexes A, B and C is represented by Figure 4.16. In the presence of parasite enriched nuclear extracts from differentiating D7 and x200 of unlabelled CAT1 oligonucleotide, binding of all complexes to the labelled probe was significantly reduced (Figure 4.16; Lane 1). Complex C was eliminated by the addition of only x20 competitor (Figure 4.16; Lane 6) while Complexes A and B survived. As shift B appears to be the most abundant of the three in the non-competed reactions it may be that it is more difficult to compete because of higher level of complex formation. However, affinity difference between the different complexes could also play a role and cannot be discounted from the obtained result.

Figure 4.14 Titration of probe concentration in CAT1 EMSA. Increasing amounts of were reacted with parasite enriched nuclear extracts prepared from D7 cells incubated at 41°C for 6 days.

Lane 1) no CAT1

Lane 2) 0.002pmoles of CAT1

Lane 3) 0.2pmoles of CAT1

Lane 4) 0.3pmoles of CAT1

Lane 5) 0.4pmoles of CAT1

Lane 6) 0.5pmoles of CAT1

Lane 7) 0.6pmoles of CAT1

Lane 8) 0.7pmoles of CAT1

Lane 9) 0.8pmoles of CAT1

Lane 10) 0.9pmoles of CAT1

Lane 11) 1pmoles of CAT1

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, C).

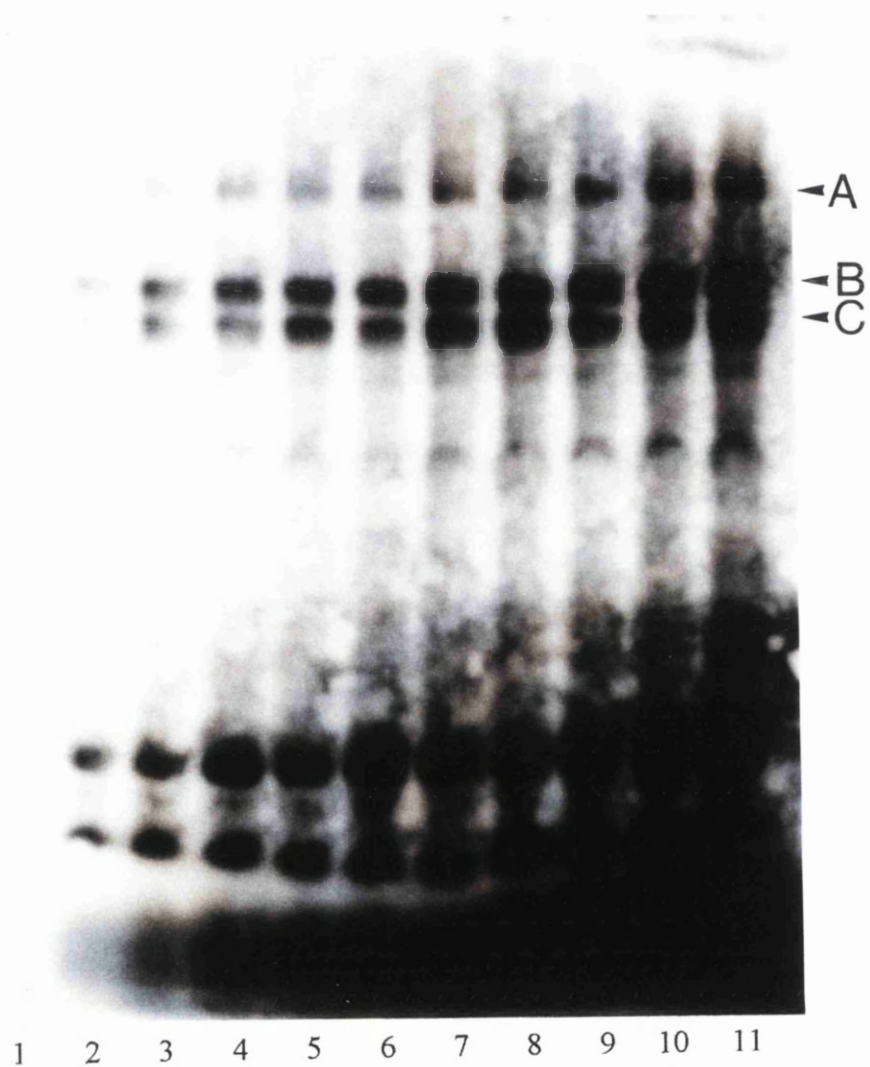
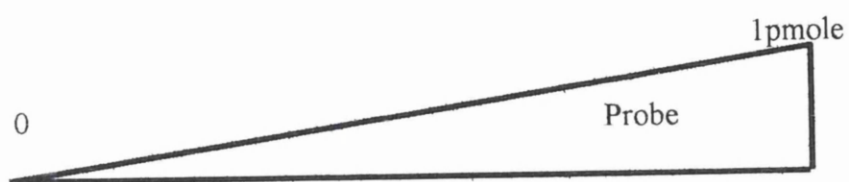


Figure 4.15 Titration of extract concentration in CAT1 EMSA. ^{32}P -labelled CAT reacted with decreasing amounts of parasite enriched nuclear extract prepared from D' incubated at 41°C for 6 days.

Lane 1) 0.48µg of parasite nuclear extract per reaction

Lane 2) 0.36µg of parasite nuclear extract per reaction

Lane 3) 0.24µg of parasite nuclear extract per reaction

Lane 4) 0.12µg of parasite nuclear extract per reaction

Lane 5) 0.048µg of parasite nuclear extract per reaction

Lane 6) 0.032µg of parasite nuclear extract per reaction

Lane 7) 0.024µg of parasite nuclear extract per reaction

Lane 8) 0.0192µg of parasite nuclear extract per reaction

Lane 9) 0.016µg of parasite nuclear extract per reaction

Lane 10) 0.0137µg of parasite nuclear extract per reaction

Lane 11) 0.012µg of parasite nuclear extract per reaction

Lane 12) Probe only

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, C).

0.48 μ g

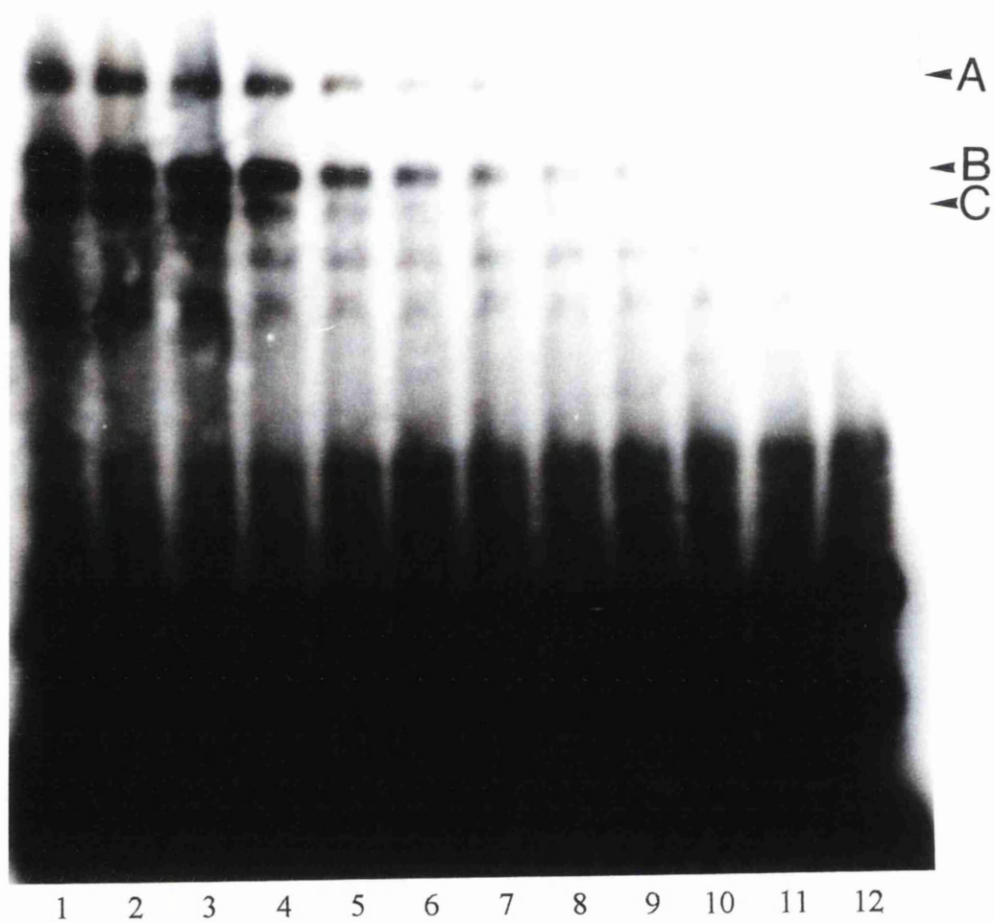


Figure 4.16 Competition experiments showing specificity of CAT1-extract complex formation. Mobility shift assay of parasite enriched extracts reacted with radiolabelled CAT1 in the absence and presence of decreasing amounts of unlabelled DNA probe.

Lane 1) 200

Lane 2) 100 fold excess

Lane 3) 80 fold excess

Lane 4) 60 fold excess

Lane 5) 40 fold excess

Lane 6) 20 fold excess

Lane 7) 10 fold excess

Lane 8) 5 fold excess

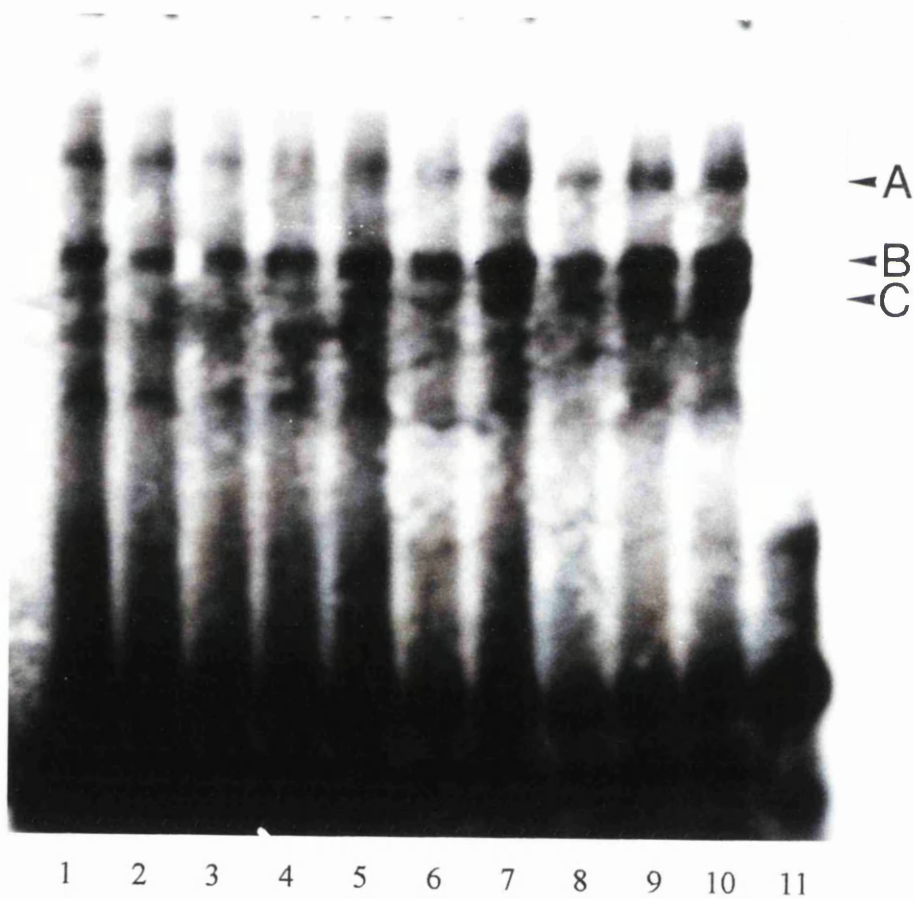
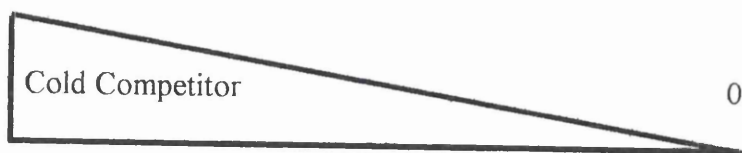
Lane 9) 1 fold excess

Lane 10) no competitor

Lane 11) Probe only

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B, C).

200x

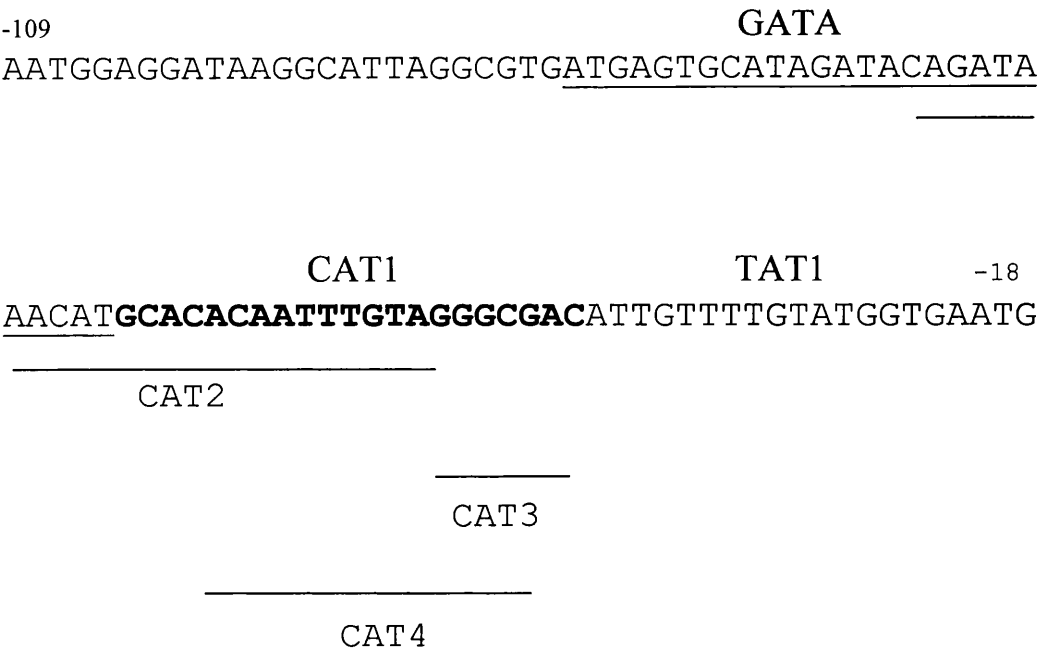


4.3.11 CAT1 Binding Elements

From the EMSA experiments performed to this point with the CAT1 probe, it was unknown whether the complexes A, B and C were related, or distinct from, each other in terms of the polypeptides which bound to them. For example, the different shift patterns could be due to the binding of single distinct polypeptides or be intermediates in the formation of a final complex of two or more different polypeptides or the probe. In the event that distinct polypeptides could recognise different sequences, one possible way to try and characterise the differences between the three complexes was to investigate whether they were conferred by different sequences.

Initially this was carried out by generating double stranded oligonucleotides representing different defined regions or variants of the CAT1 probe, CAT 2, 3, and 4. Figure 4.17 shows the sequence of the 3 oligonucleotides (CAT2, CAT3 and CAT4) and their relationship to CAT1. Oligonucleotide CAT2 consisted of sequence from the GATA and CAT1 oligonucleotides, but lacked the last 7 (3') base pairs of CAT1. CAT3 essentially encompassed these 3' nucleotides of CAT1 missing from CAT2, the A at position 14 of CAT1 being the 3' and 5' terminal nucleotides of CAT2 and CAT3 respectively. CAT4 was a truncated version of CAT1, the 5' and 3' nucleotides of CAT4 corresponding to positions 5 and 20 of CAT1.

These oligonucleotides and CAT1 were incubated with parasite enriched nuclear extracts prepared from 37°C and differentiating D7 cell cultures (Figure 4.18). As has been observed previously, the CAT1 probe formed 3 complexes A, B, and C when incubated with extracts prepared from differentiating parasite nuclei (Figure 4.18; Lane 1B). Furthermore, it also confirmed that the CAT1 probe formed complexes A and B when incubated with parasite extracts prepared from D7 cells cultured at 37°C (Figure 4.18; Lane 1A). The CAT4 probe, incubated with parasite extracts from differentiating cultures also formed complexes A, B and C (Figure 4.18; Lane 4B), with reduced efficiency and the representative shifts were not obtained when CAT4 was incubated with the 37°C parasite enriched nuclear extract. A shift of similar mobility to that representative of CAT1 complex A was also observed for the CAT2 probe (Figure 4.18; Lane 2B) but, shifts at the same



Oligonucleotides:

- CAT 1 - GCACACAATTTGTAGGGCGAC
- CAT 2 - AGATAAACATGCACACAATTTGTA
- CAT 3 - AGGGCGAC
- CAT 4 - ACAATTTGTAGGGCGA

Figure 4.17. The area depicted represents part of the Tams1pD fragment which is proximal to the RNA initiation site of the *TamS1* gene. The numbers indicate the position of the sequence with respect to the start of transcription. The composition and position of the oligonucleotides used in the mobility shift assay are shown above.

relative mobility as CAT1 complex B and C were not observed. Two additional mobility shifts were detected for the CAT2 probe when incubated with parasite nuclear extract for D7 37°C cells (Figure 4.18; Lane 2A). The CAT3 probe did not generate any detectable mobility shift with either of the extracts.

To confirm the results of the previous experiment, a competition experiment was performed, where labelled CAT1 probe was competed with x100 unlabeled double-stranded, CAT2, CAT3 and CAT4, in the presence of parasite enriched nuclear extracts, (Figure 4.19). Comparisons between competed and non-competed reactions show that only CAT4 prevented the formation of the 3 complexes (Figure 4.19; Lane 4B). These complexes were found to be stable in the presence of high concentrations of unlabelled oligonucleotides CAT 2 (Figure 4.19; Lanes 2A and 2B) and 3 (Figure 4.19; Lanes 3A and 3B). In contrast to the previous experiment, shifts representative of complexes A and B were not clearly detected with the parasite enriched nuclear extracts (Figure 4.19; Lane 1A).

From these results it appeared that neither the CAT2 or CAT3 oligonucleotides were able to reproduce complex formation to generate the A, B and C mobility shifts. The CAT2 probe did generate a mobility shift of similar size to the CAT1 A complex (Figure 4.18, Lane 2B), however, the inability of the cold CAT2 to compete the formation of this shift from the labelled CAT1 probe suggests that the two complexes are not identical (Figure 4.19, Lane 2B). In contrast the CAT4 probe reproduced all three CAT1 mobility shifts. It was concluded that the sequence of CAT1 and CAT4 which crosses the junction of CAT2 and CAT3 are necessary for complex formation. Moreover in that the binding sites for all 3 complexes are contained within the 16 nucleotides of CAT4.

4.3.12 Generation of Mutations that Abolish Specific Binding to CAT1

To analyse further the nature of the sequences motif(s) within CAT1 which confer formation of complexes A, B and C with parasite enriched nuclear extracts site specific mutant oligonucleotides were generated. It was hoped that if there were overlapping binding sites which preferentially contributed to the formation of one shift complex over another, then this would be revealed by use of these mutants. As the CAT4 oligonucleotide generated all three complexes, mutants were made within this sequence. Three mutants were

Figure 4.18 Binding activity of the CAT 1, 2, 3 and 4 probes in parasite enriched extracts prepared from (A) D7 cells cultured at 37°C or (B) 41°C for 6 days.

- 1) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT1
- 2) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT2
- 3) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT3
- 4) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT4

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B and C).

1		2		3		4	
A	B	A	B	A	B	A	B

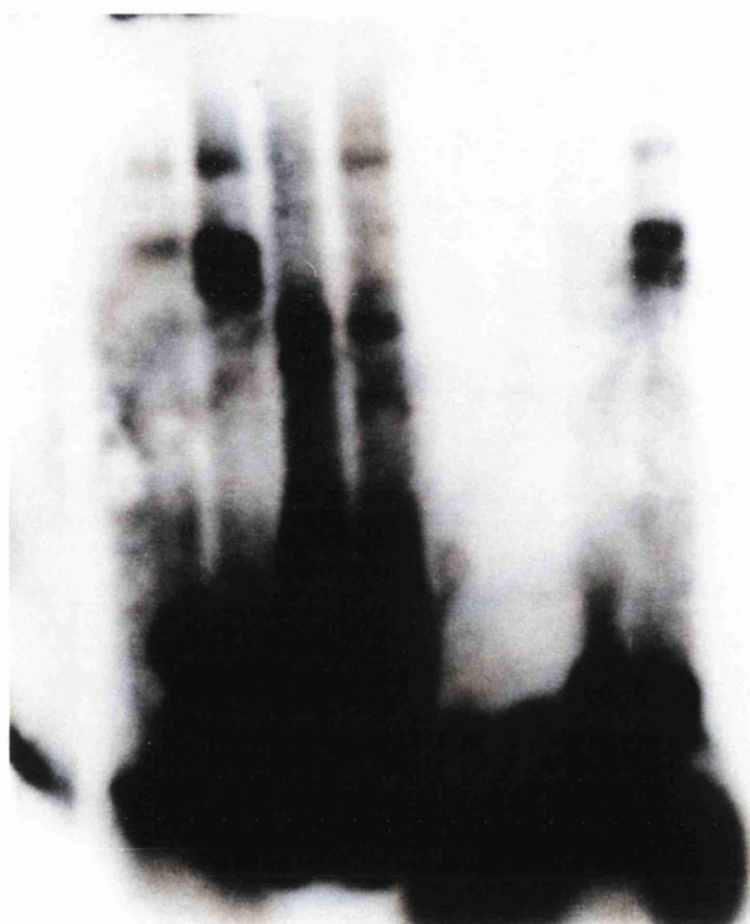
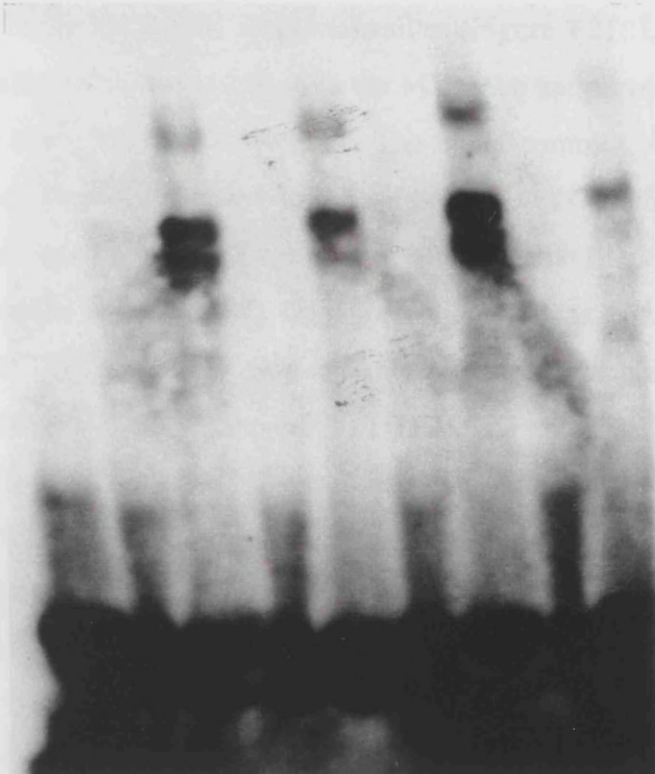


Figure 4.19 EMSA competition experiments of CAT1 complex formation. Mobility shift assay of parasite enriched nuclear extracts derived from D7 cells cultured at 37°C (A) and 41°C (B) reacted with ^{32}P -labelled CAT1 in the absence and presence of unlabeled DNA probes (100-fold excess).

- 1) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT1
- 2) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT1 in the presence of unlabelled CAT2
- 3) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT1 in the presence of unlabelled CAT3
- 4) Parasite enriched nuclear extracts reacted with ^{32}P labelled CAT1 in the presence of unlabelled CAT4

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B, C).

1	2	3	4
┌ A B ┐	┌ A B ┐	┌ A B ┐	┌ A B ┐



─A
 ─B
 ─C

made, differing from CAT1 by only 3 base changes (Figure 4.20). These triplets were chosen to span the CAT1 motif but in particular were designed to disrupt the a) ACACA pentamer, b) the TTT triplet and the GGG triplet and c) the CTA triplet which is between the TTT and GGG triplets.

The three mutant double stranded oligonucleotides and CAT1 probe were incubated in the presence of parasite enriched extracts made from D7 cells incubated at 41°C for 6 days as shown in Figure 4.21. Formation of the complexes A, B and C was limited to the CAT1 (Figure 4.21; Lane 4) and M1 oligonucleotides (Figure 4.21; Lane 3). Complexes B and C appeared unaffected by the mutation in the M1 probe and formed as effectively with the mutant as with the CAT1 oligonucleotide. The upper complex A, however showed a significant reduction in intensity relative to the wild type, CAT1. Both mutants M2 and M3 appeared to abrogate formation of all 3 complex, although faint levels of a shift at the C position were detectable with the M3 oligonucleotide. It was concluded that the core binding site for CAT1 complex formation is likely to reside after base pair 7 with critical residues likely to be contained within the TTTGTAGGG sequence. Additional shifts to those normally observed with CAT1 were detected with mutant oligonucleotides, in particular M2 and M3. It was unclear whether they represent specific shifts or are caused by non-specific binding of factors whose access to the probe is normally blocked by the specific complex formation.

To confirm the binding experiments with the mutant oligonucleotides, a competition assay using cold mutant oligonucleotides was carried out. Increasing concentrations of unlabelled mutant oligonucleotides, at 50, 100 and 200 fold excess, were added to parasite enriched nuclear extracts made from D7 cells at 41°C for 6 days prior to incubation with the labelled CAT1 probe. Formation of the complexes A, B and C, were significantly reduced in the presence of x50 excess of unlabelled M1 oligonucleotide (Figure 4.22, Lane 9) and almost were totally removed x200 competitor (Figure 4.22; Lane 7). In contrast, the three base pair alterations within the M2 and M3 mutant oligonucleotides destroyed the ability of these oligonucleotides to effectively compete for binding of the three complexes (Figure 4.22; Lanes 1 - 6), although M2 may have competed complex C relative to complex A or B, and also relative to the M3 oligonucleotides. Thus the competition experiments with the

mutants confirmed that results of the direct binding assays. These experiments seem to rule out the possession of non-overlapping multiple distinct binding sites within the CAT1 region.

4.3.13 Identification of Transcription Factors

4.3.14 UV Crosslinking of Transcription Factors to CAT1 Sequence

If factors which were quantitatively distinct were binding to generate complexes A, B, and C then one possible way to characterise them would be by their molecular masses. This information would also be useful for further studies to purify the transcription factors. To attempt to determine the number and relative molecular mass of proteins involved in recognising the CAT1 sequence, UV fixation studies were conducted using parasite enriched nuclear extracts made from D7 cells incubated at 41°C for 6 days. The assay relies on DNA-protein complexes subjected to UV radiation forming covalent bonds between the DNA and bound protein. Substituting thymidine residues with bromodeoxyuridine allows the DNA to crosslink with protein more readily and also allows a lower wavelength of UV light to be used, reducing damage to the protein. Firstly, single-stranded DNA was made from a pGem7zf clone containing an insert of concatenated CAT1. As internally radiolabelled double stranded DNA probe, incorporating, 5-bromo-2'-deoxyuridine triphosphate, was then generated as described and used in a standard binding assay. Following incubation on ice binding reactions were exposed to UV radiation for 0, 5, 10, 15, 20, 30, 40, 50 and 60 minutes. Unbound DNA was then digested with DNaseI and the samples analysed by SDS-PAGE. The mobility of radiolabelled DNA/protein complex was estimated.

The optimal time for irradiation varies from one protein-DNA complex to another. By performing a series of time points between 5 and 60 minutes it was possible that the best time for fixation of each complex could be determined. Exposure times of 30 minutes or more to UV irradiation detected 6 distinct bands whose intensity increased with time (see Figure 4.23, Lanes 6-9). In particular a significant increase was observed for bands at 45kDa and 97kDa and the irradiation time point appeared to result in a significant increase of bands with estimated mass of 45kDa and 97kDa. In order to distinguish specific from non-specific binding the UV fixation experiment was repeated in the presence of cold

CAT 1 MUTANTS

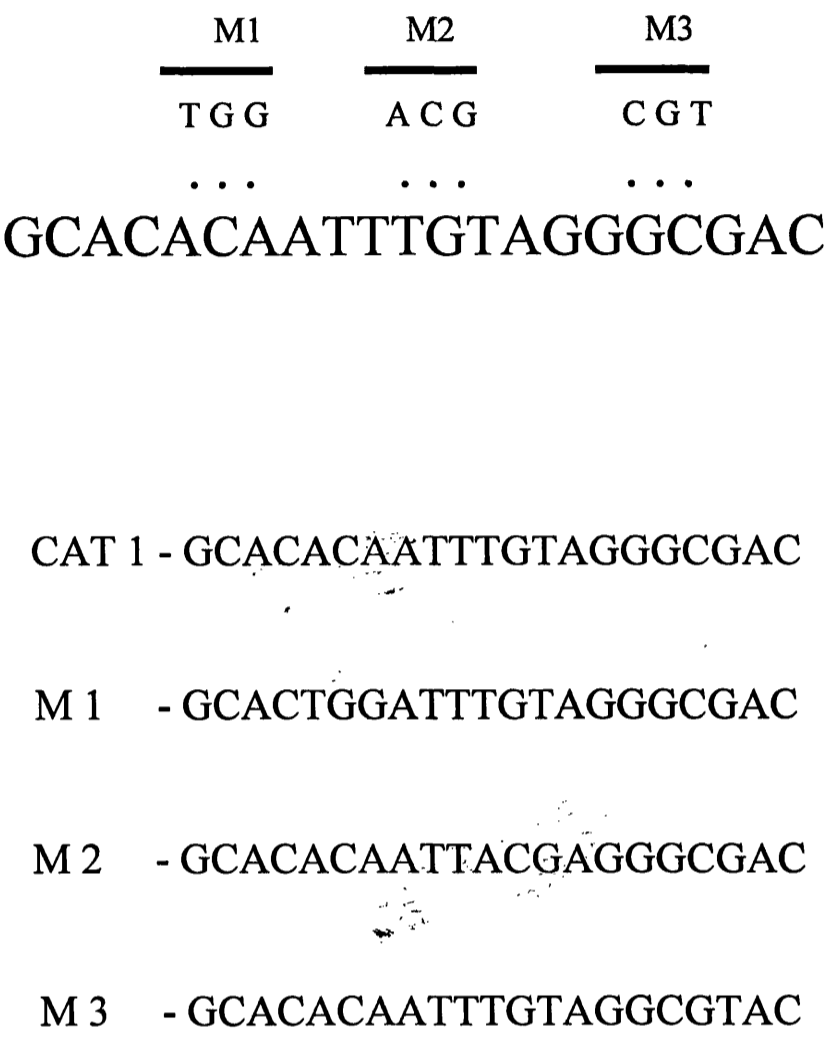


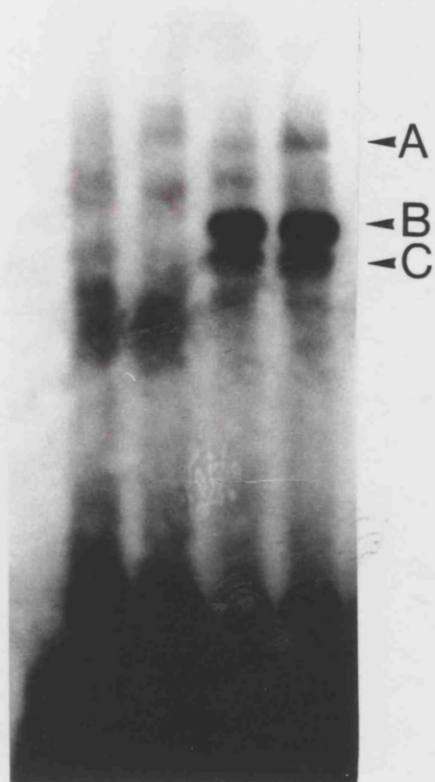
Figure 4.20. Base substitutions of CAT1. Closed circles below the nucleotides indicate the bases of CAT1 which were substituted with the nucleotides denoted above the closed circles. The complete sequences of the mutated oligonucleotides and wild type CAT1 are as indicated.

Figure 4.21 EMSA and base substitution analysis of CAT1. Mobility shift assay differentiating parasite enriched nuclear extracts reacted with ^{32}P -labelled CAT1 or M1, M2, M3.

- 1) Parasite enriched extracts reacted with ^{32}P labelled M3
- 2) Parasite enriched extracts reacted with ^{32}P labelled M2
- 3) Parasite enriched extracts reacted with ^{32}P labelled M1
- 4) Parasite enriched extracts reacted with ^{32}P labelled CAT1

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B, C).

M3 M2 M1 WT



1 2 3 4

Figure 4.22 Competition experiments using base substitution mutants to demonstrate specificity of binding to the CAT1 element. Mobility shift assay of differentiated macroschizont parasite extracts reacted with ^{32}P -labelled CAT1 in the absence and presence of increasing unlabelled DNA probes (50, 100 and 200 fold excess)

Lane 1) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 200 fold excess of unlabelled M3

Lane 2) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 100 fold excess of unlabelled M3

Lane 3) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 50 fold excess of unlabelled M3

Lane 4) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 200 fold excess of unlabelled M2

Lane 5) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 100 fold excess of unlabelled M3

Lane 6) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 50 fold excess of unlabelled M2

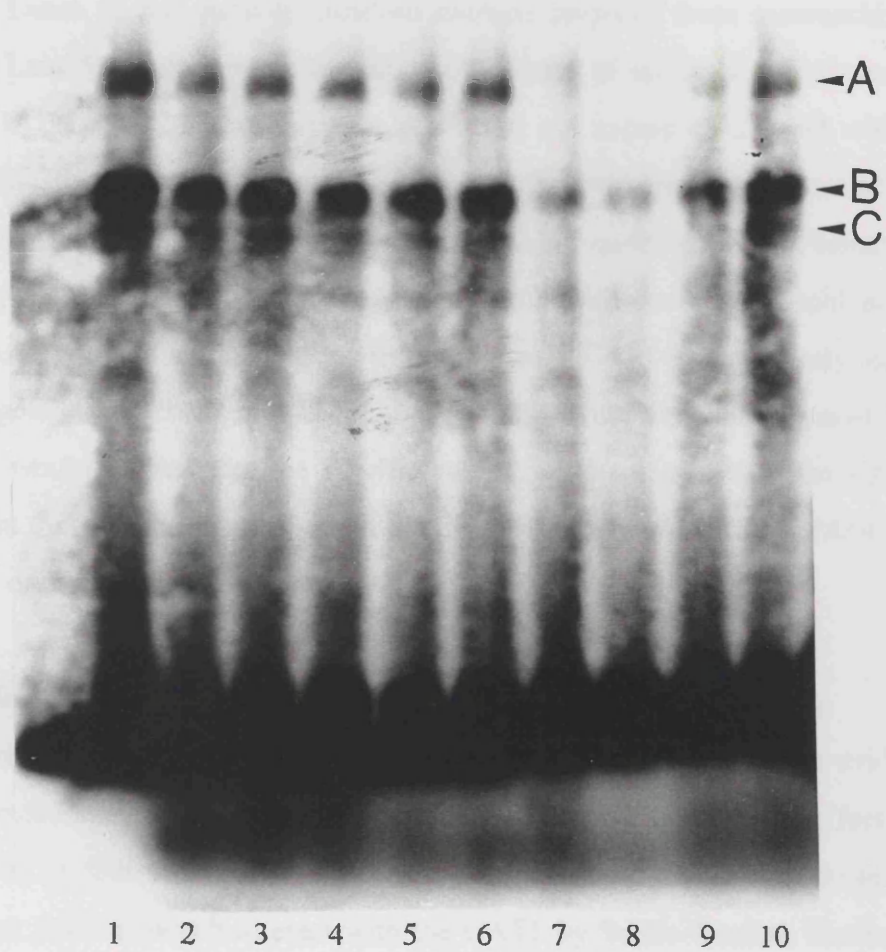
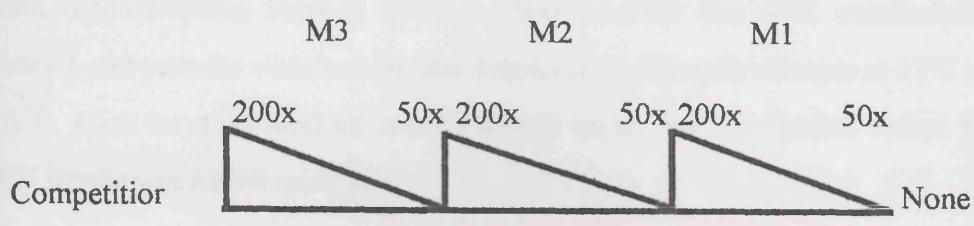
Lane 7) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 200 fold excess of unlabelled M1

Lane 8) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 100 fold excess of unlabelled M1

Lane 9) Parasite enriched extracts reacted with ^{32}P labelled CAT1 in the presence of 50 fold excess of unlabelled M1

Lane 10) Parasite enriched extracts reacted with ^{32}P labelled CAT1

Arrowheads on figures denote shifts detected in enriched parasite nuclear extracts (A, B and C).



competitor oligonucleotides known to remove the radiolabelled complexes observed in EMSA. In addition, competition reactions were also carried out with the 3 mutant CAT1 oligonucleotides: and, to try and distinguish parasite from host factors and factors associated with differentiation binding reactions were carried out with uninfected BL-20 nuclear extract and parasite enriched nuclear extract from D7 cells cultures at 37°C and for 6 days at 41°C. Each reaction mixture was incubated on ice for 10 minutes before being exposed to UV irradiation for 60 minutes.

