

**Genes encoding rhoptry proteins of the rodent  
malaria parasite *Plasmodium yoelii* YM**

**Judith Louise Green**

**PhD**

**University College London**

**University of London**

**1998**

ProQuest Number: 10016073

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 10016073

Published by ProQuest LLC(2016). 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

## **ABSTRACT**

The 235 kDa merozoite rhoptry proteins of the rodent malaria parasite *Plasmodium yoelii* are encoded by a multigene family, members of which are present on several of the parasite's chromosomes. The proteins have been implicated in the specificity of red blood cell invasion observed for different strains of the *Plasmodium yoelii* parasite. Although partial coding sequence for two of the members of the family, designated E3 and E8, has been previously described, the entire coding sequence for a member of the multigene family has not.

In this thesis, the complete coding sequence for the E8 gene is described. The sequence was obtained by the construction, and subsequent screening of, a *Plasmodium yoelii* YM genomic library. The structure of the gene is similar to that seen for many *Plasmodium* genes, with an intron separating a short signal sequence from the main body of the coding sequence. The intron is bounded by consensus eukaryotic splice sites at its 5' and 3' ends. The putative signal sequence itself contains all the features associated with a classical signal sequence, with positively charged residues followed by a largely hydrophobic core bounded by charged amino acid residues.

In addition to the E8 gene, partial sequence encoding three other members of the multigene family was obtained. Each of these had an extremely high homology to the E8 gene. All of the members of the gene family demonstrated sequence homology to a family of *Plasmodium vivax* reticulocyte binding proteins - PvRBP-1 and PvRBP-2. The significance of such homology is discussed, as well as further similarity to an integrin homologue protein of *Saccharomyces cerevisiae*.

## ACKNOWLEDGEMENTS

I would like to thank my supervisor, Tony Holder, for ideas and guidance during the development of this thesis.

I am grateful to all the members of the Division of Parasitology, NIMR for their advice. In particular, special thanks to the following: Irene Ling (for automated sequencing instruction), Sola Ogun (for help with the preparation of parasites), Ekkehard Werner (for advice on the construction of the genomic DNA library), and Don Williamson & Peter Moore (for running the pulsed field gel of *P. yoelii* chromosomes).

This work was carried out at the National Institute for Medical Research (NIMR) and financially supported by the Medical Research Council (MRC).

Thankyou to my friends and family for their support and encouragement. In particular, my parents, who have given many years of advice and tireless support.

This is especially for Sarah and John.



## ABBREVIATIONS

5' RACE	5' <u>R</u> apid <u>A</u> mplification of <u>c</u> DNA <u>E</u> nds
bp	base pairs
BFA	Brefeldin A
ddNTP	dideoxyNTP (where N is any of the four triphosphate nucleotides)
dNTP	deoxyNTP (where N is any of the four triphosphate nucleotides)
DDW	distilled, deionised water
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
IPTG	iso propyl- $\beta$ -D-thiogalactopyranoside
kb	kilobase pairs
mRNA	messenger ribonucleic acid
PFGE	pulsed field gel electrophoresis
PvRBP-1 or 2	<i>Plasmodium vivax</i> reticulocyte binding protein 1 or 2
Py235	<i>Plasmodium yoelii</i> 235 kDa rhoptry protein
Py235.E8	<i>Plasmodium yoelii</i> 235 kDa rhoptry protein encoded by the E8 gene
RT-PCR	reverse transcriptase polymerase chain reaction
UV	ultraviolet
w/v	weight/volume
X-Gal	5-bromo-4-chlor-3-indoyl- $\beta$ -D-galactosidase

## INDEX OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
ABBREVIATIONS	iv
INDEX OF CONTENTS	v
INDEX OF FIGURES	xi
 <u>CHAPTER ONE: INTRODUCTION</u>	 1
1.1 Malaria - a historical perspective	2
1.2 Malaria in the present day	3
1.3 Hypothesis of pathogenesis	4
1.4 The life cycle of <i>Plasmodium</i>	6
1.5 The merozoite	9
1.6 Invasion of red blood cells by merozoites	12
1.6.1 Initial attachment of the merozoite to the red blood cell	12
1.6.2 The role of the rhoptries and apical organelles	14

1.6.3	Internalisation of the merozoite	15
<b>1.7</b>	<b>Apical organellar proteins and their role in invasion</b>	<b>19</b>
1.7.1	Microneme proteins	19
1.7.2	Rhoptry proteins	20
1.7.2.1	Rhoptry associated proteins 1 and 2 (RAP-1 and 2)	20
1.7.2.2	RhopH1,2 and 3	22
1.7.2.3	Apical membrane antigen-1	23
1.7.2.4	The 225 kDa rhoptry protein of <i>Plasmodium falciparum</i>	24
1.7.2.5	The 235 kDa rhoptry proteins of <i>Plasmodium yoelii</i>	24
1.7.3	Reticulocyte binding proteins	28
<b>1.8</b>	<b>Multigene families in <i>Plasmodium</i></b>	<b>30</b>
1.8.1	The <i>var</i> family of genes in <i>Plasmodium falciparum</i>	30
1.8.2	The multigene family encoding the 235 kDa rhoptry proteins of <i>Plasmodium yoelii</i>	35
<b>1.9</b>	<b>The control of gene expression in <i>Plasmodium</i></b>	<b>38</b>
1.9.1	General features of promoters in eukaryotic genes	38
1.9.2	Transcription and its control in <i>Plasmodium</i>	41
1.9.3	Stage-specific expression of <i>Plasmodium</i> genes	42
1.9.4	Identification of promoters for <i>Plasmodium</i> genes	42
<b>1.10</b>	<b>Methods for the study of gene expression of <i>Plasmodium</i></b>	<b>48</b>
<b>1.11</b>	<b>Aims of this project</b>	<b>49</b>
	<b><u>CHAPTER TWO: MATERIALS AND METHODS</u></b>	<b>51</b>

<b>2.1</b>	<b>Buffers and media</b>	<b>52</b>
<b>2.2</b>	<b>Preparation of <i>Plasmodium yoelii</i> parasites</b>	<b>53</b>
<b>2.3</b>	<b>Purification and preparation of nucleic acids</b>	<b>54</b>
2.3.1	Purification of <i>Plasmodium yoelii</i> genomic DNA	54
2.3.2	Purification of <i>Plasmodium yoelii</i> total RNA	54
2.3.3	Synthesis of oligonucleotides	54
2.3.4	Preparation of plasmid DNA	55
2.3.5	Restriction endonuclease digestion of DNA	56
2.3.6	Purification of DNA from agarose gels	56
2.3.7	Preparation of vector and insert DNA for ligation	56
2.3.8	Ligation of vector and insert DNA	57
2.3.9	Transformation of DNA into competent bacterial cells	57
<b>2.4</b>	<b>Sequencing of DNA</b>	<b>58</b>
2.4.1	Manual sequencing of double stranded plasmids	58
2.4.2	Automated sequencing of DNA	58
<b>2.5</b>	<b>Polymerase chain reaction</b>	<b>59</b>
<b>2.6</b>	<b>Inverse PCR</b>	<b>63</b>
2.6.1	Digestion of <i>P. yoelii</i> YM genomic DNA with restriction endonucleases	63
2.6.2	Ligation of digested DNA	63
2.6.3	Amplification of target DNA using Touchdown PCR	64
2.6.4	Purification of inverse PCR products in preparation for sequencing	65

<b>2.7</b>	<b>5'-Rapid Amplification of cDNA Ends (5'-RACE)</b>	<b>68</b>
2.7.1	Synthesis of first strand of cDNA	68
2.7.2	Purification of cDNA	68
2.7.3	Oligo-dC tailing of cDNA using terminal deoxynucleotidyl-transferase	68
2.7.4	PCR of oligo-dC tailed cDNA	69
2.7.5	Cloning and sequencing of 5'-RACE products	70
<b>2.8</b>	<b>Reverse Transcriptase PCR (RT-PCR)</b>	<b>73</b>
2.8.1	Synthesis of cDNA using gene-specific primers	73
2.8.2	PCR from cDNA using E8-specific primers	73
2.8.3	Cloning of RT-PCR products	73
<b>2.9</b>	<b>Construction and screening of a sub-genomic DNA library</b>	<b>76</b>
2.9.1	Identification of positive fractions by hybridization	76
2.9.2	Preparation of pBluescriptII SK <sup>-</sup> plasmid	76
2.9.3	Ligation of purified insert into pBII SK <sup>-</sup> and transformation of <i>E. coli</i> SURE cells	77
2.9.4	Screening the library	77
<b>2.10</b>	<b>Construction of a partially Sau3A digested, size-fractionated genomic library in pBluescriptII KS<sup>+</sup></b>	<b>79</b>
2.10.1	Partial Sau3A digestion of <i>P. yoelii</i> genomic DNA	79
2.10.2	Digestion of pBluescriptII KS <sup>+</sup> with XhoI	79
2.10.3	Size-fractionation of insert DNA	80
2.10.4	Partial fill-in of vector and insert ends using the Klenow fragment of DNA polymerase I	80
2.10.5	Ligation and transformation	81
2.10.6	Assessment of insert size	81

2.10.7	Plating out the library	82
2.10.8	Preparing replicas of the library	82
2.10.9	Storage of the library	83
2.10.10	Preparing the replicas for screening	83
2.10.11	Screening the library	84
2.10.12	Analysis of positive clones	85
<b>2.11</b>	<b>Southern Blots</b>	<b>85</b>
2.11.1	Transfer of nucleic acid to a solid support	85
2.11.2	Preparation of radioactively labelled probes	86
2.11.3	Hybridisation of the probe to the DNA blot	86
2.11.4	Washing conditions and exposure to X-ray film	87
<b>2.12</b>	<b>Colony hybridization</b>	<b>87</b>
<b>2.13</b>	<b>Pulsed Field Gradient Gel Electrophoresis (PFGE)</b>	<b>88</b>
<b>2.14</b>	<b>Dot Blots</b>	<b>89</b>
	<b><u>CHAPTER THREE: CONSTRUCTION OF A SIZE-FRACTIONATED SUB-GENOMIC LIBRARY</u></b>	<b>90</b>
<b>3.1</b>	<b>Construction and screening of the library</b>	<b>91</b>
<b>3.2</b>	<b>Characterisation of clone D1</b>	<b>93</b>

<b><u>CHAPTER FOUR:</u> CONSTRUCTION AND SCREENING OF A</b>	109
<b><i>Plasmodium yoelii</i> YM DNA GENOMIC LIBRARY</b>	
<b>4.1</b> Construction of a <i>Plasmodium yoelii</i> YM genomic library	110
<b>4.2</b> Screening the genomic library with an E8-specific probe to obtain additional 5' sequence	116
<b>4.3</b> Inverse PCR	126
<b>4.4</b> Analysis of E8-positive clones	140
<b>4.5</b> Screen of library with the 5' end of clone 8.1.6	147
4.5.1 Analysis of clone 5 and clone 7	147
<b><u>CHAPTER FIVE:</u> STRUCTURE OF THE Py235-E8 GENE</b>	161
<b>5.1</b> RT-PCR	162
<b>5.2</b> Analysis of the putative translation start site	169
<b>5.3</b> Analysis of the proposed signal sequence of Py235-E8	171
<b>5.4</b> 5' RACE	179
<b>5.5</b> Promoter motifs	190