The results of these experiments are shown in Figure 4.24. Incubation of the bromodeoxyuridine probe with extracts derived from differentiating macroschizont cultures (Figure 4.24; Lanes 4) and parasite enriched extracts prepared from macroschizont cells (Figure 4.2.4; Lane 5) resulted in the formation of 6 bands of similar size to those observed in Figure 4.23 (Lane 9). Interestingly, the probe did not appear to interact with extracts derived from BL20 (Figure 4.24; Lane 1) and macroschizont host enriched extracts (Figure 4.24; Lane 2) suggesting that the 6 bands were specific to parasite extracts. Formation of the 6 bands using the differentiating macroschizont host enriched extracts could possibly be related to parasite contamination resulting from the parasite load. Unfortunately, none of the competing oligonucleotides (Figure 4.24; Lanes 6 - 12) inhibited the formation of any of the 6 bands. This would suggest that the specific complexes associated with the CAT1 probe did not form in these experiments and that the detected bands on the SDS-PAGE were not representative of complex A, B or C.

4.3.15 Identification of DNA-Binding Proteins by South-Western Blotting

The previous UV crosslinking experiments did not show conclusive evidence that any of the detected bands on SDS-PAGE were specific interactions of nuclear factor(s) with the CAT1 probe. A final attempt was made to try to determine the mass of proteins within parasite nuclear extracts which interact with the CAT1 by South-Western blotting. Whole cell protein samples of BL20, D7 at 37°C and differentiating D7 cultures were run on a SDS-polyacrylamide gel and blotted on to a nitro-cellulose membrane. Ponceau S staining of the membranes confirmed that protein had been transferred prior to washing and renaturing of the protein on the membrane overnight at 4°C. The nitro-cellulose membrane

Figure 4.23 UV crosslinking of mobility shift reaction. Parasites enriched nuclear extracts from differentiating D7 cells were reacted with ^{32}P labelled probe followed by UV irradiation of increasing duration.

Lane 1) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1

Lane 2) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 5 minutes

Lane 3) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 10 minutes

Lane 4) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 20 minutes

Lane 5) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 30 minutes

Lane 6) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 40 minutes

Lane 7) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 50 minutes

Lane 8) Parasite enriched extracts reacted with ^{32}P labelled concatenated CAT1 and irradiated for 60 minutes

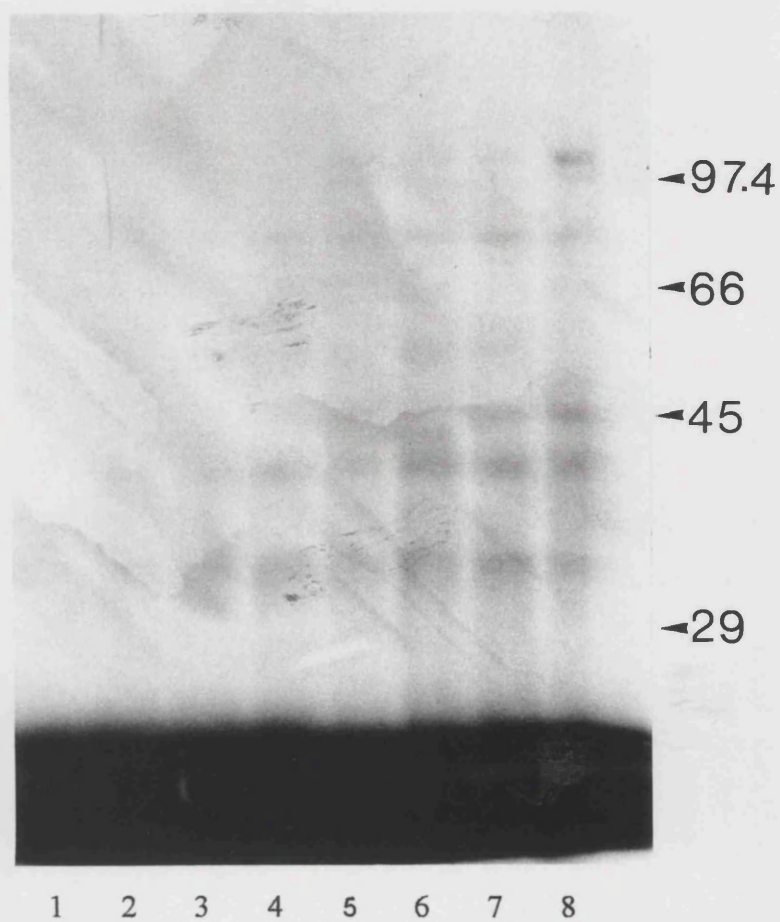
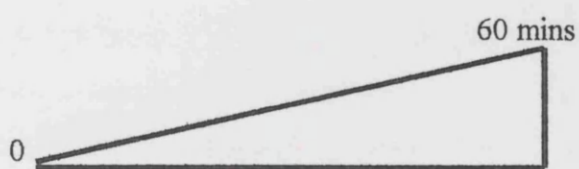


Figure 4.24 Competition analysis of UV crosslinking of mobility shift reactions. BL20 macroschizont and differentiating macroschizont extracts were reacted with concatenated CAT1 before UV irradiation for 60 minutes in the presence or absence of unlabelled protein (100 fold excess).

Lane 1) BL20 extracts reacted with ^{32}P labelled concatenated CAT1

Lane 2) Host enriched extracts prepared from D7 cells cultured at 37°C were reacted with ^{32}P labelled concatenated CAT1

Lane 3) Parasite enriched extracts prepared from D7 cells cultured at 37°C were reacted with ^{32}P labelled concatenated CAT1

Lane 4) Host enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1

Lane 5) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1

Lane 6) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled CAT1

Lane 7) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled CAT2

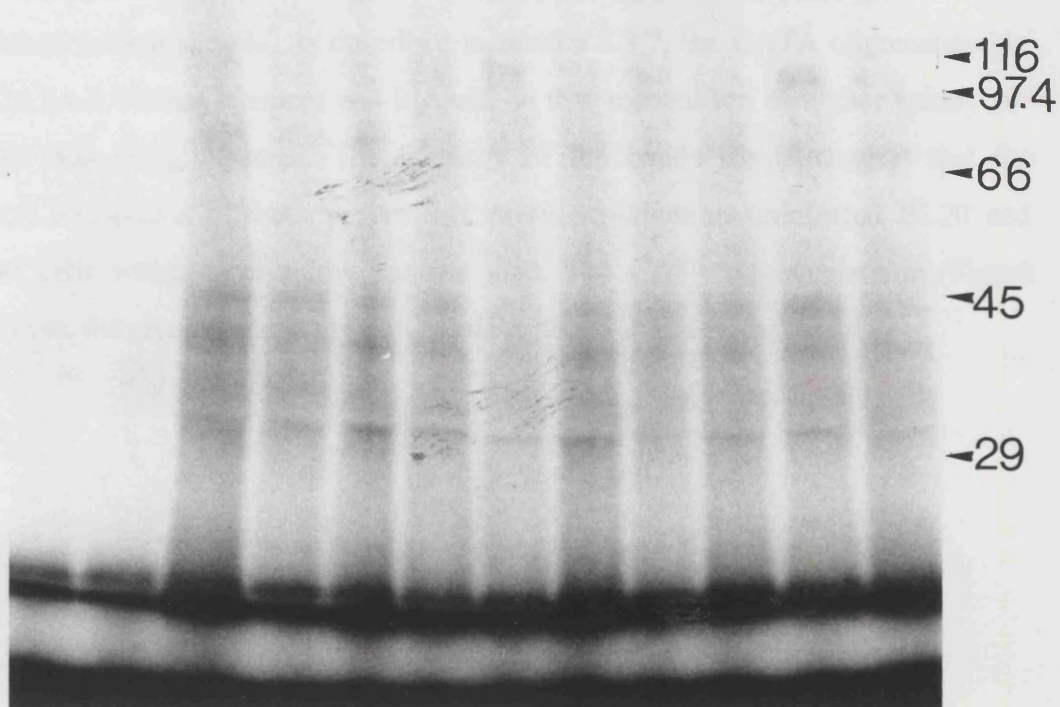
Lane 8) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled CAT3

Lane 9) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled CAT4

Lane 10) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled M1

Lane 11) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled M2

Lane 12) Parasite enriched extracts prepared from D7 cells incubated at 41°C for 6 days were reacted with ^{32}P labelled concatenated CAT1 in the presence of unlabelled M3



1 2 3 4 5 6 7 8 9 10 11 12

was incubated at room temperature with radiolabelled probes, Tams1pD, CAT1 and GATA for 3 hours washed and exposed to X-ray film (section 4.2.9).

South-Western blot analysis of the BL20 and D7 protein samples using the Tams1pD, CAT1 and GATA probes is shown in Figure 4.25. The Tams1pD (Figure 4.25b) and GATA (Figure 4.25C) probes recognized polypeptides in BL20 (Lane 4), D7 (lane 3) and differentiating D7 (Lanes 1 and 2) whole cell protein samples. The banding pattern obtained was similar for both probes and neither probe recognized any polypeptides in the parasite enriched nuclear sample. As described in section 4.3.7, the GATA oligonucleotide contains a GATA-1 binding element and it could be that recognition of higher eukaryotic GATA factor was being detected. The intensity of the bands would suggest that the polypeptide(s) recognized by both probes are more prevalent in uninfected BL20 and infected host cells which have a low parasite load. The CAT1 oligonucleotide (Figure 4.25A), however, did recognise not any polypeptides bound to the membrane.

Figure 4.25 South-Western blot analysis of CAT1, Tams1pD and GATA. Total cell extract of BL20, 37°C or 41°C D7 cultures and parasites enriched extract derived from D7 cells cultured at 41°C for 6 days were western blotted onto a nitro-cellulose membrane and hybridised with ^{32}P labelled probes (A) CAT1, (B) Tams1pD and (C) GATA.

Lane 1) Parasite enriched nuclear extract prepared from D7 cells incubated at 41°C for 6 days.

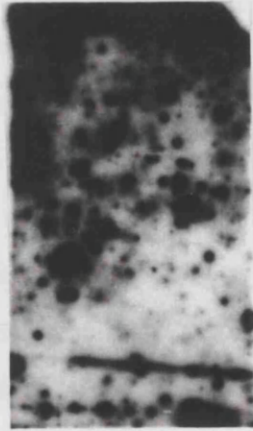
Lane 2) Parasite nuclei prepared from D7 cells incubated at 41°C for 6 days

Lane 3) Total cell extract prepared from D7 cells incubated at 37°C.

Lane 4) BL20 Total cell extracts

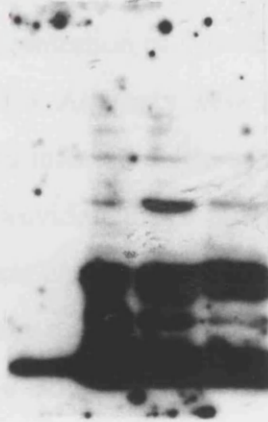
(A)

1 2 3 4



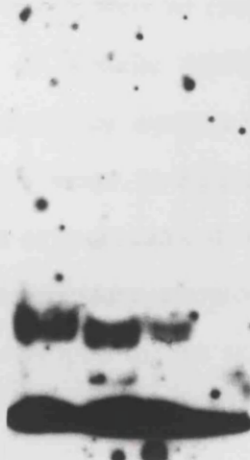
(B)

1 2 3 4



(C)

4 3 2 1



4.4 Discussion

Nuclear extracts with active transcription factors are usually obtained from relatively homogeneous cellular populations such as tissue-culture cell lines or isolated tissues (eg Liver). Macroschizont infected cells are a homogeneous population, when cultured at 37°C, but the parasite resides within the cytoplasm of the host cell. Therefore, before extracts enriched for parasite factors could be prepared, it was necessary to separate parasite from host nuclei. Attempts to isolate parasite nuclear extracts, which gave positive mobility shifts with the *TamSI* upstream region using Bionebulisation were unsuccessful. Therefore, this study adapted methodology employed by Lanzer *et al.*, (1992) to generate nuclear extracts from *Plasmodium falciparum*. The *Plasmodium* method has the advantage that the infected erythrocyte does not have a nucleus to contaminate parasite samples, therefore, to adapt this procedure to *Theileria* infected leukocytes, it was necessary to introduce a differential centrifugation step to enrich for parasite nuclei. Salt extraction of nuclear factors from parasite nuclei yielded a protein concentration of 0.1-0.2 µg/µl which was 10 fold less than that achieved by Lanzer *et al.*, (1992). Attempts were also made to use this procedure on piroplasms isolated from *T.annulata* infected cattle as 10 litre volumes of infected blood with parasitaemias of up to 80% provide a large amount of starting material for extract preparation. Unfortunately the nuclear extracts produced did not give any positive mobility shifts. As yet it is unknown why active nuclear extracts for the *TamSI* upstream region were not generated from piroplasms as isolated piroplasm nuclei have been shown to be transcriptionally active (David Swan, unpublished data) and the gene is highly expressed in this stage. Therefore, at present higher levels of extract could only be obtained by increasing the volume of tissue culture or increasing the efficiency of factor extraction.

Assessment of contamination during the nuclear preparation by both IFA and DAPI analysis of various steps in the protocol showed (see Figures 4.1, 4.2, and 4.3) that, in general, a significant enrichment of parasite nuclei was achieved by the developed methodology. The parasite enriched nuclear extracts were stored at -70°C for 3-4 months without significant loss of activity. However, as all extracts prepared were used within this period, the actual length of time that extracts can remain active has as yet to be determined. Thus, although optimisation may be necessary, a reproducible procedure for the preparation of parasite nuclear extracts was developed in this study and could be an important step

towards the further understanding of gene regulation in *Theileria annulata* using *in vitro* techniques.

To screen for factors which specifically bind to the promoter of the *TamS1* gene, 600bp of the 5' intergenic region proximal to the polypeptide open reading frame was divided into four fragments of approximately equal size using PCR amplification (see section 4.3.4). Putative cis-Regulatory elements involved in protein-DNA interactions were identified by incubation with parasite enriched nuclear extracts before separation on a non-denaturing polyacrylamide gel. The gel retardation assay located specific binding to the fragments most proximal and distal to the mapped transcription initiation site (see section 4.3.5).

Further work concentrated on the fragment Tams1pD, due to its location next to the transcription initiation site. Oligonucleotides representing species conserved sequences within Tams1pD were then used in competition and direct binding assays (see 4.3.7 and 4.3.8). This showed that one oligonucleotide, CAT1, competed the Tams1pD mobility shifts associated with parasite enriched nuclear extracts, but not host. Moreover, of the three oligonucleotides tested, CAT1 alone showed evidence of parasite associated shifts in direct binding assays. The specificity of the shifts was confirmed by competition with CAT1, whereas the GATA oligonucleotide, representing sequence flanking CAT1, failed to compete. The interaction of parasite specific factors with the CAT1 sequence was also competed with the TAT1 oligonucleotide. However, TAT1 did not completely compete the parasite associated mobility shifts obtained with Tams1pD and there was no evidence of direct binding to the TAT1 sequence. Sequence comparisons between CAT1 and TAT1 revealed that both oligonucleotides share a 6 base pair sequence, TTTGTA, which in combination with differences in flanking sequence could account for the different abilities to bind and compete for parasite specific factors. Thus, it can be hypothesised from this data that the TAT1 sequence may function to stabilise factor binding to the CAT1 motif but does not form a core domain which is essential for mobility shift formation.

The profile of complexes formed with the CAT1 element differed depending on the differentiation state of the parasite. Complexes A and B were detected in both

macroschizont and differentiating macroschizont extracts, while the presence of an extra complex (C) was only detected in differentiating macroschizont extracts. Therefore, the presence of complexes A and B in parasite enriched nuclear extracts from macroschizont cells suggested that factors bind to the 5' upstream region of *TamSI* in parasites cultured at 37°C. In addition a quantitative increase of complexes A and B was observed when parasite enriched nuclear extracts from macroschizont and differentiating macroschizont cells were compared, although this has as yet to be controlled for by a constitutive factor and complex C appeared to be specific to differentiating cells. This data could be accounted for by the model proposed by Shiels *et al.*, (1994). Here it was postulated that expression of the *TamSI* gene in the macroschizont could be due to either factors which have low functional activity after commitment, or is caused by a functional overlap of stage specific factors. In the second case it would be expected that the macroschizont would be down regulated or remain constant following commitment, while both models are consistent with a qualitative change over of factors after the commitment step. Thus it is possible that complexes A and B represent merozoite factors which function at a lower activity in the macroschizont and display increased activity up to and after commitment, and that complex C could represent a qualitative change in factor composition. These possibilities were supported by further studies analysing mobility shifts from cells attenuated for differentiation (see chapter 6). To clarify the observed shifts with the model of Shiels, it will be necessary for future studies to find an association between complex formation and functional activity.

The DNA-protein contacts of the three complexes were investigated by base substitutions within the CAT1 element. Surprisingly, all three mobility shift complexes were affected by base substitutions within the middle and the region most proximal to the RNA initiation site. Mobility shift A was also affected by the sequence mutation of ACA in the 5' region of the CAT1 sequence. This suggests that all three mobility shifts result from the recognition of either an identical or slightly overlapping sequence motifs. Future studies using saturation mutagenesis of the CAT1 sequence may further define the number and position of the motifs

As yet the relationship of each complex with the CAT1 sequence is unknown. However, there are at least three general mechanisms for interacting with a regulatory

element and it may be that a combination of these mechanisms that could account for the 3 complexes which form with the CAT1 probe. Firstly, ‘co-occupancy’ may occur, where factors from the different regulatory families interact specifically with the same region of DNA as well as with each other (Diamond *et al.*, 1990). Mutation analysis of the CAT1 probe has suggested that mobility shift A binds slightly upstream of complexes B and C providing evidence of a overlapping motif, and complexes A and B are both present in macroschizont parasite enriched nuclear extracts. The quantitative differences in the two complexes (see 4.3.10) would suggest that complex A is more likely to be the result of such a mechanism.

The second mechanism, denoted as ‘factor tethering’, where two different regulators establish protein-protein contacts on the DNA, but only one of the two actually binds to the DNA could also account for the appearance of either complex (Miner *et al.*, 1991; Liu and Green, 1990; Gaub *et al.*, 1990). ‘Factor tethering’ also has the flexibility of using proteins that specifically bind to the DNA motif; or without DNA-binding capability which could interact protein-protein with a factor that could require sequences in addition to those involved in binding to provide stabilisation to a complex. Such interactions could mediate a diverse range of functions, for example, acting as a bridging factor between the transcription factor binding to the DNA and the basal transcriptional machinery, stabilising the DNA-binding complex or changing the specificity of target sequence recognition (Calkhoven and AB, 1996). While the ‘factor tethering’ and ‘co-occupancy models’ bear obvious similarities, the co-occupancy mechanism provides greater site specificity, as functionality would only result from interaction with a motif bearing binding sites for both interacting factors. This means that within the same cell, individual regulators can still act independently at their simple response elements, or together with other factors at different composite elements.

The complexes observed in this study could also result from a ‘direct competition’ mechanism where factors from distinct regulator families compete for a motif by interacting in the absence of DNA, each factor having the potential to block the DNA-binding activity of the other. Here quantitative differences between factor levels determine which factor binds to the motif (Yang-yen *et al.*, 1990; Schule *et al.*, 1990). Alternatively, one factor may

have a greater affinity for the motif either/and preventing another factor from binding, or displacing a factor already bound; (Papazafiri *et al.*, 1991; Weissman and Singer, 1991). A modification of this last possibility is that a negatively acting factor can bind to a sequence adjacent to or overlapping the binding site and prevent positive factors binding by steric hindrance (Masquillier and Sassone, 1992; Lamph *et al.*, 1990). Therefore, direct competition could explain the binding of each of the complexes to the CAT1 probe. For example, titration of parasite enriched extracts prepared from differentiating macroschizont cells (see 4.3.10) suggested that complex A could be of higher affinity than B or C. In addition, the mobility of complex C is faster than that of mobility shifts A and B, and the formation of C occurs after that of A and B. Therefore, competition maybe required to remove complexes A and B so that complex C may form. The faster mobility of complex C could also be due either to 'factor tethering' or 'co-occupancy' which result in a structural change or alters the overall charge of the original complex.

Transcription activation occurs via protein-protein interactions, which bring the basal transcriptional machinery (or RNA polymerase and associated factors) into juxtaposition with the regulatory factors near the transcriptional start. It is currently believed that the activation domain of DNA binding factors interact directly or indirectly with co-activators or adapters which in turn bind to the basal machinery (Kelleher *et al.*, 1990; Berger *et al.*, 1990; Lewin, 1990(b)). It was demonstrated by EMSA in this study that the CAT1 sequence was the element closest to the RNA start site in *TamS1* which specifically interacted with parasite specific factors. Therefore, it is possible that at least one of the complexes described above is the result of 'factor tethering' and conceivable due to the position of the CAT1 sequence, that the CAT1 motif is required for transcription activity of *TamS1*. However, it has not been established whether the CAT1 sequence is the only motif to bind parasite specific factors and whether this motif is required for both basal transcriptional activity and/or inducible expression of *TamS1* during differentiation to the merozoite.

Attempts to employ UV DNA-protein crosslinking and South-Western blotting assays to determine the size and number of proteins binding to the CAT1 element were unsuccessful. The draw back with both assays is that a number of conditions have to be

fulfilled for positive binding to occur. The UV crosslinking assay relies on UV light to cause the formation of covalent bonds between pyrimidines and certain amino acid residues, and the correct level of irradiation required to fix individual complexes can vary significantly (Dent and Latchman, 1993). Experimental design requires that the distance from the UV source and duration of exposure is taken into consideration. An unfortunate consequence of this is that if a protein in general is in close proximity to the DNA probe and the sample is the correct distance from the UV source it can be bound non-specifically to the DNA and complicate the analysis. This can be overcome by partial purification of extracts, irradiating the complex of interest within the non-denaturing gel or by titrating the concentration of non-specific competitor in an effort to remove non-specific binding. The South-Western procedure relies on the ability of proteins to bind as monomers to the DNA probe, thus, factors which are dependent on sub-unit association for DNA binding are unlikely to be detected. In addition, this assay may rely on the renaturation of the protein bound to the nitro-cellulose membrane (Dent and Latchman, 1993). Not all proteins renature under the same conditions, to reproduce the native structure, and this can introduce further variability into the assay.

The CAT1 oligonucleotide was named because of a CAATT sequence within the probe which has similarity with the CCAAT box defined for a large number of eukaryotic genes. In higher eukaryotes, the position of the CCAAT box varies. It is most frequently 80 nucleotides upstream of the transcription start site, but can be found between position -50 to nearly -200 (Nussinov, 1990). Several proteins that specifically recognise CCAAT elements have been described (Benoist *et al.*, 1980; Efstratiadis *et al.*, 1980; McKnight and Tjian, 1986). The CAATT sequence of *TamS1*, therefore, could fulfil the role played by CCAATT in other eukaryotic genes, which can function in the absence of a consensus downstream TATA motif (Nussinov, 1990) to initiate transcription at the defined start site. However, mutation analysis of the CAT1 sequence revealed that the core motif required for protein binding was TTTGTAGGGCGAC and that the CAATT sequence is most likely required for stabilisation of binding. Furthermore, if the CAATT sequence was required for recognition of the basal transcription machinery it would be expected would be present in a similar position in other genes. From the data presented in chapter 3 this has not been shown to be the case for *SPAG-1*, and *TamR1*. Thus, taking the data together it appears that the mobility

shifts detected in this study are brought about by recognition of a motif which seems to be, at present, unique to the *TamS1* gene.

5 Comparative Analysis of Differential Gene Expression Between Enhanced and Attenuated Macroschizont Infected Cell Lines.

5.1 Introduction

Differentiation from one Life-Cycle stage to the next is crucial for parasite survival but upon *in vitro* cultivation of protozoan parasites the loss of/or reduction in ability to differentiate is a common event. For example, the loss or attenuation of differentiation has been observed for *Leishmania* (da Silva and Sacks, 1987), *T.gondi* (Gross *et al.*, 1996), *Trypanosomes* (Seed and Sechelshki, 1989), *Plasmodium* (Alano *et al.*, 1995), *Babesia* (Carson *et al.*, 1990), *Eimeria* (Matsui *et al.*, 1996), as well as *Theileria* parasites (Shiels *et al.*, 1994, Adamson and Hall, 1996; Baylis *et al.*, 1992). Moreover, a reduction in differentiation capability has also been documented in higher eukaryotic systems such as HL-60 (Tarella *et al.*, 1982).

The molecular alterations which result in attenuation are unknown. This is partly because the process which controls differentiation has not been elucidated, and determining attenuation mechanisms in isolation may be as complex as determining how differentiation occurs. By comparing competent and attenuated cell lines simultaneously, however, it may be possible to confirm whether this event is associated with differentiation and gain information on how attenuation occurs. Perhaps the most informative way attenuated lines could be used is that if attenuation is known to be linked to the alteration or loss of genetic material, it would be possible by mapping the mutation to identify key molecules involved in controlling differentiation. Caveats of these types of studies are: comparative lines of parasites should be as genetically related as possible and it should be kept in mind that the attenuation trait determined may be a secondary rather than primary event.

A number of studies on *Plasmodium* have shown that parasite isolates adapting to *in vitro* cultures can result in the loss in the ability to differentiate into the sexual forms, gametocytes. Various types of defects have been described which include a block in the progression beyond morphological stage III of gametocyte maturation (Teklehaimnot *et al.*, 1987) and deficient production of male gametes (Vaidya *et al.*, 1993). In several cases the loss of gametocyte production has been attributed to the rapid generation of deletions in

a subtelomeric portion of chromosome 9, of about 300kb. (Shirley *et al.*, 1990). Studies using parasite populations heterogeneous for chromosome 9 size have confirmed that these deletions are associated with defective gametocyte production, and suggest that the terminal portion of chromosome 9 regulates an early step in the specialisation of the sexual cell (Alano *et al.*, 1995). In many cases break point deletions in chromosome 9 cluster to a specific region which is thought to be structurally prone to undergo rearrangement (Shirley *et al.*, 1990; Barnes *et al.*, 1994). In addition, within a cloned cell line, chromosome 9 molecules of heterogeneous sizes have been observed indicating that it may be possible to study these rearrangements in a controlled fashion starting from a parasite population homogeneous for an intact chromosome 9 (Alano *et al.*, 1995) and relate these to observed effects on gametocytogenesis.

Other genetic determinants, besides those located in the subtelomeric locus of chromosome 9, appear to govern early events in sexual differentiation of *Plasmodium*. Parasite clones have been observed which conserve the full length chromosome 9 as well as the cytoadherent phenotype, but are totally defective in gametocyte production. It has been proposed that these mutants have been blocked in an early stage of gametocytogenesis, which is distinct from the defect described for the lines previously discussed (Alano *et al.*, 1995). It would therefore appear that in *Plasmodium* attenuation could result from alteration to the structure/function of a number of genes involved in different processes during gametocyte formation or different stages of gametogenesis.

In *T.gondi* the loss of ability to differentiate has not been linked to a known genetic alteration but, variability of different strains to differentiate and an association of attenuation with *in vitro* cultivation have been documented. Most bradyzoites develop into tachyzoites under standard cell culture conditions and this process has been observed in human fibroblasts and in murine macrophages, indicating that spontaneous differentiation takes place regardless of the host cell type (Gross *et al.*, 1996). However, this process is not synchronous within a cloned parasite population (Soete *et al.*, 1993). This phenomenon has as yet not been explained and it could suggest that an internal factor(s) could play a role in differentiation whereas, previously only external factors had been thought to be involved (Soete *et al.*, 1993).

As in the case of *Toxoplasma* strains, *Theileria* parasites show variable levels of differentiation potential and it has been found that long term cultivation results in the loss of differentiation (Brown, 1980; Subramanian *et al.*, 1986). The observation that these attenuated cell lines continue to afford protection from heterologous challenge has been used in the development of live attenuated vaccines (Subramanian *et al.*, 1986). Although the process of attenuation is well established and documented in *T.annulata*, little is known of the molecular mechanisms involved. It is possible that the loss of virulence is due to either the selection of pre-existing avirulent subpopulations of the parasite or due to a genetic alteration of a parasite type. Both mechanisms have been detected in apicomplexan parasites. For example in *B.bovis* avirulent subpopulations have been shown to be enriched as a result of faster growth rate of the parasite by passage in splenectomised calves (Carson *et al.*, 1990). As described above in *Plasmodium* large chromosomal deletions occur following continuous culture which adversely affect the parasite undergoing gametocytogenesis, (Alano *et al.*, 1995) but probably provides a proliferative advantage *in vitro*.

Observed alterations in *T.annulata* infected lines kept *in vitro* culture (Melrose *et al.*, 1991; Baylis *et al.*, 1992) have provided evidence that the selection of parasites with different virulence phenotypes maybe taking place. Further studies by Sutherland *et al.*, (1996) have provided evidence that attenuated *T.annulata* cell lines are the result of both genotype selection and altered gene expression during the continuous *in vitro* culture.

Shiels *et al.*, (1992) isolated cloned macroschizont-infected cells lines from a *T.annulata* (Ankara stock) parental cell line by limiting dilution. Two types of cloned lines were isolated: D7 and C9 cloned cell lines differentiated at an enhanced level and produced large quantities of merozoites upon heat induction at 41°C whereas the cell lines, D3 and E3 had a diminished ability to differentiate and produced fewer merozoites. Theses two types of clones were represented by distinct genotypes but, it was also observed for diminished clones that after a peak point of merozoite formation at 41°C the cells could adapt to culture at the higher temperature and shows an even lower level differentiation. Therefore, the degree of attenuation displayed by a cell line could either be due to the selection of different

genotypes with different abilities to differentiate or a loss in ability/or adaption of a single parasite genotype to differentiate.

Recloning of the enhanced D7 cell line resulted in the isolation of a cell line D7/B12, that showed a severely attenuated ability to differentiate (Shiels *et al.*, 1994). From RFLP analysis it was demonstrated that the isolation of the D7/B12 cell line was unlikely to be due to the presence of more than one clonal type of parasite in the parental clone as both the D7 and D7/B12 cell lines appeared to have identical genotypes. This means that the reduction in differentiation was probably caused by an alteration that occurred during the recloning procedure. The D7/B12 cell line demonstrates that attenuation could be due to a loss in the ability of a single genotype to undergo differentiation and may not increasingly depend on the selection of distinct parasite types.

A characteristic that has been observed for *Theileria* infected cloned cell lines which show enhanced differentiation levels *in vitro*, is an enlargement of the macroschizont, with a concomitant increase in parasite nuclear number (Shiels *et al.*, 1992). As noted previously (Hulliger *et al.*, 1966), this appears to be caused by a reduction in the rate of host cell proliferation, while parasite nuclear division may actually increase (Shiels *et al.*, 1992). Also, because the parasite is dependent on host cell division for division of its cytoplasm, as host cell proliferation slows down the leucocyte cytoplasm becomes completely filled with enlarged macroschizont syncytia. In contrast, the observed size increase of the diminished clones was significantly lower and this was concomitant with a higher rate of host cell division at all time points tested. Therefore, it may be that the diminished phenotype is caused by an altered relationship between parasite growth and host cell division. Indeed it has been postulated that attenuated cells are impaired in their ability to reach a putative quantitative threshold of commitment and that movement towards this threshold is mediated through the disruption between parasite growth and host/parasite cell division (Shiels *et al.*, 1997, see section 1.6). These conclusions were based on the smaller size increase of diminished cell lines and the result that attenuated D7/B12 cells show a marked reduction in *TamS1* gene expression relative to differentiation competent D7 cells over the reversible phase of the differentiation process. It is also possible that in addition to an alteration which affects growth and/or division that specific mutations which change the activity of factors

regulating merozoite gene expression are responsible for an attenuated ability to differentiate. Some of these factors must ensure the correct ordering of the sequential development events and bring about the biochemical and structural changes which culminate in merozoite formation. Therefore, mutations within regulatory factors which control gene expression or target genes which they control could generate a diminished ability to differentiate. However, from the ability of most attenuated cell lines to generate apparently normal merozoites, it seems that the majority of the mutations would be confined to regulatory polypeptides.

5.1.2 Aims

Previous studies have shown that attenuation of the D7/B12 line to differentiate, correlated with an inability to up regulate *TamS1* polypeptide production in the early phase of the differentiation process. The aim of this chapter was to carry out further comparative studies at the molecular level and search for possible mutations which could alter the ability of the D7/B12 cell line up regulate *TamS1* gene expression. To achieve these aims, comparative Northern analysis and gel shift mobility assays were carried out, in addition to a structural study of the 5' intergenic region of the *TamS1* gene of the D7/B12 cell line.

5.2 Materials and Methods

Genomic DNA was prepared from 100mls of macroschizont infected cell culture grown to a density of 10^6 cells/ml was spun down at 1500g for 10 minutes. The pellet was resuspended in 5ml of 1x SSC (section 2.2.3) plus 4.5ml of TNE (100mM Tris-HCl, pH7.5, 100mM NaCl, 10mM EDTA) and mixed with 0.5ml of 10% Sarkosyl. Proteinase K was then added to a final concentration of 100ng/ml and the cell lysate incubated for 2 hours at 55°C. Following this incubation, the lysate was extracted once with phenol:chloroform and once with chloroform. DNA contained within the resulting aqueous phase was precipitated by the addition of 1/10th volume of 3M Na Acetate, 2.5 volumes of ethanol and incubation at overnight at -20°C. The following day the precipitate was collected by centrifugation at 13,000g for 10 minutes and the DNA pellet resuspended in 600µl of sterile dH₂O.

For Southern blots, 5µg of genomic DNA was digested overnight at 37°C as described in section 2.2.2. DNA fragments were resolved by agarose gel electrophoresis and the DNA transferred to nylon membrane filters by Southern blotting (see section 2.2.3). Following transfer the DNA was fixed to the membrane by exposure to ultra-violet light and incubated with a radiolabelled probe. Hybridisation was allowed to proceed overnight at 65°C before washing and exposure of the filter to Kodak K-OMAT film overnight was carried out (section 2.2.3).

For Northern blots, Tri-Reagent (Molecular Research Centre) was used to isolate RNA from *T.annulata* infected cell lines incubated at 37°C and 41°C (see sections 2.2.8 and 2.2.9). RNA was separated by formaldehyde/agarose gel electrophoresis (1.2%) before transfer overnight onto a Hybond N membrane (Amersham), see 2.2.3. The RNA was fixed to the membrane by exposure to ultra-violet light and hybridised with probes prepared using the Random Priming DNA labelling kit (2.2.3).

For the electrophoretic mobility shift assay parasite enriched nuclear extracts were prepared from *T.annulata* infected cell lines as described in section 4.2.1. The assay was carried out (see section 4.2.5) using the double stranded oligonucleotide CAT1 end labelled with γ -³²P ATP (section 4.3.4). Samples for the protein binding reaction were run on a non-

denaturing polyacrylamide gel. Upon completion of the run the gel was dried onto filter paper and exposed to a K-OMAT film overnight.

To isolate and sequence the 5' upstream sequence of the *TamS1* gene from D7/B12 cells genomic DNA was used as a template for PCR amplification with primers TamS1DB and 2735, as described in section 3.2.3 (the sequence of each primer is shown in Table 4.1). The amplified DNA was analysed by agarose gel electrophoresis, excised from the gel and cloned into the TA cloning vector (Invitrogen) (see 2.2.4). The clones were sequenced by automated DNA polymerase sequencing as described in section 2.2.7.

5.3 Results

5.3.1 Restriction fragment polymorphism (RFLP) analysis of the macroschizont enhanced cell line D7 and diminished cell line D7/B12.

RFLP analysis of enhanced and diminished D7 lines demonstrated that both cell lines appeared to have the same genotype (Shiels *et al.*, 1994). Therefore, the reduction in differentiation observed for the D7/B12 cell line was thought to be caused by an alteration to the original D7 cell line during the cloning procedure. One possibility is that the altered differentiation of the D7/B12 phenotype is due to mutations/deletions which were not detected in the initial Southern blot analysis. To test if the intergenic regions of the *TamSI* gene had been altered, RFLP analysis was performed on genomic DNA isolated from the parental D7 cell line and its attenuated derivative, D7/B12. The DNA was digested with restriction enzymes Hind III, Sal I, Nsi I, Hinc III, Dde I, Spe I and EcoRV overnight and analysed by agarose gel electrophoresis. The gel was Southern blotted and probed with the *TamSI* cDNA probe. The result of the analysis is shown in Figure 5.1.

For all seven restriction enzymes the *TamSI* probe hybridised to an identical pattern of bands in DNA from both D7 and D7/B12. The Hind III band detected for both cell lines (Figure 5.1; Lanes 1 and 8) corresponded to the ~8kb contig cloned and sequenced in chapter 2, while the Sal I band corresponded to a 600bp fragment formed by two Sal I sites internal to the *TamSI* gene (Figure 5.1; Lanes 2 and 9) see Figure 2.14. Three Hinc II bands were detected, the ~400bp, as for the Sal I digest, corresponded to restriction sites within the protein coding sequence of the *TamSI* gene while the 1.2 kb and ~4kb fragments corresponded to the 5' and 3' intergenic regions of the gene, respectively (Figure 5.1; Lanes 4 and 11). Two bands of ~300bp and 1.4kb were observed for the Dde I digests. The 300bp fragment contained most of the 5' coding sequence of the *TamSI* gene and the 1.4kb contained the 3' coding sequence plus intergenic sequence (Figure 5.1; Lanes 5 and 12). Single bands were observed for the Nsi I (Figure 5.1; Lanes 3 and 10), EcoRV (Figure 5.1; Lanes 7 and 14) and Spe I (Figure 5.1; Lanes 6 and 13) digests.

Figure 5.1 Southern blot analysis of genomic DNA isolated from D7 cells (Lanes 1-7) and D7/B12 cells (Lanes 8-14), digested with a variety of restriction enzymes and hybridized with the *TamSI* cDNA probe. The position of each band is indicated by an arrowhead. Molecular weight markers are indicated in kilobases.

Lane 1) Hind III

Lane 2) Sal I

Lane 3) Nsi I

Lane 4) Hinc II

Lane 5) Dde I

Lane 6) Spe I

Lane 7) EcoRV

Lane 8) Hind III

Lane 9) Sal I

Lane 10) Nsi I

Lane 11) Hinc II

Lane 12) Dde I

Lane 13) Spe I

Lane 14) EcoRV

1 2 3 4 5 6 7 8 9 10 11 12 13 14



◀12kb

◀8kb

◀4kb

◀3kb

◀2kb

◀1.6kb

This experiment provided further evidence that the genotype of the attenuated parasite cell line derived from the recloning procedure is the same as the parasite represented by the D7 macroschizont infected cell line. It was concluded, therefore, that if mutations were present in the promoter/intergenic regions of the *TamS1* gene of the D7/B12 cell line, then they were present as point mutations that were not detectable by the analysis employed.