<b><u>CHAPTER SIX:</u></b>	<b>SCREEN OF GENOMIC LIBRARY WITH D1</b>	<b>194</b>
<b>6.1</b>	Screen of genomic library with D1	195
<b>6.2</b>	Analysis of clone D1.2	195
<b>6.3</b>	Analysis of clone DX	201
<b><u>CHAPTER SEVEN:</u></b>	<b>CHROMOSOMAL LOCATIONS OF Py235</b>	<b>209</b>
	<b>GENES</b>	
<b>7.1</b>	Chromosome blots	210
<b><u>CHAPTER EIGHT:</u></b>	<b>GENERAL DISCUSSION</b>	<b>214</b>
<b>8.1</b>	Homologies within the Py235 multigene family	215
<b>8.2</b>	Homologies with other proteins	219
<b>8.3</b>	Future work	228
<b><u>CHAPTER NINE:</u></b>	<b>REFERENCES</b>	<b>230</b>
<b><u>APPENDIX ONE:</u></b>	<b>ENTIRE CODING SEQUENCE OF Py235.E8</b>	<b>251</b>
<b><u>APPENDIX TWO:</u></b>	<b>SEQUENCING CONTIG OF CLONES 8.1.6 &amp; clone 7</b>	



## INDEX OF FIGURES

1.4.1	The life cycle of <i>Plasmodium</i>	7
1.5.1	Ultrastructure of the merozoite	10
1.6.1	Diagram showing the different stages of erythrocytic invasion by a malarial merozoite	17
		46
1.9.1	The upstream regions of <i>Plasmodium</i> genes	
2.5.1	Primer sequences and positions in the E8 sequence for PCR to generate a probe	61
2.6.1	Primer sequences and positions for inverse PCR experiment	66
2.7.1	Sequences and positions of E8-specific primers for 5'RACE	71
2.8.1	Position and sequences of primers for RT-PCR	74
3.1	Southern blot of restriction enzyme digested <i>Plasmodium yoelii</i> YM genomic DNA using the an E8 specific probe	97
3.2	Size fractions of <i>DpnII</i> digested <i>Plasmodium yoelii</i> YM genomic DNA probed with the first 360 bp of clone E8-END	99
3.3	DNA and deduced amino acid sequence of D1	101
3.4	Similarity of clone D1 with E8-END	103
3.5	Alignment of D1 and E8	105
3.6	Southern blot analysis	107
4.1.1	Partial <i>Sau3A</i> digests of <i>Plasmodium yoelii</i> YM genomic DNA	112
4.1.2	Construction of a genomic DNA library of <i>Plasmodium yoelii</i> YM from partially <i>Sau3A</i> digested, size-fractionated DNA	114
4.2.1	Dot blot analysis of clones isolated from the genomic library with an E8 specific probe	118
4.2.2	Sequences of primers used to screen E8-positive clones	120

4.2.3	PCR of positive clones using primers based on the 5'-end sequence of clone E8-END	122
4.2.4	DNA and deduced amino acid sequence of clone 8.2.8	124
4.3.1	Inverse PCR	129
4.3.2	Inverse PCR primers	131
4.3.3	Products of inverse PCR	133
4.3.4	Partial DNA sequence and deduced amino acid of <i>DpnII</i> inverse PCR product	135
4.3.5	Confirmation by PCR that the new sequence for E8 is correct	137
4.4.1	Restriction enzyme digest analysis of the insert sizes of clones 8.1.6, 8.2.1, and 8.2.3	141
4.4.2	DNA and deduced amino acid sequence of clone 8.1.6	143
4.5.1	Restriction enzyme digest analysis of the insert sizes of clones 5 and 7	149
4.5.2	DNA and deduced amino acid sequence of clone 7	151
4.5.3	DNA and deduced amino acid sequence of clone 5	154
4.5.4	Alignment of clone 5 with E8	157
4.5.5	Alignment of the introns of clone 5 and E8	159
5.1.1	RT-PCR of <i>P. yoelii</i> YM RNA	165
5.1.2	DNA and deduced amino acid sequence of the RNA derived RT-PCR product	167
5.3.1	Analysis of the putative signal sequence of the Py235-E8 gene	177
5.4.1	5' RACE	182
5.4.2	Primers used in 5' RACE	184
5.4.3	5' RACE product from <i>P. yoelii</i> YM RNA using E8 specific primers	186
5.4.4	DNA and deduced amino acid sequence of 5' RACE product	188
5.5.1	Identification of GC-rich elements in the non-coding region of clone E8.7	192

6.2.1	DNA and deduced amino acid sequence of clone D1.2	197
6.2.2	Alignment of clone D1.2 with Py235.E8	199
6.3.1	DNA and deduced amino acid sequence of clone DX	202
6.3.2	Alignment of clone DX with Py235.E8	205
6.3.3	Alignment of clone DX with clone D1.2	207
7.1	Chromosome blot	212
8.1.1	Alignment of overlapping regions of E8, E3, D1, DX, and clone5	217
8.2.1	Homology of Py235.E8 with PvRBP-1	222
8.2.2	Predicted secondary structure of Py235.E8	224
8.2.3	Homology of Py235.E8 with the <i>S. cerevisiae</i> integrin homologue	226

**CHAPTER ONE**

**INTRODUCTION**

## 1.1 Malaria - a historical perspective

Malaria is an infectious disease caused by parasitic protozoa of the genus *Plasmodium*. The impact of the disease on the human race cannot be overestimated. For example, summer epidemics of malaria in Europe played a significant role in shaping the history of the continent. In 413 BC, epidemics amongst soldiers of the Athenian army led to their defeat at the hands of the Spartans. This resulted in the demise of Classical Greece and the rise of Rome to become the dominant power in the region.

The disease has also resulted in genotype selection amongst a susceptible population. For example, the mutation in the gene encoding haemoglobin which results in the disease sickle cell anaemia has been positively selected due to the fact that blood cells which contain the mutated haemoglobin protein do not support development of the *Plasmodium* parasite.

The disease appears to have spread geographically following the migration of man, with its wide distribution being largely attributable to the adaptability and enormous breeding potential of its insect host, the *Anopheles* mosquito. Arguably, the most vital discovery in the history of malaria research was that of the British physician Ronald Ross. In 1897 he confirmed the theory that the malaria parasite was transmitted by mosquitoes, allowing the fight against malaria to continue on an additional front, with the eradication of mosquitoes and their breeding grounds being used as an effective method of disease control. Draining of the Italian marshes to rid the mosquito of its breeding habitat has led to the reduction and eventual absence of malaria in that country.

## 1.2 Malaria in the present day

Malaria occurs throughout the tropical world and ranks as one of the most prevalent amongst all infectious diseases. Four parasite species which cause disease in humans have been identified, *Plasmodium falciparum*, *P. vivax*, *P. malariae*, and *P. ovale*. The distribution of the species varies from one endemic area to another, as do the clinical symptoms arising from infection.

In India and Sri Lanka malaria infections are caused by both *P. falciparum* and *P. vivax*, but severe and complicated disease is relatively rare in all age groups. By contrast, in tropical Africa infection is almost universal in all age groups, with severe disease mainly occurring between six months and five years of age. Severe and complicated disease is caused almost exclusively by infection with *P. falciparum*, the symptoms of which can vary depending on the individual infection. Symptoms can include anaemia, acidosis, renal failure, convulsions, and coma. Some of the symptoms occur more frequently in infections in a particular geographical area and with a particular age group. For example, organ failure is a relatively common outcome of severe disease in adults from South America and Asia, but is seen much less often in African children (Mendis and Carter, 1995).

### 1.3 Hypotheses of pathogenesis

In attempting to explain the pathology of severe disease, two main hypotheses exist which should not be considered mutually exclusive. In the first, the observed pathology is thought to be due to parasitised erythrocyte adherence to the microvascular endothelium of the brain and other organs via interactions between proteins on the surfaces of the two cells. Vascular cell adhesion molecule-1 (VCAM-1), intracellular adhesion molecule-1 (ICAM-1), endothelial adhesion molecule-1 (ELAM-1, also known as E-selectin), and CD36 on the surface of the endothelial cell have all been implicated in the pathogenesis of *P. falciparum* cerebral malaria (Berendt et al., 1993; Roberts, 1993). The main candidates for the parasite ligands which mediate the interaction are *P. falciparum* infected erythrocyte membrane protein-1 (PfEMP-1) (Howard and Pasloske, 1993; Newbold and Marsh, 1990). This protein exhibits a high degree of clonal variation (Roberts, 1993) and has been shown to be encoded by a large gene family (Baruch et al., 1995; Smith et al., 1995; Su et al., 1995). The gene family will be discussed in further detail in section 1.9.1.

The second hypothesis concerns the induction of host cytokines during the course of infection, in particular tumour necrosis factor (TNF). The pathology observed in the brain in cases of cerebral malaria has been postulated to be caused by secondary mediators of TNF action such as nitric oxide and oxygen free radicals.

Several studies have recorded a correlation between TNF levels in the plasma of *P. falciparum* infected individuals and the severity of disease (Grau, 1989). TNF- $\alpha$  has been demonstrated to be a pyrogen in *P. falciparum* infections, as treatment with an anti-TNF- $\alpha$  monoclonal

## Chapter One

antibody led to a decrease in body temperature (Kwiatkowski, 1993). It should be noted that other cytokines, notably interleukin-1 (IL-1) and lymphotoxin, can produce the same effects as TNF- $\alpha$  (Rockett, 1994) and their involvement in malaria pathogenesis is being investigated.



## 1.4 The life cycle of *Plasmodium*

The *Plasmodium* parasite requires two hosts to complete its life cycle - an insect host (the *Anopheles* mosquito), and a vertebrate host. The parasite is transferred from one host to the other by the bite of the mosquito when it takes an infected blood meal from the parasite's vertebrate host, or when an infected mosquito injects sporozoites into the vertebrate (Figure 1.4.1).

The erythrocytic asexual life cycle is initiated when merozoites are released from infected hepatocytes. These merozoites quickly make their way into the bloodstream of the vertebrate host and invade red blood cells almost immediately. The exposure of the merozoite to the host's plasma is therefore transient. The asexual blood stage of the life cycle is responsible for the majority, if not all, of the clinical symptoms which present in cases of severe malaria.

It has long been realised that the invasion of red blood cells by merozoites and the subsequent rupture of the red blood cell to release further merozoites is a crucial sequence of events in the life cycle of all species of *Plasmodium*. If the invasion of red blood cells could be prevented or interrupted by some means, the asexual cycle of the parasite would be broken and the infection effectively terminated. A detailed understanding of the merozoite and its component proteins is therefore crucial if this aim is to be achieved.

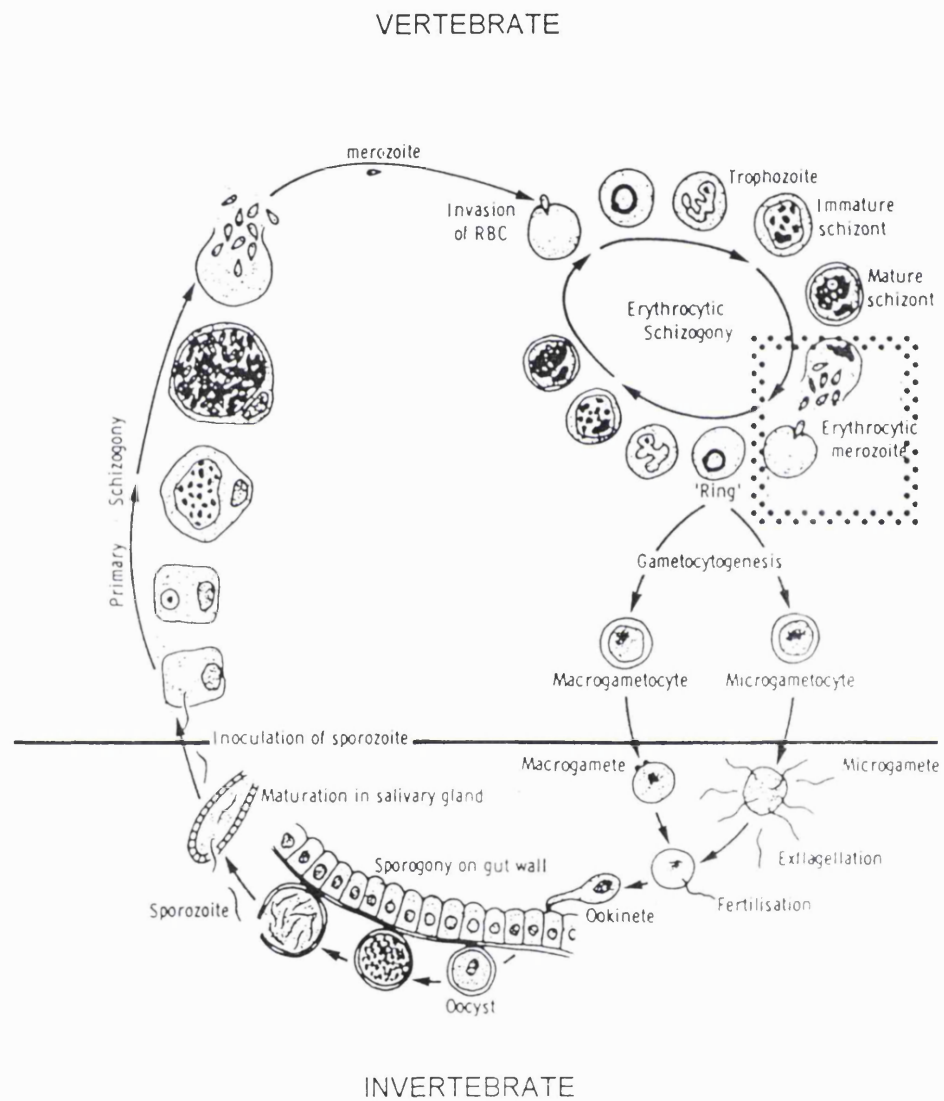
**Figure 1.4.1** The life cycle of *Plasmodium*

An infection with *Plasmodium* is initiated when a female mosquito of the genus *Anopheles* takes a blood meal from a vertebrate. The sporozoite stage of the parasite is present in the salivary glands of an infected mosquito and is transferred into the bloodstream of the vertebrate host upon feeding by *Anopheles*. Sporozoites can exist extracellularly in the bloodstream for a maximum period of one hour, and therefore quickly invade parenchymal cells of the vertebrate's liver, where they develop into multinucleate tissue schizonts. In the course of up to two weeks, each tissue schizont divides to form at least ten thousand merozoites. The intrahepatic stage of the *Plasmodium* life cycle ends when the infected hepatocyte ruptures to release the merozoites.

Merozoites invade erythrocytes in the bloodstream. Upon invasion, the parasite undergoes morphological changes as it progresses through the ring and trophozoite stages. The trophozoite stage develops into a multinucleate schizont which gives rise to merozoites which are released when the erythrocyte ruptures.

These merozoites invade further erythrocytes, to perpetuate the asexual erythrocytic cycle of the *Plasmodium* life cycle.

After invasion, some merozoites develop into male or female gametocytes which, when ingested by an *Anopheles* mosquito, develop into gametes in the mosquito midgut. The sexual stage of the parasite's life cycle occurs in the mosquito's gut where haploid gametes unite to form a diploid unicellular zygote. The zygote develops into an ookinete which penetrates the intestinal wall of the mosquito and develops into an oocyst. The oocyst undergoes sporogony, which results in the release of ten to twenty thousand sporozoites. The sporozoites migrate to the salivary glands of the mosquito where they are transferred to the vertebrate host upon feeding by the mosquito.



Source: Critical reviews in oncology/hematology 8(4) 255-311; G.H. Mitchell, and L.H. Bannister.

(1988)

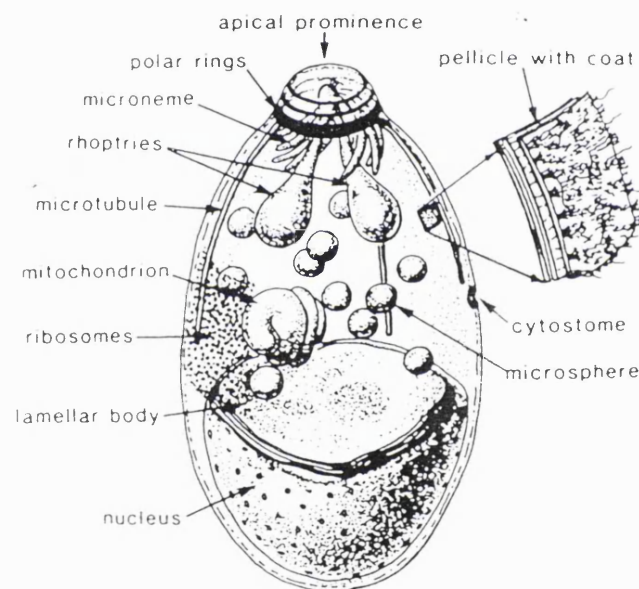
## 1.5 The merozoite

Merozoites are released from the hepatocyte, and on invasion of red blood cells are immediately transformed into the ring stage of the parasite. The merozoite is oval in shape and measures approximately  $1\mu\text{m}$  in diameter and  $1.5\mu\text{m}$  in length. It contains many organelles in common with most eukaryotic cells, such as a nucleus, mitochondrion and ribosomes. In addition, it contains rhoptry organelles, micronemes, and dense granules at its apical end. These are specialised organelles which are thought to play a vital role in invasion of red blood cells by merozoites. These apical organelles are present in the invasive stages of other *Apicomplexa* parasites, such as tachyzoites of *Toxoplasma* (Nichols et al., 1983). The rhoptry organelles of *Plasmodium* are club-shaped and approximately 300 nm in diameter (Etzion et al., 1991). The rhoptry organelles are synthesised *de novo* in each cycle of intraerythrocytic development of the parasite, but most rhoptry proteins are synthesised at least 12-16 hours before they are compartmentalised in the organelle (Etzion et al., 1991; Jaikaria et al., 1993). An organelle which is unique to *Apicomplexan* parasites is a plastid-like organelle whose function is currently unclear. It contains nucleic acid which appears to be related to that possessed by green algae, suggesting a possible endosymbiotic origin for this organelle (Wilson et al., 1996). The merozoite possesses an electron dense surface coat which is lost upon invasion of a red blood cell. Microtubules beneath the plasma membrane of the merozoite are thought to be responsible for maintaining cell shape, but they may also be involved in providing the propulsive forces required for invasion. An ultrastructural view of the merozoite is shown in figure 1.5.1.

**Figure 1.5** Ultrastructure of the merozoite

The *Plasmodium* merozoite is oval and approximately 1µm in diameter. It is surrounded by a double lipid bilayer and a thick surface coat. Beneath the inner plasma membrane is an array of microtubules, possibly functioning as a cytoskeleton. The merozoite has mitochondria, a nucleus, and specialised apical organelles called rhoptries and micronemes.

Part of the cell surface has been enlarged to show the the coat and pellicular membranes.



Source: Critical reviews in oncology/hematology 8(4) 255-311; G.H. Mitchell, and L.H. Bannister.

(1988)

## 1.6 Invasion of red blood cells by merozoites

The dynamics of invasion of erythrocytes by *Plasmodium* merozoites was first studied using an electro-optical system to observe the interactions of the two cells (Dvorak et al., 1975). From the analysis of these images, a theoretical model of invasion could be formulated. This suggested that free merozoites come into contact with erythrocytes by chance, and correctly oriented contact leads to the attachment of the merozoite to the erythrocyte surface by interaction of specific parasite receptors on the apical end of the merozoite and erythrocyte membrane ligands. Deformation of the erythrocyte membrane occurs and the merozoite enters the erythrocyte by producing a localised invagination of the erythrocyte membrane. Following invasion, the erythrocyte membrane reseals.

This hypothesis was proposed in 1975 and is still generally accepted, although more recent discoveries have enabled more details to be added with regard to the molecular mechanisms involved at each step of the invasion process.

### 1.6.1 Initial attachment of the merozoite to the red blood cell

The merozoite initially attaches to the surface of the red blood cell via fibrils on its surface coat. This can occur with the merozoite in any orientation. The merozoite then reorientates so that its apical end is against the red blood cell surface. Entry of the merozoite into the red blood cell can only occur once this reorientation has occurred. Identification of the molecules on the merozoite surface responsible for the initial attachment has focused on polypeptides which are uniformly distributed on the merozoite surface. Most of the polypeptides are derived from

merozoite surface protein-1 (MSP-1) (Holder et al., 1987; McBride and Heidrich, 1987) although other proteins have been identified, for example merozoite surface protein-2 (MSP-2) (Clark et al., 1989), MSP-3 and MSP-4.