5.3.2 Analysis of *TamS1* expression in D7 (enhanced) and D7/B12 (diminished) cell lines

Shiels *et al.*, 1994 demonstrated that the diminished cell line D7/B12 was associated with an inability to express differentiation markers, specifically the *TamS1* and rhoptry 117kDa (*TamR1*) antigen. In addition it was demonstrated by IFA analysis that the reversible elevation of the *TamS1* polypeptide occurred at a significantly lower level in the D7/B12 cell line. To determine whether the changes to *TamS1* expression were due to alterations either at the post-transcriptional or transcriptional level Northern blot analysis was performed on RNA isolated from a differentiation time course of D7 and D7/B12 cell lines. Figure 5.2A shows the result of the analysis when the blot was hybridised with the *TamS1* cDNA probe. The probe detected a message of 1.1kb at day 0 (Figure 5.2A; Lane 1) of the D7 time course, which was observed to progressively increase from days 0 (Figure 5.2A; Lane 1) to day 6 (Figure 5.2A; Lane 4). 1.1kb band was detected in the Day 0 time point of the D7/B12 time course (Figure 5.2A; Lane 5) and this band was observed to be of similar intensity throughout the time course (Figure 5.2A; Lanes 5 to 10). Thus, the inability of the D7/B12 cell line to express the *TamS1* polypeptide during induction of differentiation is the result of changes in the production of the corresponding mRNA.

In addition to the 1.1kb band, the cDNA probe detect a larger band of ~1.5kb when hybridised with RNA prepared from the Day 4 (Figure 5.2A; Lane 3) and Day 6 (Lane 4) time points of the D7 differentiation time course. It is possible that either this detected species is due to differential transcription start sites, intron splicing, this is unlikely as no introns have been found when genomic sequences were compared with cDNA sequence (Katzner, unpublished data), or homology to a related gene. To

Figure 5.2. Analysis of RNA levels during a differentiation time course of cell lines D7/B12 by Northern blotting. RNA was isolated from the cells at 37°C and every following incubation of the culture at 41°C. The size of the bands estimated in kilobases relative to RNA markers are indicated by arrowheads. (A) Hybridisation with the 5' (i) and 3' (ii) cDNA probe; Lane 1) D7 Day 0; Lane 2) D7 Day 2; Lane 3) D7 Day 4; Lane 4) D7 Day 6; Lane 5) D7/B12 Day 0; Lane 6) D7/B12 Day 2; Lane 7) D7/B12 Day 4; Lane 8) D7/B12 Day 6; Lane 9) D7/B12 Day 8; Lane 10) D7/B12 Day 10. (B) Hybridisation with the 5' (i) and 3' (ii) cDNA *TamSI* probes (see 2.3.1); Lane 1) D7 Day 4 RNA and Lane 2) D7 Day 6 RNA.

(A)

1 2 3 4 5 6 7 8 9 10



←1.5

←1.1

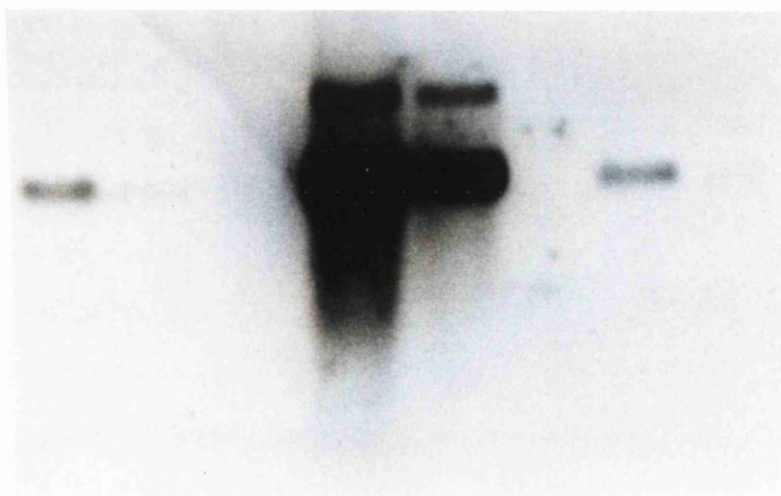
(B)

i

ii

iii

1 2 1 2 1 2



←1.5

←1.1

determine the origin of this extra band a Northern blot of differentiating macroschizont (days 4 and day 6) RNA was hybridised with the 5', 3' and centre cDNA probes of the *TamS1* gene (see section 2.3.1). The centre probe detected the abundant 1.1kb message of the *TamS1* gene at days 4 and 6 of the D7 time course and the additional 1.5kb band (Figure 5.2B, Lanes (ii) 1 + 2). When the blot was re-washed at increased stringency it was observed that the 1.5kb band decreased in intensity until it disappeared (results not shown). In addition only the 1.1kb band was detected by either the 5' or 3' cDNA probes (Figure 5.2B; Lanes (ii) 1 + 2 and (iii) 1 + 2). From these results, it appears that the hybridisation of the *TamS1* cDNA probe to the 1.5kb band is due to homology with a related gene.

To determine whether changes of gene expression within the diminished cell line D7/B12 were specific to the *TamS1* gene during differentiation or were a global event affecting all macroschizont and merozoite genes, Northern blots of macroschizont RNA were hybridised with probes representing macroschizont stage specifically expressed gene, CL12, and a constitutive expressed gene, the large SU ribosomal 2P3 probe (Swan *et al.*, 1996). A 3.2kb band was detected by the 2P3 probe in RNA isolated from D7 and D7/B12 macroschizont cells. The transcript of the large ribosomal unit was observed to gradually increase from day 0 to day 6 of the D7 time course (Figure 5.3A; Lanes 1 to 4). This progressive increase in the ribosomal unit appeared to correspond with the amplitude in parasite load of differentiating D7 cells (Shiels *et al.*, 1992, 1994). A much lower increase in the expression of the LSU ribosomal gene was observed in the D7/B12 cell line (Figure 5.3A; Lanes 5 to 8), although the levels at 37°C between the D7 and D7/B12 samples were comparable. It was concluded that, as previously described, the increase in ribosomal RNA relative to *TamS1* RNA showed that within the D7 cell the expression of the *TamS1* gene is elevated during differentiation. For the D7/B12 line it appears that expression of *TamS1* does not increase relative to the LSU rRNA and that the levels of both RNAs are lower in this cell line. This is likely to be due to the inability of the parasite load to increase at 41°C. Interestingly it appeared from the analysis that, relative to the rRNA control, the expression level of *TamS1* at 37°C in D7/B12 cells was lower than D7 cells. This confirms previous observations at the polypeptide level and implies that basal expression of *TamS1* in the macroschizont stage is lower in attenuated cell lines.

Figure 5.3 Analysis of RNA levels during a differentiation time course of cell lines D7 and D7/B12. RNA was isolated from the cells at 37°C and every 2 days following incubation of the culture at 41°C. The size of the bands estimated in kilobases relative to RNA markers are indicated by arrowheads. The Northern blots were hybridised with (A) large subunit ribosomal, 2P3 probe, (B) macroschizont stage CL12 probe.

Lane 1) D7 Day 0

Lane 2) D7 Day 2

Lane 3) D7 Day 4

Lane 4) D7 Day 6

Lane 5) D7/B12 Day 0

Lane 6) D7/B12 Day 2

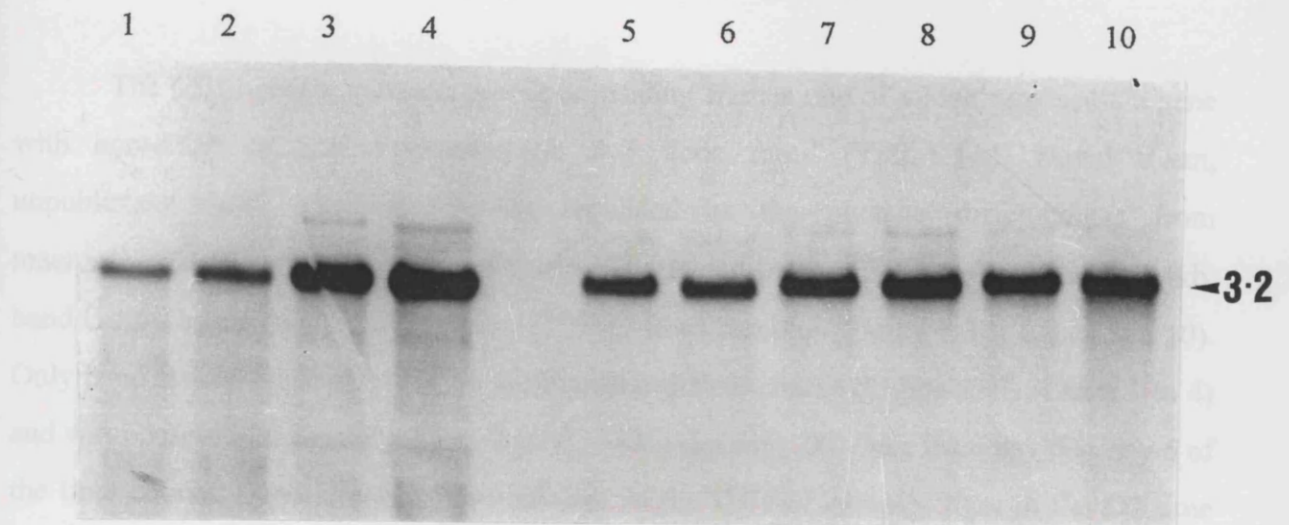
Lane 7) D7/B12 Day 4

Lane 8) D7/B12 Day 6

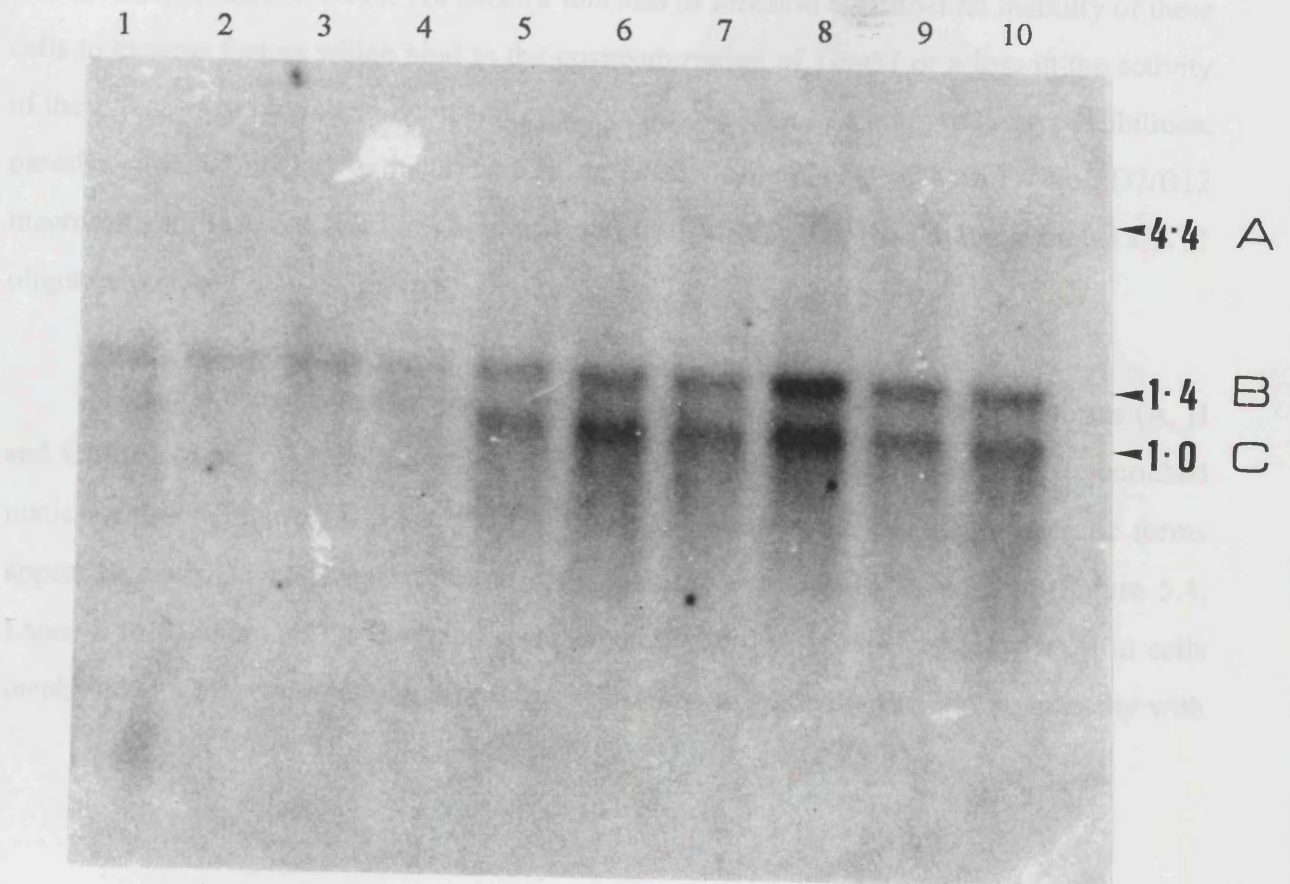
Lane 9) D7/B12 Day 8

Lane 10) D7/B12 Day 10

(A)



(B)



The CL12 probe, contains two open reading frames one of which represents a gene with homology to genes encoding an A.T hook motif (TashA.Th1, David Swan, unpublished data) which are down regulated as the parasite differentiates from macroschizont to merozoite. Three bands were detected by the CL12 probe, band A 4.4kb, band B 1.4 kb, and band C 1.0kb in the D7/B12 RNA samples (Figure 5.3B; Lanes 5 to 10). Only band B was detected in the D7 differentiating time course (Figure 5.3B; Lanes 1 to 4) and was observed (as described previously) to progressively decrease from day 0 to day 6 of the time course. Band B was more abundant in the D7/B12 samples than in the D7 time course. Therefore, it is possible that all of the bands detected in the D7/B12 time course are present in the D7 samples but that the levels of RNA which represent species A and C are so low that the bands are not visibly detectable. Thus there appears to be the possibility of differential expression of the gene(s) detected by the CL12 probe.

5.3.3 Mobility shift analysis of nuclear extracts derived from D7 and D7/B12 cell line

The Northern analysis described above demonstrated that the ability of attenuated cells to up regulate *TamS1* mRNA production during differentiation induction conditions was severely limited. Possible reasons for this loss of function are either an inability of these cells to express factors which bind to the upstream region of *TamS1* or a loss in the activity of these factors to bind this region and/or activate expression. To analyse these possibilities, parasite enriched nuclear extracts were made from a time course of both D7 and D7/B12 macroschizont infected cell lines and analysed by EMSA using the double stranded CAT1 oligonucleotide.

Figure 5.4 shows that as previously demonstrated, three electrophoretic forms (A, B and C) bind to the CAT1 oligonucleotide when incubated with D7 day 8 parasite enriched nuclear extract (Figure 5.4; Lane 6). The upper (A) and middle (B) electrophoretic forms appear in both D7 macroschizont and differentiating macroschizont extracts (Figure 5.4; Lanes 2 to 8) although the complex represented by shift A is at a very low level in cells incubated at 37°C. Complexes A and B were observed to gradually increase in intensity with

Figure 5.4 Interaction of the CAT1 sequence element with parasite enriched nuclear extracts prepared from differentiation time courses of D7 and D7/B12 macroschizont infected cell lines. End labelled oligonucleotide CAT1 was incubated with the crude nuclear extracts under standard conditions.

Lane 1) Probe alone

Lane 2) Nuclear extract prepared from D7 day 0

Lane 3) Nuclear extract prepared from D7 day 2

Lane 4) Nuclear extract prepared from D7 day 4

Lane 5) Nuclear extract prepared from D7 day 6

Lane 6) Nuclear extract prepared from D7 day 8

Lane 7) Nuclear extract prepared from D7 day 10

Lane 8) Nuclear extract prepared from D7 day 12

Lane 9) Nuclear extract prepared from D7/B12 day 0

Lane 10) Nuclear extract prepared from D7/B12 day 2

Lane 11) Nuclear extract prepared from D7/B12 day 4

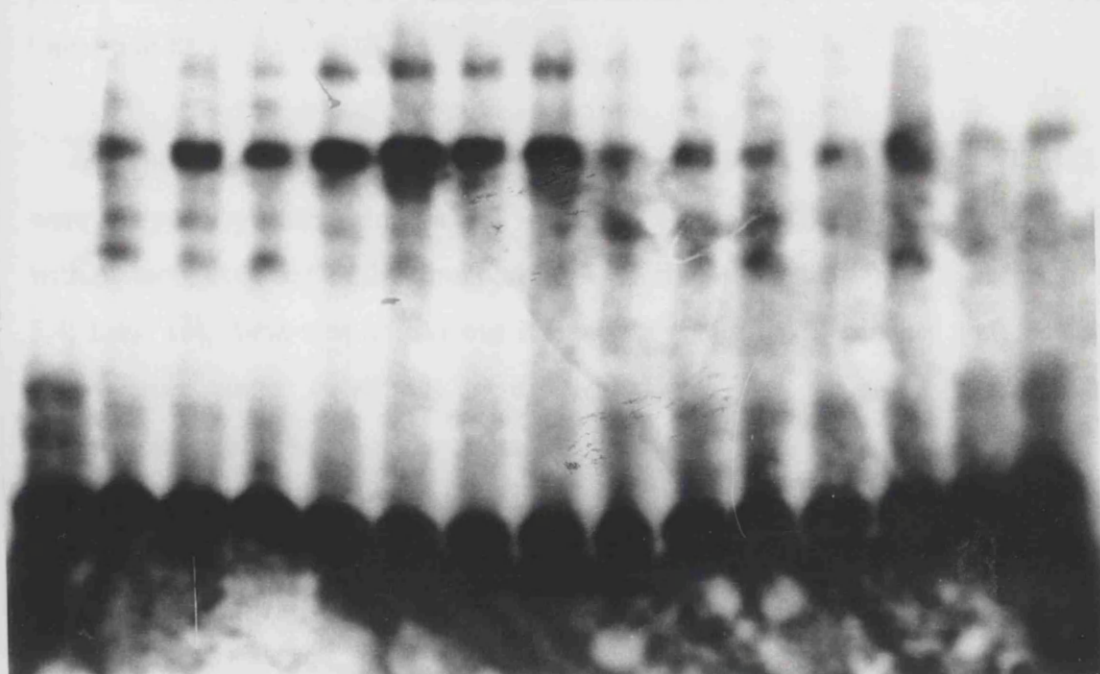
Lane 12) Nuclear extract prepared from D7/B12 day 6

Lane 13) Nuclear extract prepared from D7/B12 day 8

Lane 14) Nuclear extract prepared from D7/B12 day 10

Lane 15) Nuclear extract prepared from D7/B12 day 12

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15



A
B
C

Day 0 to Day 8 (Figure 5.4; Lanes 2 to 6) extracts with the major elevation in band intensity occurring between day 4 and 6 (Figure 5.4; Lanes 4 and 5). It is possible that elevation of shift A was greater relative to shift B over these time points, but this would need further verification by a quantitative assessment of band intensity. A further increase in mobility shift activity appeared to occur at day 8 (Figure 5.4; Lane 6) and appeared to be approximately, maintained until day 12 (Figure 5.4; Lane 8). In contrast to complexes A and B, the C mobility shift was not detected in the early time points of the time course, up to day 6 (Figure 5.4; Lane 5). Indeed the presence of this shift appeared to coincide with the time points where the majority of merozoite production took place, day 8 – day 12 (Figure 5.4; Lanes 6 to 8).

The pattern of complex formation was drastically altered when D7/B12 extracts were analysed by EMSA. In this case only the major mobility shift B was clearly detectable with extracts from any of the time points. Furthermore, apart from the day 8 extract (Figure 5.4; Lane 13), there was no obvious increase in the level of mobility shift when the cells were placed under conditions to differentiate. It can be concluded, therefore, that differentiation of the D7 cell line correlated with increases in the level/activity of factor(s) which specifically bind to the CAT1 motif in the upstream region of the *TamS1* gene and that the C mobility shift appears specific to differentiating cultures. This conclusion is strengthened by the observation that the attenuated cells are unable to elevate expression/activity of these factor(s).

5.3.4 Sequence comparison of the 5' untranslated region of the D7 and D7/B12 alleles of *TamS1*

Although the EMSA analysis appeared to show either loss of function of factor activity or expression in cells attenuated for differentiation, it is possible lower *TamS1* expression levels were also due to an alteration in the CAT1 binding domain of the D7/B12 cell line. To test this possibility the upstream region of *TamS1* DNA was cloned from D7/B12 genomic DNA and sequenced.

DNA was purified from a culture of D7/B12 cells that had previously been shown not to differentiate. Primers 2735 and Tams1DB were used to amplify a 771bp D7/B12

fragment from genomic DNA, and the PCR product was cloned into pGem-T and sequenced using methodology described in section 5.2. The sequence comparison over the first -608bp upstream of the ATG start codon is shown in Figure 5.5. Only one base difference was noted between the obtained sequence with that of the homologous sequence determined for non-attenuated D7 cells. Therefore, the diminished ability of the D7/B12 cell line to express the *TamS1* gene was not found to be associated with sequence mutations within the CAT1 motif. Furthermore, the sequence alteration which was observed is unlikely to influence *TamS1* expression levels as this was located to the junction of a conserved and non-conserved regions of sequence which were previously shown to be conserved and non conserved across *Theileria* species.

D7B12	GAGACTGATTAAACTCAAAGACATCACACTCTGATGAAAGTGAAGGCGGGTAAGGACAGCT	-546
D7	GAGACTGATTAAACTCAAAGACATCACACTCTGATGAAAGTGAAGGCGGGTAAGGACAGCT	
D7B12	CATAACGACAAATTCAGTATAAAAACATAATGAATGATTGTTTTAAGGAAATGATGTGTGG	-485
D7	CATAACGACAAATTCAGTATAAAAACATAATGAATGATTGTTTTAAGGAAATGATGTGTGG	
D7B12	AATCACGGTTATTTGTGCGGGAAATTATGCATGGATATTATCTATTTAACTTATAATAATG	-424
D7	AATCACGGTTATTTGTGCGGGAAATTATGCATGGATATTATCTATTTAACTTATAATAATG	
D7B12	TTATATTAAAATATGACAAATTAATGTGTGAAATAAATAAAGTAAAGTGGAGGGAAATTCT	-363
D7	TTATATTAAAATATGACAAATTAATGTGTGAAATAAATAAAGTAAAGTGGAGGGAAATTCT	
D7B12	GTCCTAATATTTAACTGGGTCTAGTGACTT-AAAAAGAAAGCAGGTATTGGTTTGAAAAA	-302
D7	GTCCTAATATTTAACTGGGTCTAGTGACTTAAAAAGAAAGCAGGTATTGGTTTGAAAAA	
D7B12	ATAATAAATTCATCTTTTTATAGTCATCAAAAAATGAAAAATGGATAAATATACATGTTAT	-241
D7	ATAATAAATTCATCTTTTTATAGTCATCAAAAAATGAAAAATGGATAAATATACATGTTAT	
D7B12	TTCATTAAGATGGACTTCAATGGAGGATAAGGCATTAGGCGTGATGAGTGCATAGATACAG	-180
D7	TTCATTAAGATGGACTTCAATGGAGGATAAGGCATTAGGCGTGATGAGTGCATAGATACAG	
D7B12	ATAAACATGCACACAATTTGTAGGGCGACATTGTTTTGTATGGTGAATGCATATCCATCTC	-119
D7	ATAAACATGCACACAATTTGTAGGGCGACATTGTTTTGTATGGTGAATGCATATCCATCTC	
D7B12	ACTTCACTTATAGTTATTGGTCTTTTTTTCCATAACATCCACCCAATTAGTTAATTTTTAA	-58
D7	ACTTCACTTATAGTTATTGGTCTTTTTTTCCATAACATCCACCCAATTAGTTAATTTTTAA	
D7B12	TATTTAAATCGCTCACTAGTCTGCCCTTCTTATCTTTTTATAATATAATTATTTGAG	
D7	TATTTAAATCGCTCACTAGTCTGCCCTTCTTATCTTTTTATAATATAATTATTTGAG	

Figure 5.5 Sequence comparison of the 5' untranslated sequence of *TamS1* from the cell lines D7 and D7/B12. The numbers at the end of each row represents the sequence position with reference to the ATG start codon. A line indicates that the bases compared are identical and a space indicates that they are different.

5.4 Discussion

Studies carried out *in vitro* have demonstrated that cloned macroschizont infected cell lines can be isolated with enhanced or diminished/attenuated ability to differentiate (Shiels *et al.*, 1992). By comparing cloned cell lines that are attenuated or competent for differentiation, it was investigated which molecular mechanisms are associated with the ability of a cell line to differentiate. The mechanism(s) which bring about the attenuation of differentiation are unknown. In this study four possibilities were considered. These are: 1) the selection of parasite genotypes which are inherently unable to differentiate at a high level, 2) alteration of the ability through major genotypic rearrangement or 3) point mutations which confer quantitative differences in the control of gene expression and 4) an epigenetic alteration arising from the adaptation of cells *in vitro*.

Recloning of the enhanced D7 cell line by Shiels *et al.*, (1994) resulted in the isolation of a cell line that showed an attenuated ability to differentiate. RFLP analysis (Figure 5.1) confirmed that the isolation of the D7/B12 cell line was unlikely to be due to the presence of more than one clonal type of parasite in the parental cell line as both the D7 and D7/B12 cell lines appeared to have identical genotypes. This would mean that the reduction in differentiation potential was caused by an alteration to the original cell line that occurred during the cloning procedure. In addition, comparison of the intergenic region 5' to the *TamS1* gene for both cell lines demonstrated that the attenuated phenotype was most likely not the result of major genomic rearrangements to this region of the genome. It cannot be discounted, from this work, that a detectable alteration has occurred elsewhere in the genome and has been missed in the analysis carried out to date. However, as pulse field gel electrophoresis studies also showed no obvious differences between the two cell line genotypes (Tait, unpublished) it appears that such alterations are not a common event, unlike the situation described for *Plasmodium* (Alano *et al.*, 1995). Thus it would appear that attenuation of differentiation in the D7/B12 may not arise from selection of parasites which have undergone major genomic alterations.

To extend previous studies examining the association of *TamS1* antigen production with attenuation a comparison of the ability of the cell lines to express the *TamS1* gene at the RNA level was carried out by Northern blot analysis. As expected both cell lines

showed a very low level of *TamSI* mRNA at day 0 (Figures 5.2A, 5.3). This level of expression was maintained for the D7/B12 cell line under differentiation while a progressive increase in *TamSI* RNA was observed for the D7 cell line (Figure 5.2A). Therefore, it can be concluded that the changes previously detected at the polypeptide level between these cell lines (Shiels *et al.*, 1994) are a result of an alteration to the production of mRNA.

Although a difference in *TamSI* transcription has still to be confirmed it is possible, due to the findings outlined in the previous chapter, that an alteration to the control of *TamSI* transcription has occurred through modulation of factor binding the upstream region of the genes. Theoretically, altered transcriptional control of the 30kDa merozoite polypeptide in the diminished cell line D7/B12 could occur as the result of a mutation in the *TamSI* promoter, a mutation within the transcription factor which recognises the *TamSI* promoter, or a mutation within a more primary regulator which is necessary to initiate or commit the parasite to the differentiation process. It is also possible that attenuation is a result of a general alteration which could cause an inability to elevate the levels of factors which control *TamSI* expression, for example, an inability to respond to an elevated temperature.

Cloning of the *TamSI* 5' intergenic region from the D7/B12 cell line revealed that except for one base alteration, the *TamSI* promoter region for both enhanced and diminished cell lines were identical. The single base alteration was located at the junction between a conserved and non-conserved sequence region, which were distinct from CAT1 sequence shown previously to specifically interact with parasite nuclear factors. The diminished ability of the D7/B12 cell line to express the *TamSI* gene is therefore, unlikely, to result from the sequence alteration.

Gel mobility shift analysis with the CAT1 oligonucleotide and parasite enriched nuclear extracts prepared from enhanced and diminished cell lines revealed that the changes in *TamSI* regulation correlated directly with changes in complex formation. Two complexes (A and B) were shown to bind to the CAT1 sequence when incubated with parasite enriched nuclear extracts prepared from the D7 cell line grown at 37°C. As the D7 cells differentiate there was a progressive increase in intensity of the complexes A and B which correlates

with the detected elevation in *TamSI* mRNA. By day 8 of the differentiation time course an additional complex (C) was observed. Thus, it is possible that the increase in the factor(s) forming complexes A and B could be associated with the differentiation process and may function in its initial phase, while the appearance of a third complex (C) could be associated with a later point in the process after it has been initiated. These correlations were supported by the data obtained with the D7/B12 cell line where only complex B was observed to form with the CAT1. Moreover, when under differentiation induction conditions there was no observable increase or reduction in factor binding to CAT1, which correlated with the constant level of *TamSI*. Therefore, it is possible that the detection of the upper (A) and lower (C) complexes detected in the D7 day 8 parasite enriched extracts are directly associated with high level expression of the *TamSI* 30kDa merozoite antigen, while a quantitative change in complex B may also relate to increased mRNA production. Indeed the lower level of this complex in D7/B12 could account from the lower level of *TamSI* mRNA detected in the early part of the differentiation time course (see Figure 5.2A).

It is important to note that the inability of the gel retardation assay to detect complexes A and C in the D7/B12 parasite enriched extracts may not be related to whether the factors forming these complexes are present. It is possible that mutations which affect the ability of the factors to interact with DNA sequences would prevent them from forming the complexes and regulating the expression of the *TamSI* gene. Thus experiments need to be conducted to determine whether those factors which form complexes A and C are present in the extracts prepared from the D7/B12 cells. This will require either purification of the relevant factors and generation of antibody reagents or the cloning of the gene(s) which encode the factor(s). Detection of the factors at approximately equal levels in both cell lines would then suggest an impairment of binding affinity and would be followed by sequence comparison of the relevant site. Although unlikely, if attenuation is associated with a lack of specific production of *TamSI* antigen, which prevents the parasite from differentiating without its surface coat, this could be the cause.

The results discussed above show that an inability to upregulate is clearly associated with attenuation of differentiation *TamSI* gene expression while it is possible that attenuation is specifically linked to this one gene this seems unlikely, unless the block was

due to an inability to differentiate without a surface coat. It seems more probable that the failure to upregulate *TamS1* gene expression is caused by a general failure to carry out differentiation. This reasoning was supported by the Northern analysis with probes representing a number of other parasite genes. Thus the LSU rRNA probe revealed that size changes in the parasite which bring about an increase in parasite RNA load do not occur in the D7/B12 cells. Furthermore, down regulation of the macroschizont gene TashA.Th1 was not observed in D7/B12 cells at 41°C, but clearly occurred in the differentiation competent D7 cell line. Thus a number of changes associated with progression of or commitment to differentiation are not observed for the D7/B12 cell line, indicating that the attenuation alteration is more primary than at the level of target gene expression.

Interestingly, the CL12 probe detected three bands in the D7/B12 RNA samples but only one band in the D7 samples. The CL12 probe hybridises with a family of genes (TashA.Th1) which have been shown to have homology to the high mobility group (HMG) proteins (David Swan, unpublished data). In eukaryotes HMG-like proteins have been shown to switch a transcriptional activator to a repressor (Lehming *et al.*, 1994). Thus, the ability of HMG proteins to repress transcription coupled with the possibility that a putative homologue is up regulated within the D7/B12 cell line suggests that a mechanism could be present where genes controlling differentiation are preferentially repressed in attenuated cells. However, it should be pointed out that the upregulated mRNA has not been conclusively shown to be related to the TashA.Th1 gene as another open reading frame is present on the isolated gene clone, and the ability to down regulate expression may simply be related to an ability to carry out the differentiation process. Thus, like the *TamS1* gene the observed changes could easily be due to the loss of differentiation rather than causing attenuation of this process.

If an epigenetic model were to exist which determined whether or not the parasite were to differentiate it would be expected that the process would be reversible. Shiels *et al.* (1994) have demonstrated that when induced to differentiate, by placing the cultures at elevated temperatures, a small number of the D7/B12 population can differentiate to the merozoite but evidence for reversal of the population as a whole has not been found *in vitro* or upon transfer of attenuated cell lines into bovines (Darghouth *et al.*, 1996). In contrast

evidence for reversible attenuation has been found for HL-60 (Tarella *et al.*, 1982). It would appear, therefore, that the underlying basis that determines whether the parasite has an enhanced or diminished ability in D7/B12 to differentiate is either a point mutation within a factor crucial to the regulation of the differentiation mechanism or/and an epigenetic phenomenon. Both mechanisms would explain global changes to gene expression observed between the D7 and D7/B12 cell lines. They would also account for the low level of differentiation detected in the D7/B12 cell line by Shiels *et al.*, (1994) as neither mechanism is absolute and could be overcome by changes to culture conditions which alter the probability of a cell reaching a quantitative threshold which determine commitment.

6 General Discussion

One of the primary aims of the work presented in this thesis was to identify the regulatory elements controlling the expression of the *TamS1* gene and to characterise the DNA binding protein(s) which recognise these elements. This was on the basis that expression of *TamS1* reflects, in a temporal manner, important events which occur during differentiation from the macroschizont. Moreover, the finding that a direct correlation exists between the amplitude of *TamS1* gene expression during the initial stages of differentiation and the potential of cell lines to form merozoites indicated that the events which control *TamS1* expression are linked to the control of parasite differentiation. Thus, identification of the regulatory process and factors would, ultimately, aid in understanding the mechanisms which lead to and commit the parasite to merozoite production.

Nuclear run on analysis indicated that the *TamS1* and *TamR1* genes are independent transcription units, with discrete initiation and termination sites. This conclusion was supported further by the findings that the open reading frames identified within the 8.0kb *T.annulata* contig are not ordered in the same orientation. In addition, initial investigations into the expression of each open reading frame has demonstrated that each gene is either expressed in different parasite Life-Cycle stages or in a different temporal order during differentiation to the merozoite. Similar studies by Nene *et al.*, (1992) also suggest that several genes within *T.parva* are expressed monocistronically and that stage specific genes are not clustered together, although, monocistronic expression has not been demonstrated experimentally. Therefore, it is possible that many, if not all, stage specifically regulated genes are randomly scattered across the *Theileria* genome, rather than being clustered together, which is indicative of monocistronic gene expression (Latchman, 1990).

All higher eukaryotic protein coding genes are transcribed by RNA polymerase II and initiation of transcription is usually very precise, beginning at one particular nucleotide which is most frequently, though not invariably an A residue. Upstream from this initiation site is the promoter region containing the TATA box, and 3' to the protein coding sequence of the gene is a Poly A signal (Hawkins, 1996). The transcription initiation site of the *TamS1* gene was mapped to an A residue and the flanking sequence to this residue was in

accordance with the transcription initiation site consensus sequence (Cohen *et al.*, 1980). However, upstream of the transcription initiation site evidence for a functional putative TATA box in the correct position was not found and no consensus Poly A signal could be observed in the sequence 3' to the protein coding sequence of the *TamSI* gene. Putative TATA box sequences have been detected for the 117kDa rhoptry (*TamRI*) and *SPAG-1* genes (Katzer, 1995; Fraser McDonald, personal communication) in the conventional manner but, due to their position in relative to the transcription initiation site and the ATG start codon, it is unlikely that they are involved in the regulation of these genes. Thus, it appears that the classical TATA motif, in a position approximately 20-30bp upstream of the transcription initiation site, is not necessary for the correct positioning of the RNA polymerase in the *Theileria* genes studied so far.

So called "TATA-less" promoters have been identified in both *Toxoplasma* (Soldati *et al.*, 1994) and *Plasmodium* (Lanzer *et al.*, 1992a, b). A recent study by McAndrew *et al.*, 1993, demonstrated that the TATA-binding protein (TBP) is highly divergent from its eukaryotic counterpart and was postulated that it recognises a different motif to that of the TATA-box. An alteration believed to be due to the high A+T content of the *Plasmodium* genome. Therefore, motifs which bind TBP probably exist in *Toxoplasma*, *Plasmodium*, and *Theileria*, but as yet the motif which the TBP in apicomplexan parasites recognises has yet to be identified. In conclusion it appears that the *TamSI* gene is expressed in a similar manner to that of many other eukaryotic genes but, the signals for positioning the RNA polymerase and where polyadenylation occurs will be novel and probably relate to that of other apicomplexan parasites

Palindromic and pentamer motifs were identified in the *TamSI* 5' intergenic region which were found to be conserved between several *Theileria* genes and had previously been identified as potential promoter motifs for other protozoan parasites. These included the AGTGA and ACACA pentamers which resembled, an element (T/AGTGTAC) reported in the upstream region of the *Plasmodium* GBP130 gene (Lanzer *et al.*, 1992a) and the pentamer TGTGT found in a number of stage regulated genes in *Toxoplasma gondii* (Mercier *et al.*, 1996). In addition, a CTTTTT motif showed conservation across several intergenic regions of *Theileria* genes and was repeated several times in the upstream region

of the *TamR1* gene. Therefore, if functional, these motifs must be involved in the general transcriptional control of *Theileria* genes. It would be expected that genes from related protozoan parasites or species of parasite would contain common or related cis-acting regulatory elements in their promoter regions but they are apparently very difficult to detect. In particular, to date, no TBP sequence has been detected in apicomplexan parasites. This may be a consequence of the genomic composition of these parasites, therefore, the A+T content of the parasites may preclude dependence upon TATA recognition and compel them to utilise motifs which vary according to the A+T content of each genome.

The nuclear run on data also showed that differential expression of *TamS1* was mediated, at least in part, at the transcriptional level. This led to studies investigating the 5' upstream region of *TamS1* for motifs that bind polypeptide factors which could be involved in this control. DNA protein interactions between the 5' intergenic region of the *TamS1* gene and parasite nuclear factors were investigated using the mobility shift assay. Specific binding by parasite nuclear factors was located to DNA fragments most proximal and distal to the mapped transcriptional start site. However, it has not been established whether these motifs are the only elements required for controlling *TamS1* gene expression. Most eukaryotic genes have several sequences which regulate transcription (Polyanovsky and Stepchenko, 1990). Therefore, the conditions used for binding factors to motifs 5' to the *TamS1* gene may inhibit the formation of additional complexes required for gene regulation. This postulation underscores the necessity for future functional studies specifically designed to dissect the upstream region of the *TamS1* gene, by deletion analysis, to define important additional regulatory motifs.