MSP-1 in *P. falciparum* has a molecular mass of approximately 200 kDa, and has been demonstrated to bind to red blood cells (Perkins and Rocco, 1988). The protein has been identified in all species of *Plasmodium* examined. The large precursor molecule is processed by proteolytic cleavage on the merozoite surface to form a complex of four non-covalently linked polypeptides (Holder et al., 1987; McBride and Heidrich, 1987). This processing is referred to as primary processing. Secondary processing of the 42 kDa component of the MSP-1 complex into 33 kDa and 19 kDa fragments occurs on the merozoite surface. After this processing occurs, the entire complex is shed from the merozoite, excepting the 19 kDa polypeptide, which remains attached to the merozoite surface via a glycosylphosphatidylinositol (GPI) anchor to the plasma membrane (Blackman et al., 1991).

It has been suggested that secondary processing of MSP-1 is an absolute requirement for invasion of red blood cells by merozoites. The 19 kDa polypeptide of MSP-1 contains two epidermal growth factor (EGF)-like modules which may be exposed upon processing, allowing interaction with red blood cell membrane ligands. It has been shown that an antibody which recognises the first EGF-module can block the binding of MSP-1 to red blood cells (Chappel and Holder, 1993) suggesting a role for the EGF-modules in red cell binding. Further evidence for this came from experiments in which mice were immunised with recombinant protein containing the two EGF-modules of *P. yoelii* MSP-1. These mice were significantly protected against subsequent challenge with *P. yoelii* parasites (Daly and Long, 1993; Ling et al., 1994) again emphasising the importance of



the EGF-modules of MSP-1 in the process of invasion of red blood cells by merozoites. However, monoclonal antibodies that block invasion appear to do so by preventing the secondary processing of MSP-1 (Blackman et al., 1994), suggesting an alternative mode of action of invasion-inhibiting antibodies.

### 1.6.2 The role of the rhoptries and other apical organelles

After recognition of specific erythrocyte ligands by receptors on the surface of the merozoite, some kind of trigger is thought to cause the exocytosis of the contents of the rhoptry organelles situated at the apical end of the merozoite. The nature of such a trigger is unknown, but it is possible that an increase/decrease in the concentration of a signalling molecule such as cAMP, inositol triphosphate, or calcium ions within the merozoite may be involved. Release of the contents of the rhoptries and/or micronemes is thought to lead to the irreversible attachment of the merozoite to the erythrocyte membrane via a tight junction. Release of the rhoptry contents may also lead to localised displacement/dissociation/disaggregation of certain erythrocyte cytoskeleton components such as spectrin, band 4.1 and band 3. Such changes in the erythrocyte have been postulated to increase erythrocyte membrane deformability and allow invasion of the red blood cell by the merozoite. It has been shown that erythrocytes with mutations in spectrin and band 4.1 show differing sensitivities to invasion by *P. falciparum* merozoites (Facer, 1995).

How the contents of the rhoptries bring about this change in the erythrocyte membrane is unclear. It is possible that the proteins are inserted into the erythrocyte membrane and may enhance its fluidity. The rhoptries have also been shown to contain lipids. These may be integrated

into the erythrocyte membrane and could bring about changes in the nature of the membrane.

It has been suggested that lipid components of the rhoptry organelles may generate the parasitophorous vacuole membrane (PVM) (Bannister and Mitchell, 1995). It may be the case that the PVM is not formed by a simple invagination of the erythrocyte membrane. The newly forming PVM has a much lower density of integral membrane proteins than the adjacent erythrocyte membrane (Aikawa et al., 1981), supporting the hypothesis that lipid components of the rhoptry organelles are inserted into the erythrocyte membrane to generate the PVM.

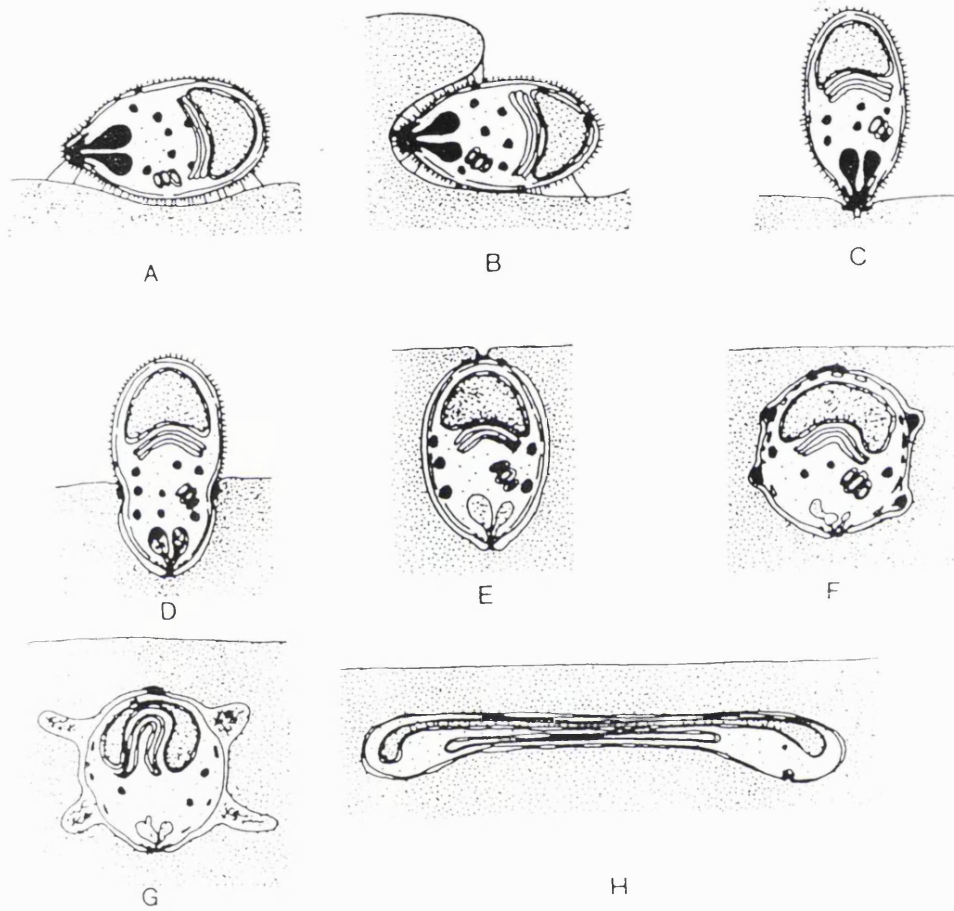
### 1.6.3 Internalisation of the merozoite

Internalisation of the merozoite occurs within a membrane bound vacuole. The surface coat of the merozoite is shed and remains outside the red blood cell. Elements in the cytoskeleton of the merozoite are thought to play a role in the internalisation process. Merozoites of *P. falciparum* have been shown to contain two actin genes (Wesseling et al., 1988). One of these is transcribed in the erythrocytic stage of the parasite life cycle and has a sequence which resembles muscle actin (Wesseling et al., 1988; Wesseling et al., 1989; Wesseling et al., 1988). Invasion of red blood cells by merozoites can be inhibited *in vitro* by treatment with cytochalasins B and D. These drugs act by inactivating actin-myosin based motility by binding to F-actin. This has been demonstrated for *P. knowlesi* (Ward et al., 1994) and *P. falciparum* (Dluzewski et al., 1992) merozoites. Another chemical which is known to block actin based motility is Staurosporine, an inhibitor of protein kinase. This drug has been demonstrated to inhibit invasion of red blood cells by *P. knowlesi* merozoites (Ward et al., 1994).

In the treatments with both cytochalasins and Staurosporine, merozoites still attach to red blood cells and discharge the contents of the rhoptry organelles, however there is no internalisation of the merozoite into the red blood cell. Therefore, components of the merozoite cytoskeleton, in particular actin, play an important role in providing the propulsive forces required in the internalisation process.

**Figure 1.6** Diagram showing the different stages of erythrocytic invasion by a malarial merozoite

- A: Initial attachment of the merozoite to the red blood cell surface
- B: Reorientation of the merozoite
- C: Contact of the apical end of the merozoite with the red blood cell surface
- D: Invagination of the invasion pit
- E: Closure of the parasitophorous vacuole
- F&G: Expansion of the parasitophorous vacuole membrane
- H: Final transformation into the ring stage



Source: Critical reviews in oncology/hematology 8(4) 255-311; G.H. Mitchell, and L.H. Bannister.  
(1988)

## 1.7 Apical organellar proteins and their role in merozoite invasion of erythrocytes

The micronemes and rhoptry organelles are situated at the apical end of the merozoite. Similar structures have been observed in all of the invasive stages of the parasite, for example the sporozoite. Several proteins have been localised to these organelles which have been proposed to play a role in the invasion process.

### 1.7.1 Microneme proteins

*P. knowlesi* in primates and *P. vivax* in humans both have stringent requirements in terms of the subset of red blood cells which their merozoites are able to invade. Both of these species of parasite have been demonstrated to invade red blood cells only if they possess the Duffy blood group antigen. Red cell-binding proteins with a molecular mass of 135 kDa have been identified in the culture supernatants of these parasites (Wertheimer and Barnwell, 1989). The proteins were found to bind only red blood cells which expressed the Duffy blood group antigen. The 135 kDa proteins were localised to the micronemes of the parasites (Adams et al., 1990). It has been suggested that these so-called Duffy binding proteins (DBP) play a role in determining host cell specificity of the parasite. *P. knowlesi* merozoites will bind to Duffy negative red blood cells but cannot invade them, as junction formation between the two cells does not occur. The release of these, and possibly other, proteins from the apical organelles of the merozoite upon initial binding and apical reorientation of the merozoite may act as a final component in the determination of red blood cell specificity for the parasite (Chitnis and Miller, 1994).

A 175 kDa protein, designated erythrocyte binding protein-175 (EBA-175), localises to the micronemes of *P. falciparum* merozoites and has been shown to have a similar overall structure to the Duffy binding proteins of *P. knowlesi* and *P. vivax*. This too has been identified as an erythrocyte binding protein, its ligand being the sialic acid-containing glycoprotein Glycophorin A (Adams et al., 1992).

### 1.7.2 Rhoptry proteins

#### 1.7.2.1 Rhoptry associated proteins 1 and 2 (RAP-1 and 2)

RAP-1 and RAP-2 are found as a complex of at least three major proteins (p80, p65 and p40) in the rhoptries of *Plasmodium falciparum* merozoites. Immunisation of *Saimiri* monkeys with the purified complex led to the protection of these animals against a lethal infection of *P. falciparum* (Ridley et al., 1990). Further investigations showed that the 80 kDa and 65 kDa proteins of the complex were the products of a single gene, *rap1*, with the 65 kDa protein being a cleavage product of the larger 80 kDa protein (Ridley et al., 1991; Ridley et al., 1990). RAP-1 is in fact synthesised as an 84 kDa precursor protein which is rapidly processed to the 80 kDa protein. Analysis of the putative amino acid sequence of the *rap1* gene demonstrates the presence of a putative N-terminal signal peptide, suggesting that the protein is transported from the endoplasmic reticulum to the rhoptry organelles. There are also stretches of amino acids which could give rise to the formation of amphiphilic helices (Ridley et al., 1990). Such structures occur in many proteins which interact with membranes, for example  $\delta$ -haemolysin (Eisenberg et al., 1984). An interesting

hypothesis suggested by such a finding is that, on release from the rhoptries, RAP-1 could penetrate and disrupt the erythrocyte membrane to allow invasion of the merozoite.

A possible dual role for RAP-1 has been suggested. Cleavage of a fraction of the 80 kDa protein into the 65 kDa product occurs in late schizogony, just prior to merozoite release from the infected red blood cell. Monoclonal antibodies directed against epitopes adjacent to the p80-p65 cleavage site have been demonstrated to inhibit merozoite release *in vitro* (Harnyuttanakorn et al., 1992). The 65 kDa product of RAP-1 is entirely absent from merozoites released naturally from intraerythrocytic schizonts, is seen only in late schizogony and not in ring stages (Bushell et al., 1988; Clark et al., 1987; Howard et al., 1984). It has been proposed that the fraction of the 80 kDa RAP-1 which is processed into the 65 kDa protein has a role in merozoite release from infected red blood cells, while the unprocessed 80 kDa protein has a role in the invasion of red blood cells by free merozoites (Harnyuttanakorn et al., 1992).

RAP-1 may be non-covalently associated with a number of proteins including RAP-2, a 76 kDa serine protease (Braun-Breton and Silva, 1988) and a family of metalloproteases (Locher et al., 1992).

RAP-2 is a 42 kDa protein located in the rhoptry organelles. Like RAP-1, it contains a putative signal sequence at its N-terminus. It has hydrophobic domains which may be responsible for its association into the complex with RAP-1 and in the lipid-rich environment of the rhoptries (Ridley et al., 1991; Saul et al., 1992). Little is known about its function in invasion of red blood cells or merozoite release.



## 1.7.2.2 RhopH1, 2, and 3

The high molecular weight complex (HMWC) of *P. falciparum* rhoptries contains three distinct polypeptides in equimolar amounts. These are RhopH1 (150 kDa), RhopH2 (135-140 kDa) and RhopH3 (105-110 kDa). Synthesis of the components of the complex occurs at the late trophozoite/schizont stage of the parasite life cycle, and the HMWC is formed within the cell long before organelle formation in the segmented schizont occurs. The HMWC is released from the rhoptries of the merozoite at the time of invasion of red blood cells (Cooper et al., 1988; Coppel et al., 1987; Holder et al., 1985; Lustigman et al., 1988; Sam-Yellowe et al., 1988). All three proteins of the HMWC have been demonstrated to bind to erythrocyte membranes and preferentially to vesicles made of phosphatidylinositol (PI) and phosphatidylserine (PS). Both of these phospholipids are localised on the cytoplasmic face of the erythrocyte plasma membrane (Perkins and Ziefer, 1994). The authors have suggested that proteins of the HMWC may disrupt interactions between the erythrocyte plasma membrane and transmembrane proteins, by translocation of the rhoptry proteins across the lipid bilayer during merozoite invasion (Sam-Yellowe and Perkins, 1991).

The gene which encodes RhopH3 has been cloned and shown to possess seven exons (Brown and Coppel, 1991). Such a complex gene structure is almost unprecedented in *Plasmodium*, although it is not uncommon for a gene to possess more than one intron. It is not known whether the exons represent functional domains of the protein, although this is not thought to be the case. The N-terminus of the protein contains a signal sequence which has been shown to be cleaved between positions 24 and 25 of the amino acid sequence of the protein (Cooper et al., 1989). The protein

contains thirteen cysteine residues which are all situated in the N-terminal half of the protein and appear to form intramolecular disulphide bonds (Brown and Coppel, 1991).

### 1.7.2.3 Apical membrane antigen-1

Apical membrane antigen-1 (AMA-1) is a rhoptry protein of unknown function. Analysis of the amino acid sequence shows that it contains two very hydrophobic regions, one at the N-terminus of the protein which is proposed to act as a signal sequence, and the second near to the C-terminus which displays a sequence characteristic of a membrane-spanning domain (Peterson et al., 1989). Synthesis of the protein begins in mature trophozoites and reaches a peak in segmenting schizonts. The 80 kDa molecule is immediately processed to a 62 kDa form. Immunoelectron microscopy studies have shown that AMA-1 is localised in the neck of the rhoptry organelles of mature merozoites within segmenting schizonts and in free merozoites, and subsequently appears to be associated with the merozoite surface (Crewther et al., 1990). It is not carried into the host cell on invasion. The redistribution of AMA-1 onto the merozoite surface has been suggested to occur by release of the protein from the rhoptries at the time of schizont rupture and integration into the merozoite membrane (Peterson et al., 1989). The fact that the antigen is released from the rhoptries prior to merozoite release from infected red blood cells has led to suggestions that AMA-1 may play a role in this process (Crewther et al., 1990).

#### 1.7.2.4 The 225 kDa rhoptry protein of *Plasmodium falciparum*

The 225 kDa protein of *P. falciparum* was identified by Roger *et al.* A monoclonal antibody which was raised against *P. falciparum* culture supernatants detected both a 225 and a 240 kDa protein on Western blots, with subsequent experiments showing that the 225 kDa protein is a processed form of the 240 kDa precursor. The same antibody gave a punctate pattern when used in immunofluorescence experiments on *P. falciparum* erythrocytic stages. This pattern is indicative of the protein being located in the rhoptry organelles of the merozoite. Immunoelectron microscopy demonstrated that the protein was detected only in a particular region of the rhoptry organelles. This could mean that the protein is located only in this particular region, and suggests different functions for different parts of the rhoptry organelles, depending upon the proteins located there. This result underlines the fact that the rhoptry organelles do not contain a homogeneous mixture of components, but suggests a heterogeneous distribution of rhoptry contents (Roger *et al.*, 1988). The function of the 225 kDa protein is not known.

#### 1.7.2.5 The 235 kDa rhoptry proteins of *Plasmodium yoelii*

The 235 kDa rhoptry proteins of *Plasmodium yoelii* were identified in 1981 (Holder and Freeman, 1981). Screening of a panel of monoclonal antibodies which had been raised against *P. yoelii* merozoites showed that several of them reacted with the rhoptry organelles, as indirect immunofluorescence assays with the antibodies on fixed merozoites gave a double dot pattern at the apical end of the cell, a pattern characteristic of the rhoptries. High resolution immunoelectron microscopy confirmed that the protein is

indeed localised in the rhoptry organelles (Oka et al., 1984). The monoclonal antibodies were demonstrated to immunoprecipitate a protein with an apparent molecular mass of 235 kDa from parasite cultures (Holder and Freeman, 1981). *Plasmodium yoelii* clones are lethal or non-lethal. The YM line of the parasite will normally infect all types of red blood cells, leads to a high parasitaemia in infected animals and is usually lethal. The 17X line of the parasite, by contrast, is restricted to reticulocytes and does not usually cause death (Yoeli et al., 1975). Passive transfer of two of the monoclonal antibodies into mice, followed by challenge with the YM strain of parasite, led to the infection mimicking that of an avirulent strain. Parasitaemia was shown to be restricted to reticulocytes and the animals survived the infection. The 235 kDa protein was purified and used to immunise mice, which were subsequently challenged with YM strain parasites. Again parasites were restricted to reticulocytes, with no observable infection in mature red blood cells (Freeman et al., 1980; Holder and Freeman, 1981). These observations suggest a role for the protein in determining the specificity of red blood cell invasion by the merozoite.

Experiments to investigate the fate of the 235 kDa protein within the merozoite, from translation in the cytoplasm to its appearance in the rhoptry organelles have been undertaken (Ogun and Holder, 1994). Brefeldin A is a fungal metabolite which has been shown to cause the disassembly of the Golgi apparatus in mammalian cells (Klausner et al., 1992). Treatment of cells with Brefeldin A thus prevents post-endoplasmic reticulum processing events. Brefeldin A has also been demonstrated to block protein export in *Plasmodium falciparum* (Das et al., 1994). The treatment of *P. yoelii* -parasitised erythrocytes with Brefeldin A resulted in a noticeable decrease in mobility of the 235 kDa rhoptry protein on polyacrylamide gels suggesting that the protein undergoes some form of processing, probably proteolytic cleavage, and this processing is a post-

endoplasmic reticulum event (Ogun and Holder, 1994). Other rhoptry proteins in *Plasmodium* are known to be processed after synthesis of a larger precursor molecule. RAP-1 in *P. falciparum* is processed from an 84 kDa precursor to an 80 kDa protein found in the rhoptries (Bushell et al., 1988) and a 225 kDa *P. falciparum* rhoptry protein has been reported to be derived from a 240 kDa precursor (Roger et al., 1988). These processing events have been demonstrated to be reversibly blocked by treatment of the parasites with Brefeldin A (Howard and Schmidt, 1995).

It has been demonstrated that the 235 kDa protein binds to the surface of red blood cells (Ogun and Holder, 1996) but the ligand on the surface of the red blood cell has yet to be identified.

DNA clones encoding a small part of the C-terminus of the protein have been identified (Keen et al., 1990). The clones exhibited a high degree of conservation of sequence, but were not identical. This suggested that there was more than one gene in the *P. yoelii* genome which encoded the 235 kDa rhoptry protein, or highly related proteins. Further work has suggested that there is a multigene family of at least eleven individual genes (Borre et al., 1995) on at least six of the parasite's chromosomes and situated at nine individual loci (C.A. Owen, unpublished). The significance and possible implications of such a gene family are discussed in detail in section 1.9.2.

The isolation of two large genomic clones, designated E3 and E8, and the deduced amino acid sequences of these, has enabled a putative structure of the polypeptide to be suggested (Keen et al., 1994). The amino acid sequence has indicated that the 235 kDa protein may adopt the structure of a classical type I membrane protein, with a large hydrophilic ectodomain, a hydrophobic membrane-spanning region, and a short hydrophilic tail which extends into the cytoplasm of the merozoite.

The 235 kDa protein first observed by Holder and Freeman has more recently been resolved on low percentage polyacrylamide gels into a group of polypeptides. Different polypeptides in the group were observed to have differing reactivities with monoclonal antibody reagents (Ogun and Holder, 1994). Whether this is because they represent products of different members of the multigene family which may contain epitopes of differing reactivity, or are modified products of a single precursor has not been determined.