A motif proximal to the transcriptional start site of the *TamS1* gene was identified which bound factors in a sequence specific manner. Three distinct complexes (A, B, and C) of different mobilities were observed when the motif was incubated with parasite enriched nuclear extracts prepared from differentiating macroschizont infected cells and two of the complexes (A and B) were also identified in extracts prepared from infected cells cultured at 37°C. The formation of the two complexes A and B in extracts derived from macroschizont infected cells may correspond to the basal transcriptional machinery of *Theileria* necessary for RNA polymerase binding so that transcription can occur. Evidence for the involvement

of complexes A and B in the expression of the *TamS1* gene was provided by the gel mobility time course experiment in section 5.5.3 which demonstrated a correlation between mRNA expression of the *TamS1* gene and binding of complexes A and B to the CAT1 sequence element. In addition, the CAT1 sequence is the closest element to the transcriptional start site shown to bind parasite specific factors and only complexes A and B have been shown to be present in both macroschizont and differentiating macroschizont extracts. On the other hand, the CAT1 sequence was not unambiguously identified in a similar position in any of the additional genes analysed. Therefore, if the factors forming complex A and B are part of the basal transcriptional machinery (i.e. like an apicomplexan TBP) they or it must have the ability to recognise distinct structural motifs in different genes or be directed by another unidentified factor or motif.

Previous studies have shown that the expression of the *TamS1* gene correlates with the timing of differentiation and it is thought that at around commitment point an irreversible elevation of *TamS1* expression occurs (Shiels *et al.*, 1994). Complex (C) is the only band which specifically interacts with the CAT1 motif during differentiation to the merozoite. Formation of the complex was observed to occur between days 6 and 8 of the differentiation time course, corresponding with the most evident increase in the expression of the *TamS1* mRNA levels. It is, therefore, possible that the factors forming the lower complex (C) could be involved either in determining or occur following the commitment to differentiate.

Investigations into how each complex bound to the CAT1 probe and their relationship to one another were inconclusive. The upper band (A) was believed to be either the result of dimerisation by the factor which formed complex (B) or due to co-operative binding of two factors, one of which formed the middle complex. Alternatively both complexes could be due to distinct factors. If this is the case then the binding sites must either overlap or be identical. It could be predicted that if the site was identified then the factors may compete with one another under probe saturation conditions, while at lower levels of factor relative to probe a complex would be more likely to dissociate. Unfortunately it is not clear from this study, which of these possibilities operate, although the experiment titrating extract concentration relative to CAT1 indicated that the factors(s)

forming mobility shift A have a greater affinity, suggesting a competition model. The situation outlined for A and B shifts is also pertinent to complex C. As this complex appears to occur at a different time point and migrates faster under electrophoresis conditions it might be predicted that this shift is more likely to be due to the production of a novel factor. However, if this is the case its binding site appears to be almost identical to that of the factor associated with shift B. It is also possible that the C shift could be due to the formation of a complex but, if this complex involved the factor(s) which generate shifts A and B, the C specific factor would have to either bring about a conformational change in the DNA probe or have a significantly higher negative charge to account for the migration of the complex. In conclusion it is possible that expression of the *TamSI* gene is controlled by either a number of distinct factors or complexes of factor(s) which bind to a core sequence of 14 nucleotides. Identifying these factors and determining how they interact is clearly a necessary goal for future research.

The work conducted on the CAT1 motif in this study used crude extracts prepared from macroschizont infected cell lines. To understand the exact relationship between the different CAT1 mobility shifts it will be necessary to use purified or recombinant factors. Initial attempts have been made to clone each factor by direct screening of a merozoite expression library with the CAT1 motif (McKellar and Shiels, unpublished data). However, this was not successful, possibly because the factors are part of a multi-protein complex and more than one protein is required for DNA binding. The successful isolation of protein also requires that the molecular weight of the factor has been determined for monitoring the final stages of the purification step. Initial attempts using UV cross-linking and South-Western blot analysis to determine the molecular weight of those factors binding to the CAT1 motif were unsuccessful. From the results obtained it was unclear why the experiments failed, but this assay is known to give variable results with different probe combinations. One possibility would be to repeat the experiments with partially purified extracts in an attempt to remove some of the non-specific binding detected in this study.

In addition to identification of the factors, it will also be necessary to associate complex formation with function. Ideally this will involve analysis by *in vitro* transcription using purified factors in combination with deletion analysis of transiently transfected *TamSI*

promoter constructs. Unfortunately, neither of these techniques are currently available. Transfection technology has recently been developed for other apicomplexans and barring some *Theileria* specific impediment to the introduction of DNA into the parasite should be achievable. To date *in vitro* transcription is not a standard method for analysis of apicomplexan gene function but has been developed for other protozoan parasites (Bennett *et al.*, 1997). The *Theileria* system may allow this assay to be developed as production of litre volumes of differentiating parasites can be easily achieved. Therefore the establishment of methods for the isolation of nuclear extracts and demonstration of specific factor binding in this study may be an important step forward for future examination of gene expression *in vitro*. Moreover, the ability to correlate complex formation with transcriptional activity and the molecular and cellular events previously mapped for differentiation to the merozoite should provide a greater understanding of how an apicomplexan parasite can respond to, or be programmed for, events which necessitate a change in biological form.

7 References

- Adamson, R.E., Hall, F.R., (1996) Matrix metalloproteinases mediate the metastatic phenotype of *Theileria annulata*-transformed cells. *Parasitology*, 113, 449-455.
- Alano, P., Roca, L., Smith, D., Read, D., Carter, R., Day, K., (1995) *Plasmodium falciparum*: Parasites defective in early stages of gametocytogenesis. *Experimental Parasitology*. 81, 227-235
- Aldritt, S.M., Joseph, J.T., Wirth, D.F. (1989) Sequence identification of cytochrome *b* in *Plasmodium gallinaceum*. *Molecular and Cellular Biology*, 9, 3614-3620
- Alexandre, S., Guyaux, M., Murphy, N.B., Coquelet, H., Pays, A., Steinert, M., Pays, E., (1988) Putative genes of a variant-specific antigen gene transcription unit in *Trypanosoma brucei*. *Molecular and Cellular Biology*. 8, 2367-2378
- Aller, P., Rius, C., Mata, F., Zorrilla, A., Cabanas, C., Bello, T., Benabeu, C., (1992) Camptothecin, induces differentiation and stimulates the expression of differentiation-related genes in U-937 human promonocytic leukemia cells. *Cancer Research* 52, 1245-1251
- Allsopp, B.A., Allsopp, M.T.E.P (1988) *Theileria parva*: genomic DNA studies reveal intra-specific sequence diversity. *Molecular and Biochemical Parasitology* 28, 77-84
- Alt, F.W., Blackwell, T.K. Yancopoulos, G.D., (1987) Development of the primary antibody repertoire. *Science* 238, 1079-1087
- Arcangioli, B., Lescure, B., (1985) Identification of proteins involved in the regulation of yeast iso-cytochrome c expression by oxygen. *EMBO Journal* 4:2627-2633
- Arndt, K., Fink, G.R., (1986) GCN4 protein a positive transcription factor in yeast, binds genera; control promoters at all 5' TGACTC 3' sequences. *Proceedings of the National Academy of Sciences, USA* 83, 8516-20
- Ausiello, C.M., Antonelli, G., Spagnoli, G.C., Gentile, M., Urbani, F., Gitto, M., (1992) A macrophage derived blastogenic factor-MBF-induces cytokines and cytotoxic effectors in human peripheral blood mononuclear cells. *Journal. of Biological Regulation of Homeostatic Agents* 6. 8-14.
- Autelitano, D.J., (1994) Glucocorticoid regulation of c-fos, c-jun and transcription factor Ap-1 in the AtT-20 corticotrope cell. *Journal of Neuroendocrinology* 6, 627-37
- Ayoubi, T.A.Y., and Van de Ven, W.J.M. (1996) Regulation of gene expression by alternative promoters. *FASEB Journal*. 10, 453-460
- Barahona, I., Soares, H., Cyrne, L., Penque, D., Denoulet, P., Rodrigues-Pousada, C., (1988) Sequence of one α - and two β -tubulin genes of *Tetrahymena pyriformis*. *Journal of Molecular Biology*. 202, 365-382.
- Barnett, S.F., (1977) In: *Parasitic Protozoa*, Vol IV. Kreier, J.P. (ed) Academic Press, New York and London, pp 77-113

Barnes, A.D., Thomson, J., Triglia, T., Day, K., Kemp, D.J., (1994) Mapping the genetic locus implicated in cytoadherence of *Plasmodium falciparum* to melanoma cells. *Molecular and Biochemical Parasitology* 66, 21-29

Barta, J.R., Jenkins, M.C., Danforth, H.D., (1991) Evolutionary relationships of avian *Eimeria* species among other Apicomplexan protozoa: monophyly of the Apicomplexa is supported. *Molecular Biology of Evolution* 8, 345-355.

Bass, K.E., Wang, C.C., (1992) Transient inhibition of protein synthesis accompanies differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms. *Molecular and Biochemical Parasitology* 56, 129-140

Bates, P.A., Cobertson, C.D., Tetley, L., Coombs, G.H., (1992). Axenic cultivation and characterization of *Leishmania mexicana* amastigote-like forms. *Parasitology* 105, 193-202

Bates, P.A., (1993). Axenic culture of *Leishmania* amastigotes. *Parasitology. Today*, 9,143-46

Baylis, H.A., Sohal, S.K., Carrington, M., Bishop, R.P., Allsopp, B.A., (1991) An unusual repetitive gene family in *Theileria parva* which is stage-specifically transcribed. *Molecular and Biochemical Parasitology*. 49, 133-142

Baylis, H.A., Megson, A., Brown, C.G.D., Wilkie, G.F., Hall.R., (1992) *Theileria annulata*-infected cells produce abundant proteases whose activity is reduced by long-term cell culture. *Parasitology* 105, 417-423.

Baylis, H.A., Megson, A., Mottram, J.C., Hall, R., (1992) Characterisation of a gene for a cysteine protease from *Theileria annulata*. *Molecular and Biochemical Parasitology*, 54, 105-108

Baylis, H.A., Megson, A., Hall, R., (1995) Infection with *Theileria annulata* induces expression of matrix metalloproteinase 9 and transcription factor AP-1 in bovine leukocytes. *Molecular and Biochemical Parasitology*. 69, 211-222.

Bayne, R.A., Kilbride, E.A., Lainson, F.A., Tetley, L., Barry, J.D., (1993). A major surface antigen of procyclic stage *Trypanosoma congolense*. *Molecular and Biochemical Parasitology*. 61, 295-310

Ben Miled, L., Dellagi, K., Bernardi, G., Melrose, T.M., Darghouth, M., Bouttour, A., Kinnaird, J., Shiels, B., Tait, A., Brown, C.G.D., (1994) Genomic and phenotypic diversity of tunisian *Theileria annulata* isolates. *Parasitology* 108, 51-60

Benne, R., (1994) RNA editing in typanosomes. *European Journal Biochemistry*. 221, 9-23

Bennett, B.J., Mohandas, N., Coppel, R.L., (1997) Defining the minimal domain of the *Plasmodium falciparum* protein MESA involved in the interaction with the red cell membrane skeletal protein 4.1. *Journal Biological Chemistry* 272, 15299-15306

Benoist C., Chambon, P., (1981) *In vivo* sequence requirements of the SV40 early promoter region. *Nature*, 290, 304-310

Berger, S.L., Cress, W.D., Cress, A., Triezenberg, S.J., Guarante, L., (1990). Selective inhibition of activated but not basal transcription by the acidic activation domain of VP16: evidence for transcriptional adapters. *Cell* 61, 1199-1208.

- Beutler, B., Cerami, A., (1986) Cachectin and tumour necrosis factor as two sides of the same biological coin. *Nature*, 320, 584-588
- Blake, M.C., Jambou, R.C., Swick, A.G., Kahn, J., Azizkhan, J.C., (1990) Transcription initiation is controlled by upstream GC-box interactions in a TATAA-less promoter. *Molecular Cellular Biology*. 10, 6632-6641
- Boam D.S., Clark, A.R., Docherty, K., (1990) Positive and negative regulation of the human insulin gene by multiple trans-acting factors. *Journal of Biological Chemistry* 265, 8285-96.
- Boer, P.H., Gray, M.W., (1988) Scrambled ribosomal RNA gene pieces in *Chlamydomonas reinhardtii* mitochondrial DNA. *Cell* 55, 399-411
- Bogler, O., Wren, D., Barnett, S.C., Land, H., Noble, M., (1990) Co-operation between two growth factors promotes extended self-renewal and inhibits differentiation of oligodendrocyte-type-2 astrocyte (O-2A) progenitor cells. *Proceedings of the National Academy of Sciences, USA* 87, 6368-6372.
- Bohne, W., Hesseemann, J., Gross, U., (1993) Induction of bradyzoite-specific *Toxoplasma gondii* antigens in gamma interferon-treated mouse macrophages. *Infection Immunity*. 61, 1141-1145
- Bohne, W., Heesemann, J., Gross, U., (1994) Reduced replication of *Toxoplasma gondii* is necessary for induction of bradyzoite-specific antigens; a possible role for nitric oxide in triggering stage conversion. *Infectious Immunology* 62, 1761-1767
- Borst, P., (1991) Molecular genetics and antigenic variation. *Immunology Today*. 12, A29-A33
- Bouattour, A., Darghouth, M.A., Ben Miled, L., (1994) Epidemiological investigation on tropical theileriosis in the subhumid bioclimate area in Tunisia: study of the *hyalomma* vector in Tunisia. In: European Union third Coordination Meeting on Tropical Theileriosis. Spooner, R. and Campbell, J., (eds) The Roslin Institute (Edinburgh), Roslin, Midlothian, Scotland, pp 3-6.
- Boulter, N., Knight, P.A., Hunt, P.D., Hennessey, E.S., Katzer, F., Tait, A., Williamson, S., Brown, D., Baylis, H.A., Hall, R., (1994) *Theileria annulata* sporozoite surface (SPAG-1) contains neutralising determinants in the C-terminus. *Parasite immunology*. 16, 97-104.
- Boulter, N.R., Glass, E.J., Knight, P.A., Bell-Sakyi, L., Brown, C.G.D., Hall, R., (1995) *Theileria annulata* sporozoite antigen fused to hepatitis B core antigen used in a vaccination trial. *Vaccine* 13, 1152-1160
- Boyer, T.G., Krug, J.R., Maquat, L.E., (1989) Transcriptional regulatory sequences of the housekeeping gene human triosephosphate isomerase. *Journal of Biological Chemistry* 264, 5177-87
- Breathnach, R., Chambon, P., (1981) Organisation and expression of eukaryotic split genes coding for proteins. *Annual Review of Biochemistry* 50, 349-383.
- Brinkman A., Groffen, C., Kortleve, D.J., Geurts van Kessel, A., Drop, S.L., (1988) Isolation and characterisation of a cDNA encoding the low molecular weight insulin-like growth factor binding protein (IBP-1). *EMBO Journal* 7, 2417-23.

- Brown, C.G.D. (1990) Control of Tropical Theileriosis (*Theileria annulata* infection) of cattle. *Parasitologia*. 32, 23-31
- Brown, C.G.D., (1990) *In vitro* cultivation of *Theileria annulata*. In "The *in vitro* Cultivation of the Pathogen of tropical Diseases" D., S. Rowe and h. Hirumi. Eds), Tropical Disease Research Series No. 3, pp 127-144. Schwabe, Basel Switzerland.
- Bruce, M.C., Alano, P., Duthie, S., Carter, R., (1990) Commitment of the malaria parasite *Plasmodium falciparum* to sexual and asexual development. *Parasitology*, 100, 191-200
- Brun R., Schonenberger, M., (1981) Stimulating effect of citrate and cis-aconitate on the transformation of *Trypanosoma brucei* bloodstream forms to procyclic forms *in vitro*. *Zeitschrift Fur Parasitenkunde* 66, 17-24
- Buratowski, S., Hahn, S., Sharp, P.A., Guarente, L., (1988) Function of a yeast TATA element-binding protein in a mammalian transcription system. *Nature* 334, 37-42.
- Buratowski, S., (1994) The basics of basal transcription by RNA polymerase II. *Cell*, 77, 1-3
- Burridge, M.J., Brown, C.G.D., Kimber, C.D., (1974) *Theileria annulata*: cross reactions between a cell culture schizont antigen and antigens of East African *Theileria* species in the indirect fluorescent antibody test. *Experimental Parasitology*. 35, 374-380.
- Calkhoven, C.F, Ab, G., (1996) Multiple steps in the regulation of transcription-factor level and activity. *Biochemical Journal*. 317, 326-342
- Campbell, J. D., Brown, D. J., Glass, E. J., Hall, F. R. and Spooner, R. L. (1994). *Theileria annulata* sporozoite targets. *Parasite Immunology* 16: 501-505.
- Carrington, M., Allsopp, B., Baylis, H., Malu, N. M., Shochat, Y., Sohal, S. (1995). Lymphoproliferation caused by *Theileria annulata* and *Theileria parva*. In *Molecular Approaches to parasitology* eds J. Boothroyd and R. Komuniecki (Wiley -Liss) p43-46
- Carson, C.A., Timms, P., Cowman, A.F., Stewart, N.P., (1990) *Babesia bovis*: Evidence for selection of subpopulations during attenuation. *Experimental Parasitology* 70, 404-410
- Carthew, R.W., Chodosh, L.A., Sharp, P.A., (1985) An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. *Cell* 43, 439-448
- Chodosh, L.A., Carthew, R.W., Sharp, P.A., (1986) A single polypeptide possesses the binding transcription activities of the adenovirus major late transcription factor. *Molecular and Cellular Biology*, 6, 4723-4733
- Chang, K.P., Dwyer, D.M., (1976) Multiplication of a human parasite (*Leshmania donovani*) in phagolysosomes of hamster macrophage *in vitro*, *Science* 193, 678-680
- Chang, P.K., Enrich, K.C., Linz, J.E., Bhatnagar, D., Cleveland, T.E., Bennett, J.W., (1996) Characterisation of the *Aspergillus parasiticus* niaD and niiA gene cluster. *Current Genetics* 30, 68-75

Chen, R., Ingraham, H.A., Treacy, M.N., Albert, V.R., Wilson, L., Rosenfeld, M.G., (1990) Auto-regulation of Pit-1 gene expression mediated by two cis-active promoter elements. *Nature*, 346, 583-586.

Chiba, T., IKawa, Y., Todokoro, K., (1991) GATA-1 transactivates erythropoietin receptor gene, and erythropoietin receptor-mediated signals enhance GATA-1 gene expression. *Nucleic Acids Research* 19, 3843-8

Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proceedings of the National Academy of Sciences, USA* 81: 1991-1995.

Clayton, D.A., (1984) Transcription of the mammalian mitochondrial genome. *Annual Review of Biochemistry* 53, 573-594

Clayton, C.E., Fueri, J.P., Itzakhi, J.E., Bellofatto, V., Sherman, D.P., Wisdom, G.S., Vijayasarathy, S., Mowatt, M.R., (1990) Transcription of the procyclic acidic repetitive protein genes of *Trypanosoma brucei*. *Molecular and Cellular Biology*. 10, 3036-3047

Cohen, L., Lacoste, J., Parniak, M., Daigneault, L., Skup, D., Hiscott, J., (1991) Stimulation of interferon beta gene transcription *in vitro* by purified NF-kappa B and a novel TH protein. *Cell Growth Differentiation*, 2, 323-333

Conrad, P.A., Kelley, B.G., Brown, C.G.D., (1985) Intraerythrocytic schizogony of *Theileria annulata*. *Parasitology*, 91, 67-82.

Corden, J., Wasylyl, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C., Chambon, P., (1980) Promoter sequences of eukaryotic protein-coding genes. *Science* 209, 1406-1414.

Cowman, A.F., Bernard, O., Stewart, N., Kemp, D.J., (1984) Genes of the protozoan parasite *Babesia bovis* that rearrange to RNA species with different sequences. *Cell*, 37, 653-660

Cox, F.E.G. (1982) Cell mediated immunity in theileriosis. *Nature*, 295 14

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

Creemers, P., (1982) Lack of reactivity of sera from *Theileria parva*-infected and recovered cattle against cell membrane antigens of *Theileria parva* transformed cell lines. *Veterinary Immunology and Immunopathology*. 3, 427-438.

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

Czichos, J., Nonnengaesser, C., Overath, P., (1986) *Trypanosoma brucei*: cis-aconitate and temperature reduction as triggers of synchronous transformation of bloodstream to procyclic trypomastigotes *in vitro*. *Experimental Parasitology*. 62, 283-291

Dalrymple, B., (1990) Cloning and characterisation of the rRNA genes and flanking regions from *Babesia bovis*: use of the genes as strain discriminating probes. *Molecular and Biochemical Parasitology*. 43, 117-124

Danot, O., Raibaud, O., (1994) Multiple protein-DNA and protein-protein interactions are involved in transcriptional activation by MalT Molecular Microbiology 4, 335-46

Darde, M.L., Bouteille, B., Leboutet, M.J., Loubet, A., Pestre_alexandre, M., (1989) *Toxoplasma gondii*: Etude ultrastructurale des formations kystiques observees en culture de fibroblastes humains. Annales de Parasitologie Humaine et comparee, 64, 403-411.

Darghouth, M.A., Ben Miled, L., Bouattour, A., Melrose, T.R., Brown, C.G.D., Kilani, M., (1996) A preliminary study on the attenuation of Tunisian schizont-infected cell lines of *Theileria annulata*. Parasitology Research 82, 647-655

Darling, T.N., Bloon, J.J., (1990) Changes in the shape of *Leishmania major* promastigotes in response to hexoses, proline and hypo-osmotic stress. Journal Protozoology. 37, 267-72

De Gee, A.L.W., Causten, P.H.B., McCann, P.P., Mansfield, J.M., (1984) Morphological changes in *Trypanosoma brucei rhodesiense* following inhibition of polyamine biosynthesis *in vivo*. Tissue and Cell 16, 731-738

Dent C.L., Latchman, D.S., (1993) The DNA mobility shift assay. Chapter 1, Transcription factors: A practical approach. Edited D.S. Latchman, Oxford University Press.

Dickson, J., Shiels, B.R., (1993) Antigenic diversity of a major merozoite surface molecule in *Theileria annulata*. Molecular and Biochemical Parasitology. 57, 55-64.

Diamond, M., Miner, J.N., Yoshinaga, S.K., Yamamoto, K.R., (1990) Transcription factor interactions: selectors of positive or negative regulation from a single DNA element. Science 249, 1266-1272

Dolan, T.T., (1989) Theileriosis: a comprehensive review. Revue Scientifique et Technique (Paris) 8, 11-36

Dolan, T. T. (1992). Recent Developments in the Research and Control of *Theileria annulata*. Proceedings of a Workshop held at ILRAD, Nairobi, Kenya, 17-19 September 1990. (T.T. Dolan, editor). ILRAD, Nairobi.

d'Oliveira, C., van der Weide, M., Habela, M.A., Jacquet P., Jongejan, F., (1995) Detection of *Theileria annulata* in blood samples of carrier cattle by PCR Journal of Clinical Microbiology, 33, 2665-2669

d'Oliveira, C., Feenstra, A., Vos, H., Osterhaus, A.D.M.E., Shiels, B.R., Cornelissen, W.C.A., Jongejan, F., (1997) Induction of protective immunity to *Theileria annulata* using two major merozoite surface antigens presented by different delivery systems. Vaccine, 15, 1796-1804

d' Oliveira, C., Tijhaar, E.J., Shiels, B., Van der Weide, M., Jongejan, F., (1996) Expression of genes encoding two major *Theileria annulata* merozoite surface antigens in *Escherichia coli* and a *Salmonella typhimurium aroA* vaccine strain. Gene, 172, 33-39

Dobbelaere, D.A.E., Spooner, P.R., Barry, W.C., Irvin, A.D., (1984) Monoclonal antibody neutralises the sporozoite stage of different *Theileria parva* stocks. Parasite Immunology 6, 361-370

Dorn, P.L., Aman, R.A., Boothroyd, J.C., (1991) Inhibition of protein synthesis results in super-induction of procyclin (PARP) RNA levels. *Molecular and Biochemical Parasitology*

Drummond, R.O., (1976) Tick-borne livestock diseases and their vector. 4. Chemical control of ticks. *World Animal Review*, 19, 28-33

Dschunkowsky, E., Luhs, J., (1904) Die *Piroplasmosen* der Rinder Zentralblatt für bakteriologie, Parasitenkunde, Infektionskrankheiten und hygiene. 35, 486-492

Duffus, W.P.H., Wagner, G.G., Preston, P.M., (1978) Initial studies on the properties of a bovine lymphoid cell culture line infected with *Theileria parva*. *Clinical. Experimental. Immunology*. 34,347-353.

Duszenko, M., (1990) Differentiation of *Trypanosoma brucei* bloodstream trypomastigotes from long slender to short stumpy-like forms in axenic culture. *Molecular and Biochemical Parasitology* 40, 105-112.

Dynan, W.S., Tjian, R., (1985) Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. *Nature*, 316, 774-778

Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissmann, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C., Proudfoot, N.J., (1980) The structure and evolution of the human β -globin gene family. *Cell* 21, 653-668

Egui, E.M., Emery, D.L., (1981) Genetically restricted cell mediate cytotoxicity in cattle immune to *Theileria parva*. *Nature* 290, 251-254

Ehlers, B., Czichos, J., Overath, P., (1987) Repression and reactivation of the variant surface glycoprotein gene in *Trypanosoma brucei*. *FEBS Letters*, 225, 53-58.

Ellis, J., Hefford, C., Baverstock, P.R., Dalrymple, B.P., Johnson, A.M., (1992) Ribosomal DNA sequence comparison of *Babesia* and *Theileria*. *Molecular and Biochemical Parasitology*, 54, 87-96.

Elsholtz, H.P., Albert, V.R., Treacy, M.N., Rosenfeld, M.G., (1990) A two-base change in a POU factor-binding site switches pituitary-specific to lymphoid-specific gene expression. *Genes and Development* 4, 43-51.

Eperson, S., McMahon-Pratt, D., (1989) I. Extracellular cultivation and morphological characterisation of amastigote like forms of *Leishmania panamensis* and *L.braziliensis*. *Journal of Protozoology*. 36, 502-10

Evans, T., Reitman, M., Felsenfeld, G., (1988) An erythrocyte-specific DNA binding factor *Proceedings of the National Academy of Sciences, USA*, 85, 5976-5980

Fawcett, D.W., Doxsey, S., Stagg, D.A., Young, A.S., (1982)(b) The entry of sporozoites of *Theileria parva* into bovine lymphocytes *in vitro*. Electron microscope observations. *European Journal of Cell. Biology*. 27, 10-21

Fawcett, D.W., Buscher, G., Doxsey, S., (1982) Salivary gland of the tick vector of East Coast Fever. III. The ultrastructure of sporogony in *Theileria parva*. *Tissue and Cell* 14, 183-206.

Fawcett, D., Musoke, A., Voigt, W., (1984) Interaction of sporozoites of *Theileria parva* with bovine lymphocytes *in vitro*. I. Early events after invasion. *Tissue and Cell* 16, 873-884.

Feagin, J.E., Stuart, K., (1985) Differential expression of mitochondrial genes between life cycle stages of *Trypanosoma brucei*. *Proceedings of the National Academy of Sciences, USA* 82, 3380-3384

Feagin, J.E., Jasmer, D.P., Stuart, K., (1985) Apocytochrome *b* and other mitochondrial DNA sequences are differentially expressed during the life cycle of *Trypanosoma brucei*. *Nucleic Acids Research*. 13, 4577-4596

Feagin, J.E., Werner, E., Gardener, M.J., Williamson, D.H., Wilson, R.J.M. (1992) Homologies between the contiguous and fragmented rRNAs of the two *Plasmodium falciparum* extrachromosomal DNAs are limited to core sequences *Nucleic Acids Research* 20, 879-887

Figueirido, M.S., Brownlee, G.G., (1995) Cis-acting elements and transcription factors involved in the promoter activity of the human factor VIII gene. *Journal of Biological Chemistry* 27, 11828-11838.

Flach, E.J., Ouhelli, H., Waddington, D., oudich, M., Spooner, R.L (1994) Factors influencing the transmission and incidence of tropical theileriosis (*Theileria annulata* infection of cattle) in Morocco. In: European Union third Coordination Meeting on tropical Theileriosis. Spooner, R., and Campbell, J.(eds). The Roslin Institute (Edinburgh), Roslin, Midlothian, Scotland, pp 7-9.

Forsyth, K.P., Karamalis, F., Biggs, B., Wycherley, K., Culvenor, J., Brown, G.V., Wilkinson, D., Boyd, A.A., Kemp, D.J., (1990) Characterisation of the cytoadherence phenotype of cloned lines of *Plasmodium falciparum*. VII. International congress of Parasitologie. *Bulletin de la Societe Francaise de Parasitologie* (Suppl. 1) 102

Fried, M.G., Crothers, D.M., (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Research* 9, 6505-6524

Fried, M., (1989) Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay. *Electrophoresis*, 10, 366

Friend, C., Scher, W., Holland, J.G., Sato, T., (1971) Hemoglobin synthesis in murine virus-induced leukemic cells *in vitro*: stimulation of erythroid differentiation by dimethyl sulfoxide. *Proceedings of the National Academy of Sciences, USA* 68, 378-382.

Galas, J.G., Schmitz, A., (1978) DNase footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Research*, 5, 3157-3170

Gale, M., Carter, V., Parsons, M., (1994) Translational control mediates the development regulation of the *Trypanosoma brucei* Nrk protein kinase. *Journal of Biological Chemistry* 269, 31659-31665

Gardner, M.J., Williamson, D.H., Wilso, R.J.M. (1991) A circular DNA in malaria parasites encodes an RNA polymerase like that of prokaryotes and chloroplasts. *Molecular and Biochemical Parasitology*. 44, 115-124

Gajadhar, A.A., Maequardt, W.C., Hall, R., Gunderson, J., Ariztia-Carmona, E.V., Sogin, M.L., (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Cryptosporidium parvum* reveal evolutionary relationships among apicomplexans, dinoflagellates, and ciliates. *Molecular and Biochemical Parasitology* 45, 147-154.

Gaub, M.P., Bellard, M., Scheuer, I., Chambon, P., Sassone-Corsi, P., (1990) Activation of the ovalbumin gene by the estrogen receptor involves the fos-jun complex. *Cell* 63, 1267-1276

Gauer, M., Mackenstedt, U., Melhorn, H., Schein, E., Zapf, F., Njenga, E., Young, A., & Morzaria, S. (1995). DNA measurements and ploidy determination of developmental stages in the life cycles of *Theileria annulata* and *Theileria parva*. *Parasitology Research*, 81, 565-574.

Gazzinelli, R.T., Eltoun, I., Wynn, T.A., Sher, A., (1993) Acute cerebral toxoplasmosis is induced by *in vivo* neutralisation of TNF- α and correlates with down regulated expression of inducible nitric oxidase synthase and other markers of macrophage activation. *Journal of Immunology*, 151, 3672-3681

Gearing, A.J.H., Beckett, P., Christodoulou, M., Churchill, M., Clements, J.M., Crimmin, M., Davidson, A.H., Drummond, A.H., Galloway, W.A., Gilbert, R., Gordon, J.J., Leber, T.M., Mangan, M., Miler, K., Nayee, P., Owen, K., Patel, S., Thomas, W., Wells, G., Wood, L.M., Woolley, K., (1995) Matrix metalloproteinases and processing of pro-TNF- α . *Journal of Leukocyte Biology* 57, 774-777

Gerhards, J., Ebel, T., Dobbelaere, D.D.A.E., Morzaria, S.P., Musoke, A.J., Williams, R.O., Lipp, J., (1994) Sequence and expression of a 90-Kilodalton heat-shock protein family member of *Theileria parva*. *Molecular and Biochemical Parasitology*. 68, 235-246

Gidoni, D., Kadonaga, J.T., Barrera-Saldana, H., Takahashi, K., Chambon, P., Tjian, R., (1985) Bidirectional SV40 transcription mediated by tandem Sp1 binding interactions. *Science* 230, 511-7

Giffin, B.F., McCann, P.P., (1989) Physiological activation of the mitochondrion and the transformation capacity of DFMO induced intermediate and short stumpy bloodstream form *Trypanosomes*. *American Journal of Tropical Medicine and Hygiene*. 40, 487-493

Gill, B.S., Bhattacharyulu, Y., Kaur, D., and Singh, A., (1978) Chemoprophylaxis with tetracycline drugs in the immunisation of cattle against *Theileria annulata* infection. *International Journal of Parasitology*. 8, 467-469

Gill, B.S., Bansal, G.C., Bhattacharyulu, Y., Kaur, D., Singh, A., (1980) Immunological relationship between strains of I Dschunkowsky and Luhs 1904. *Research in Veterinary Science* 29, 93-97.

Gill, B.S., Bhattacharyulu, Y., Kaur, D., Singh, A., (1981) Immunisation of cattle against *Theileria annulata*: a resume of work done at Punjab Agricultural University, Ludhiana. In "Advances in the control of Theileriosis". Eds A.D. Irvine, M.P. Cunningham and A.S. Young. Martinus Nijhoff Publishers, The Hague. pp. 259-261.

Gimour, R.S., (1994) The implication of insulin-like growth factor mRNA heterogeneity. *Journal of Endocrinology*. 140, 1-3.

Glascodine, J., Tetley, L., Tait, A., Brown, D., Shiels, B.R., (1990) Developmental expression of a *Theileria annulata* merozoite surface antigen. *Molecular and Biochemical Parasitology*. 40, 105-112.

Glass, E.J., Innes, E.A., Spooner, R.L., Brown, C.G.D. (1989) Infection of bovine monocyte/macrophage populations with *Theileria annulata* and *Theileria parva*. *Veterinary Immunology and Immunopathology* 22, 355-368.

Glass, E.J., Spooner, R.L., (1990) Parasite-accessory cell interactions in Theileriosis. Antigen presentation by *Theileria annulata* -infected macrophages and production of continuously growing antigen-presenting cell lines. *The European Journal for Immunology* 20, 2491-2497.

Ghosh, P.K., Lebowitz, P., Frisque, F.J., Gluzman, Y., (1981) Identification of a promoter component involved in positioning the 5' termini of simian virus 40 early mRNAs. *Proceedings of the National Academy of Sciences, USA* 78, 100-104

Gill, B.S., Bansal, G.C., Bhattacharyulu, Y., Kaur, D., Sing, A., (1980) Immunological relationship between strains of I Dschunkowsky and Luhs 1904. *Research in Veterinary Science* 29, 93-97

Gray, M. A., Brown, C.G.D., (1981) in *Advances in the Control of theileriosis*, eds Irvin, A.D., Cunningham, M.P., and Young, A.S., (Nijhoff, The Hague), pp 127-131

Gray, M.W., Origin and evolution of mitochondrial DNA. *Annals Review of Cell Biology* 5, 25-50

Grewal, A.S., Singh A., Hondal, J.K., Kapur, J., (1994) *Theileria annulata* vaccination field trials at village level in young crossbred calves in the Punjab State, India (1991-93) In: *European union Third Co-ordination Meeting on tropical theileriosis*. Spooner, R., and Campbell, J., (eds), The Roslin Institute (Edinburgh), Roslin, Midlothian, Scotland, pp 30-36.

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

Gunderson, J.H., Sogin, M.L., Woollett, G., Hollingdale, M., de la Cruz, V.F., Walters, A.P. McCutchan, T.F. (1987) Structurally distinct stage specific ribosomes occur in *Plasmodium*. *Science* 238, 933-937.

Guntaka, R.V., Kandala, J.C., Reddy, V.D. (1992) Cloning and characterisation of a highly conserved HMG-like protein (Pf16) gene from *Plasmodium falciparum*. *Biochemical and Biophysical Research Communications* 182, 412-419

Gusella, J., Geller, R., Clark, B., Weeks, V., Housman, D., (1976) Commitment to erythroid differentiation by Friend erythroleukemia cells: A stochastic analysis. *Cell*, 9, 221-229.

Hagenbuchle, O., Wellauer, P.K., Cribbs, D.L., Schibler, U., (1984) Termination of transcription in the mouse α -amylase Amy-2^a occurs at multiple sites down stream of the polyadenylation site. *Cell*, 38, 737-744

Hall, F.R., (1988) Antigens and immunity in *Theileria annulata*. *Parasitology Today* 4, 257-261

Hall, R., Coggins, L., McKellar, S., Sheils, B., Tait, A., (1990) Characterisation of an extrachromosomal DNA element for *Theileria annulata*. Molecular and Biochemical Parasitology, 38, 253-260

Hall, R., Hunt, P.D., Carrington, M., Simmons, D., Williamson, S., Mecham, R.P., Tait, A., (1992) Mimicry of elastin repetitive motifs by *Theileria annulata* sporozoite surface antigen. Molecular and Biochemical Parasitology 53, 105-112

Hashemi-Fesharki, R., (1988) Control of *Theileria annulata* in Iran. Parasitology Today, 4, 36-40.

Hashemi-Fesharki, R. (1992). Theileriosis due to *Theileria annulata* in Iran. In "Recent Developments in the Research and Control of *Theileria annulata*: Proceedings of a Workshop held at ILRAD 17-19 September 1990" (T.T. Dolan, editor) pp 15-18. Nairobi, ILRAD.

Hawkins, J.D., (1996) Gene Structure and expression. Cambridge University Press, Third edition.