There is relatively convincing evidence to suggest the existence of proteins analogous to Py235 in distantly related species of *Plasmodium*. *P. vivax* reticulocyte binding protein-2 (PvRBP-2) is a high molecular mass protein found in *P. vivax* merozoites (Galinski et al., 1992). Analysis of the deduced amino acid sequence of the partial gene sequence of PvRBP-2 has shown that there is a region of 500 amino acids which shows significant homology to the sequence of Py235. This region shows an identity of 29.6% and a similarity of 51.6% (Keen et al., 1994). Further analysis of the two sequences has suggested a structural homology. Both proteins possess a putative transmembrane region preceded by a set of short amino acid repeats (Galinski and Barnwell, 1996). From this analysis, it would appear that these proteins from highly diverged *Plasmodium* species belong to the same family. The proteins are structurally and functionally similar, both being proposed to play a role in determining the specificity of red blood cell invasion for their respective parasite merozoites (Galinski and Barnwell, 1996; Galinski et al., 1992; Holder and Freeman, 1981).

### 1.7.3 Reticulocyte binding proteins

In *P. vivax* the Duffy binding protein (DBP) forms part of the system which determines host cell specificity. The Duffy blood group antigen is present on mature as well as immature red blood cells, so the interaction of the parasite DBP with this red cell ligand does not explain *P. vivax* merozoites' preferential invasion of reticulocytes.

Two *P. vivax* proteins which bind to reticulocytes have been identified (Galinski et al., 1992). The binding of the so-called reticulocyte binding proteins (RBPs) to the reticulocytes of other primate species correlates with the susceptibility of these species to *P. vivax* infection. The PvRBPs bind to *Saimiri* monkey reticulocytes, which are invaded by *P. vivax* merozoites, whereas they do not bind to rhesus monkey reticulocytes, which are not (Galinski et al., 1992). They are relatively large proteins, with molecular masses of 275-280 kDa, and are predicted to be largely  $\alpha$ -helical in their secondary structure. The localisation of the proteins has not been determined. By indirect immunofluorescence assays both reticulocyte binding proteins 1 and 2 (PvRBP-1 and 2) give a fluorescence pattern of a cap or crescent which appears to cover the apical pole of the merozoite (Galinski and Barnwell, 1996)

The entire gene encoding PvRBP-1 has been sequenced (Galinski et al., 1992). In common with PvRBP-2 it also possesses a putative transmembrane region and a short cytoplasmic tail. Interestingly, the protein also has two RGD-motifs, which have been shown to be integrin-binding sites for a number of proteins (Galinski et al., 1992; Hynes, 1987). Whether these motifs are adhesive domains for PvRBP-1 has yet to be determined. It is thought that the reticulocyte binding proteins form a multimeric complex on the surface of the merozoite. RBP-1 contains sixteen cysteine residues clustered in four regions, which could be

responsible for the formation of intra- and intermolecular disulphide bonds. It has been proposed that PvRBP-1 is a homodimeric membrane bound protein and the monomers associate via interchain disulphide bonds. PvRBP-2 is thought to associate with PvRBP-1 through non-covalent interactions between the two molecules (Galinski and Barnwell, 1996)

It has been proposed that binding of *P. vivax* merozoites to reticulocytes via the RBP complex of proteins leads to the release of the Duffy binding protein (DBP) from the micronemes of the merozoite, and therefore commits the parasite to invasion of the reticulocyte (Galinski and Barnwell, 1996).



## 1.8 Multigene families in *Plasmodium*

Several multigene families have been identified in species of *Plasmodium*. The role of such families is becoming clearer, with roles in antigenic variation of the parasite and the generation of receptor heterogeneity.

### 1.8.1 The *var* family of genes in *Plasmodium falciparum*

There is considerable evidence that malaria parasites exhibit antigenic variation, and that such parasite variability is a major factor in the survival strategy of the parasite. Antigenic variation in the strictest sense is defined as the regular replacement of the antigens which are exposed to the host immune system. Antigenic variation at the surface of the parasitised erythrocyte is associated with prolonged and chronic infection, and also the capacity to re-establish infection in previously infected animals (Baruch et al., 1995).

The proteins which are involved in antigenic variation of *P. falciparum* are of diverse molecular masses, ranging from 200 to 350 kDa. They are expressed, not on the parasite surface, but are transported from the intra-erythrocytic parasite to the surface of the infected erythrocyte itself, hence they are called *P. falciparum*-infected erythrocyte membrane protein-1 (PfEMP-1). Interactions between PfEMP-1 and various ligands on the surface of endothelial cells are responsible for the adhesion of parasitised erythrocytes to the walls of vessels in the brain and other organs. Such adhesion is responsible for many of the symptoms of severe and cerebral malaria.

The evidence that PfEMP-1 is responsible for antigenic variation in the parasite has been accumulating. Selection of parasites for changes in cytoadherence characteristics is accompanied by changes in PfEMP-1 expression (Biggs et al., 1992; Magowan et al., 1988). Switches in the antigenic type of a parasite were shown to be always accompanied by a change in molecular mass of the PfEMP-1 protein, and a change in the binding properties of the molecule (Roberts et al., 1992). Therefore, the antigenic diversity of PfEMP-1 is correlated with functional diversity. Different populations of infected red blood cells may adhere to a variety of endothelial cell receptors including VCAM-1, ICAM-1, CD36 and E-selectin. The structure of PfEMP-1 proteins which allow interaction of the parasite protein with those of the host cell would be expected to be equally diverse.

PfEMP-1 proteins are encoded by a large gene family named the *var* gene family. An upper estimate of 150 *var* genes per haploid genome has been made (Su et al., 1995). Evidence that *var* genes encode the diverse PfEMP-1 proteins has come from a variety of independent sources. Four *var* genes have been sequenced by Su et al. (1995). Analysis of the deduced amino acid sequence of these genes showed that they possess structures characteristic of cellular adhesion molecules, that is they have large and variable extracellular domains, a single transmembrane segment, and a conserved intracellular domain. Baruch *et al* (1995) expressed fusion proteins of parts of a *var* gene from the Malaysian Camp (MC) strain of *P. falciparum*. Antisera made against these fusion proteins recognised PfEMP-1 in the knobs of parasitised erythrocytes. Interestingly, these antibodies also blocked adherence of parasitised erythrocytes to the CD36 receptor, but had no effect on their binding to other receptors, for example thrombospondin. This suggests that MC strain parasites express multiple forms of PfEMP-1 with different cytoadherence properties. Further

evidence for the differential expression of *var* genes has come from studies into the switching of these genes (Smith et al., 1995). This work has demonstrated that the switching rate of *var* genes is as high as 2.4% per generation, and that expression of distinct PfEMP-1 antigens on the surface of infected erythrocytes correlates with expression of distinct *var* genes. The authors have also demonstrated that such differential expression of *var* genes is the basis of antigenic variation in *Plasmodium falciparum*.

The mechanism by which the parasite effects switching in expression of its array of *var* genes is not entirely clear. From studies in other organisms, there are three basic mechanisms which are used for switching in large families of surface antigens (Borst et al., 1995). The first is at the pretranscriptional level. In this mechanism, the antigen gene which is transcribed is physically replaced by another. This can occur either by reciprocal recombination or gene conversion. The promoter of the gene remains unaltered. Such events occur in the switching of expression of the variable surface glycoprotein (VSG) gene family of African trypanosomes, with previously silent VSG genes being translocated into active expression sites which are situated near the telomeres of the organism's chromosomes.

A second method for switching, which again is utilised by trypanosomes, is at the level of transcription. The promoter of a silent gene is activated and the promoter of the previously active gene is switched off (Borst and Rudenko, 1994). The precise mechanisms by which such events occur are yet to be elucidated.

A final mechanism is at the posttranscriptional level. The *opa* genes of *Neisseria* contain pentanucleotide CTCTT repeats. Depending on the number of repeats a given transcript contains, the translation products will either be full length protein or truncated N-terminal fragments. It is

thought that variation in the number of repeats contained in a given transcript is a result of "slippage" of the RNA polymerase during transcription of this region of the gene (Borst et al., 1995; Stern and Meyer, 1987). Other posttranscriptional controls could also be feasible. It is possible that differential degradation of mRNAs of multiple genes could occur to allow expression of different gene products. Such a strategy would appear wasteful in the case of the *var* genes, where estimates of the copy number in the parasite genome have been put at 150, which would constitute 6% of the genome if all the genes represent full-length genes (Su et al., 1995). However, it is not known what percentage of the genes is functional in the sense of being capable of being expressed in the first place, or whether all *var* genes are capable of being expressed simultaneously.

The evidence for the mechanism of *var* gene expression switching in *Plasmodium* is limited. From what is known, it does not appear that expressed *var* genes are invariably located at telomeric expression sites like the VSG genes of *Trypanosoma brucei*. Nor are there any indications for the activation of silent *var* genes by duplicative transposition into an expression site. This leaves us with the transfer of *var* genes to expression sites by means of a non-duplicative mechanism, for example reciprocal recombination, or control of gene expression by transcriptional or post-transcriptional mechanisms.

Su et al (1995) noticed a broad 1.8-2.4 kb band on RNA blots which had been probed with exon II of a *var* gene. These products lack a consensus translation start site and some also aligned with the intron of a *var* gene, suggesting that these transcripts do not themselves encode protein products. The authors have speculated that the transcripts may participate in the expression or rearrangements of *var* genes. They have suggested

that this family of transcripts could be involved in *trans*-splicing events which have regulatory importance. It is also possible that the transcripts arise from the initiation of transcription from internal promoters, which in turn could affect transcription from regions upstream of the *var* gene, i.e. the promoter whose activity gives rise to full length transcripts of the *var* gene. The authors have also speculated that transcription of the small mRNAs may correlate with accessibility of *var* genes to DNA rearrangements. Such a scenario is not unprecedented and has been observed in other systems. Transcriptional activity stimulates recombination in certain yeast and mammalian genes (Nickoloff and Reynolds, 1990; Thomas and Rothstein, 1989) and is also thought to correlate with the rearrangement of immunoglobulin genes in lymphocytes (Alt et al., 1987; Okada and Alt, 1994).

It is clear that the expression and switching of the *var* gene family must be tightly controlled. While it is desirable for a repertoire of proteins to be expressed on the surface of parasitised red blood cells in order for the cells to adhere to endothelial cells and thus avoid destruction in the host's spleen, if each parasite in a single infection expressed a different *var* gene then antigenic variation would not work. The immune system of the host would experience the parasite's entire repertoire of surface antigens and the infection would be cleared, with no chance of a "successful" infection subsequent to this. It is more than likely that the order of expression of *var* genes is determined by a complex combination of both genetic and external factors, as is thought to be the case for the expression of VSG genes in trypanosomes (Borst et al., 1995)

Due to the diversity and high frequency of switching in the *var* gene family, it is unlikely that they will form the basis of a vaccine strategy

against malaria. The genes would only be useful if an antigenically conserved region of the PfEMP-1 proteins was found. It is possible, however, that the study of PfEMP-1 and the *var* genes could lead to the design of chemotherapeutic strategies to reverse the adhesion of parasitised erythrocytes to the endothelium of vessels in the brain to treat cerebral malaria.