Heath, S., Hieny, S., Sher, A., (1990) A cyclic AMP inducible gene expressed during the development of infective stages of *Trypanosoma cruzi*. Molecular and Biochemical Parasitology. 43, 133-142

Hesse, F., Selzer, P.M., Muhlstadt, K., Duszenko, M. (1995) A novel cultivation technique for long term maintenance of bloodstream form *Trypanosomes in vitro*. Molecular and Biochemical Parasitology, 70, 157-166

Hightower, R.C., Wong, M.L., Ruiz-Perez, L., Santi D.V., (1987) Electron microscopy of amplified DNA forms in antifolate-resistant *Leishmania*. Journal of Biological Chemistry 262, 14618-14624

Hodgkin, J., (1990) Sex determination compared in *Drosophila* and *Caenorhabditis*. Nature, 344, 721-728

Hoffmann, I, Birnstiel, M.L., (1990) Cell cycle-dependent regulation of histone precursor mRNA processing by modulation of U7 SnRNA accessibility. Nature, 346, 665-668

Hooshmand-Rad, P., Hawa, N.J., (1973) Transmission of *Theileria hirci* in sheep by *Hyalomma anatolicum anatolicum*. Tropical Animal Health and Production. 5, 103-108

Hooshmand-Rad, P., (1976) The pathogenesis of anaemia in *Theileria annulata* infection. Research in Veterinary Science. 20, 324-329

Hulliger, L., (1965) Cultivation of three species of *Theileria* in Lymphoid cell *in vitro*. Journal of Protozoology. 12, 649-655

Hulliger, L., Brown, C.G.D., Wilde, J.K.H. (1966) Transition of developmental states of *Theileria parva in vitro* at high temperature. Nature 211, 328-329

Hunt, C., Morimoto, R.I., (1985) Conserved features of eukaryotic hsp70 genes revealed by comparison with the nucleotide sequence of human hsp70. Proceedings of the National Academy of Sciences, USA 82, 6455-9

Iams, K.P., Hall, R., Webster, P., Musoke, A.J., (1990) Identification of λ gt11 clones encoding the major antigenic determinants expressed by *Theileria parva* sporozoites. *Infection and Immunity* 58, 1828-1834

Irvin, A.D., Ocama, J., Spooner, P.R., (1982) Cycle of bovine lymphoblastoid cells parasitised by *Theileria parva*. *Research in Veterinary Science*, 33, 298-304.

Irvin, A.D., (1987) Characterisation of species and strains of *Theileria*. *Advances in Parasitology* 26, 145-197

Ito, E., Toki, T., Ishihara, H., Ohtani, H., Gu, L., Yokoyama, M., Engel, J.D., Yamamoto, M., (1993) Erythroid transcription factor GATA-1 is abundantly transcribed in mouse testis. *Nature*, 362, 466-468

Jacquiet, P., Colas, F., Cheikh, D., Thiam, E., Ly, B.A., (1994) Epidemiologie descriptive de la theileriose bovine a *Theileria annulata* en Mauritanie, Afrique de l'Ouest subsaharienne. *Revue D Elevage et de Medecine Veterinaire des pays Tropicaux.*, 47, 147-155.

Janse, C.J., Ramesar, J., van den Berg, F.M., Mons, B., (1992) *Plasmodium berhei*: *In vivo* generation and selection of karyotype mutants and non-gametocyte producer mutants. *Experimental Parasitology*, 74, 1-10.

Jarrett, W.F.H., Crichton, G.W., Pirie, H.M., (1969) *Theileria parva*: Kinetics of Replication. *Experimental Parasitology*. 24, 9-25.

Jasmer, D.P., Feagin, J.E., Stuart, K., (1985) Diverse patterns of expression of the cytochrome *c* oxidase subunit I gene and unassigned reading frames 4 and 5 during the life cycle of *Trypanosoma brucei*. *Molecular and Cellular Biology*. 5, 3041-3047.

Jasmer, D.P., Reduker, D.W., Goff, W.L., Stiller, D., McGuire, T.C., (1990) DNA probes distinguish geographical isolates and identify a novel DNA molecule of *Babesia bovis*. *Journal of Parasitology* 6, 834-841

Jefferies, D., Tebabi, P., Pays, E., (1991) Transient activity assays of the *Trypanosoma brucei* variant surface glycoprotein gene promoter: control of gene expression in the post-transcriptional level. *Molecular and Cellular Biology*. 11, 338-343

Ji, Y.E., Mericle, B.L., Rehkopf, D.H., Anderson, J.D., Feagin J.E., (1996) The *Plasmodium falciparum* 6kb element is polycistronically transcribed. *Molecular and biochemical Parasitology* 81, 211-223.

Johnston, A.M., Illana, S., Dubey, J.P., Dame, J.B. (1987) *Toxoplasma gondii* and *Hammondia hammondii*: DNA comparison using cloned rRNA gene probes. *Experimental Parasitology*. 63, 272-278

Jones, K.A., Kadonaga, J.T., rosenfeld, P.J., Kelley, T.J., Tjian, R., (1987) A cellular DNA binding protein that activates eukaryotic transcription and DNA replication. *Cell* 48, 79-89.

Jongejan, F., Uilenberg, G., (1994) Ticks and control methods. *Rev. Sci. tech. Off. int. Epiz.*, 13, 1201-1226

Joseph, J.Y., Aldritt, S.M., Unnasch, T., Pujalon, O., Wirth, D.F., (1989) Characterisation of a conserved extra-chromosomal element isolated from the avian malarial parasite *Plasmodium gallinaceum* *Molecular and Cellular Biology*. 9, 3621-3629

- Joshi, M., Dwyer, D.M., Nakhashi, H.L., (1993) Cloning and characterisation of differentially expressed genes from *in vitro*-grown "amastigotes" of *Leishmania donovani*. *Molecular and Biochemical Parasitology*. 58, 345-54
- Jura, W.G.Z.O., Brown, C.G.D., Kelley, B., (1983) Fine structure and invasive behaviour of the early developmental stages of *Theileria annulata in vitro*. *Veterinary Parasitology* 12, 31-44.
- Jura, W.G.Z.O. (1984) Factors affecting the capacity of *Theileria annulata* sporozoites to invade bovine peripheral blood lymphocytes. *Veterinary Parasitology* 16, 215-223.
- Kachani, M., Spooner, R.L., Rae, P., Bell-Sakyi, L., Brown, C.G.D. (1992) Stage-specific responses following infection with *Theileria annulata* as evaluated using ELISA. *Parasitology Research.*, 78, 43-47.
- Kachani, M., Flach, E., Williamson, S., McDonald, F., Shiels, B., Spooner, R.L., Ouhelli, H., (1994) Use of ELISA in theileriosis studies in Morocco. In: European Union Third Co-ordination Meeting on tropical Theileriosis. Spooner, R. and Campbell, J. (eds), The Roslin Institute (Edinburgh), Roslin, Midlothian, Scotland, pp 49-51.
- Kadonaga, J.T., Jones, K.A., Tjian, R., (1986) Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends in Biochemical Science*. 11, 20-23.
- Kadonaga, J.T., Carner, K.R., Masiarz, F.R., Tjian, R., (1987) Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell*, 51, 1079-1090.
- Kairo, A., Fairlamb, A.H., Gobright, E., Nene, V., (1994) A 7.1kb linear DNA molecule of *Theileria parva* has scrambled rDNA sequences and open reading frames for mitochondrially encoded proteins. *EMBO Journal* 13, 898-905
- Katende, J.M., Coddeeris, B.M., Mozaria, S.P., Nkonge, C.G., Musoke, A.J., (1990) Identification of a *Theileria mutans*-specific antigen for use in an antibody antigen detection ELISA. *Parasite. Immunology*. 12, 419-433.
- Katzer, F., Carrington, M., Knight, P., Williamson, S., Tait, A., Morrison, I.W., Hall, R., (1994) Polymorphism of *SPAG-1*, a candidate antigen for inclusion in a subunit vaccine against *Theileria annulata*. *Molecular and Biochemical Parasitology*. 67, 1-10
- Katzer, F., (1995) Molecular genetic and functional analyses of surface molecules of *Theileria annulata* sporozoites. DPhil. Thesis. University of York
- Kawazu, S., Sugimoto, C., Kamio, T., Fujisaki, K., (1992) Molecular cloning and immunological analysis of immunodominant piroplasm surface protein of *Theileria sergenti* and *T. buffeli*. *Journal of Veterinary Medical. Science*. 54, 305-311.
- Kawazu, S., Sugimoto, C., Kamio, T., Fujisaki, K., (1992b) Antigenic difference between Japanese *Theileria sergenti* and other benign *Theileria* species of cattle from Australia (*T. buffeli*) and Britain (*T. orientalis*) *Parasitology Research* 78, 130-135
- Kawamoto, S., Takahashi, K., Onuma, M., Kurosawa, T., Sonoda, M., (1990) Invasion of Bovine erythrocytes by *Theileria sergenti* piroplasms *in vitro*. *Japanese Journal for Veterinary Science* 52, 1261-1263.

Kelleher, R.J. III, Flanagan, P.M., Kornberg, R.D., (1990). A novel mediator between activator proteins and RNA polymerase II transcription apparatus. *Cell* 61, 1209-1215

Kibe, M.K., ole-MoiYoi O.K. Nene, V., Khan, B., Allsopp, B.A. Coliins, N.E. Morzaria, S.P. Gobright E.I., Bishop, R.P. (1994) Evidence for two single copy units in *Theileria parva* ribosomal RNA genes. *Molecular and Biochemical Parasitology*. 66, 249-259

Kiltz, H.H., Uilenberg, G., Franssen, F.F.J., Perie, N.M., (1986) *Theileria orientalis* occurs in Central Africa. *Research in Veterinary Sciences*., 40, 197-200

Kimber, C.D., Purnell, R.E., Sellwood, S.A. (1973) The use of fluorescent antibody techniques to detect *Theileria parva* in salivary glands of the tick *Rhipicephalus appendiculatus*. *Research in Veterinary Sciences* 14, 126-127.

Kinnaird, J.H., Logan, M., Kirvar, E., Tait, A., Carrington, M., (1996) The isolation and characterisation of genomic and cDNA clones coding for a cdc2-related kinase (ThCRK2) from the bovine protozoan *Theileria*. *Molecular Microbiology*. 22, 293-302

Knepel, W., Jepeal, L., Habener, J.F., (1990) A pancreatic islet cell-specific enhancer like element in the glucagon gene contains two domains binding distinct cellular proteins. *Journal of Biological Chemistry* 265, 8725

Knight, P.A., (1993) The characterisation of *Theileria annulata* sporozoite surface antigens. Ph.D. Thesis, University of Glasgow.

Knight, P., Musoke, A.J., Gachanja, J.N., nene, V., Katzer, F., Boulter, N., R.Hall, Brown, C.G.D., Williamson, S., Kirvar, E., Bell-Sakyi, L., Hussain, K., Tait, A., (1996) Conservation of neutralising Determinants between the sporozoite surface antigens of *Theileria annulata* and *Theileria parva*. *Experimental Parasitology*. 82, 229-241.

Knowles, D.P., (1996) Control of *Babesia equi* Parasitemia. *Parasitology Today*. 12, 195-198

Kollmar, R., Farnham, P.J., (1993) Site-specific initiation of transcription by RNA polymerase II. *Proceeding of the Society of Experimental Biology Medicine* 203, 127-139

Kwok, S., Kellog, D.E., McKinney, N., Spasic, D., Goda, L., Levenson, C., Sninsky, J.J., (1990) Effects of primer-template mis-matches on the polymerase chain reaction:- human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*, 18, 999-1005

Labbe, S., Prevost, J., Remondelli, P., Leone, A., Seguin, C., (1991) A nuclear factor binds to the metal regulatory elements of the mouse gene encoding metallothionein-I. *Nucleic Acids Research* 19, 4225-31.

Lamph, W.W., Dwarki, V.J., Ofir, R., Montminy, M., Verma, I., (1990) Negative and positive regulation by transcription factor cAMP response element-binding protein is modulated by phosphorylation. *Proceedings of the National Academy of Sciences, USA* 87, 4320-4324.

Lane, D., Prentki, P., Chandler, M., (1992) Use of gel retardation to analyse protein-nucleic acid interactions. *Microbiological Reviews* 56, 509-528

Laney, J.D., Biggin, M.D., (1992) Zeate, a nonessential gene, potently activates Ultrabithorax transcription in the *Drosophila* embryo. *Genes and Development* 6, 1531-41

Lanzer, M., de bruin, D., Ravetch, J.V., (1992a) A sequence element associated with the *Plasmodium falciparum* KAHRP gene is the site of developmentally regulated protein-DNA interactions. *Nucleic Acids Research*. 20, 3052-3056

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

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

Larkin, N.D., (1993) The DNA mobility shift assay. Chapter 2, Transcription factors: A practical approach. Edited D.S. Latchman, Oxford University Press.

Latchman, D.S., (1990) Gene Regulation. A eukaryotic perspective. Unwin Hyman, London.

Lechn, D.A., Bustin, M., (1993) Evolutionarily conserved motifs and protein binding elements in the 5' region of the chromosomal protein HMG-14 gene. *DNA Cell Biology* 12, 753-761

Lehming, N., Thanos, D., Brickman J.M., Ma J., Maniatis, T., Ptashne, M., (1994) An HMG-like protein that can switch a transcriptional activator to a repressor. *Nature* 371, 175-179.

Levine (1988) The Protozoan Phylum Apicomplexa Volume II. CRC Press Boca Raton, Florida.

Lewin, B., (1990(a)) Commitment and activation at pol II promoters: a tail of protein-protein interactions. *Cell* 61, 1161-1164

Lewin, B., (1990(b)) *Genes IV* Oxford University Press

Li, S., Crenshaw III, E.B., Rawson, E.J., Simmons, D.M., Swanson, L.W., Rosenfeld, M.G. (1990) Dwarf locus mutants lacking three pituitary cell types result from mutations in the POU-domain gene *pit1*. *Nature*, 347, 528-533.

Liu, F., Green, M.R., (1990) A specific member of the ATF transcription factor family can mediate transcription activation by the adenovirus E1a protein. *Cell* 61, 1217-1224.

Lucibello, F.C., Slater EP, Jooss, K.U., Beato, M., Muller, R., (1990) Mutual transrepression of Fos and the glucocorticoid receptor: involvement of a function domain in Fos which is absent in FosB. *EMBO J* 9, 2827-2843

Maniatis, T., Goodbourn, S., Fischer, J.A., (1987) Regulation of inducible and tissue-specific gene expression. *Science* 236, 1237-45

Martin, D.I.K., Tsai, S.F., Orkin, S.H., (1989) Increased gamma-globin expression in a nondeletion HPFH mediated by an erythroid-specific DNA-binding factor *Nature*, 338, 435-438

- Martin, D.I., Orkin, S.H., (1990) Transcriptional activation and DNA binding by the erythroid factor GF-1/NF-E1/Eryf 1. *Genes and Development* 4, 1886-98
- Mason, P.J., Elkington, J.A., Malgorzatta, M.L., Jones, M.B., Williams, J.G., (1986) Mutations downstream of the polyadenylation site of a *Xenopus* β -globin mRNA affect the position but not the efficiency of 3' processing. *Cell* 46 263-270
- Mason, P.J., Shiels, B., Tait, A., Beck, P., Hall, R., (1989) Sequence and expression of a gene from *Theileria annulata* coding for a 70-kilodalton heat-shock protein. *Molecular and biochemical Parasitology* 37, 27-36
- Masquilier, D., Sassone-Corsi, P., (1992) Transcriptional cross-talk: nuclear factors CREM and CREB bind to AP-1 sites and inhibit activation by Jun. *Journal of Biological Chemistry* 267, 22460-22466
- Matsui, T., Fujino, T., Kobayashi, F., Tsutsumi, Y., Tsuji, M., (1996) Attenuation of *Eimeria caviae* by selection for precocious development. *International Journal of Parasitology*, 26, 1243-1248
- McAndrew, M.B., Read, M., Sims, P.F., Hyde, J.E., (1993) Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T rich human malaria parasite *Plasmodium falciparum*. *Gene*, 124,165-171
- McCutchan, T.F., (1986) The ribosomal gene of *Plasmodium*. *International Review Cytology* 99, 295-309
- McGinnis, W., Krumlauf, R., (1992) Homeobox genes and axial patterning. *Cell*. 68, 283-302
- McHardy, N., (1978) *In vitro* studies on the action of menoctone and other compounds on *Theileria parva* and *Theileria annulata*. *Annals of Tropical Medicine and Parasitology*. 72, 501-511
- McHardy, N., Hudson, A. T., Morgan, D. W. T., Rae, D. G. and Dolan, T. T. (1983). Activity of 10 naphthoquinones, including parvaquone (993C) and menoctone, in cattle artificially infected with *Theileria parva*. *Research in Veterinary Science* 35: 29-33.
- McHardy, N., Wekesa, L.S., Hudson, A.T., Randall, A.W., (1985) Antitheilerial activity of BW720C (buparvaquone): a comparison with parvaquone. *Research in Veterinary Sciences* 39, 29-33.
- McKnight, S., Tjian, R., (1986) Transcriptional selectivity of viral genes in mammalian cells. *Cell* 46, 795-805
- Megson, A., Inman, G.J., Hunt, P.D., Baylis, H.A., Hall, R. (1991) The gene for apocytochrome *B* of *Theileria annulata* resides on a small linear extrachromosomal element. *Molecular and Biochemical Parasitology*, 48, 113-116
- Mehlhorn, H., Schein, E., (1984) The Piroplasm: Life cycle and sexual stages. *Advances in Parasitology* 23, 37-103

Melrose, T.M., Brown, C.G.D., Morzaria, S.P., Ocama, J.G.R., Trvin, A.D. (1984) Gluscoe Phosphate Isomerase polymorphism in *Theileria annulata* and *T.parva*. Tropical Animal Health Production 16, 239-245

Melrose, T.R., Wilkie, G.M., Brown, C.G.D., (1991) Studies on cloned populations of *Theileria annulata*. In "Orientation and Coordination of Research on Tropical theileriosis. Proceedings of the Second International Workshop Sponsored by the EC science and technology for Development Programme" (D.K., Singh and B.C Varshney, Eds), pp71-75

Mercurio, F., Karin, M., (1989) Transcription factors AP-3 and Ap-2 interact with the SV40 enhancer in a mutually exclusive manner. EMBO Journal 8, 1455-60.

Minami, T., Spooner, P.R., Irvin, A.D., Ocama, J.G.R., Dobbelaere, D.A.E., Fujinaga, T., (1983). Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. Research in Veterinary Science. 35, 334-340

Michaelis, G., Vahrenholz, C., Pratje, E., (1990) Mitochondrial DNA of *Chlamydomonas reinhardtis*: the gene for apocytochrome *b* and the complete functional map of the 15.8kb DNA. Molecular and General Genetics. 223, 211-216.

Miner, J.N., Diamond, M.I., Yamamoto, K.R., (1991) Joints in the regulatory lattice: composite regulation by steroid receptor-Ap1 complexes. Cell Growth and Differentiation, 2, 525-530

Mitchell, P.J., Tjian, R., (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245, 371-378

Molan, A., Sergura, C., Guzman, F., Lozada, D., Patarroyo, M.E. (1992) In human malaria protective antibodies are directed mainly against the Lys-Glu ion pair within the Lys-Glu-Lys motif of the synthetic vaccine spf 66. Parasite Immunology 14, 111-124.

Moltman, U.G., Mehlhorn, H., Schein, E., Rehbein, G., Voigt, W.P., Zwegarth, E., (1983) Fine structure of *Babesia equi* Laveran 1901 within lymphocytes and erythrocytes of horses: an *in vivo* and *in vitro* study. Journal of Parasitology, 69, 111-126.

Morimoto, R.I., Tissieres, A., Georgopoulos, C., (1990) Stress Proteins in Biology and Medicine. Cold Spring Harbor monograph Series, vol19. Cold Spring Harbor laboratory press. Cold Spring Harbor. NY

Morzaria, S. P., Roeder, P. L., Roberts, D. H., Chasey, D. and Drew, T. W. (1982). Characteristics of a continuous suspension cell line (BL20) derived from a calf with sporadic bovine leukosis. In Fifth International Symposium on Bovine Leukosis, Tubingen, 1982. Straub, O.C. (editor), Commission of the European Communities Report Eur 8471EN pp519-528

Morzaria, S.P., Spooner, P.R., Bishop, R.P., Musoke, A.J., Young, J.R., (1990) Sfi I and Not I polymorphisms in *Theileria* stocks detected by pulse field gel electrophoresis. Molecular and Biochemical Parasitology. 40, 203-212.

Morzaria, S.P., Nene, V., (1990) Bovine theileriosis: progress in immunisation methods. International Journal on Animal Science 5, 1-14.

- Morzaria, S. P. and Young, J. R. (1992). Restriction mapping of the genome of the protozoan parasite *Theileria parva*. Proceedings of the National Academy of Sciences, USA 89: 5241-5245.
- Mullner, E.W., Kuhun, L.C., (1988) A stem-loop in the 3' untranslated region mediates iron dependent regulation of transferrin receptor mRNA stability in the cytoplasm. Cell 53, 815-25
- Musisi, F.L., (1990) Methods currently used for the control of East Coast fever: their validity and proposals for future control strategies. Parasitologia. 32, 15-22.
- Musoke, A.J., Nantulya, V.M., Buscher, G., Masake, R.A., Otim, B., (1982) Bovine immune responses to *Theileria parva*: neutralising antibodies to sporozoites. Immunology 45, 663-668.
- Musoke, A.J., Nantulya, V.M., Rurangirwa, F.R., Buscher, G., (1985) Evidence for a common protective antigen determinant on sporozoites of several *Theileria parva* strains. Immunology 52, 231-238
- Musoke, A., Morzaria, S., Nkonge, C., Jones, E., Nenes, V., (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proceedings of the National Academy of Sciences U.S.A. 89, 514-518.
- Musoke, A., Nene, V., Morzaria, S.P., (1993) A sporozoite-based vaccine for *Theileria parva*. Parasitology Today. 9, 385-388
- Nachtgal, M.W., Nickel, B.E., Cattini, P.A., (1993) Pituitary-specific repression of placental members of the human growth hormone gene family. A possible mechanism for locus regulation. Journal of Biological Chemistry 268, 8473-9
- Neitz, W.O., (1957) Theileriosis, gonderiosis and cytauxzoonosis, The Onderstepoort Journal of Veterinary Research 27, 275-430.
- Nene, V., Iams, K.P., Gobright, E., Musoke, A.J., (1992) Characterisation of a gene encoding a candidate vaccine antigen of *Theileria parva* sporozoite. Molecular and Biochemical Parasitology 51, 17-28
- Nene, V., Inumaru, S., McKeever, D., Morzaria, S., Shaw, M., Musoke, A., (1995) Characterisation of an insect cell-derived *Theileria parva* sporozoite vaccine antigen and immunogenicity in cattle. Infectious Immunology. 63, 503-508
- Oliver, R. A., Williams, J. L. (1996). Altered expression of Class I major histocompatibility antigens on bovine cells infected with the protozoan parasite *Theileria annulata*. Veterinary Immunology and Immunopathology, 50, 173-179.
- Orkin, S.H., Harosi, F., Leder, P., (1975) Differentiation in erythroleukemic cells and their somatic hybrids. Proceedings of the National Academy of Sciences, USA 72, 98-102
- Ossorio, P.N., Sibley, L.D., Boothroyd, J.C., (1991) Mitochondrial-like DNA sequences flanked by direct and inverted repeats in the nuclear genome of *Toxoplasma gondii*. Journal of Molecular Parasitology 222, 525-536.

- Parsons, M., (1990) Current concepts in stage-regulated gene expression in Kinetoplastida. *Parasitology Today* 6, 244-245
- Pan, A.A., (1984) *Leishmania mexicana*: serial cultivation of intracellular stages in a cell-free medium. *Experimental Parasitology*. 58, 72-80
- Pang, K.M., Knecht, D.A., (1997) Partial inverse PCR: a technique for cloning flanking sequences. *Biotechniques* 22, 1046-1048
- Papadopoulos, B., Perie, N. M., & Uilenberg, G. (1996). Piroplasms of domestic animals in the macedonia region of Greece. *Veterinary Parasitology*, 63, 41-56.
- Papazafiri, P., Ogami, K., Ramji, D.P., Nicosia, A., Monasi, P., Cladaras, C., Zannis, V.I., (1991) Promoter elements and factors involved in hepatic trascription of the human ApoA-I gene positive and negative regulators bind to overlapping sites. *Journal Biological Chemistry* 266, 5790-5797
- Paton, M.G., Barker, G.C., Matsuoka, H., Ramesar. J., Janse, C.J., Waters, A.P., Sinden, R.E., (1993) Structure and expression of a post-transcriptionally regulated malaria gene encoding a surface protein from the sexual stages of *Plasmodium berghei*. *Molecular and Biochemical Parasitology*, 59, 263-276.
- Pave-Preux, M., Aggerbeck, M., Veyssier, C., Bousquet-Lemerceir, B., Hanoune, J., Barouki, R., (1990) Hormonal discrimination among transcription start sites of aspartate amino-transferase. *Journal of Biological Chemistry*, 265, 4444-4448.
- Pays, E., Coquelet, H., Pays, A., Tebabi, P., Steinert, M., (1990) *Trypanosoma brucei*: constitutive activity of the VSG and procyclin promoters. *EMBO Jouranl* 9, 3145-3151.
- Pays, E., Coquelet, H., Tebabi, P., Pays, A., Jefferies, D., Steinert, M., Koenig, E., Williams, R.O., Roditi, I., (1990) *Trypanosoma brucei*: constitutive activity of the VSG and procyclin gene promoters. *EMBO Journal* 9, 3145-3151
- Pays, E., Hanocq-Quertier, J., Hanocq, F., Van Assel, S., Nolan, D., Rolin, S. (1993) Abrupt RNA changes precede the first cell division during the differentiation of *Trypanosoma brucei* bloodstream forms into procyclic forms *in vitro*. *Molecular and Biochemical Parasitology* 61, 107-114
- Pearson, T.W., Dolan, T.T., Stagg, D.A., Lundin, L.B., (1979) Cell mediated immunity to *Theileria*-transformed cell lines. *Nature* 281, 678-680
- Peers, B., Voz, M.L., Monget, P., Mathy-Hatert, M., Brwaer, M., Belayew, A., Martial, J.A., (1990) Regulatory elements controlling pituitary-specific expression of the human prolactin gene. *Molecular and Cellular Biology* 10, 4690-700.
- Pevny, L., Simon, M.C., Robertson, E., Klein, W.H., Tsai, S-F., Agati, V.D., Orkin, S.H., Constantini, F., (1990) Erythroid differentiation in chimaeric mice blocked by a targeted mutation in the gene for transcription factor GATA-1. *Nature* 349, 257-260
- Pipano. E., (1981) Schizonts and tick stages in immunisation against *Theileria annulata* infection. In "Advances in the control of Theilerosis". Eds A.D. Irvin, M.P. Cunningham and A.S. Young. Martinus Nijhoff, The Hague. pp 242-252.

- Pipano, E., (1989) Bovine theileriosis in Israel. *Revue Scientifique et Technique* (Paris) 8, 79-87.
- Plumb, M., Framptin, J., Wainwright, H., Walker, M., Macleod, K., Goodwin, G., Harrison, P., (1989) GATAAG: a cis-control region binding an erythroid-specific nuclear factor with a role in globin and non-globin gene expression. *Nucleic Acids Research*. 17, 73-92.
- Pollack Y., Katzen A.L., Spira D.T., Golenser J., (1982) The genome of *Plasmodium falciparum*: DNA base composition *Nucleic Acids Research*, 10, 539-546
- Polyanovsky, O.L., Stepchenko, A.G., (1990) Eukaryotic transcription factors. *BioEssays* 12, 205-210
- Preston, P.M., Brown, C.G.D., Spooner, R.L. (1983) Cell mediated cytotoxicity in *Theileria annulata* infection of cattle with evidence for BoLA restriction. *Clinical and Experimental Immunology* 53, 88-100.
- Preston, P.M., Brown, C.G.D., (1985) Inhibition of lymphocyte invasion by sporozoites and the transformation of trophozoite infected lymphocytes in vitro by serum from immune cattle. *Parasite Immunology*. 7, 301-314
- Preston, P.M., Brown, C.G.D., (1988) Macrophage-mediated cytostasis and lymphocyte cytotoxicity in cattle immunised with *Theileria annulata* sporozoites or macroschizont-infected cell-lines. *Parasite Immunology*. 10, 631-647.
- Preston, P.M., Brown, C.G.D., Richardson, W., (1992) Cytokines inhibit the development of trophozoite-infected cells of *Theileria annulata* and *Theileria parva* but enhance the proliferation of macroschizont-infected cell lines. *Parasite Immunology*. 14, 125-141.
- Preston, P.M., Brown, C.G.D., Entrican, G., Richardson, W., Boid, R., (1993) Synthesis of tumour necrosis factor-alpha and interferons by mononuclear cells from *Theileria annulata*-infected cattle. *Parasite Immunology*. 15, 525-534
- Pugh, B.F., Tjian, R., (1992) Diverse transcriptional functions of the multi-subunit eukaryotic TFIID complex. *Journal of Biological Chemistry*. 267, 679-682
- Ram, N., Rao, M.V., Prabhu, L., Nirodi, C.S., Sultana, S., Vatsala, P.G., Padmananban., (1995) Characterisation of a negative cis-acting DNA element regulating the transcription of CYP2B1/B2. *Archives of Biochemistry and Biophysics* 317, 39-45
- Radley, D.E., Brown, C.D., Burridge, M.J., Cunningham, M.P., Peirce, M.A., Purnell, R.E., East coast fever: quantitative studies of *Theileria parva* in cattle. *Experimental Parasitology*. 1974, 36, 278-287.
- Reddy, G.R., Chakrabati, G., Yowell, C.A., Dame, J.B. (1991) Sequence microheterogeneity of three small subunit ribosomal RNA genes of *Babesia bigemina*: Expression in erythrocyte culture. *Nucleic Acids Research*. 19, 3641-3645
- Remacle, J.E., Holmberg, S., (1992) A REB1-binding site is required for GCN4-independent ILV1 basal level transcription and can be functionally replaced by an ABF1-binding site. *Molecular and Cellular Biology* 12, 5516-26
- Revzin, A., (1987) Gel electrophoresis assays for DNA-protein interactions. *Biotechniques*, 7, 346-355

Rolin S., Paindavoine, P., Hanocq-Quertier, J., Hanocq, F., Claes, Y., LeRay, D., Overath, P., Pays, E., (1993) Transient adenylate cyclase activation accompanies differentiation of *Trypanosoma brucei* from bloodstream to procyclic forms. *Molecular and Biochemical Parasitology*. 61, 115-126

Rorth P., Nerlov, C., Blasi, F., Johnsen, M., (1990) transcription factor PEA3 participates in the induction of urokinase plasminogen activator transcription in murine keratinocytes stimulated with epidermal growth factor or phorbol-ester. *Nucleic Acids Research* 18, 5009-17

Rouet, P., Raguenez, G., Tronche, F., Mfou'ou V., Salier, J.P., (1995) Hierarchy and positive/negative interplays of the hepatocyte nuclear factors HNF-1, -3 and -4 in the liver specific enhancer for the human alpha-1-microglobulin/bikunin precursor. *Nucleic Acids Research* 23, 395-404

Rueben, R.C.Wife, R.L., Breslow, R., Rifkind, R.A., Marks, P.A., (1976) A new group of potent inducers of differentiation in murine erythroleukemia cells. *Proceedings of the National Academy of Sciences, USA* 73, 862-866

Ruiz i Altaba, A., Ozaki, L.S., Gwadz, R.W., Godson, G.N., (1987) Organisation and expression of the *Plasmodium knowlesi* circumsporozoite antigen gene. *Molecular and Biochemical Parasitology*. 23, 233-245.

Sabanshiev, M., (1994) Theileriosis of cattle in kazakhstan. In: European union Third Co-ordination Meeting on tropical theileriosis. Spooner, R., and Campbell, J., (eds). The Roslin Institute (Edinburgh), Roslin, Midlothian, Scotland, pp 10-12.

Sacks, D.L., Perkins, P.V., (1984) Identification of an infective stage of *Leishmania* promastigotes. *Science* 233, 1417-1419

Sacks, D.L., Perkins, P.V., (1985) Development of infective stage *Leishmania* promastigote within phlebotomine sandflies. *American Journal of Tropical Medicine and Hygiene*, 34, 456-459

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY), 2nd Ed.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* 74: 5463-5467.

Schatt, M.D., Rusconi, S., Schaffner, W., (1990) A single DNA-binding transcription factor is sufficient for activation from a distant enhancer and/or from a promoter position. *EMBO Journal*. 9, 481-487.

Schein, E., (1975) On the life cycle of *Theileria annulata* in the midgut and hemolymph of *Hyalomma anatolicum excavatum*. *Zeitschrift fur Parasitenkunde* 47, 165-167

Schein, E., Voigt, W.P., (1979) Chemotherapy of bovine theileriosis with halofuginone. *Acta Tropica*, 36, 391-394

Scherf, A., Hilbich, C., Sieg, K., Mattei, D., Mercereau-Pulijalon, O., Muller-Hill, B., (1988). The 11-1 gene of *Plasmodium falciparum* codes for distinct fact evolving repeats. *EMBO Journal* 7, 1129-1137

- Schmidt, M.C., Kao, C.C., Pei, R., Berk, A.J., (1989) Yeast TATA-box transcription factor gene. *Proceedings of the National Academy of Sciences, USA* 86, 7785-7789.
- Schule, R., Rangarajan, P., Kliewer, S., Ransone, L.J., Bolado, J., Yang, N., Verma, I.M., Evans, R.M., (1990) Functional antagonism between oncoprotein c-Jun and the glucocorticoid receptor. *Cell* 62, 1217-1226.
- Sealey, L., Chalkley, R., (1987) At least two nuclear proteins bind specifically to the Rous sarcoma virus long terminal repeat enhancer. *Molecular and Cellular Biology* 7, 787-98
- Seed, J.R., Sechelski, J.B., (1989) Mechanism of long slender (LS) to short stumpy (SS) transformation in the African trypanosomes. *Journal of Protozoology*. 36, 572-577
- Serfling, E., (1989) Autoregulation- a common property of eukaryotic transcription factors? *Trends in Genetics*, 5, 131-133
- Sentenac, A., (1985) Eukaryotic RNA polymerases. *CRC Critical Reviews in Biochemistry* 18, 31-91
- Shadel, G.S. Clayton, D.A., (1993) Mitochondrial transcription initiation. Variation and conservation. *Journal of Biological Chemistry*. 268, 16083-16086.
- Shapiro, S.Z., Fujisaki, K., Mozaria, S.P. Webster, P., Fujinaga, T., Spooner, P.R., Irvin, A.D., (1987) A Life-cycle stage-specific antigen of *Theileria parva* recognised by anti-macroschizont monoclonal antibodies. *Parasitology* 94, 29-37.
- Shapira, M., McEwen, J.G., Jaffe, C.L., (1988). Temperature affects on molecular processes which lead to stage differentiation in *Leishmania*. *EMBO Journal*, 7, 2895-2901.
- Shaw, M.K., Tilney, L.G., Musoke, A.J., (1991) The entry of *Theileria parva* sporozoites into bovine lymphocytes: Evidence for MHC class I involvement. *The Journal of Cell Biology* 113, 87-101.
- Shaw, M.K., Tilney, L.G., (1992) How individual cells develop from a syncytium: merogony in *Theileria parva* (Apicomplexa). *Journal of Cell Science*. 101, 109-123
- Shaw, M.K., Tilney, L.G., Musoke, A.J. Teale, A.J. (1995) MHC class I molecules are an essential cell surface component involved in *Theileria parva* sporozoite binding to bovine lymphocytes. *Journal of Cell Science*. 108, 1587-1596.
- Shaw, M.K., (1997) The same but different: the biology of *Theileria* sporozoite entry into bovine cells. *International journal for Parasitology*, 27, 457-474
- Sheffery, M., Marks, P.A., Rifkind, R.A. (1984) Gene expression in murine erythroleukemia cells: transcriptional control and chromatin structure of the $\alpha 1$ globin gene. *Journal of Molecular Biology*., 172, 417
- Shiels, B.R., McDougall, C., Tait, A., Brown, C.G.D. (1986) Antigenic diversity of *Theileria annulata* macroschizonts. *Veterinary Parasitology*, 21 1-10.
- Shiels, B., Hall, F.R., Glascodine, J., McDougall, C., Harrison, C., Taracha, E., Brown, D., Tait, A., (1989) Characterisation of surface polypeptides on different life-cycle stages of *Theileria annulata*. *Molecular and Biochemical Parasitology*. 34, 209-220.

Shiels, B., Kinnaird, J., McKellar, S., Dickson, J., Ben Miled, L., Melrose, R., Brown, D. and Tait, A. (1992). Disruption of synchrony between parasite growth and host cell division is a determinant of differentiation to the merozoite in *Theileria annulata*. *Journal of Cell Science* 101: 99-107.

Shiels, B., Smyth, A., Dickson, J., McKellar, S., Tetley, L., Fujisaki, K., Hutchinson, B., Kinnaird, J.H., (1994) A stoichiometric model of stage differentiation in the protozoan parasite *Theileria annulata*. *Molecular and Cellular Differentiation*, 2, 101-125

Shiels, B., d'Oliveira, C., McKellar, S., Ben-Miled, L., Kawazu, S., Hide, G., (1995) Selection of diversity at putative glycosylation sites in the immunodominant merozoite/piroplasms surface antigen of *Theileria* parasites. *Molecular and Biochemical Parasitology*. 72, 149-162

Shiels, B., Aslam, N., McKellar, S., Smyth, A., Kinnaird, J., (1997) Modulation of protein synthesis relative to DNA synthesis alters the timing of differentiation in the protozoan parasite *Theileria annulata*. *Journal of cell Science* 110, 1441-1451

Shirley, M.V., Biggs, B.A., Forsyth, K.P., Brown, H.J., Thompson, J.K., Brown, G.V., Kemp, D.J., (1990) Chromosome 9 from independent clones and isolates of *Plasmodium falciparum* undergoes subtelomeric deletions with similar breakpoints in vitro. *Molecular and Biochemical Parasitology* 40, 137-146.

da Silva, R., Sacks, D.L., (1987) Metacyclogenesis is a major determinant of *Leishmania* promastigote virulence and attenuation. *Infectious. Immunology*. 55, 2802-2807

Singh, D.K., Jagdish, S., Gautam, O.P. (1977) Cell mediated immunity in tropical theileriosis (*Theileria annulata* infection). *Research in Veterinary. Science*. 23, 391-392

Smale, S.T., Baltimore, D., (1989) The initiator as a transcription control element. *Cell*, 57, 103-113.

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

Soete, M., Fortier, B., Camus, D., Dubremetz, J.F., (1993) *Toxoplasma gondii*: Kinetics of bradyzoite-tachyzoite interconversion *in vitro*. *Experimental Parasitology*. 76, 249-264.