### 1.8.2 The multigene family encoding the 235 kDa rhoptry proteins of *Plasmodium yoelii*

The 235 kDa rhoptry proteins of *Plasmodium yoelii* merozoites play a role in the invasion of erythrocytes by the parasite. As has been mentioned in section 1.7.2.4, immunisation of mice with purified Py235 protein, or passive transfer of certain monoclonal antibodies raised against the proteins, results in a subsequent *P. yoelii* YM infection being restricted to immature red blood cells, rather than all erythrocytes as one would expect. A cloned DNA fragment of a gene encoding a Py235 protein was isolated from a DNA expression library by screening the library with serum from mice hyperimmune to *P. yoelii*. This DNA fragment was demonstrated to hybridise to multiple fragments in Southern blots of *P. yoelii* genomic DNA (Keen et al., 1990). This result suggested that the sequence, or closely related ones, was represented more than once in the parasite genome.

The number of genes and their variability was examined by PCR amplification of a region which had been shown to be highly conserved in the two gene sequences previously identified (Borre et al., 1995; Keen et al., 1994). Amplification of the region followed by restriction fragment length polymorphism (RFLP) analysis of the products, showed that the genes fell into seven distinct categories, with distinct RFLP patterns for each group

(Borre et al., 1995). Further variation within the gene family was evident upon sequencing the PCR generated fragments. Of sixteen clones analysed, eleven different sequences were present (Borre et al., 1995). Further work on the location of the genes revealed that the family is distributed on six of the parasite's chromosomes, with at least nine loci. A true estimate of the number of genes encoding Py235 is not available, but the varying frequency of the RFLP patterns observed by Borre *et al* has suggested that there may be as many as fifty copies, if indeed the frequency of the clones is indicative of the number of genes in the genome.

Little is known about the expression of members of the gene family. Pertinent questions such as whether all of the genes are expressed or merely a subset, do individual parasites express only one of the Py235 genes, and whether the gene which is expressed affects the virulence of the parasite are being addressed. Preliminary data suggest that the repertoire of genes which are expressed is smaller than the repertoire in the genome. A recent publication has shown that the repertoire of Py235 genes which are expressed depends upon the parasite line (Dr. P. Preiser, Experimental Parasitology, in press). Using apparently conserved primers which flank a variable repetitive sequence near the 3' end of the genes, reverse transcriptase PCR was carried out on mRNA from a lethal (YM) and a non-lethal (17X) line of *P. yoelii*. They found that they could detect transcripts for the Py235 gene family in both lines, but the 17X line expressed a larger subset of genes than the YM line. It is interesting at this point to consider the origin of the YM line of *P. yoelii*. The virulent YM line arose spontaneously from an avirulent strain (Yoeli et al., 1975). Both the YM line and the avirulent 17X line of the parasite possess the multigene family encoding Py235, with evidence for an additional locus in the 17X line (Dr. C.A. Owen, unpublished observation). As immunisation of animals with

Py235 leads to a YM infection resembling that of the avirulent 17X line in being reticulocyte-restricted, it is not unreasonable to suggest a role for the Py235 protein in the invasion of mature red blood cells by YM line merozoites. However, when the repertoire of genes expressed by the YM line was investigated while the parasite was restricted to reticulocytes in the presence of anti-Py235 monoclonal antibodies, no difference in the pattern of transcription of the Py235 genes was observed (Dr. P. Preiser, *Experimental Parasitology*, in press). The primers used to amplify the Py235 genes came from a region that had been shown to be conserved in several genomic clones isolated previously. The assumption has been made that using these primers to screen the transcription pattern of parasite lines results in amplification of all the members of the Py235 genes. However, because of the specificity of PCR, it may be the case that only a subset of the family is being amplified.

These results raise many further questions about the role of Py235 proteins in red blood cell invasion. There may be multiple pathways for red blood cell invasion, only one of which involves Py235. It is also possible that different Py235 proteins recognise different ligands on the surface of the red blood cell, enabling invasion of different cell types. Receptor heterogeneity has been proposed as a possible explanation for the parasite possessing such a large gene family, however there is no evidence for this at present and it remains speculative.



## 1.9 The control of gene expression in *Plasmodium*

### 1.9.1 General features of promoters in eukaryotic genes

The large majority of genes which encode cellular proteins in eukaryotes are transcribed by RNA polymerase II. The elements which control the start site and frequency of transcription from any given gene are contained in a region known as the promoter. These elements are located upstream of the transcription start site.

Promoters of eukaryotic genes are modular. The elements which constitute the promoter can be classified according to their proximity to the transcription start site. There are three regions: the near region, the middle region, and the distal region. The regulatory sequences contained in these regions are very short, usually less than ten base pairs in length.

The near region of the vast majority of promoters contains a TATA box (also known as the Hogness box). The TATA box is located approximately thirty base pairs upstream of the transcription start site, and has been demonstrated to have a consensus sequence of

$$T_{82}A_{97}T_{93}A_{85}(A_{63}/T_{37})A_{83}(A_{50}/T_{33})$$

where the numbers in subscript represent the frequency of the given base in percentage (Breathnach and Chambon, 1981).

The TATA box is responsible for the choice of start point for transcription, by correctly positioning RNA polymerase II on the gene. If the TATA box of a gene is deleted, the transcription start site becomes erratic, however the frequency of transcription initiation is not affected. Conversely, if

sequences downstream of the TATA box are replaced, transcription of the gene still initiates approximately thirty base pairs downstream of the TATA box. Some promoters lack a TATA box, for example the human epidermal growth factor (EGF) receptor gene. In this gene, initiation of transcription does not occur at a unique start point but occurs at any of a cluster of start points (Ishii et al., 1985).

The CAAT box is a feature of the middle region of eukaryotic promoters, often situated approximately eighty base pairs upstream of the transcription start site. Comparison of the CAAT boxes of several genes has shown a recognisable consensus sequence, however the conservation is not as impressive as that seen for the TATA box. The consensus sequence is

**5'-GG(C/T)CAATCT-3'**

The CAAT box is conserved in several, but not all, known promoters. This motif can be in either orientation and still be functional.

GC-rich sequences are a common feature of eukaryotic promoters. A GC box with a consensus sequence of

**5'-GGGCGG-3'**

has been identified in the middle and distal regions of several gene promoters, and again these elements are functional in either orientation.

Elements of the distal and middle regions of promoters are thought to play a role in controlling the frequency of initiation of transcription. They are

probably involved in the initial binding of RNA polymerase II to the promoter.

Specific promoter sequences interact with protein transcription factors. The first sequence specific transcription factor to be identified was specificity protein-1 (sp-1), which has been demonstrated to bind to the GC-box sequence motif of promoters (Dyana and Tjian, 1985; Dyana and Tjian, 1983; Tjian, 1995).

The TATA box is bound by a transcription factor known as TATA box binding protein (TBP). Other factors which interact with TBP, known as TBP associated factors (TAFs) are thought to convey molecular signals from activators such as sp-1 to the basal transcription apparatus of which TBP is a component (Tjian, 1995). The TBP complex therefore acts as a way of integrating the regulatory signals from DNA-bound transcription factors. Although the elements which play a role in the control of transcription are separated by tens and sometimes hundreds of bases on the DNA molecule, it is thought that they interact by inducing looping of the DNA to allow them to come into close proximity.

The factors described above are required for gene activity. Other factors which regulate the transcription of specific genes have also been identified. The genes which encode heat-shock proteins are transcribed only in response to certain environmental changes. Eukaryotic heat shock genes contain a conserved sequence located approximately fifteen base pairs upstream of the transcription start site, and a heat shock transcription factor (HSTF) has been identified which binds specifically to this sequence. HSTF is only active in heat shocked cells, and activation of this factor leads to the transcription of the group of genes which contain the consensus sequence for HSTF binding in their genes (Topol et al., 1985).

Thus, eukaryotic gene promoters contain basic modules such as the TATA box and CAAT box which are involved in general interactions with the RNA polymerase II complex and influence the frequency of initiation and the start point of transcription. They also contain a series of specific sequences which allow RNA polymerase II to transcribe the gene under particular conditions.

### 1.9.2 Transcription and its control in *Plasmodium*

*Plasmodium* is not a particularly good organism for the study of molecular biology. The complete life cycle of the parasite is difficult to maintain, or impossible for some species. Classical genetic approaches are very difficult to apply to the parasite because of its complex life cycle, and techniques for the stable transfection of genes into the parasite are only in their infancy (Crabb and Cowman, 1996; van-der-Wel et al., 1997; Wu et al., 1996). However, despite these obstacles, it is vital to understand the molecular mechanisms involved in the control of gene expression as these will provide insights into the biology of the parasite. Of considerable importance in this respect is understanding how the parasite evades the immune system of the host by switching expression of specific surface antigens such as PfEMP-1 (Baruch et al., 1995; Borst et al., 1995; Smith et al., 1995; Su et al., 1995) and discovering the amount of natural genetic diversity possessed by the parasite. It is also possible that certain aspects of gene expression in *Plasmodium* parasites will lend themselves to targeting with chemotherapeutic agents.

### 1.9.3 Stage-specific expression of *Plasmodium* genes

Initial evidence for the controlled expression of parasite genes has come from observations that the patterns of biosynthetically labelled parasite proteins differ between the parasite stages (Deans et al., 1983; Falanga et al., 1982; Myler et al., 1983; Newbold et al., 1982). Further evidence of stage-specific RNA accumulation was obtained using Northern blot analysis, nuclear run-on analysis and the study of promoter activity (Lanzer et al., 1992a; Lanzer et al., 1992b).

How is this developmentally regulated gene expression achieved by the parasite? Observations have indicated that regulation of the genes can occur transcriptionally or post-transcriptionally (Waters et al., 1989). Expression of merozoite surface protein-1 (MSP-1) of *P. falciparum* appears to be regulated in a post-transcriptional manner. RNA encoding the protein is detectable only during the late erythrocytic stages of the parasite, but it is possible to detect promoter activity throughout the erythrocytic cycle. Expression of other *P. falciparum* genes appears to be at the level of transcription, for example knob associated histidine rich protein (KAHRP), circumsporozoite protein (CSP), and histidine rich protein II (HRPII) (Lanzer et al., 1992a).

### 1.9.4 Identification of promoters for *Plasmodium* genes

The upstream regions of several *Plasmodium* genes have been cloned and characterised. The most completely studied upstream region is that of the glycophorin binding protein (GBP) gene of *P. falciparum*. The region between the GBP gene and the 3.8 gene preceding it is relatively short, only 2 kb. Transcription of the two genes has been shown to be discontinuous and monocistronic, so the intergenic region must contain signals for the

termination of transcription of the 3.8 gene and also for the regulation of transcription initiation of the GBP gene (Lanzer et al., 1992b). Transcription of the GBP gene has been shown to be sensitive to  $\alpha$ -amanitin, a compound which inhibits the activity of eukaryotic RNA polymerase II.