Soete, M., Camus, D., Dubremetz, J.F., (1994) Experimental induction of bradyzoite-specific antigen expression and cyst formation by the RH strain of *Toxoplasma gondii* in vitro. *Experimental Parasitology* 78, 361-370

Soldati, D., Boothroyd, J.C., (1995) A selector of transcription initiation in the protozoan parasite *Toxoplasma gondii*. *Molecular and Cellular Biology*, 15, 87-93

Spieth, J., Shim, Y.H., Lea, K., Conrad, R., Blumenthal, T., (1991) elt-1, an embryonically expressed *Caenorhabditis elegans* gene homologous to the GATA transcription factor family. *Molecular and Cellular Biology*, 11, 4651-4659.

Stamler, J.S., Singel, D.J. Loscalzo, J., (1992) Biochemistry of nitric oxide and its redox-activated forms. *Science* 258, 1898-1902.

- Strauss, F., Varshavsky, A., (1984) A protein binds to a satellite DNA repeat at three specific sites that would be brought into mutual proximity by DNA folding in the nucleosome. *Cell* 37, 889-901
- Subramanian, G., Ray, D., Naithani, R.C., (1986) *In vitro* culture and attenuation of macroschizonts of *Theileria annulata* (Dschunowsky and Luhs, 1904) and *in vivo* use as a vaccine. *Indian Journal of animal Science* 56, 174-182.
- Sugimoto, C., Kawazu, S., Kamio, T., Fujisaki, K., (1991) Protein analysis of *Theileria sergenti/buffeli/orientalis* piroplasms by two-dimensional polyacrylamide gel electrophoresis. *Parasitology* 102, 341-346
- Suplick, K., Morrissey, J., Vaidya, A.B., (1990) Complex transcription from the extrachromosomal DNA encoding mitochondrial functions of *Plasmodium yoelli*. *Molecular and Cellular Biology*, 10, 6381-6388
- Sutherland, I.A., Shiels, B.R., Jackson, L., Brown, C.G.D., Brown, D.J., Preston, P.M., (1996) *Theileria annulata*: Altered gene expression and clonal selection during continuous *in vitro* culture. *Experimental Parasitology*, 83, 125-133.
- Swan, D.G., McDonald, F., Shiels, B.R., (1996) A stage regulated cyclophin gene from *Theileria annulata*. *Molecular and Biochemical Parasitology* 81, 111-114
- Tait A., Hall, F.R., (1990) *Theileria annulata*: control measures, diagnosis and the potential use of subunit vaccines. *Revue Scientifique et Technique (Paris)* 9, 387-403
- Tamura, T., Sumita, K., Hirose, S., Mikoshiba, K., (1990) Core promoter of the mouse myelin basic protein gene governs brain specific transcription *in vitro*. *EMBO Journal* 9, 3101-3108
- Tanaka, M., Ohgitani, T., Okabe, T., Kawamoto, S., Takahashi, K., Onuma, M., Kawakami, Y., Sasahi, N., (1990). Protective effect against intraerythrocytic merozoites of *Theileria sergenti* infection in calves by transfer of monoclonal antibody. *Japan. Journal of Veterinary Science* 52, 631-633
- Tarella, C., Ferrero, D., Gallo, E., Pagliardi, G.L. Ruscetti, F.W., (1982) Induction of differentiation by dimethyl sulfoxide: Evidence for a stochastic model not linked to cell division cycle. *Cancer Research*. 42, 445-449.
- Teklehaimanot, A., Collins, E.W., Nguyen-Dinh, P., Cambell, C.C., Bhasin, V., (1987) Characterisation of *Plasmodium falciparum* cloned cell lines with respect to gametocyte production *in vitro*, infectivity to *Anopheles* mosquitoes and transmission to *Aotus* monkeys. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81, 885-887.
- Temple, S., Raff, M.C., (1986) Clonal analysis of oligodendrocyte development in culture: evidence for a developmental clock that counts cell divisions. *Cell*. 44, 773-779.
- Tjian, R., Maniatis, T., (1994) Transcriptional activation: A complex puzzle with few easy pieces. *Cell* 77, 5-8.
- Tomavo, S., Boothroyd, J.C., (1995) Interconnection between Organellar functions. Development and drug resistance in the protozoan parasite, *Toxoplasma gondii*. *International Journal of Parasitology*, 23, 1293-1299

- Toye, P.G., Goddeeris, B.M., Iams, K., Musoke, A.J., Morrison, W.I., (1991) Characterization of a polymorphic immunodominant molecule in sporozoites and schizonts of *Theileria parva*. *Parasite Immunology* 13, 49-62
- Truong, A.T., Duez, C., Belayew, A., Renard, A., Pictet, R., Bell, G.I., Martial, J.A., (1984) Isolation and characterisation of the human prolactin gene. *EMBO Journal* 3, 429-37.
- Tsur, I., (1945) Multiplication in vitro of Koch bodies of *Theileria annulata*. *Nature*, 156, 391
- Tully, D.B., Cidlowski, J.A., (1993) Protein-blotting procedures to evaluate interaction of steroid receptors with DNA. *Methods of Enzymology* 218, 535-551
- Uilenberg, G. (1981) *Theileria* infections other than East Coast fever. In *Diseases of Cattle in the Tropics*". Eds. M. Ristic and I. McIntyre. Martinus Nijhoff, The Hague. pp 411-427
- Ulich, T.R., del Castillo, J., Keys, M., Granger, G.A., Ni, R., (1987) Kinetics and mechanisms of recombinant human interleukin-1 and tumour necrosis factor- α -induced changes in circulating numbers of neutrophils and lymphocytes. *Journal of Immunology.*, 139, 3406-3415
- Urquart, G.M., Duncan, J.L., Dunn, A.M., and Jennings, F.W., (1987) *Veterinary Parasitology*. Longman U.K.
- Vaidya, A., Arasu, P., (1987) Tandemly arranged gene clusters of malarial parasites that are highly conserved and transcribed. *Molecular and Biochemical Parasitology* 22, 249-257
- Vaidya, A.B., Akella, R., Suplick, K. (1989) Sequences similar to genes for two mitochondrial proteins and portions of ribosomal RNA in tandemly arrayed 6-Kilobase-pair DNA of a malarial parasite. *Molecular and Biochemical Parasitology*, 35, 97-108
- Vaidya, A.B., Morrissey, J., Plowe, C.V., Kaslow, D.C., Wellems, T.E., (1993) Unidirectional dominance of cytoplasmic inheritance in two genetic crosses of *Plasmodium falciparum*. *Molecular and Cellular biology*, 13, 7349-7357
- Van der Hoorn, F.A., (1987) c-mos upstream sequence exhibits species-specific enhancer activity and binds murine-specific nuclear proteins. *Journal of Molecular Biology* 193, 255-266
- Van der Ploeg, L.H.T., Gianni, S.H., Cantor, C.R., (1985) Heat shock genes: Regulatory role for differentiation in parasitic protozoa. *Science*. 228, 1443-1446
- Van Dijk, M.A., Rodenburg, R.J.T., Holthuisen, P., Sussenbach, J.S., (1992) The liver-specific promoter of the human insulin-like growth factor II genes is activated by CCAAT/enhancer binding protein (C/EBP). *Nucleic Acid Research*. 20, 3099-3104.
- Vanhamme, L., Pays, E., (1995) Control of gene expression in trypanosomes. *Microbiological Reviews*. 59, 223-240
- Vickerman, K., (1985) Developmental cycles and biology of pathogenic trypanosomes. *British Medical Bulletin* 41, 105-114

Visser, A.E., Abraham, A., Bell Sakyi, L.J. Brown, C.G.D., Prestron, P.M., (1995) Nitric oxide inhibits establishment of macroschizont-infected cell lines and is produced by macrophages of calves undergoing bovine tropical theileriosis or East Coast Fever. *Parasite Immunology*, 17, 91-102

Walker, A.R., McKellar, S.B.W., (1983) The maturation of *Theileria annulata* in *Hyalomma anatolicum anatolicum* stimulated by incubation or feeding to produce sporozoites. *Veterinary Parasitology*. 13, 13-21.

Wall, L., deBoer, E., grosveld, F., (1988) The human beta-globin gene 3' enhancer contains multiple binding sites for an erythroid-specific protein. *Genes and Development* 2., 1089-1100.

Waters, A.P., Syin, C., McCutchan, T.F. (1989) Developmental regulation of stage specific ribosome populations in *Plasmodium*. *Nature* 342, 438-440.

Weber, J.L., 1988. Molecular biology of malaria parasites. *Experimental Parasitology*. 66, 143-170

Weiher, H., Konig, M., Gruss, P., (1983) Multiple point mutations affecting the simina virus 40 enhancer. *Science*. 219, 626-631.

Weissman, J.D., Singer, D.S., (1991) A complex regulatory DNA element associated with a major histocompatibility complex I gene consists of both a silencer and an enhancer. *Molecular Cellular Biology* 11, 4217-4227.

Wenshun, L., Hong, Y., (1994) Bovine and ovine theileriosis in China and its immune prophylaxis. in: European Union Third Co-ordination Meeting on Tropical Theileriosis. Spooner, r., and Campbell, J. (eds) the Roslin Institute (Edinburgh) Roslin, Midloatioan, Scotland, pp 13-17.

Wharton, R.H. (1976) Tick-borne livestock diseases and their vectors. 5. Acaricide resistance and alternative methods of tick control. *World Animal Review* 20, 8-15.

Wieland, S., Schatt, M.D., Rusconi, S., (1990) Role of TATA element in transcription from glucocorticoid receptor-responsive model promoters. *Nucleic Acids Research* 18, 5113-5118.

Wilkie, G.M., Melrose, T.M., Fletcher, J.D., Brown, C.G.D. (1986) GPI Isoenzyme identification of cloned populations of *Theileria annulata* (Ankara) after passages through the vertebrate and invertebrate host. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 80, 987-988

Williams, J.F., (1989) Optimisation strategies for the polymerase chain reaction. *Biotechniques* 7, 762-769

Williamson, S., Tait, A., Brown, D., Walker, A., Beck, P., Shiels, B., Fletcher, J. Hall, R., (1989) *Theileria annulata* sporozoite surface antigen expressed in *Escherichia coli* elicits neutralising antibodies. *Proceedings of the National Academy of Sciences, U.S.A.* 86, 4639-4643

Wu, Y., Sifri, C.D., Lei, H-H., Su, X-S., Wellems, T.E., (1995) Transfection of *Plasmodium falciparum* within human red blood cells. *Proceedings of the National Academy of Sciences, USA*, 92, 973-977

- Yamada, H., Muramatsu, S., Mizuno, T., (1990) An *Escherichia coli* protein that preferentially binds to sharply curved DNA. *Journal of Biochemistry*, 108, 420-425.
- Yang-yen, H.F., Chambard, J.C., Sun, Y.L., Smeal, T., Schmidt, T.J., Drouin, J., Karin, M., (1990) Transcriptional interference between c-Jun and the glucocorticoid receptor: mutual inhibition of DNA binding due to direct protein-protein interaction. *Cell* 62, 1205-1215.
- Zilberstein, D., Blumenfeld, N., Liveanu, V., Gepstein, A., Jaffe, C.L., (1991). Growth at acidic pH induces an amastigote stage specific protein in *Leishmania* promastigotes. *Molecular and Biochemical Parasitology*, 45, 174-178.
- Zinkernagel, R.M. (1979) Cellular immune responses to intracellular parasites: role of the major histocompatibility gene complex and thymus in determining immune responsiveness and susceptibility to disease. *Parasite Immunology* 1, 91
- Zomerdijs, J.C.B.M., Ouellette, M., Ten Asbroek, A.L.M.A. Kieft, R., Bommer, M.M.A., Clayton, C.E., Borst, P., (1990) The promoter for a variant surface glycoprotein gene expression site in *Trypanosoma brucei*. *EMBO Journal* 9, 2791-2801.
- Zon, L.I., Gurish, M.F., Stevens, R.L., Mather, C., Reynolds, D.S., Austen, K.F., Orkin, S.H., (1991) GATA-binding transcription factors in mast cells regulate the promoter of the mast cell carboxypeptidase A gene. *Journal of Biological Chemistry* 266, 22948-53

Appendix

Appendix I. The nucleotide sequence of the *TamS1* 8.0kb contig, showing restriction enzymes sites for *Ava* I, *Dde* I, *Dra* I, *EcoRV*, *Hha* I, *Hinc* II, *Hind* III, *Hpa* I, *Kpn* I, *Nde* I, *Nhe* I, *Nsi* I, *Sal* I, and *Xho* I. The numbers represent the sequence position with reference to the most 5' base. The *TamS1* open reading frames is located between 3079-3922, ORF-1 between 5816-4906, ORF-2 between 8068-7692 and ORF-3 1-188. Many of the oligonucleotides used for PCR and mobility shift reactions are highlighted in bold with an arrowhead showing the orientation of each oligonucleotide 5' to 3'.

HindIII | DdeI

1 AAGCTTGCGAACAGTCTCCAAGGGCTTCTCTGAGACGAAGTTCGCTCAATCAATGTTGTC 60
 -----+-----+-----+-----+-----+
 TTCGAACGCTTGTCTCAGAGGTTCCCGAAGAGACTCTGCTTCAAGCGAGTTAGTTACAACAG

61 CAGGTCTGAAGTTTTACGACTGGATGGTTAAGATTTACCGGGATCAGAAGGGTAGTAATC 120
 -----+-----+-----+-----+-----+
 GTCCAGACTTCAAAATGCTGACCTACCAATTCTAAATGGCCCTAGTCTTCCCATCATTAG

121 CTCAATTATCTACCTTCGATTATGATAACTACGTCTACACCTATAACCCTATCGCAAAGA 180
 -----+-----+-----+-----+-----+
 GAGTTAATAGATGGAAGCTAATACTATTGATGCAGATGTGGATATTGGGATAGCGTTTCT

DdeI |

181 AGAGCTAGTTCAGCTTATTCTAAGAATCCGTTCTCCTATTTCGTATAAATCTTTTTTCTT 240
 -----+-----+-----+-----+-----+
 TCTCGATCAAGTCGAATAAGATTCTTAGGCAAGAGGATAAGCATATTTAGAAAAAAGAA

DdeI |

241 TTTACTTTTATAGGTTTTCTTCTCATATTTGTTCCCCAGGTTCTAAGAAGGACGTTATT 300
 -----+-----+-----+-----+-----+
 AAATGAAAATATCCAAAAGGAAGAGTATAAACAAGGGGTCCAAGATTCTTCTGCAATAA

DraI |

301 TGTGATGATTCTTTTTCTTCTGGGACACTGAAATGCAGTTTAAATATTTGAGAACAAAT 360
 -----+-----+-----+-----+-----+
 ACACTACTAAAGAAAAAGAAGACCCTGTGACTTTACGTCAAATTTATAAACTCTTGTTTA

DraI |

361 TGGCTTTAAAATTAATGAAGTATTGTTATTTATTGAGTTTAAGATTTTGTTATTTTAAAT 420
 -----+-----+-----+-----+-----+
 ACCGAAATTTTAATTACTTCATAACAATAAATAACTCAAATTCTAAACAATAAAAATTA

DraI |

421 TATTTGGCAAATTTGGTTTAATTGAGTCATTTTAAATTTAATTATTCACCTACACATTTA 480
 -----+-----+-----+-----+-----+
 ATAAACCGTTTAAACCAAATTAACCTCAGTAAAATTTAAATTAATAAGTGAATGTGTAAAT

DraI |

481 TTATTCTCACCTCTATAAATAAATTACTTCTGGTTTTTTAAAACCCCATATTTATTTGAA 540
 -----+-----+-----+-----+-----+
 AATAAGAGTGAGATATTTATTTAATGAAGACCAAAAAATTTGGGGTATAAATAAACTT

DraI |

541 ATTTTAAAAATGTTTATACGCAAATTATATTATCTAATTTAGTTTCACGAAAAACAAGAAA 600
 -----+-----+-----+-----+-----+
 TAAAATTTTACAAATATGCGTTTAATATAATAGATTAAAGTCAAAGTGCTTTTGTTCTTT

DraI |

601 TAATAAGCCCTACACATGAATTTATAATGTAATAATCAACACATTAAATTTTAAAGTTA 660
 -----+-----+-----+-----+-----+
 ATTATTCGGGATGTGTACTTAAATATTACATTATTAGTTGTGTAATTTAAATTTTCAAT

661 GTTAAATTTACAAATCACTGTTACAACCTTTATCTTCATTTTTTGGTTTACTAAATATGAT 720
 -----+-----+-----+-----+-----+
 CAATTTAAATGTTTAGTGACAATGTTGAAATAGAAGTAAAAAACCAATGATTTATACTA

DdeI
|
GTGTAGGAATCCACCCAACCTTAGGTCCAACCTGTGTTATGAAATTAGAATCCTAAAACCTT
721-----+-----+-----+-----+-----+-----+ 780
CACATCCTTAGGTGGGTGAATCCAGGTTGACACAATACTTTAATCTTAGGATTTTGAAA

ACCTTTTGAAAGTGACTCACATCCAACCTTCTTCCCATTTATATAGTACAATGATATTAT
781-----+-----+-----+-----+-----+-----+ 840
TGGAAAACCTTTCCTGAGTGTAGGTTGGAAGAAGGGTAAATATATCATGTTACTATAATA

DdeI
|
TACAAGTCCCACCTCCCTCAGCAACACCCTAATGAATATCTTATTTATATATTTATTCAT
841-----+-----+-----+-----+-----+-----+ 900
ATGTTTCAGGGTGGAGGGAGTCGTTGTGGGATTACTTATAGAATAAATATATAAATAAGTA

ACCATGCTTCCTTTAACAGATTAAATGAATACTCGTTTGATATTATACCATTATTTAATT
901-----+-----+-----+-----+-----+-----+ 960
TGGTACGAAGGAAATTGTCTAATTTACTTATGAGCAAACCTATAATATGGTAATAAATTAA

TACCTTTTGAAAGTGACTCACATCCAACCTTCTTCCCATTTATATAGTACAATGATATTA
961-----+-----+-----+-----+-----+-----+ 1020
ATGGAAAACCTTTCCTGAGTGTAGGTTGGAAGAAGGGTAAATATATCATGTTACTATAAT

DdeI
|
TTACAAGTCCCACCTCCCTCAGCAACACCCTAATGAATATCTTATTTATATATTTATTCA
1021-----+-----+-----+-----+-----+-----+ 1080
AATGTTTCAGGGTGGAGGGAGTCGTTGTGGGATTACTTATAGAATAAATATATAAATAAGT

TACCATGCTTCCTTTAACAGATTAAATGAATACTCGTTTGATATTATACCATTATTTAAT
1081-----+-----+-----+-----+-----+-----+ 1140
ATGGTACGAAGGAAATTGTCTAATTTACTTATGAGCAAACCTATAATATGGTAATAAATTA

HincII
HpaI
NdeI
|
TTCAACAACCTCATATGTTTCTGTGTTGTTAACTGCGTTTTGTTGATCGGTATCGTTTGAA
1141-----+-----+-----+-----+-----+-----+ 1200
AAGTTGTTGAGTATACAAAGACACAACAATTGACGCAAAACAACCTAGCCATAGCAAACCT

TTAGAGAAATTACAGGAGCCAACCTTGAGACTTTCGAGGTTTATGTGATCTATGAAAGAT
1201-----+-----+-----+-----+-----+-----+ 1260
AATCTCTTTAATGTCCTCGGTTGGAACCTCTGAAAGCTCCAAATACACTAGATACTTTCTA

DraI
|
ATAAGGTTGTGTATTGGTGTTTTCTTTCTGAAAATTTCGTTTTAAATTCTTCCTATTCAGT
1261-----+-----+-----+-----+-----+-----+ 1320
TATTCCAACACATAACCACAAAAGAAAAGACTTTTAAGCAAATTTAAGAAGGATAAGTCA

TCGTGGATCATGTCCTTAATCACTTGCTTTTAAACGGTTTCTTTTTCAATGTGCTA
1321-----+-----+-----+-----+-----+-----+ 1380
AGCACCTAGTACAGGAATTAGTGAACGAACGAAAAATTGCCAAAGAAAAAGTTACACGAT

AATACCCTTTACCAAAATCTCTCTGACTTTCCATTTTCTCATATCTATGTTCCAAATC
1381-----+-----+-----+-----+-----+-----+ 1440
TTATGGGAAATGGTTTTAGAGAGAGACTGAAAGGTAAAAGAGTATAGATACAAGGTTTAG

ATCGTTATCTAGTATCAGTTTTAAGACGTTTTCCACAACCTCCGTTTAATTCATGGCTCT
1441-----+-----+-----+-----+-----+-----+ 1500
TAGCAATAGATCATAGTCAAAATTCTGCAAAGGTGTTGAAGGCAAATTAAGTACCGAGA

GTTCTTCAACTAACTTCTCAATGTTTTCACTAAAGAACTTTAGAGAGTATTCCTTGTGGT
1501-----+-----+-----+-----+-----+-----+ 1560
CAAGAAGTTGATTGAAGAGTTACAAAAGTGATTTCTTGAAATCTCTCATAAGGAACACCA

DdeI
|
TCAGGGGGACCTAAGATGAAACTTTTGGGATCTCAAAATTACCCTAAAATGCTCAACATT
1561-----+-----+-----+-----+-----+-----+ 1620
AGTCCCCCTGGATTCTACTTTGAAAACCCTAGAGTTTAAATGGGATTTTACGAGTTGTAA

DdeI
|
AGTACGACAGGGAGAAATTGTAACCTCTTGACCCGCCTTTAGAGATGTTCTAAGCTCGTT
1621-----+-----+-----+-----+-----+-----+ 1680
TCATGCTGTCCCTCTTTAACATTGGGAGAACTGGGCGGAAATCTCTACAAGATTTCGAGCAA

DraI EcoRV
| |
CAAATCCATAGAGTTTAAATTTCCGATAAAATGATCACAGTTGATGGCGATATCTTTGA
1681-----+-----+-----+-----+-----+-----+ 1740
GTTTAGGTATCTCAAATTTTAAAGGCTATTTTACTAGTGTCAACTACCGCTATAGAACT

TAAACTAAATAGTAAACGTTTCGTAAAGCCTCGAATGAGGTTTCAAATACGATGTGTCCAG
1741-----+-----+-----+-----+-----+-----+ 1800
ATTTGATTTATCATTTGCAAGCATTTTCGGAGCTTACTCCAAAGTTTATGCTACACAGGTC

TAAATTTATTTCCAAAGTCACAGACTCATCTAAAGAAGTTATTACCACTAAATCAAACCT
1801-----+-----+-----+-----+-----+-----+ 1860
ATTTAAATAAAGGTTTCAGTGTCTGAGTAGATTTCTTCAATAATGGTGATTTAGTTTGGA

TCGTTGCGATGTGTTATATTACAGTTTGACAATCCTATAGTTCTGAAAACCTTCCAGATCA
1861-----+-----+-----+-----+-----+-----+ 1920
AGCAACGCTACACAATATAATGTCAAACCTGTTAGGATATCAAGACTTTTGAAGGTCTAGT

HincII
HpaI
|
AGCTATAAATTATTATAAAATATAACAAATGAAATTATTAAACATACATTGATTGTTAAC
1921-----+-----+-----+-----+-----+-----+ 1980
TCGATATTTAATAATATTTTATATTGTTTACTTTAATAATTTGTATGTAACATAACAATTG

DraI
|
ACCCCTTTGGAGCTTATTATGAAATTGATTTTCGATTTATAATTATAATATATATTTAAA
1981-----+-----+-----+-----+-----+-----+ 2040
TGGGGAAACCTCGAATAATACTTTAACTAAAAGCTAAATATTAATATTATATATAAATTT

ACGGAGAGTCTTTTGGGAACCAAATAATTTACAATAAAGTCCAAATTAAGCAAAGACGTC
2041-----+-----+-----+-----+-----+-----+ 2100
TGCTCTCAGAAAACCTTTGGTTTATTAAATGTTATTTTCAGGTTTAATTCGTTTCTGCAG

DraI
|
AGAGTAATGTTTCGAAACAACAAAGCATGTGTCCTGTTTAACTCAAATTAGTCATTTTAAA
2101-----+-----+-----+-----+-----+-----+ 2160
TCTCATTACAAGCTTTGTTGTTTCGTACACAGGACAAATTGAGTTTAAATCAGTAAAATTT

DdeI
|
TAAATTAAGGGGTAAAATCAAATACACATCAAAGTATAAATATATAAAATACCTTAGAA
2161-----+-----+-----+-----+-----+-----+ 2220
ATTTAATTCCTTCATTTTAGTTTATGTGTAGTTTTCATATTTATATATTTTATGGAATCTT

TTAAAAAGATCTCTCTCTAAAAGAATCAATGCAAGGATTCTCTTGGGTAGGGATGAGAGGCA
2221-----+-----+-----+-----+-----+-----+ 2280
AATTTTTCTAGAGAGATTTTCTTAGTTACGTTCCCTAAGAGAACCCATCCCTACTCTCCGT

CCGTTGCCTATACGACAGATATAGAGTGTTAGGATCGACAAGCACAGTAGTTTAATTACA
2281-----+-----+-----+-----+-----+-----+ 2340
GGCAACGGATATGCTGTCTATATCTCACAACTCTAGCTGTTTCGTGTCATCAAATTAATGT

TATTACAATCTGGATAGATGTATAGAAATAAAAAGAGCATACACACCTCTATCTGTAACCT
2341-----+-----+-----+-----+-----+-----+ 2400
ATAATGTTAGACCTATCTACATATCTTTATTTTCTCGTATGTGTGGAGATAGACATTGA

1945 →
CTCTGGAGAATAAATAGGGTGTACTATTTGCTGAAGAAAATTTTCAAATCCGGGCCGGT
2401-----+-----+-----+-----+-----+-----+ 2460
GAGACCTCTTATTTATCCACATGATAAACGACTTCTTTTAAAGTTTTAGGCCCGGCCA

AATATCTGAAGGAGACTGATTAAACTCAAAGACATCACACTCTGATGAAAGTGAAGGCGG
2461-----+-----+-----+-----+-----+-----+ 2520
TTATAGACTTCCTCTGACTAATTTGAGTTTCTGTAGTGTGAGACTACTTTCACCTCCGCC

1963 →
GTAAGGACAGCTCATAACGACAAATTCAGTATAAAAACATAATGAATGATTGTTTTAAGG
2521-----+-----+-----+-----+-----+-----+ 2580
CATTCTGTGCGAGTATTGCTGTTTAAGTCATATTTTTGTATTACTTACTAACAAAATTCC

← 1944 NsiI
|
AAATGATGTGTGGAATCACGGTTATTTGTGCGGGAAATTATGCATGGATATTATCTATTT
2581-----+-----+-----+-----+-----+-----+ 2640
TTTACTACACACCTTAGTGCCAATAAACACGCCCTTTAATACGTACCTATAATAGATAAA

1947 →
AACTTATAATAATGTTATATTAAAATATGACAAATTAATGTGTGAAATAAATAAAGTAAA
2641-----+-----+-----+-----+-----+-----+ 2700
TTGAATATTATTACAATATAATTTTATACTGTTTAATTACACACTTTATTTATTTCATTT

GTGGAGGGAAATTCTGTCCTAATATTTAACTGGGTCTAGTGACTTAAAAAAGAAAGCAGG
2701-----+-----+-----+-----+-----+-----+ 2760
CACCTCCCTTTAAGACAGGATTATAAATTGACCCAGATCACTGAATTTTTTCTTTCGTCC

← 1963
TTATTGGTTTGAAAAAATAATAAATTCATCTTTTTTATAGTCATCAAAAAATGAAAAATGG
2761-----+-----+-----+-----+-----+-----+ 2820
AATAACCAAACCTTTTTTATTATTTAAGTAGAAAAATATCAGTAGTTTTTTACTTTTTTACC

TA3051 →
ATAAATATACATGTTATTTTCATTAAGATGGACTTCAATGGAGGATAAGGCATTAGGCGTG
2821-----+-----+-----+-----+-----+-----+ 2880
TATTTATATGTACAATAAAGTAATTCTACCTGAAGTTACCTCCTATTCCGTAATCCGCAC

← 1946

GATA35a → CAT 1a → TAT1
ATGAGTGCATAGATACAGATAAACATGCACACAATTTGTAGGGCGACATTGTTTTGTATG
2881-----+-----+-----+-----+-----+-----+ 2940
TACTCAGTATCTATGTCTATTTGTACGTGTGTTAAACATCCCGCTGTAACAAAACATAC

← GATA35b ← CAT 1b

NsiI
 → |
 GTGAATGCATATCCATCTCACTTCACTTATAGTTATTGGTCTTTTTTCCATAACATCCA
 2941-----+-----+-----+-----+-----+ 3000
CACTTACGTATAGGTAGAGTGAAGTGAATATCAATAACCAGAAAAAAGGTATTGTAGGT

 ← TA3052 DraI
 |
 CCCAATTAGTTAATTTTTTAATATTTAAATCGCTCACTAGTCTGCCCTTTCTTATCTTTTT
 3001-----+-----+-----+-----+-----+ 3060
 GGGTTAATCAATTAAAAATTATAAATTTAGCGAGTGATCAGACGGGAAAGAATAGAAAAA

 ATAATATAATTATTTGAGATGTTGTCCAGGACCACCCTCAAGTTCTTATATTTGAGCTTC
 3061-----+-----+-----+-----+-----+ 3120
 TATTATATTAATAAACTCTACAACAGGTCCTGGTGGGAGTTCAAGAATATAAACTCGAAG

 TTCGTTATCTCATCCGTTAATGCTGCAAATGAGGATGAAAAGAAAAAGGAGGAAAAAAA
 3121-----+-----+-----+-----+-----+ 3180
 AAGCAATAGAGTAGGCAATTACGACGTTTACTCCTACTTTTCTTTTTCTCCTTTTTTTT

 HincII
 DraI SalI |
 | | |
 GATGTTGTTCTTGATGTTACTCTCACTTCATGTGAGAATGTAACCTTTAAAAACGTCGAC
 3181-----+-----+-----+-----+-----+ 3240
 CTACAACAAGAACTACAATGAGAGTGAAGTACACTCTTACATTGGAAATTTTGCAGCTG

 HincII
 DdeI HpaI
 | |
 TCTAACACCACTGAGTTAACTGTGCGGGATGGCTACCGTTTCAAGACCCTTAAGGTCGGA
 3241-----+-----+-----+-----+-----+ 3300
 AGATTGTGGTGACTCAATT**GACAGCGCCTACCGATGGCAAAGTTCTGGGAATTCCAGCCT**
 ← GSP 2 ← Tam1DB

 HincII
 |
 GACAAGACCTTGTTCAATGTTGACACCTCAAACATACCCCAGTACAGGCATTCAAACCTT
 3301-----+-----+-----+-----+-----+ 3360
CTGTTCTGGAACAAGTTACAACTGTGGAGTTTTGTATGGGGTCATGTCCGTAAGTTTGAA

 AAGCATGAATCCGATGAGTGGTTCAGACTTAATCTTCACCCTGCCAGCCAAAGATGTTC
 3361-----+-----+-----+-----+-----+ 3420
 TTCGTACTTAGGCTACTACCAAGTCTGAATTAGAAGTGGGACGGGTCGGTTTCTACAAG

 DdeI
 |
 AAGAAGAAGGGAGACAAGGAATATTCTGAGGTCAAATTCGAGACCTACTACGATGATGTC
 3421-----+-----+-----+-----+-----+ 3480
 TTCTTCTTCCCTCTGTTCCCTTATAAGACTCCAGTTTAAGCTCTGGATGATGCTACTACAG

 TTGTTCAAGGGAAAATCCGCCAAGGAAGTCTTCCAAGTTCTGAAGATACATCTTTG
 3481-----+-----+-----+-----+-----+ 3540
 AACAAAGTTCCTTTT**AGGCGGTTTCCTTGATCTACGAAGGTTCAAGCTTCTATGTAGAAAC**
 ← GSP 1

 DraI DraI
 | |
 TTCACCTCCTCCGCCTTCGGCACTGGAAAGATGTACACCTTTAAAAAGGAATTTAAACCT
 3541-----+-----+-----+-----+-----+ 3600
 AAGTGGAGGAGGCGGAAGCCGTGACCTTTCTACATGTGGAATTTTCTTAAATTTGGA

 TCCAAAGTCACCTTCGACAAGAAAGAAGTCGAAAACCAAACAATGCCAAGTATCTTGAA
 3601-----+-----+-----+-----+-----+ 3660
 AGGTTTCAGTGAAGCTGTTCTTTCTTCAGCCTTTTGGTTTGTTACGGTTCATAGAACTT

361-----+-----+-----+-----+-----+-----+ 3720
GTTGTTGTTTTTGTGGTTCTGATTCCAAGAAGTTCGTCAAACCTCTACTACTTCTATACC
CAACAACAAAAACAACCAAGACTAAGGTTCTTCAAGCAGTTTGAGATGATGAAGATATGG
3721-----+-----+-----+-----+-----+-----+ 3780
GGTGACTCAAGGTTGAAGGAGACCTACTTCGAGCTTAAGGACGATAAGTGGGTTCAAATG
CCACTGAGTTCCAACCTCCTCTGGATGAAGCTCGAATTCCTGCTATTACCCAAGTTTAC
HincII
|
3781-----+-----+-----+-----+-----+-----+ 3840
ACACAGGCAGATGCAAACAAGGCCTTGAATGCCATGAACTCATCCTGGTCAACCGATTAC
TGTGTCCGTCTACGTTTGTTCGGAACCTACGGTACTTGAGTAGGACCAGTTGGCTAATG
HincII
SalI | DdeI
| | |
3841-----+-----+-----+-----+-----+-----+ 3900
AAACCAGTTGTGCAAGTTCTCCCCCTTGCAGTCTTCGCCTCAGTACTCATCGTCTTC
TTTGGTCAACAGCTGTTCAAGAGGGGGGAACGTCAGAAGCGGAGTCATGAGTAGCAGAAG
DraI DraI
| |
3901-----+-----+-----+-----+-----+-----+ 3960
TCATCAGTCCTTTACTTCCTTTAAACCCATGTTGTAACAACCTTATCAACTTTTAAAC
AGTAGTCAGGAAATGAAGGAAATTTTGGGTACAAGCATTGTTGAATAGTTGAAAATTTTG
AATTTTGATAATTTGTATACAATTGCAGAACTAAATAACTAGCTTAAGTCATTATATGC
3961-----+-----+-----+-----+-----+-----+ 4020
TTAAACTATTAAACATATGTTAACGTCTTTGATTTATTGATCGAATTCAGTAATATACG
DraI
|
4021-----+-----+-----+-----+-----+-----+ 4080
CACTTAATTTTATACTTTAAACTTATATGTTTAGATATAACTTCAACAGCTATAGGTCAT
GTGAATTAAAATATGAAATTTGAATATACAAATCTATATTGAAGTTGTCGATATCCAGTA
CAATATCCAAGTCACGGTAAATCGCGGCCCTCACGCTAACCAGCTAAAAGCAGAACAGC
4081-----+-----+-----+-----+-----+-----+ 4140
GTTATAGGTTCAAGTCCATTTTAGCGCCGGGAGTGCGATTGGTCGATTTTCGTCTTGTCG
AAACTCTGGTTATCGCAGCTTCCTCTTCACTTTTCGTATCTCTTTCCACAGTGTCGAATG
4141-----+-----+-----+-----+-----+-----+ 4200
TTTGAGACCAATAGCGTCGAAGGAGAAGTGAAAGCAGTAGAGAAAGGTGTCACACGTTAC
AACCTGTAAAAATTAAGTGAATTGTAAAAATGAAATTATTAATCTTACTGAAAAGACGGT
4201-----+-----+-----+-----+-----+-----+ 4260
TTGGACATTTTAAATCACTTAACATTTTACTTTAATAATTAGAATGACTTTTCTGCCA
DdeI
|
4261-----+-----+-----+-----+-----+-----+ 4320
CGAACTCAGAATAGGCATCGTAATCGTCCTCAACAACCCCTAAAGCCTCGTTAATAGGAC
GCTTGAGTCTTATCCGTAGCATTAGCAGGAGTTGTTGGGGATTTCGGAGCAATTATCCTG
DdeI DdeI
| |
4321-----+-----+-----+-----+-----+-----+ 4380
GAGAATAGAGCCGTAGTCCCTAAGCAAATTCTGGTACTGAATGAACTAGTCCTCAGCCA
CTCTTATCTCGGCATCAGGGATTGTTTTAAGACCATGACTTACTTTGATCAGGAGTCGGT

HhaI
|

4381-----+-----+-----+-----+-----+-----+ 4440
GCACCTGCGAGTGAACACGAAATGGTAGTCCTTGCGCACATCGTCCTGATTCACAGAGTA
CGTGGACGCTCACTTGTGCTTTACCATCAGGAACGCGTGTAGCAGGACTAAGTGTCTCAT

HhaI
|

4441-----+-----+-----+-----+-----+-----+ 4500
TATCTCCATCATGAGGTGGTTTCCATTCTTCTCACGCAAAAAATCTCGACGACACCCTC
ATAGAGGTAGTACTCCACCAAAGGTAAGAAGAGTGC GTTTTTTTTAGAGCTGCTGTGGGAG

HhaI
|

4501-----+-----+-----+-----+-----+-----+ 4560
ACTGCCGGCGCCAGTCCAAACAACGATGCAGTTGTCCAAGATTCTGTTAATTTTGTCCCT
TGACGGCCGCGGT CAGGTTTGTGCTACGTCAACAGGTTCTAAGACAATTAAACAGGGA

HhaI
|

4561-----+-----+-----+-----+-----+-----+ 4620
GCCAATCGAAGTGATAATGGTAGTGTCCAAAGTGTACCCTAGGTGTTTCCTGACCCTTAT
CGGTTAGCTTCACTATTACCATCACAGGTTTCACATGGGATCCACAAAGGACTGGGAATA

HhaI
|

4621-----+-----+-----+-----+-----+-----+ 4680
GTTTGAATTAAACTCCAGGTTGTGGAGATTAATCTCAAAGTAGCAGAGAGTCCAGGACGA
CAAACCTTAATTTGAGGTCCAACACCTCTAATTAGAGTTTCATCGTCTCTCAGGTCTGCT

DdeI
|

4681-----+-----+-----+-----+-----+-----+ 4740
ACTTAGATAAAAAATTACAAGTAACAAATATTTCAATTTGAAAAGACTAACTGCATAAATA
TGAATCTATTTTTTAATGTTTCATTGTTTATAAAGTAACTTTTCTGATTGACGTATTTAT

NheI
|

4741-----+-----+-----+-----+-----+-----+ 4800
TTTTTACTAGCTAAATTATTTACTGAAATGTACAAAATGGAAATAAAATTAACAAGCT
AAAAATGATCGATTTAATAAATGACTTTACATGTTTTACCTTTATTTTAATTTTGTTTGA

NheI
|

4801-----+-----+-----+-----+-----+-----+ 4860
AGCTGGAGAACATGGTCTGACAAAAAGCTAGGCAGATAATAAGTTACAAAAACAACCTT
TCGACCTCTTGTACCAGACTGTTTTTCGATCCGTCTATTATTCAATGTTTTTTGTTGAAA

DraI DdeI DdeI
| | | |

4861-----+-----+-----+-----+-----+-----+ 4920
ACAAAATTTTTTAAACAACCTAGTTTCCTGAGGCTCAGACAATTTAATTTAATGTGGAAA
TGTTTTTAAAAAATTTTGTGATCAAAGGACTCCGAGTCTGTAAATTAATTAACACCTT

DdeI
|

4921-----+-----+-----+-----+-----+-----+ 4980
TAATATCGAAGAGATTAATCAACTTTAGTGGTATTCGGCTCAGTGTCTGCTGATGTTTA
ATTATAGCTTCTCTAATTAGTTGAAATCACCATAAGCCGAGTCACAAAACGACTACAAAT

DdeI
|

4981-----+-----+-----+-----+-----+-----+ 5040
TGCTTAGGCTTACTTTTGAAATATTTCCGTTTCGTGTGTATTTTCTCTGAATTGACTGCG
ACGAATCCGAATGAAAACCTTTATAAAGGCAAGCACACATAAAAAGAGACTTAACTGACGC

DdeI
|

5041-----+-----+-----+-----+-----+-----+ 5100
TTATTTTGTGGTTTTTCGTTTGCGGAGTTCCTCCAGTTCATCCTCGTCCCTGTGGTGG
AATAAACACCAAAAAGCAAACGCCTCAAGGAGGTCAAGAGTAGGAGCAGGGACACCACC

AvaI
|

5101-----+-----+-----+-----+-----+-----+ 5160
GACTTGGGCCTGTGACGGGATCTACGAGACCTACCGGAGCCTGTTCCCGAGGTGGGAAGT
CTGAACCCGGACACTGCCCTAGATGCTCTGGATGGCCTCGGACAAGGGCTCCACCCTTCA

HhaI
|

5161-----+-----+-----+-----+-----+-----+ 5220
TTATCATTTTTTAGGGTCAGGAGGCTGTTTCAGAAACACTAGAAGTAGCGGGAGGCGCTGTA
AATAGTAAAAATCCCAGTCCTCCGACAAGTCTTTGTGATCTTCATCGCCCTCCGCGACAT

5221-----+-----+-----+-----+-----+-----+ 5280
GAGTCTGGAGGTTGTGAAGTTGGAACGTCATTGGGGTTTTAGTTGGATCAGAATTAGTGA
CTCAGACCTCCAACACTTCAACCTTGCAGTAACCCCAAATCAACCTAGTCTTAATCACT

5281-----+-----+-----+-----+-----+-----+ 5340
ATAACCTGTCCTTTTTATCCTTTTGAGATTTAGCATTAATGTTTGGTTGATTAGTATTTA
TATTGGACAGGAAAAATAGGAAAACCTCTAAATCGTAATTACAAACCAACTAATCATAAAT

5341-----+-----+-----+-----+-----+-----+ 5400
CAGTTTGAGCGTTAGAAGATGTACTAGTGGGTTTTTTCGCTCTTGGATAACGATCCTTTT
GTCAAACCTCGCAATCTTCTACATGATCACCCAAAAAACGCAGAACCTATTGCTAGGAAAA

5401-----+-----+-----+-----+-----+-----+ 5460
TACGCTTCTCTGCTAAAGGTGATAAATCACTTGTAAGAGTTTCATGAATTGTGGGTTCCTA
ATGCGAAGAGACGATTTCCACTATTTAGTGAACATTCTCAAAGTACTTAACACCCAAGGT

5461-----+-----+-----+-----+-----+-----+ 5520
GCTGGTCTGACAGTGTATTTTCCAGCTGATTTGATAATGTGTTTTCTAACTGATTTGTTA
CGACCAGACTGTCACATAAAAGGTGCTAACTATTACACAAAAGATTGACTAAACAAT

5521-----+-----+-----+-----+-----+-----+ 5580
GTCTACTTTCTAGCTGTCCGACGAGTCTTTCATCTAGTTTGTGGTTAATGACTCTTGCA
CAGATGAAAGATCGACAGGCTGCTCAGAAAGTAGATCAAACAACCAATTACTGAGAACGT

5581-----+-----+-----+-----+-----+-----+ 5640
ATAAAAGAGAGTTGTCTTTATAGTTGGTTAAATCCAGAAAAGGCTGCGACAGCTTATCAG
TATTTTCTCTCAACAGAAATATCAACCAATTTAGGTCTTTTCCGACGCTGTCTGAATAGTC

5641-----+-----+-----+-----+-----+-----+ 5700
ATACATACTGCTTAATTGTAAGTCTATCGTCTCTTGTTCTATCTGATAGAATTGAGTGGT
TATGTATGACGAATTAACATTGAGATAGCAGAGAACAAGATAGACTATCTTAACCTACCA

5701-----+-----+-----+-----+-----+-----+ 5760
TAAGAGTGTCAATTGCTGTTGTCAGTTCAGCTTCGTTATATCCATCACTTTTGTGTCCA
ATTCTCACAGTTAACGACAACAGTCAAGTCGAAGCAATATAGGTAGTGAAAACAACAGGT

DdeI
|

DraI
|

5761-----+-----+-----+-----+-----+-----+ 5820
TCCAATTAGACTCTGAGACGGTAGGCGACTCCACCATAACGTTTGGTTTAAACATTTGAG
AGGTTAATCTGAGACTCTGCCATCCGCTGAGGTGGTATTGCAAACCAAATTTGTAAACTC

5821-----+-----+-----+-----+-----+-----+ 5880
GTCTCGAATTACTCTTCAACCTTGTTGCATACTGCTGAAAACCTTGGTAGATCCCGTTAG
CAGAGCTTAATGAGAAGTTGGAACAACGTATGACGACTTTTGGAACCATCTAGGGCAATC

5881-----+-----+-----+-----+-----+-----+ 5940
GTTGGCTGGGTACGGCACTGAAGGTGGAGGTGGAGGTGGTGACTCTACACCTCTTCCGAT
CAACCGACCCATGCCGTGACTTCCACCTCCACCTCCACCACTGAGATGTGGAGAAGGCTA

5941-----+-----+-----+-----+-----+-----+ 6000
GAAGTTGGTAAACCCTAGTGGGTAGTTGCCCTTCATTTAGGGGCATCAGGAAGTCGGTAC
CTTCAACCATTTGGGATCACCCATCAACGGGAAGTAAAGTCCCGTAGTCCTTCAGCCATG

KpnI
|

6001-----+-----+-----+-----+-----+-----+ 6060
TCCGTACCCCAGGATATTATCAGTGTCATCGCTGTTGGAATTTCTGAAAGATAACGAGTT
AGCCATGGGGTCCTATAATAGTCAAGTAGCGACAACCTTAAAGACTTTCTATTGCTCAA

6061-----+-----+-----+-----+-----+-----+ 6120
GGAGTAGCTGCTGTCACGACTACTATGGAATCCTGGGCTGGCAACTGATTTAACGCTTGA
CCTCATCGACGACAGTGCTGATGATACCTTAGGACCCGACCGTTGACTAAATTGCGAACT

6121-----+-----+-----+-----+-----+-----+ 6180
AATTCTATGTTTGAACCTGTATCTGGTACTTTCGAGGTTCCCTGGAGAGGCATCTTGGTC
TTAAGATACAAACTTGAACATAGACCATGAAGCTCCAAGGAACCTCTCCGCTAGAACCAG

AvaI
XhoI
|

6181-----+-----+-----+-----+-----+-----+ 6240
GTTGGTGACGCTTGAGGTTTCCGAGATCCGATTATTTACACTAACATCTCGGTTACTCGA
CAACCACTGCGAACTCCAAAAGGCTCTAGGCTAATAAATGTGATTGTAGAGCCAATGAGCT

6241-----+-----+-----+-----+-----+-----+ 6300
GCCCAGTAAATCTGGGAATGAAACATCGTCCAGTGTTCCGCTACATATTTTTATTTAATT
CGGGTCATTTAGACCCCTACTTTGTAGCAGGTCACAAGGCATGTATAAAAAATAAATTAA

HhaI
|

6301-----+-----+-----+-----+-----+-----+ 6360
AGTTAGCTTACGTTTTGCTTTCGACGTTTGATTTCGTACTTTACTTCTGATTTCCAAGCGC
TCAATCGAATGCAAAACGAAAGCTGCAAACTAAGCATGAAATGAAGACTAAAGGTTTCGCG

HincII
HpaI
|

6361-----+-----+-----+-----+-----+-----+ 6420
TTCCCGCTGTCTTAATTTGGCTGGTTGTGTTAACTGCTATACCTCGTACGCTTACACTTA
AAGGGCGACAGAATTAAACCGACCAACACAATTGACGATATGGAGCATGCGAATGTGAAT

6421-----+-----+-----+-----+-----+-----+ 6480
TGCCTGAATAAAATATTAATTTTTATTTATATGTTTATTTGCCTACCGGTCTTTTTTAAG
ACGGAATTATTTTATAATTAAAAATAAATATACAAATAAACGGATGGCCAGAAAAAATTC

DdeI
|

6481-----+-----+-----+-----+-----+-----+ 6540
CACTTAGATAATAGCCTTTGTCCCAGTGATAGGGATTGAGGCTCATATTAGACAAATTT
GTGAATCTATTATCGGAAACAGGGTCACTATCCCTAAGCTCCGAGTATAATCTGTTTAAA

DdeI
|

6541-----+-----+-----+-----+-----+-----+ 6600
TCCTCAGAATTGGTTTCATTATTCCAATTTTTATCCATTTTTATAAATGAGTGGTTTGAT
AGGAGTCTTAACCAAAGTAATAAGGTTAAAAATAGGTAAAAATATTTACTCACCAAAC TA

6601-----+-----+-----+-----+-----+-----+ 6660
CAGGCAAACCTGGTGAAAATATACAAAACAACCAATATACCTTAAATAAAAAATCCAGGT
GTCCGTTTGAACCACTTTTATATGTTTTGTTGTTTATATGGGAATTTATTTTAGGGTCCA

DraI
|

6661-----+-----+-----+-----+-----+-----+ 6720
TAAGGCCAAAATTATGAATCGTATATGTTAAATAACGAAAAGTTGGAAGTTTTGGTTTAAA
ATTCCGTTTAAATACTTAGCATATACAATTTTATGCTTTTCAACCTTCAAAACCAAATTT

DdeI
|

6721 AATGCTATGAGTTGGATTTGGTAGGGAATTAGCCTTAGACCAGAAGATACCCAAACAACA 780
-----+-----+-----+-----+-----+-----+-----+
TTACGATACTCAACCTAACCATCCCTTAATCGGAATCTGGTCTTCTATGGGTTTGTGT

6781 AAGTTTTTGGATTTAAGGGTAACAGTTTACCAATTTAGGCAAAAGCCTAGTGGGAACGAT 840
-----+-----+-----+-----+-----+-----+-----+
TTCAAAACCTAAATTTCCATTGTCAAATGGTTAAATCCGTTTTTCGGATCACCTTGCTA

6841 AAGAAAATAAAATATAGATGTATAATTATAACAAGGGGCAAACATGAAGGTATCTGGGA 900
-----+-----+-----+-----+-----+-----+-----+
TTCTTTTATTTTATATCTACATATTAATATTTGTTCCCCGTTTGTACTTCCATAGACCT

6901 TGTTTTGGGATCTTTTACTCCTTCACGAGCCCAACAACCGACCCAAGGATTACTTAATAG 960
-----+-----+-----+-----+-----+-----+-----+
ACAAAACCCTAGAAAATGAGGAAGTGCTCGGGTTGTTGGCTGGGTTCTAATGAATTATC

DdeI DraI
| |

6961 CCAAGGGTAACGGGAGTACATATCAACTATATCACTCAGAGAAGTTTAAATAGTCTGGGT 7020
-----+-----+-----+-----+-----+-----+-----+
GGTTCCTTGCCTCATGTATAGTTGATATAGTGAGTCTCTTCAAATTTATCAGACCCA

7021 AAGTATATGATAAGAAATTAGTTCAGTTCATATTTAAATATTGTACACATCCAAGTGTC 7080
-----+-----+-----+-----+-----+-----+-----+
TTCATATACTATTCTTTAATCAAGTCAAGGTATAATTTTATAACATGTGTAGGTTACAG

7081 ACATCCGAGCCACCTCTCCGTACAGTTCCTGAAAATTCTTTCAAGTTACTTATTATACCT 7140
-----+-----+-----+-----+-----+-----+-----+
TGTAGGCTCGGTGGAGAGGCATGTCAAGGACTTTTAAGAAAGTTCAATGAATAATATGGA

7141 GTAATTTTCTGTTTCATTGATTTTTCTTCTGTAATTGGCTCAACTGGAATGATATTCTCT 7200
-----+-----+-----+-----+-----+-----+-----+
CATTAAGACAAGTAACTAAAAGGAAGACATTAACCGAGTTGACCTTACTATAAGAGA

7201 CCTCCGATGATCTTCTTTCCACCTGTTCCATTTCTTGAATGAGTTGCAGTCATTTGTTG 7260
-----+-----+-----+-----+-----+-----+-----+
GGAGGCTACTAGAAGAAAGGTGGACAAGGTAAAGGAACCTACTCAACGTCAGTAAACAAC

7261 ATTGTATAATGGAATCATACATTTCTCCTATTACAATTTTATTATTTCTTCTACTTACCA 7320
-----+-----+-----+-----+-----+-----+-----+
TAACATATTACCTTAGTATGTAAAGAGGATAATGTTAAATAATAAAGAAGATGAATGGT

HincII
|

7321 ATTTCTTTAAGTATTTTTCTTCTCTACTGCGATACCCTCCAGTTGACACATTATACTTA 7380
-----+-----+-----+-----+-----+-----+-----+
TAAAGAAATTCATAAAAAAGAAGAGATGACGCTATGGGAGGTCAACTGTGTAATATGAAT

DdeI
|

7381 TTTTCCAGGCTTTTTATTCTCTTAGATAGATTTTCCAGGTGTTTCTCGTATTCTTTACAT 7440
-----+-----+-----+-----+-----+-----+-----+
AAAAGGTCCGAAAATAAGAGAATCTATCTAAAAGGTCCACAAAGAGCATAAGAAATGTA

7441 AACATTTCTTCCGCTAACTTTTGTCTTGCAGTGTTAGCTTTTCACATGGCATATATCTC 7500
-----+-----+-----+-----+-----+-----+-----+
TTGTAAAGAAGGCGATTGAAAACAAGACGTCACAATCGAAAAGTGACCGTATATAGAG

7501 TTTTGGAGGTGAGAATACTTAAAGTTGCTCGTTTTCTATTAAATGTGTTGATTTTGGTA 7560
-----+-----+-----+-----+-----+-----+-----+
AAAACCTCCACTCTTATGAATTTCAACGAGCAAAGGATAATTTACACAATAAAACCAT

GAGAACGTACCATTTCTAGCAACAAATTTGAATCTTCTAGGCCTTTTGACAAGTTTTCCGC
75 61-----+-----+-----+-----+-----+-----+ 7620
CTCTTGCATGGTAAGATCGTTGTTTAACTTAGAAGATCCGGAAAAGTGTTCAAAAGGCG
TTCGGAATCGATTAGTTTCCTTGATTCTACAGATAATTTCAATTGGAGTTAGTTACCTC
76 21-----+-----+-----+-----+-----+-----+ 7680
AAGCCTTAGCTAATCAAAGGAACTAAGGATGTCTATTAAAGTTAACCTCAATCAATGGAG
DdeI
|
TATTAGTTTCAAATCAAATACTCCTTAGGTTTGTACTGGTCCGCAGTCTTCCGTGAGGA
76 81-----+-----+-----+-----+-----+-----+ 7740
ATAATCAAAGTTTAGTTTTATGAGGAATCCAAACATGACCAGGCGTCAGAAGGCACTCCT
ATAGGATGTTTCACACTGTCGTATGCCAGCAGGTTTACGAACTCTTCATTTATTAAGTGG
77 41-----+-----+-----+-----+-----+-----+ 7800
TATCCTACAAAGTGTGACAGCATACGGTCGTCCAATGCTTGAGAAGTAAATAATTGACC
TTATACTTGGACTGCGATTCATTAAATGTATTTTCTAGGTCATTTACAAAACTATTGAA
78 01-----+-----+-----+-----+-----+-----+ 7860
AATATGAACCTGACGCTAAGTAATTTACATAAAAGATCCAGTAAATGTTTTTGATAACTT
TTGAAGACTACTGGTCTTGGCAGCTTTCTCTGGATGACTTGTGTCTCCAGGAGTTCTCTT
78 61-----+-----+-----+-----+-----+-----+ 7920
AACTTCTGATGACCAGAACCGTCGAAAGAGACCTACTGAACACAGAGGTCCTCAAGAGAA
TCTTCTTGTTTTTTCTTTTCCAATTCCTGTTTACGTCTTTCTATCTCTTCCATATCCAAA
79 21-----+-----+-----+-----+-----+-----+ 7980
AGAAGAACAAAAAGAAAAGGTTAAGGACAAATGCAGAAAGATAGAGAAGGTATAGGTTT
TCTTGTTCAACTTCTTCCTCTTCCATGGTTTCAATATCTGGAATTGTAACCTTCAACATTC
79 81-----+-----+-----+-----+-----+-----+ 8040
AGAACAAGTTGAAGAAGGAGAAGGTACCAAAGTTATAGACCTTAACATTGAAGTTGTAAG
CCTTCTGGCTCTGGGAGGTTTGACAAGC
80 41-----+-----+-----+-----+-----+-----+ 8068
GGAAGACCGAGACCCTCCAAACTGTTTCG

Appendix II. The nucleotide sequence of the *TpmS1* ~5.5kb contig, showing restriction enzymes sites for Ava I, Dde I, Dra I, EcoRV, Hha I, Hinc II, Hind III, Hpa I, Kpn I, Nde I, Nhe I, Nsi I, Sal I, and Xho I. The numbers represent the sequence position with reference to the most 5' base. The *TpmS1* open reading frames is located between 2079-2974, and ORF-1 between 5005-4053,

HincII
 SalI |
 | |
 TCGAGGTCGACACTAGTGGATCCAAGAATTCAAAAAGCTTCGTCCCGTTTACAAAATACA
 1 -----+-----+-----+-----+-----+ 60
 AGCTCCAGCTGTGATCACCTAGGTTCTTAAGTTTTTTCGAAGCAGGGCAAATGTTTTATGT

 ATGACATTATTACGAGTCCTACTTCCTTCAGCAATACCCTGAAATAATTTATACACACTT
 61 -----+-----+-----+-----+-----+ 120
 TACTGTAATAATGCTCAGGATGAAGGAAGTCGTTATGGGACTTTATTAAATATGTGTGAA

 DraI
 |
 GTTTACATCCACGCCTCCTTTAAAAGATTAAATGTGATCTCATTGGATATTATACCATTA
 121 -----+-----+-----+-----+-----+ 180
 CAAATGTAGGTGCGGAGGAAATTTTCTAATTTACACTAGAGTAACCTATAATATGGTAAT

 TTTAATTTCAACAACCTCGTATGTTTCCGAGTTATCAACTGCATTTCCTCGGATCGTGCCCG
 181 -----+-----+-----+-----+-----+ 240
 AAATTAAGTTGTTGAGCATACAAAGGCTCAATAGTTGACGTAAAGGGCCTAGCACGGGC

 DraI
 |
 TTTAAATTAGATATATTACACGAGCCAACTTTAAGACTCTCGAGGTTATATGATCTATGA
 241 -----+-----+-----+-----+-----+ 300
 AAATTTAATCTATATAATGTGCTCGGTTGAAATTCTGAGAGCTCCAATATACTAGATACT

 AvaI
 XhoI
 |
 AAGATAAAAGGTTGTGTATTGGAGTTTTCTTCTCGGAAAATTCATTTAAATTCTTCTTAT
 301 -----+-----+-----+-----+-----+ 360
 TTCTATTTTCCAACACATAACCTCAAAAAGAGGCCTTTTAAGTAAATTTAAGAAGAATA

 TCAATTCGTGGATCATGTCTTGATCACTTGCTTATTTTTTATCGGTTTTTTCTTCAATG
 361 -----+-----+-----+-----+-----+ 420
 AGTTAAGCACCTAGTACAGGAACTAGTGAACGAATAAAAAATAGCCAAAAAAGAAGTTAC

 TGTTAAATATCTTTACCAAAATCTCTCTGACTTTCCATCTTCTCATATCTATGCTCCA
 421 -----+-----+-----+-----+-----+ 480
 ACAATTTATAGAAATGGTTTTAGAGAGAGACTGAAAGGTAGAAGAGTATAGATACGAGGT

 DdeI
 |
 AATCATCACTGTTTGTCTCAGTTTTGTGCCGCTTTTCTGATGTTCCATGTACTTCGAGGT
 481 -----+-----+-----+-----+-----+ 540
 TTAGTAGTGACAAACAGAGTCAAAACACGGCGAAAAGACTACAAGGTACATGAAGCTCCA

 TCTGTTCTGAACTAACTTCTTGATGTTTTCATTAAGAAATTAACCGAGTATTCTTTGG
 541 -----+-----+-----+-----+-----+ 600
 AGACAAGGACTTGATTGAAGAACTACAAAAGTAATTTCTTTAATTGGCTCATAAGAAACC

 DdeI
 |
 AATTCACAGGAACTAAGGAGTAATTTTTAGACTTAAATATTACCTTAAATGCTCAAAT
 601 -----+-----+-----+-----+-----+ 660
 TTAAGTGTCTTTGATTCCTCATTAATAATCTGAATTTATAATGGAATTTTACGAGTTTA

 DdeI
 |
 TTAGTATAACAAGGTGAAATAGTGACCTGATCACCTGGTTTCAGCCATGCTCTAAGCTCT
 661 -----+-----+-----+-----+-----+ 720
 AATCATATTGTTCCACTTTATCACTGGACTAGTGGACCAAAGTCGGTACGAGATTCGAGA

 GTCAAGTCTAGAGAGTTGAAGATCTCTGATGAAATGATCACAGTGACGGGTGTTATCTTT
 721 -----+-----+-----+-----+-----+ 780
 CAGTTCAGATCTCTCAACTTCTAGAGACTACTTTACTAGTGTCACTGCCACAATAGAAA

 GATAAATTAGTTAGTAACCGGTCGTAAAGCCTAGATTGAGGTTTCAAATATGATGTGTCG
 781 -----+-----+-----+-----+-----+ 840
 CTATTTAATCAATCATTGGCCAGCATTTTCGGATCTAACTCCAAAGTTTATACTACACAGC

```

                                DdeI
                                |
841  AGTAGATTTATTTCCAAAGTAACAGACTGGTCTGAGAATGTTATTATTACTTAATCAAAC
-----+-----+-----+-----+-----+-----+-----+
TCATCTAAATAAAGGTTTCATTGTCTGACCAGACTCTTACAATAATAATGAATTAGTTTG

CTTCATTGCGATGTACTATATTACAGTTTGATAATCCCAAAGTTCTAAAAACCTCCAAAT
901  -----+-----+-----+-----+-----+-----+-----+
GAAGTAACGCTACATGATATAATGTCAAACCTATTAGGGTTTCAAGATTTTGGAGGTTTA

                                DraI
                                |
                                HincII
                                |
961  CAAGCTGTAAATCATTATTGAATGCACTAAATGAGATATTTAAACATACATTGAGTGTTA
-----+-----+-----+-----+-----+-----+-----+
GTTTCGACATTTAGTAATAACTTACGTGATTTACTCTATAAAATTTGTATGTAACTCACAAT

                                DraI
                                |
1021 ACACCCCTTTGGAATTTATCACAAAATTGATCTTTGATTTATAATTATAATATATATTTA
-----+-----+-----+-----+-----+-----+-----+
TGTGGGGAAACCTTAAATAGTGTTTTAACTAGAACTAAATATTAATATTATATATAAAT

AAATTGACAGTCTTTTAGGAACCAAATAATTTACAATAAAGTCCAGATTGACCAAAGATT
1081 -----+-----+-----+-----+-----+-----+-----+
TTTAACTGTCAGAAAATCCTTGGTTTATTAAATGTTATTTTCAGGTCTAACTGGTTTTCTAA

CAAGAGTAATGTTTCGATGCAACAAAGCATGTGTCCTGTGTAAAGTAAATTAGTCGTTTTT
1141 -----+-----+-----+-----+-----+-----+-----+
GTTCTCATTACAAGCTACGTTGTTTCGTACACAGGACACATTTTCATTTAATCAGCAAAAA

AATAAATAACAACAAAAGTCGAATACACACTAAAAGTATATATGTAAAAATACCCTAGAA
1201 -----+-----+-----+-----+-----+-----+-----+
TTATTTATTGTTGTTTTTCAGCTTATGTGTGATTTTCATATATACATTTTTTATGGGATCTT

TTAAAAAGTTTCTCTCTGAAGGCATTAGCGAAAGGGGTCTCCTGGTAGGGATAGAGGACG
1261 -----+-----+-----+-----+-----+-----+-----+
AATTTTTTCAAAGAGAGACTTCCGTAATCGCTTTCCCCAGAGGACCATCCCCTATCTCCTGC

                                DdeI
                                |
1321 GTCGCCTAAAGTGTAATGTTAGGAATCTCAGCAGATTATTAAATCATTGAATTCAGAAT
-----+-----+-----+-----+-----+-----+-----+
CAGCGGATTTACATTTACAATCCTTAGAGTCGTCTAATAATTTAGTAACTTAAGTCTTA

AAAAAACATACGCTCTTGACAAAATTTGGAGAAAAAATAACAGGGAGTATTTGAGGCAAA
1381 -----+-----+-----+-----+-----+-----+-----+
TTTTTTGTATGCGAGAACTGTTTTAAACCTCTTTTTTATTGTCCCTCATAAACTCCGTTT

AAATTTTCAAATTTGGACCCAGAAAATCTAAATGAGACTGACTAAACTCAAAGACATCA
1441 -----+-----+-----+-----+-----+-----+-----+
TTTAAAAGTTTTAAACCTGGGTCTTTTAGATTTACTCTGACTGATTTGAGTTTCTGTAGT

CACTCTGATGGAAGTGACAATAGATGAGTACGAACCATGATTACAAATTACATATTAAAA
1501 -----+-----+-----+-----+-----+-----+-----+
GTGAGACTACCTTCACTGTTATCTACTCATGCTTGGTACTAATGTTTAATGTATAATTTT

CACGTTAGATGAGTGTTTTACGGAAAAAGATGTGTGGGACCGTTTTAACTTGTGCGGAAAA
1561 -----+-----+-----+-----+-----+-----+-----+
GTGCAATCTACTCACAAAATGCCTTTTCTACACACCCTGGCAAAATTGAACACGCCTTTT

                                DdeI   DdeI
                                |       |
1621 TTGATTAAAGTTATTCTCTATCTAAGTCTGAGTATTGTTATATGAAAATATAATAAATTA
-----+-----+-----+-----+-----+-----+-----+
AACTAATTTCAATAAGAGATAGATTCAGACTCATAACAATATACTTTTATATTATTTAAT

```

```

      ATGTGTGGAATAAATGAAGAAAAGTAGGTAGGAATCGTGTCTTAATATTTAACTGGGTCT
1681-----+-----+-----+-----+-----+-----+-----+ 1740
      TACACACCTTATTTACTTCTTTTCATCCATCCTTAGCACAGGATTATAAATTGACCCAGA
                                     DdeI
                                     |
      AGTGAATAAAAAATAAAGCAGGTTATTTGTTTGAAAAAACAACGAATTTTCTTTTTTCTA
1741-----+-----+-----+-----+-----+-----+-----+ 1800
      TCACTTGATTTTTTATTTTCGTCCAATAAACAACTTTTTTGTGCTTAAAAGAAAAAGAT

      AGCGTCAAAAAATAAAAAATAGATAAATATATATGTTATGCAATTATGATGATGTGAAAT
1801-----+-----+-----+-----+-----+-----+-----+ 1860
      TCGCAGTTTTTTATTTTTTATCTATTTATATATACAATACGTTAATACTACTACACTTTA

      TGAGCGTTAGGGGTTAGGCGTGACGAGTGCAATGATACAGATGAACATGCACACAATTTG
1861-----+-----+-----+-----+-----+-----+-----+ 1920
      ACTCGCAATCCCCAATCCGCACTGCTCACGTTACTATGTCTACTTGTACGTGTGTTAAAC

      TTGGGCGAGAACGCTTTGTATGGTGAATCTTCCTCCATCTCACTTCAATTATAGTTATTG
1921-----+-----+-----+-----+-----+-----+-----+ 1980
      AACCCGCTCTTGCGAAACATACCACTTAGAAGGAGGTAGAGTGAAGTTAATATCAATAAC

      GTCTTTTTCCATAACATTTATCCGGTTATTTAATTTTTAATATTTGGATCGCTAACTAGT
1981-----+-----+-----+-----+-----+-----+-----+ 2040
      CAGAAAAAGGTATTGTAAATAGGCCAATAAATTAAAAATTATAAACCTAGCGATTGATCA

      CTGCCTTTTGTTTCGGTTTTTATAATATAATTATTTGAGATGTTGTCCAGAAATACCCTCA
2041-----+-----+-----+-----+-----+-----+-----+ 2100
      GACGGAAAACAAGCCAAAAATATTATATTAATAAACTCTACAACAGGTCTTTATGGGAGT

      AGTTCTTATATTTGAGTTTCTTCGTTATCTCTTGCCTTAATGCCGAAAAGAAGAAGAGA
2101-----+-----+-----+-----+-----+-----+-----+ 2160
      TCAAGAATATAAACTCAAAGAAGCAATAGAGAACGCAATTACGGCGTTTTCTTCTTCTCT

      AGAAGAAGGAGAAAAAGGAGGATCTTACAGTTGATGTTACCCTTTCGTATGGGAAAATG
2161-----+-----+-----+-----+-----+-----+-----+ 2220
      TCTTCTTCTCTTTTTTCTCCTAGAAATGTCAACTACAATGGGAAAGCAGTACCCTTTTAC
      AvaI
      |
      TTACTTCTACTCCCAGGCTGGCGGTACATTATTGAAAGCCAATGAAGGTTACCGTTTCA
2221-----+-----+-----+-----+-----+-----+-----+ 2280
      AATGAAGATGAGGGCTCCGACCGCCATGTAATAACTTTTCGGTTACTTCCAATGGCAAAGT
                                     HincII
                                     |
      AGACACTTAAGGTTCGGCGACAAGACTTTGTACAACGTTGACACCTCAAAATACGATGCAG
2281-----+-----+-----+-----+-----+-----+-----+ 2340
      TCTGTGAATTCCAGCCGCTGTTCTGAAACATGTTGCAACTGTGGAGTTTTATGCTACGTC

      TACACCTATACAACTTACCCATGATTCTGATGAATGGCTAAAGCTCCTTCTCCACCCAG
2341-----+-----+-----+-----+-----+-----+-----+ 2400
      ATGTGGATATGTTTGAATGGGTACTAAGACTACTTACCGATTTCGAGGAAGAGGTGGGTC

      CCAAGCCAGTGATGTTCAAGAAGAAGTCAGACAAGGAATATTCCGAAGTCAAATTCGAAA
2401-----+-----+-----+-----+-----+-----+-----+ 2460
      GGTTTCGGTCACTACAAGTTCTTCTTCAGTCTGTTCCCTTATAAGGCTTCAGTTTAAGCTTT

      CCTACTATGATGATGTCTTGTTCAAGGGGAAATCAGCCAAGGAACTCGATGCTTCCAAGG
2461-----+-----+-----+-----+-----+-----+-----+ 2520
      GGATGATACTACTACAGAACAAGTTCCCCCTTTAGTCGGTTCCTTGAGCTACGAAGGTTCC

      TCACTGATACTGGCTTGTTTTACCCAAGAGAGCTTCGGCACTGGGAAGAAGTACACCTTCA
2521-----+-----+-----+-----+-----+-----+-----+ 2580
      AGTGACTATGACCGAACAAATGGGTTCTCTCGAAGCCGTGACCCCTCTTCATGTGGAAGT

      ACAATAGCTTCAAACCTTCCAAGGTCTCATTCGACAAAAAAGATGTTGGAAAGCCCGACA
2581-----+-----+-----+-----+-----+-----+-----+ 2640
      TGTTATCGAAGTTTGAAGGTTCCAGAGTAAGCTGTTTTTTCTACAACCTTTCGGGCTGT

```

AGGCCAAGTTCCTCGACGTTTTTCGTCTATGTGGCTCTGATGACAAGAAGGTTGTTAGGC
2641-----+-----+-----+-----+-----+-----+ 2700
TCCGGTTCAAGGAGCTGCAAAAGCAGATACAGCCGAGACTACTGTTCTTCCAACAATCCG

TCGACTACTTCTTTGGTGGTGAAGGTTGAAGGAGGTCTACTTCGAGCTTAAAGACG
2701-----+-----+-----+-----+-----+-----+ 2760
AGCTGATGAAGAAACCACCACTGAGTTCCAACCTCCTCCAGATGAAGCTCGAATTTCTGC

ACAAGTGGGTCAAAATGGAACAGAATGACGCAAACAAGGCATTGCATGCCATGAGCGATT
2761-----+-----+-----+-----+-----+-----+ 2820
TGTTCACCCAGTTTTACCTTGTCTTACTGCGTTTGTTCGGTAACGTACGGTACTCGCTAA
HincII
SalI | DdeI
| |
CATGGAAATTGGACTACAAACCAGTCGTCGACAAGTTCTCTCCCCTTGCAGTCCTCGCCT
2821-----+-----+-----+-----+-----+-----+ 2880
GTACCTTTAACCTGATGTTTGGTCAGCAGCTGTTCAAGAGAGGGGAACGTCAGGAGCGGA
DraI
|
CAGTACTCATCGTCGCCGCTTCAGTCTTTTACAACCTTTAAAACCCATGTGCGTAACAAC
2881-----+-----+-----+-----+-----+-----+ 2940
GTCATGAGTAGCAGCGGCGAAGTCAGAAAATGTTGGAAATTTGGGTACACGCATTGTTG
DraI
|
TTATCAACGTTTAAAACAATATTGATAATTTGTATACAATTACAGCAGCTAACTAACTTA
2941-----+-----+-----+-----+-----+-----+ 3000
AATAGTTGCAAATTTTGTATACTATTAAACATATGTTAATGTCGTCGATTGATTGAAT

AGTCATTTTATTGACCACTTAATTTCATACCCTTCGACATATCTATCTAGTTATTACCAG
3001-----+-----+-----+-----+-----+-----+ 3060
TCAGTAAAATAACTGGTGAATTAAAGTATGGGAAGCTGTATAGATAGATCAATAATGGTC

TTTACAGTTTTGATTTCTACGTTTTTATATCTCTTGAGAATAGTTATAGTAAGCAGGTGA
3061-----+-----+-----+-----+-----+-----+ 3120
AAATGTCAAACTAAAGATGCAAAATATAGAGAACTCTTATCAATATCATTTCGTCCACT
HhaI
|
TAGCGCCGCAACAGATTTTAGTAACCGTAGATTTCCTCTTCTACCTCCTCCGGTACGTTGT
3121-----+-----+-----+-----+-----+-----+ 3180
ATCGCGGCGTTGTCTAAAATCATTGGCATCTAAGGAGAAGATGGAGGAGGCCATGCAACA
HincII
HpaI
|
TGGCACCTGTAGCATTTAATTTGACAATTGTACTAATATGTTAACTAGAATTTATATAAA
3181-----+-----+-----+-----+-----+-----+ 3240
ACCGTGGACATCGTAAATTAACTGTTAACATGATTATACAATTGATCTTAAATATATTT

GAAAATTACTGAAGAGAAGGTGCAATTTCGGAGTAGGAATTGAATTCCTCTTCAGCGACTT
3241-----+-----+-----+-----+-----+-----+ 3300
CTTTTAATGACTTCTCTTCAGCTTAAGCCTCATCCTTAACTTAAGGAGAAGTCGCTGAA
DdeI DdeI
| |
CCGCATAAAAATCTCAGACATAGGCTTGGAACGAAGCCTTAAGTCCCTAAGGAGATTACA
3301-----+-----+-----+-----+-----+-----+ 3360
GGCGTATTTTTAGAGTCTGTATCCGAACCTTGCTTCGGAATTCAGGGATTCTCTAATGT

GTACTGAAGGAAAGTCGTCTTGACCCAGAAGTTGGAAGTGAAGACAAAGTAATAATCTTT
3361-----+-----+-----+-----+-----+-----+ 3420
CATGACTTCCTTTCAGCAGAACTGGGTCTTCAACCTTCACTTCTGTTTCATTATTAGAAA

ACGAATTTGATATTCATTCACAGAATAAATCTCAAGCATGAGGTGTTTCCCATTCTTTTC
3421-----+-----+-----+-----+-----+-----+ 3480
TGCTTAAACTATAAGTAAGTGTCTTATTTAGAGTTCGTACTCCACAAAGGGTAAGAAAAG

CTGAAGATGTATCTCAATGACACCCTCGCCTCCGACTCCATTCCAGATAAACCACACAGTT
3481-----+-----+-----+-----+-----+-----+ 3540
GACTTCTACATAGAGTTACTGTGGGAGCGGAGGCTGAGGTAAGGTCTATTGGTGTGTCAA
GTCGAAGATCTTGTTGATTTTCTCACTGCCAACAGATGTGATCATTAAACGTCTCCAAAGT
3541-----+-----+-----+-----+-----+-----+ 3600
CAGCTTCTAGAACAACTAAAAGAGTGACGGTTGTCTACACTAGTAATTGCAGAGGTTTCA
DdeI
|
GTACCCTGAGTATTTCTGACTGTTATGTTTTGATTAACTCCAGGTTGTGAAGGTTAAT
3601-----+-----+-----+-----+-----+-----+ 3660
CATGGGACTCATAAAGGACTGACAATACAAAATAATTTGAGGTCCAACACTTCCAATTA
DdeI
|
CTCGAAGTTACACAGGGACTGAGATACACACAAGCAAAAAAGTGGGAGGAATAAACGTTT
3661-----+-----+-----+-----+-----+-----+ 3720
GAGCTTCAATGTGTCCCTGACTCTATGTGTGTTTCGTTTTTTCACCCTCCTTATTTGCAAA
CATTTCTAATATTAAGTATATAAGTTATTTGGTCTGAATGTTTTTATAAATAGTTTTA
3721-----+-----+-----+-----+-----+-----+ 3780
GTAAAGGATTATAATTGATATATTCAATAAACAGACTTACAAAAATATTTATCAAAAT
NheI
|
TATAAATATTTTAGATCAAATGTATAAATCCAGAAACAAAATTAGAGAAAGCTAGCTCA
3781-----+-----+-----+-----+-----+-----+ 3840
ATATTTTATAAATCTAGTTTACATATTTAGGTCTTTGTTTTAATCTCTTTTCGATCGAGT
AAAGCCTTGGAATTTAAATGTTGGTTGAGATATAAATTAGGAAAAACCAATTTATGAGA
3841-----+-----+-----+-----+-----+-----+ 3900
TTTCGGAACCTTTAATTTTACAACCAACTCTATATTTAATCCTTTTTTGGTTAAATACTCT
ATTTTATGGAATGTTTTACTGACAGATTATTATAAAAAATTTTGAGAAATTA AAACTTT
3901-----+-----+-----+-----+-----+-----+ 3960
TAAAATACCTTTACAAAATGACTGTCTAATAATATTTTTAAAACTCTTTAATTTTGAAA
CATCGAAAAATTAAAAAGGAACAGCGACATATTAATAAAGCGATTCACTTAAAAGTAAAA
3961-----+-----+-----+-----+-----+-----+ 4020
GTAGCTTTTTAATTTTTCTTGTGCTGTATAATTATTTTCGCTAAGTGAATTTTCATTTT
DraI
|
GTAAACAAATTGATTAATGTTGATAATAGTTAAAATTTAAATGTAAAAAATGTCATTAAA
4021-----+-----+-----+-----+-----+-----+ 4080
CATTTGTTTAACTAATTACAACCTATTATCAATTTTAAATTTACATTTTTTACAGTAATTT
AGAGATTAATCAACTTTAATTAGTGGTGGCTGATGAGTCAGTATTTTGCTGCTGCTTATG
4081-----+-----+-----+-----+-----+-----+ 4140
TCTCTAATTAGTTGAAATTAATCACCACCGACTACTCAGTCATAAAACGACGACGAATAC
DdeI
|
CTTAGATTTACTTTTGAAATATTTCTTTTCGTGTGTTTTATTCTCTGAATTGACTGCGTT
4141-----+-----+-----+-----+-----+-----+ 4200
GAATCTAAATGAAAACCTTTATAAAGGAAAGCACACAAAATAAGAGACTTAACTGACGCAA
ATTTTGTGGCTTTTCATTTCGCGGAGTTCCCTCCAGCTCTCATCATCGTCCCTAGGACGTGG
4201-----+-----+-----+-----+-----+-----+ 4260
TAAACACCGAAAAGTAAGCGCCTCAAGGAGGTCGAGAGTAGTAGCAGGGATCCTGCACC
CTTGGCCCTGTGACGAGATCTACGAGACCTACCAGAGCCACTGCCTGAAGTGGCAGGCTT
4261-----+-----+-----+-----+-----+-----+ 4320
GAACCGGGACACTGCTCTAGATGCTCTGGATGGTCTCGGTGACGGACTTCACCGTCCGAA
DdeI
|
CTCAGTTTTGGGGACAGAAGGTTGTGGATCAGAAACGCCAGTTGTAGGTGGGGGATCTGT
4321-----+-----+-----+-----+-----+-----+ 4380
GAGTCAAAACCCCTGTCTTCCAACACCTAGTCTTTGCGGTCAACATCCACCCCTAGACA

AGAAGGTGTGGCGTTTGGAGGTTGAGAAGTTGGAACCTCATTGGAATTTATTTGGGTCAG
4381-----+-----+-----+-----+-----+-----+ 4440
TCTTCCACACCGCAAACCTCCAACCTTTCAACCTTGGAGTAACCTTAAATAAACCCAGTC
DdeI
|
GCTTAGTGAATAACCTGTCCTTTTTATCCTTTTGAGGTTTCGCATTTGTGTTTGGTTGATT
4441-----+-----+-----+-----+-----+-----+ 4500
CGAATCACTTATTGGACAGGAAAAATAGGAAACTCCAAGCGTAAACACAAACCAACTAA

AGAATTTGCATATTTGAGCATTAGAAGACGAACTAGTGTATTTTCTACGTCTTGAATAAC
4501-----+-----+-----+-----+-----+-----+ 4560
TCTTAAACGTATAAACTCGTAATCTTCTGCTTGATCACATAAAAGATGCAGAACTTATTG

GATCCTTTTTACGCTTCTCTACTCCAGCTGGTAAAGTAGTTGTGAGGGTTTCACGAATTG
4561-----+-----+-----+-----+-----+-----+ 4620
CTAGGAAAAATGCGAAGAGATGAGGTCGACCATTTTCATCAACACTCCCAAAGTGCTTAAC

TAGTGTCAAGCTGGTCTGAAAGTGTATTTTCCAGCTGATTTGATAGTGTGTTGTCGATTT
4621-----+-----+-----+-----+-----+-----+ 4680
ATCACAGTTCGACCAGACTTTCACATAAAAGGTCGACTAAACTATCACACAACAGCTAAA

GATCTGATAATGTGTTTTCTAGCTGGTTTTGTAAGCCTATTCTCAAGCTGGCCGACAAGTC
4681-----+-----+-----+-----+-----+-----+ 4740
CTAGACTATTACACAAAAGATCGACCAAACATTTCGGATAAGAGTTTCGACCGGCTGTTTCAG

TTTCATCCAACCTGTTTGTGTTGTGAATCTTGTAAGAAGTGGAGTTGTCTTTGAAGTTGG
4741-----+-----+-----+-----+-----+-----+ 4800
AAAGTAGGTTGAACAAACAAACACTTAGAACATTTTCTTACCTCAACAGAACTTCAACC

TTAAATCCAGAAAAGGCTGGGACAGCTTATCAGATACATACCCCTTAATTGTTAGACCAT
4801-----+-----+-----+-----+-----+-----+ 4860
AATTTAGGTCTTTTCCGACCCTGTGCAATAGTCTATGTATGGGGAATTAACAATCTGGTA

CGTCCGGTGCTCTATCTGACAGAATTGAGTGATTAAGGGTGTCAATTGCAGTTGTCAGTT
4861-----+-----+-----+-----+-----+-----+ 4920
GCAGGCCACGAGATAGACTGTCTTAACTCACTAATTCCCACAGTTAACGTCAACAGTCAA

CTGCCTCATTGTATCCTTCACTTTTGTATCAAGCCAATTAGAGTCCGATACAGTAGGCG
4921-----+-----+-----+-----+-----+-----+ 4980
GACGGAGTAACATAGGAAGTGAAAACAATAGTTTCGGTTAATCTCAGGCTATGTCATCCGC
AvaI
XhoI
|
ACCCTACCATGACGTTTGGCTTAAACATTTGAGGTCTCGAGCTCTTCAGCCTTGTGGCAT
4981-----+-----+-----+-----+-----+-----+ 5040
TGGGATGGTACTGCAAACCGAATTTGTAACTCCAGAGCTCGAGAAGTCGGAACACCGTA
DdeI
|
ACTGCTGAAAACCTTGGTAAATGCCGTTAGGGTTAGCTGGATATGGAACCTGAGGGTGGAG
5041-----+-----+-----+-----+-----+-----+ 5100
TGACGACTTTTGAACCATTTACGGCAATCCCAATCGACCTATACCTTGACTCCACCTC

GTGGAGGTGGAGACTCTACGCCTCTTCCAATGAAGTTGGTAAATCCCAGGGGGTAGTTAC
5101-----+-----+-----+-----+-----+-----+ 5160
CACCTCCACCTCTGAGATGCGGAGAAGGTTACTTCAACCATTTAGGGTCCCCCATCAATG

CCTTCATTCCAGGCATCAGGAAGTCAGTACCCAGGATATTGTCAGAGTCTTCGTTGCTGG
5161-----+-----+-----+-----+-----+-----+ 5220
GGAAGTAAGGTCCGTAGTCCTTCAGTCATGGGTCCTATAACAGTCTCAGAAGCAACGACC

AATTCCTGAAAGACAAAGAGTTTGAATAGCTGCTGTCACGACTACTATGGAATCCTGGGC
5221-----+-----+-----+-----+-----+-----+ 5280
TTAAGGACTTTCTGTTTCTCAAACCTTATCGACGACAGTGCTGATGATACCTTAGGACCCG

TTGCAACCGACTTAACACTCGAAATTCTGTGCTTGAAC TTGTACCTAGTACTTCGAGGTT
5281-----+-----+-----+-----+-----+-----+ 5340
AACGTTGGCTGAATTGTGAGCTTTAAGACACGAACTTGAACATGGATCATGAAGCTCCAA

CTTTAGAAAGGCGATCTTGGTCATTTGTAACGCTGGAGGCTTCAGAGATACGATTATTTA
5341-----+-----+-----+-----+-----+-----+ 5400
GAAATCTTCCGCTAGAACCACTTAAACATTGCGACCTCCGAAGTCTCTATGCTAATAAAT
AvaI
|
CACTAACATCGCGGTTACTTGACCCGAGTAGATCTGGGAATGACACGTCGTCCAGTGTTTC
5401-----+-----+-----+-----+-----+-----+ 5460
GTGATTGTAGCGCCAATGAACCTGGGCTCATCTAGACCCTTACTGTGCAGCAGGTCACAAG
HincII
HpaI
|
CACTGCATATCTTTTATTTAATTAGTTAACTTACGTTTTAGTTTCAACGTTTGAGTCGTA
5461-----+-----+-----+-----+-----+-----+ 5520
GTGACGTATAGAAAATAAATTAATCAATTGAATGCAAAATCAAAGTTGCAAACCTCAGCAT
HincII
HpaI
HhaI
|
CTTTGCTTCTGACTCCCAGGCGCTTCCGGCAGTCTTAATCTGGCTGGTCGTGTTAACTGC
5521-----+-----+-----+-----+-----+-----+ 5580
GAAACGAAGACTGAGGGTCCGCGAAGGCCGTGAGAATTAGACCGACCAGCACAAATTGACG
HincII
HpaI
|
AATAGCCCCGGACACTAACACTTATACCTGAATAAAATGTTAACCTTTGGTTGTATATAT
5581-----+-----+-----+-----+-----+-----+ 5640
TTATCGGGGCCTGTGATTGTGAATATGGACTTATTTTACAATTGGAAACCAACATATATA

HindIII
|
AAGCTTGAGTA
5641-----+-- 5651
TTCGAACTCAT

Appendix III Comparison of the *TamS1* contig with the *TpmS1* contig over the 5.0kb. The *T.annulata* (*TamS1*) and *T.parva* (*TpmS1*) sequences were derived from the genomic clones from chapters 2 and 3 respectively. The numbers represent the sequence position with reference to the most 5' base. All alignments are made to the *TamS1* sequence and dash represents identity, dot represents a gap and any substitutions are shown by the relevant letter. The protein coding start site for the 30kDa polypeptide is located at base 2127 (ATG sequence is boxed) and terminates at 3026 (highlighted in bold print). The protein coding sequence for the Ta-ORF-1 gene starts are position 5063 (ATG sequence is boxed) and terminates at 4109 (highlighted in bold print).

1 50
TamS1 ~~~~~~T
TpmS1 TCGAGGTCGA CACTAGTGGA TCCAAGAATT CAAAAAGCTT CGTCCCGTTT

51 100
TamS1 ATATAGTACA ATGATATTAT TACAAGTCCC ACCTCCCTCA GCAACACCCCT
TpmS1 -C-A-A----- ----C----- ---G-----T --T---T--- ----T-----

101 150
TamS1 AATGAATATC TTATTTATAT ATTTATTCAT ACCATGCTTC CTTTAACAGA
TpmS1 G-..---AAT ----AC-C-C T-G-T-A--- .---C--C-- -----A---

151 200
TamS1 TTAAATGAAT ACTCGTTTGA TATTATACCA TTATTTAATT TCAACAACCTC
TpmS1 -----TGA T---A--G-- -----

201 250
TamS1 ATATGTTTCT GTGTTGTTAA CTGCGTTTTG TTGATCGGTA TCGTTTGAAT
TpmS1 G-----C -A---A-C-- ----A---CC -G-----TGC C-----A---

251 300
TamS1 TAGAGAAATT ACAGGAGCCA ACCTTGAGAC TTTCGAGGTT TATGTGATCT
TpmS1 ----T-T--- ---C----- ----A----- -C-----.- --A-----

301 350
TamS1 ATGAAAGATA TAAGGTTGTG TATTGGTGTT TTCTTTTCTG AAAATTCGTT
TpmS1 ----- A----- ----A--- ----C-G-- -----A--

351 400
TamS1 TAAATTCTTC CTATTCAGTT CGTGGATCAT GTCCTTAATC ACTTGCTTGC
TpmS1 ----- T-----A-- ----- ----G--- -----AT

401 450
TamS1 TTTTTAACGG TTTCTTTTTC AATGTGCTAA ATACCCTTTA CCAAAATCTC
TpmS1 -----T--- ---T--C--- ----T--- ---.T-----

451 500
TamS1 TCTCTGACTT TCCATTTTCT CATATCTATG TTCCAAATCA TCGTTATCTA
TpmS1 ----- ----C----- ----C----- --AC-G-. -T

501 550
TamS1 GTATCAGTTT TAAGACGTTT TCCACAACCTT CCGTTTAATT CATGGCTCTG
TpmS1 --C----- -GTGC--C-- -T-TG-TG-- --A-G--C-- -GA--T----

551 600
TamS1 TTCTTCAACT AACTTCTCAA TGTTTTCACT AAAGAACTTT AGAGAGTATT
TpmS1 ---C-G----- ----TG- ----T- ----A--A -CC-----

601 650
TamS1 CCTTGTGGTT CAGGGGGACC TAAGATGAAA CTTTTGGGAT CTCAAAATTA
TpmS1 -T---GAA-- --CA--A-A- ----GA-T-- .----TAG-C T-A--T----

651 700
TamS1 CCCTAAAATG CTCAACATTA GTACGACAGG GAGAAATTGT AACCTCTTGA
TpmS1 --T----- ----AT--- ---TA---A- -T-----A-- G-----GA-C-

701 750
TamS1 CCCGCCTTTA GAGATGTTCT AAGCTCGTTC AAATCCATAG AGTTTAAAT
TpmS1 --T-GT--C- -CC---C--- ----TG-- --G--T-G-- ----G--G--

	751				800
<i>TamS1</i>	TTCCGATAAA	ATGATCACAG	TTGATGGCGA	TATCTTTGAT	AAACTAAATA
<i>TpmS1</i>	C--T---G--	-----	-GACG--T-T	-----	---T--GT--
	801				850
<i>TamS1</i>	GTAAACG TTC	GTAAAGCCTC	GAATGAGGTT	TCAAATACGA	TGTGTCCAGT
<i>TpmS1</i>	----C--G--	-----A	--T-----	-----T--	-----G---
	851				900
<i>TamS1</i>	AAATTTATTT	CCAAAGTCAC	AGACTCATCT	AAAGAAGTTA	TTACCACTAA
<i>TpmS1</i>	-G-----	-----A--	-----GG---	G-GA-T----	---TT---T-
	901				950
<i>TamS1</i>	ATCAAACCTT	CGTTGCGATG	TGTTATATTA	CAGTTTGACA	ATCCTATAGT
<i>TpmS1</i>	-----	-A-----	-AC-----	-----T-	----C-A---
	951				1000
<i>TamS1</i>	TCTGAAAACT	TCCAGATCAA	GCTATAAATT	ATTATAAAAT	ATAACAAATG
<i>TpmS1</i>	---A-----C	----A-----	---G-----C	-----TG---	GC-CT-----
	1001				1050
<i>TamS1</i>	AAATTATTAA	ACATACATTG	ATTGTTAACA	CCCCTTTGGA	GCTTATTATG
<i>TpmS1</i>	-G--A-----	-----	-G-----	-----	AT----C-CA
	1051				1100
<i>TamS1</i>	AAATTGATTT	TCGATTTATA	ATTATAATAT	ATATTTAAAA	CGGAGAGTCT
<i>TpmS1</i>	-----C-	-T-----	-----	-----	TT--C-----
	1101				1150
<i>TamS1</i>	TTTGGGAACC	AAATAATTTA	CAATAAAGTC	CAAATTAAGC	AAAGACGTCA
<i>TpmS1</i>	---A-----	-----	-----	--G---G-C-	-----TTCA-
	1151				1200
<i>TamS1</i>	GAGTAATGTT	CGAAACAACA	AAGCATGTGT	CCTGTTTAAC	TCAAATTAGT
<i>TpmS1</i>	-----	---TG-----	-----	-----G---A	GT-----
	1201				1250
<i>TamS1</i>	CATTTTAAAT	AAATTAAGGG	GTAAAATCAA	ATACACATCA	AAAGTATAAA
<i>TpmS1</i>	CG----T---	---.---CAA	CA---G-CG-	-----CT-	-----T-
	1251				1300
<i>TamS1</i>	TATATAAAAT	ACCTTAGAAT	TAAAAAGATC	TCTCTAAAAG	AATCAATGCA
<i>TpmS1</i>	-G--A---TA	C---AGA-T-	A----GT-T-	-----G--G-	C--T-GCGA-
	1301				1350
<i>TamS1</i>	AGGATTCTCT	TGGGTAGGGA	TGAGAGGCAC	CGTTGCCTAT	ACGACAGATA
<i>TpmS1</i>	---GG----C	---TAG--AT	A---GACGGT	--.....CC
	1351				1400
<i>TamS1</i>	TAGAGTGTTA	GGATCGACAA	GCACAGTAGT	TTAATTACAT	ATTACAATCT
<i>TpmS1</i>	--A-----A-	ATG-TAGG--	T-T---C--A	---T-A-ATC	---GA-T--A
	1401				1450
<i>TamS1</i>	GGATAGATGT	ATAGAAATAA	AAAGAGCATA	CACACCTCTA	TCTGTAAC TC
<i>TpmS1</i>	-A---.....-A-	A---TACGCT	CT--AC-AAA
	1451				1500
<i>TamS1</i>	TCTGGAGAAT	AAATAGGGTG	TACTATTTGC	TGAAGAAAAT	TTTCAAAATC
<i>TpmS1</i>	-T-----A	-----ACAG-	G-G-----A	G-C-A-----	-----T

	1501				1550
<i>TamS1</i>	CGGGCCGGTA	ATATCTGAAG	GAGACTGATT	AAACTCAAAG	ACATCACACT
<i>TpmS1</i>	T--A--CAG-	-A----A--T	-----C-	-----	-----
	1551				1600
<i>TamS1</i>	CTGATGAAAG	TGAAGGCGGG	TAAGGACAGC	TCATAACGAC	AAATTCAGTA
<i>TpmS1</i>	-----G---	---CAATA-A	-G--T--GAA	C---G-TT--	-----ACA--
	1601				1650
<i>TamS1</i>	TAAAAACATA	ATGAATGATT	GTTTTAAGGA	AATGATGTGT	GGAATCACGG
<i>TpmS1</i>	TT-----CG	T-AG----G-	-----C---	--A-----	--G-C-GTTT
	1651				1700
<i>TamS1</i>	TTATTTGTGC	GGGAAATTAT	GCATGGATAT	TATCTATTTA	ACTTATAATA
<i>TpmS1</i>	-A-C-----	--A-----GA	TT-AA-T---	-C-----C--	-G-CTG-G--
	1701				1750
<i>TamS1</i>	ATGTTATATT	AAAATATGAC	AAATTAATGT	GTGAAATAAA	TAAAGTAAAG
<i>TpmS1</i>	T-----G-	-----A-T	-----	---G-----	-G---A----
	1751				1800
<i>TamS1</i>	TGGAGGGAAA	TTCTGTCCTA	ATATTTAACT	GGGTCTAGTG	ACTTAAAAAA
<i>TpmS1</i>	-A-GTA-G--	-CG-----	-----	-----	-AC-.-----
	1801				1850
<i>TamS1</i>	GAAAGCAGGT	TATTGGTTTG	AAAAAATAAT	AAATTCATCT	TTTTATAGTC
<i>TpmS1</i>	T-----	----T-----	-----C--C	G----TTCT-	----C--AGC
	1851				1900
<i>TamS1</i>	ATCAAAAAAT	GAAAAATGGA	TAAATATACA	TGTTATTTCA	TTAAGATGGA
<i>TpmS1</i>	G-----	A-----A--	-----T-	-----GCA-	---T----AT
	1901				1950
<i>TamS1</i>	CTTCAATGGA	GGATAAGGCA	TTAGGCGTGA	TGAGTGCATA	GATACAGATA
<i>TpmS1</i>	G-GA---T--	-CG-----GG	-----	C-----AT	-----G
	1951				2000
<i>TamS1</i>	AACATGCACA	CAATTTGTAG	GGCGACATTG	TTTGTATGG	TGAATGCATA
<i>TpmS1</i>	-----	-----T-	-----G-AC-	C-----	-----CTTC
	2001				2050
<i>TamS1</i>	TCCATCTCAC	TTCACTTATA	GTTATTGGTC	TTTTTTTCCA	TAACATCCAC
<i>TpmS1</i>	-----	----A-----	-----	..-----	-----TT-T
	2051				2100
<i>TamS1</i>	CCAATTAGTT	AATTTTAAAT	ATTTAAATCG	CTCACTAGTC	TGCCCTTTCT
<i>TpmS1</i>	--GG---T--	-----	----GG----	--A-----	----T---G-
	2101				2150
<i>TamS1</i>	TATCTTTTTA	TAATATAATT	ATTTGAGATC	TTGTCCAGGA	CCACCCTCAA
<i>TpmS1</i>	-CGG-----	-----	-----	-----A-	AT-----
	2151				2200
<i>TamS1</i>	GTTCTTATAT	TTGAGCTTCT	TCGTTATCTC	ATCCGTTAAT	GCTGCAAATG
<i>TpmS1</i>	-----	----T-----	-----	T-G-----	--C-----A-
	2201				2250
<i>TamS1</i>	AGGATGAAAA	GAAAAAGGAG	GAAAAAAAAG	ATGTTGTTCT	TGATGTTACT
<i>TpmS1</i>	-AG-A--G--	---G-----	A---GG-G-	--C--ACAG-	-----C

	2251				2300
<i>TamS1</i>	CTCACTTCAT	GTGAGAATGT	AACCTTTAAA	AACGTCGACT	CTAACACCAC
<i>TpmS1</i>	--TT-G----	-G--A-----	T-CT-C-...	-CTCC---GG	--GG-GGT--
	2301				2350
<i>TamS1</i>	TGAGTTAACT	GTCGCGGATG	GCTACCGTTT	CAAGACCCTT	AAGGTCGGAG
<i>TpmS1</i>	ATTA--G-AA	-C-AAT--A-	-T-----	-----A---	-----C-
	2351				2400
<i>TamS1</i>	ACAAGACCTT	GTTCAATGTT	GACACCTCAA	AACATACCCC	AGTACAGGCA
<i>TpmS1</i>	-----T--	--A---C---	-----	--T-CGATG-	-----CCT-
	2401				2450
<i>TamS1</i>	TTCAAACCTTA	AGCATGAATC	CGATGAGTGG	TTCAGACTTA	ATCTTCACCC
<i>TpmS1</i>	TA-----	CC-----T--	T-----A---	C-A-AG--CC	T---C-----
	2451				2500
<i>TamS1</i>	TGCCCAGCCA	AAGATGTTCA	AGAAGAAGGG	AGACAAGGAA	TATTCTGAGG
<i>TpmS1</i>	A---A-----	GT-----	-----TC	-----	-----C--A-
	2501				2550
<i>TamS1</i>	TCAAATTCGA	GACCTACTAC	GATGATGTCT	TGTTCAAGGG	AAAATCCGCC
<i>TpmS1</i>	-----	A-----T	-----	-----	G-----A---
	2551				2600
<i>TamS1</i>	AAGGAACTAG	ATGCTTCCAA	GTTCGAAGAT	ACATCTTTGT	TCACCTCCTC
<i>TpmS1</i>	-----C-	-----	GG--ACT---	--TGGC----	TT---CAAGA
	2601				2650
<i>TamS1</i>	CGCCTTCGGC	ACTGGAAAGA	TGTACACCTT	TAAAAAGGAA	TTTAAACCTT
<i>TpmS1</i>	GAG-----	-----G----	A-----	C--C--TAGC	--C-----
	2651				2700
<i>TamS1</i>	CCAAAGTCAC	CTTCGACAAG	AAAGAAGTCG	GAAAAACCAA	CAATGCCAAG
<i>TpmS1</i>	----G---T-	A-----A	-----T--T-	----G--CG-	---G-----
	2701				2750
<i>TamS1</i>	TATCTTGAAG	TTGTTGTTTT	TGTTGGTTCT	GATTCCAAGA	AGTTTCGTCAA
<i>TpmS1</i>	T-C--C--C-	--T-C--C-A	---C--C---	---GA-----	--G-T--T-G
	2751				2800
<i>TamS1</i>	ACTCTACTAC	TTCTATACCG	GTGACTCAAG	GTTGAAGGAG	ACCTACTTCG
<i>TpmS1</i>	G---G-----	----T-GGT-	-----	-----	GT-----
	2801				2850
<i>TamS1</i>	AGCTTAAGGA	CGATAAGTGG	GTTCAAATGA	CACAGGCAGA	TGCAAACAAG
<i>TpmS1</i>	-----A--	---C-----	--CA-----G	A----AAT--	C-----
	2851				2900
<i>TamS1</i>	GCCTTGAATG	CCATGAACTC	ATCCTGGTCA	ACCGATTACA	AACCAGTTGT
<i>TpmS1</i>	--A---C---	-----G-GA	T--A---AA-	TTG--C----	-----C---
	2901				2950
<i>TamS1</i>	CGACAAGTTC	TCCCCCCTTG	CAGTCTTCGC	CTCAGTACTC	ATCGTCTTCT
<i>TpmS1</i>	-----	--T-----	-----C----	-----	-----GC-G
	2951				3000
<i>TamS1</i>	CATCAGTCCT	TTACTTCCTT	TAAAACCCAT	GTTTCGTAACA	ACTTATCAAC
<i>TpmS1</i>	-T-----T-	----AA-----	-----	--G-----	-----

	3001				3050
<i>TamS1</i>	TTTTAAAAACA	ATTTTGATAA	TTTGT ATACA	ATTGCAGAAA	CTAAATAACT
<i>TpmS1</i>	G-----	-TA-----	-----	---A---C-G	----C-----
	3051				3100
<i>TamS1</i>	AGCTTAAGTC	ATTATATGCC	ACTTAATTTT	ATA.CTTTAA	ACTTATATGT
<i>TpmS1</i>	TAAG-C-T-.	.-----TGA--	-----C	---C-C--CG	--A---C-A-
	3101				3150
<i>TamS1</i>	TTAGATATAA	C..TTCAACA	GCTATAGGTC	ATCAATATCC	AAGTCACGGT
<i>TpmS1</i>	C--GT---T-	-CAG-TT---	-T-T-GAT-T	C-ACG-T-TT	-TA--T-TTG
	3151				3200
<i>TamS1</i>	AAAATCG...	CGGCCCTCAC	GCTAACCAGC	TAAAAGCAGA	ACAGCAAAC
<i>TpmS1</i>	-G---A-TTA	TA-TAAG--G	-TG-TAGC--	CGC--CAGAT	TTTAGT---C
	3201				3250
<i>TamS1</i>	CTGGTTATCG	CAGCTTCCTC	TTCACTTTCG	TCATCTCTTT	CCACAGTGTG
<i>TpmS1</i>	G-A-A-TC-T	-TT--A----	C--CGG-A--	-TG-TGGCAC	-TGT--CA-T
	3251				3300
<i>TamS1</i>	CAATGAACCT	GTAAAAATTA	AGTGAATTGT	AAAAATGAAA	TTATTAATCT
<i>TpmS1</i>	T---TTGA-A	A-TGT-C-A-	TA--TTAAC-	-G--T-T-T-	-A-AG--AA-
	3301				3350
<i>TamS1</i>	TACTGAAAAG	ACGGTCGAAC	TCAGAATAGG	CATCGTAATC	GTCCTCAACA
<i>TpmS1</i>	-----G--	-A-----T	--G--G----	A--T-A-T--	C--T---G-G
	3351				3400
<i>TamS1</i>	AC.....CCC	TAAAGCCTCG	TTAATAGGAC	GAGAATAGAG	CCGT.AGTCC
<i>TpmS1</i>	--TTCCG-AT	A---AT---A	GAC-----CT	TG---CGA--	--T-A-----
	3401				3450
<i>TamS1</i>	CTAAGCAAAT	TCTGGTACTG	AATGAAACTA	GTCCTCAGCC	AGCACCTGCG
<i>TpmS1</i>	-----G-G--	-ACA-----	--G----G-C	--T-G-C--	--A-GT--GA
	3451				3500
<i>TamS1</i>	AGTGAACACG	AAATGGTAGT	CCTTGCGCAC	ATCGTCCTGA	TTCACAGAGT
<i>TpmS1</i>	-----G--A	--G-AA--A-	-T--A--A-T	T-GA-AT-C-	-----A-
	3501				3550
<i>TamS1</i>	ATATCTCCAT	CATGAGGTGG	TTTCCATTCT	TCTCACGCAA	AAAAATCTCG
<i>TpmS1</i>	-A-----A-G	-----T	--C-----	-T--CT-A-G	-TGT-----A
	3551				3600
<i>TamS1</i>	ACGACACCCT	CACTGCCGGC	GCCAGTCCAA	ACAACGATGC	AGTTGTCCAA
<i>TpmS1</i>	-T-----	-G-CT---A-	T---T----G	-T---C-CA-	-----G--
	3601				3650
<i>TamS1</i>	GATTCTGTTA	ATTTTGTCCC	TGCCAATCGA	AGTGATAATG	GTAGTGTCCA
<i>TpmS1</i>	---CT----G	-----C--A-	-----CA--	T-----C-T-	AAC--C----
	3651				3700
<i>TamS1</i>	AAGTGTAACC	TAGGTGTTTC	CTGACCCTTA	TGTTTGAATT	AAACTCCAGG
<i>TpmS1</i>	-----	-GA--A----	-----TG---	-----TG---	-----
	3701				3750
<i>TamS1</i>	TTGTGGAGAT	TAATCTCAAA	GTAGCAGAGA	GTCCAGGACG	AACTTAGATA
<i>TpmS1</i>	-----A--G-	-----G--	--TA--C--G	-A--GA--TA	C--AC-AGC-

	3751				3800
<i>TamS1</i>	AAAAATTACA	AGTAACAAAT	ATTTCATTTG	AAAAGACTAA	C.....
<i>TpmS1</i>	-----G-GGG	--G--T---C	G-----C	CT--T-T---	-TATATAAGT
	3801				3850
<i>TamS1</i>TG	CATAAATATT	TTTACTAGCT	AAATTATTTA
<i>TpmS1</i>	TATTTGGTCT	GAATGTTT-T	T-----G-	----TATAAA	-T---T-AG-
	3851				3900
<i>TamS1</i>	CTGAAATGTA	CAAAATGGAA	ATAAAATTAA	AACAAGCTAG	CTGGAGAACA
<i>TpmS1</i>	TCA--TGTAT	A--TCCA---	-C-----G	-GA-----	--CA-A-G-C
	3901				3950
<i>TamS1</i>	TGGTCTGACA	AAAAGCTAGG	CAGATAATAA	GTTACAAAAA	ACAACTTTAC
<i>TpmS1</i>	-T-GAAATT-	---T-T-G-T	TGAGAT----	A---GG----	--C-A----T
	3951				4000
<i>TamS1</i>	AAAATTTTT.
<i>TpmS1</i>	G-G-A----A	TGGAAATGTT	TTACTGACAG	ATTATTATAA	AAATTTTGAG
	4001				4050
<i>TamS1</i>
<i>TpmS1</i>	AAATTAAAAA	CTTTCATCGA	AAAATTAAAA	AGGAACAGCG	ACATATTAAT
	4051				4100
<i>TamS1</i>TAAAAC	AACTAGTTTC	CTGAGGCTCA
<i>TpmS1</i>	AAAGCGATTC	ACTTAAAAGT	AAAAGT----	--A-T-A--A	A--TT-ATA-
	4101				4150
<i>TamS1</i>	GACAATT TAA	TTTAATGTGG	AAATAAT..A	TCGAAGAGAT	TAATCAACTT
<i>TpmS1</i>	T-GTTAAA-T	--A-----AA	---ATG-CAT	-AA-----	-----
	4151				4200
<i>TamS1</i>	TA.....G	TGGTATTCGG	CTCAGTGTTT	TGCTGATGTT	TATGCTTAGG
<i>TpmS1</i>	--ATTAGTG-	---CTGAT-A	G-----A---	-----CT-C-	-----A
	4201				4250
<i>TamS1</i>	CTTACTTTTG	AAATATTTCC	GTTCGTGTGT	ATTTTTCTCT	GAATTGACTG
<i>TpmS1</i>	T-----	-----	T-----	T--A-----	-----
	4251				4300
<i>TamS1</i>	CGTTATTTTG	TGGTTTTTCG	TTTGCGGAGT	TCCTCCAGTT	CTCATCCTCG
<i>TpmS1</i>	-----	---C-----A	--C-----	-----C-	-----CA---
	4301				4350
<i>TamS1</i>	TCCCTGTGGT	GGGACTTGGG	CCTGTGACGG	GATCTACGAG	ACCTACCGGA
<i>TpmS1</i>	-----AG-AC	-T-G-----	-----A	-----	-----A--
	4351				4400
<i>TamS1</i>	GCCTGTTCCC	GAGGTGGGAA	GTTTATCATT	TTTAGGGTCA	GGAGGCTGT.
<i>TpmS1</i>	---AC-G--T	--A----C-G	-C--C---G-	---G---A--	-A---T---G
	4401				4450
<i>TamS1</i>	..TCAGAAAC	ACTAGAAGTA	GCGGGAGGCG	CTGTAGA...GTCT
<i>TpmS1</i>	GA-----	G-C--TT---	-GT--G--AT	-----AGG	TGTGGC-T-
	4451				4500
<i>TamS1</i>	GGAGGTTGTG	AAGTTGGAAC	GTCATTGGGG	TTTTAGTTGG	ATCAGAATTA
<i>TpmS1</i>	-----A-	-----	C-----.-A	A---T----	G----GC---

	4501				4550
TamS1	GTGAATAACC	TGTCCTTTTT	ATCCTTTTGA	GATTTAGCAT	TAATGTTTGG
TpmS1	-----	-----	-----	-.G--C----	-TG-----
	4551				4600
TamS1	TTGATTAGTA	TTTACA.GTT	TGAGCGTTAG	AAGATGTACT	AGTGGGTTTT
TpmS1	-----A-	---G--TA--	-----A----	----C-A---	----TA----
	4601				4650
TamS1	TTGCGTCTTG	GATAACGATC	CTTTTTACGC	TTCTCTGCTA	AAGGTGATAA
TpmS1	C-A-----	A-----	-----	-----A--C	C--C--G---
	4651				4700
TamS1	ATCACTTGTA	AGAGTTTCAT	GAATTGTGGG	TTCCAGCTGG	TCTGACAGTG
TpmS1	-GT-G----G	--G-----C	-----A-T	G--A-----	-----A----
	4701				4750
TamS1	TATTTTCCAG	CTGATT....	TGATAATGTG
TpmS1	-----	-----TGAT	AGTGTGTTGT	CGATTTGATC	-----
	4751				4800
TamS1	TTTTCTAACT	GATTTGTTAG	TCTACTTTCT	AGCTGTCCGA	CGAGTCTTTC
TpmS1	-----G--	-G-----A--	C---T-C--A	-----G----	-A-----
	4801				4850
TamS1	ATCTAGTTTG	TTGGTTAATG	ACTCTTGCAA	TAAAAGAGAG	TTGTCTTTAT
TpmS1	---C-AC---	--T---TG--	A-----T--	A-G-T-G---	-----GA
	4851				4900
TamS1	AGTTGGTTAA	ATCCAGAAAA	GGCTGCGACA	GCTTATCAGA	TACATACTGC
TpmS1	-----	-----	-----G----	-----	-----CC-
	4901				4950
TamS1	TTAATTGTAA	GTCTATCGTC	TCTTGTTCTA	TCTGATAGAA	TTGAGTGGTT
TpmS1	-----T-	-A-C-----	CGG--C----	-----C----	-----A--
	4951				5000
TamS1	AAGAGTGTCA	ATTGCTGTTG	TCAGTTCAGC	TTCGTTATAT	CCATCACTTT
TpmS1	---G-----	-----A----	-----T--	C--A--G---	--T-----
	5001				5050
TamS1	TGTTGTCCAT	CCAATTAGAC	TCTGAGACGG	TAGGCGACTC	CACCATAACG
TpmS1	----A--A-G	-----G	--C--T--A-	-----C---	T-----G---
	5051				5100
TamS1	TTTGTTTAA	ACATTGAGG	TCTCGAATTA	CTCTTCAACC	TTGTTGCATA
TpmS1	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
	5101				5150
TamS1	CTGCTGAAAA	CCTTGGTAGA	TCCCGTTAGG	TTGGCTGGGT	ACGGCACTGA
TpmS1	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
	5151				5200
TamS1	AGGTGGAGGT	GGAGGTGGTG	ACTCTACACC	TCTTCCGATG	AAGTTGGTAA
TpmS1	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~
	5201				5246
TamS1	ACCCTAGTGG	GTAGTTGCCC	TTCATTTTCA	GGCATCAGGA	AGTCGG
TpmS1	~~~~~	~~~~~	~~~~~	~~~~~	~~~~~