Analysis of the DNA sequence immediately upstream of the transcription initiation site of the GBP gene has revealed certain features which are characteristic of eukaryotic promoters (Lanzer et al., 1992b). The region contains many GC-rich islands, one of which is highly homologous to the SV40 core enhancer region (Weiher et al., 1983). A similar element is present in the upstream region of the gene encoding the circumsporozoite protein of *P. knowlesi* (Ruiz-i-Altaba et al., 1987). The sequence present in the *P. falciparum* GBP gene has been demonstrated to bind nuclear proteins derived from erythrocytic stage parasites, suggesting that it interacts with parasite transcription factors (Lanzer et al., 1992b). However, as a similar sequence element is present in the gene of the sporozoite stage circumsporozoite protein in *P. knowlesi*, and GBP is expressed by blood stage parasites this element would not necessarily be expected to be involved in controlling the stage-specific expression of the gene. It is more likely that it is involved in general transcriptional control, for example controlling the frequency of transcription initiation.

Analysis of the upstream sequence has also revealed a large direct duplication of 305 bp (Lanzer et al., 1992b). The authors have speculated that this motif may mediate stage-specific regulation of GBP gene expression.

Studies into the transcription of the merozoite surface protein-1 (MSP-1) gene of *P. yoelii* has shown apparently heterogeneous transcription

initiation sites (Lewis, 1990). There are five initiation sites located between 377 and 417 bp upstream of the initiating methionine codon. Whether these are real or an artefact due to the pausing of reverse transcriptase at A+T-rich sequences during primer extension experiments has not been determined. However, such heterogeneity has also been observed for *P. falciparum* MSP-1 (Myler, 1989) and *P. knowlesi* CS (Ruiz-i-Altaba et al., 1987) genes. There are several potential TATA-boxes located thirty to forty base pairs upstream of the RNA start sites of the gene. This is in keeping with the distance restrictions observed for other eukaryotic genes, suggesting that at least one of these could be a functional TATA-box (Lewis, 1990). Two sequence elements with 100% homology to the immunoglobulin octamer sequence ATTTGCAT, or its complement, were also present in the upstream region of the gene. Such sequences have been found in enhancers and promoters, and the sequence has been shown to bind two transcription factors with high specificity (Staudt et al., 1988; Staudt et al., 1986). Whether the sequences in the MSP-1 promoter bind to parasite nuclear extracts has not been determined. It has also been noted that several regions of the MSP-1 gene's upstream region are capable of forming Z-DNA, and that one of the sequences with homology to the immunoglobulin octamer motif lies within one of these regions (Lewis, 1990). It has been suggested that Z-DNA formation may be related to transcriptional activation, with a role in enhancer function (Nordheim and Rich, 1983). A diagram representing the upstream regions of several *Plasmodium* genes is shown in figure 1.10.1.

The function of specific upstream promoter elements in *Plasmodium* has not been demonstrated. Sequences which fit the consensus described for TATA boxes and CAAT box elements have been described but there is no evidence that these are indeed functional. One piece of evidence that there

is a *Plasmodium* equivalent of at least one of these elements, the TATA box, is the identification of a TATA-box binding protein (TBP) in *P. falciparum* (McAndrew et al., 1993). The intergenic regions of *P. falciparum* contain extremely A+T regions, many of which resemble the consensus sequence for the TATA box. This has led to suggestions that the initiation of transcription in *P. falciparum* may differ from that of other eukaryotes (McAndrew et al., 1993).

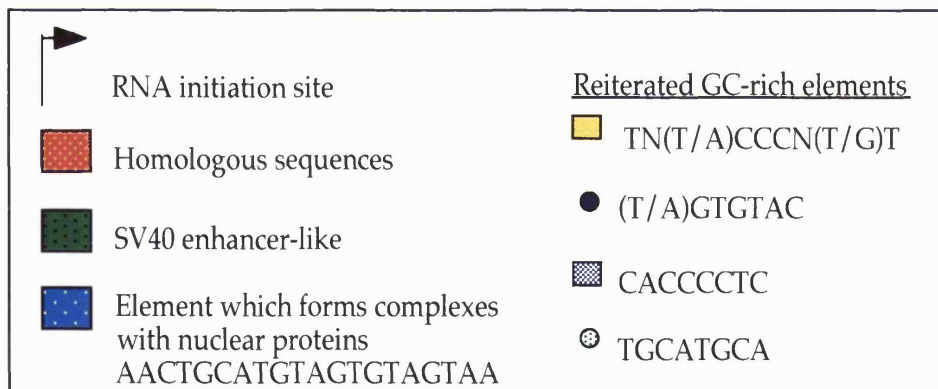
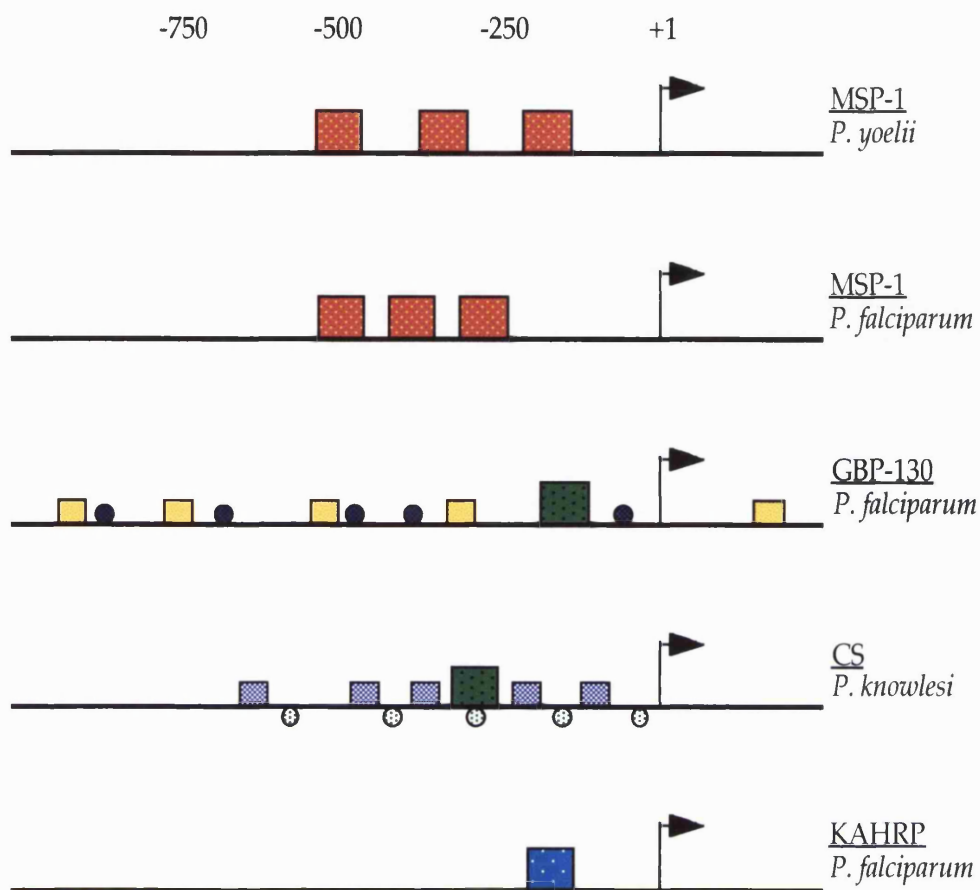
Polymeric A+T rich regions in *Saccharomyces cerevisiae* have been shown to be involved in gene regulation (Iyer and Struhl, 1995). These regions are thought to alter DNA structure in that they may make transcription factor binding sites more accessible to the relevant proteins. Whether there is a similar phenomenon in *Plasmodium* is not known, however it should be noted that the whole *Plasmodium* genome is extremely A+T rich and, although the intergenic regions have an even higher A+T content, such mechanisms may not be feasible in *Plasmodium*.



**Figure 1.9.1** The upstream regions of *Plasmodium* genes

The structural organisation of several *Plasmodium* promoter regions has been analysed. Most notable features are the presence of GC-rich reiterated sequences and the conservation of sequence motifs in the same gene of different *Plasmodium* species.

Figure adapted from (Lanzer et al., 1993).



### 1.10 Methods for the study of gene expression of *Plasmodium*

The ability to transfect *Plasmodium* parasites with exogenous DNA has recently emerged as a tool with which it is possible to study the promoter regions of parasite genes. A construct containing a reporter gene linked to the proposed control region can be used to determine the minimum region of the promoter for maximal expression of the gene (Crabb and Cowman, 1996; Horrocks and Kilbey, 1996).

For example, *P. falciparum* erythrocytic asexual stages have been successfully transfected to produce chloramphenicol acetyltransferase (CAT). Parasitised red blood cells were transfected with plasmids that contained CAT-encoding DNA flanked by 5' and 3' untranslated sequences of the *P. falciparum* hsp86, hrp3, and hrp2 genes. These flanking sequences were required for expression as their excision abolished CAT activity in transfected parasites (Wu et al., 1995). Such transfection techniques have increased the opportunities to follow genetic analysis of the malaria parasite.

### 1.11 Aims of this project

The 235 kDa rhoptry proteins of *Plasmodium yoelii* have been shown to be encoded by a multigene family. However, the entire coding region has not been identified for any of the members of the gene family. Comparison of the sequences for the clones designated E3 and E8 have shown that they lack the 5'-end of the genes, with none of the features one would expect for a non-coding region of *Plasmodium* DNA, that is the presence of in-frame stop codons and an increase in A+T content of the DNA.

The aims of this project were to complete the gene sequence of at least one of the members of the 235 kDa rhoptry protein family of *Plasmodium yoelii*.

Once the complete sequence of one of the genes had been determined, analysis of its structural features could be undertaken with a view to identifying regions of functional significance. It was also hoped to undertake an analysis of putative promoter region of one or more members of the multigene family in order to gain an insight into how the parasite controls expression of such a large family of related genes.

During the course of the project, several other sequences were cloned which were fragments located at the 5' end of other genes of the rhoptry protein family. Comparison of these clones gives an impression of the variation present between the members of the family at the sequence level. From previous work (Borre et al., 1995; Keen et al., 1990; Keen et al., 1994; Sinha et al., 1996) it had been determined that the C-terminus of the proteins showed relatively little sequence variation, whereas the N-termini of the proteins, although still significantly similar, showed more

variation. This may reflect differences in the proteins' affinity for ligands on the erythrocyte surface.

**CHAPTER TWO**

**MATERIALS AND METHODS**

## 2.1 Buffers and Media

KS	Krebs' Saline: 150mM NaCl, 10mM glucose
LB	10.00g bacto-tryptone, 5.00g bacto-yeast extract, 10.00g NaCl in 1L, pH 7.0
PBS	Phosphate buffered saline - 140mM NaCl, 3mM KCl, 3mM $\text{KH}_2\text{PO}_4$ , 140mM $\text{Na}_2\text{PO}_4$ , pH 7.4
SOC	20.00g bacto-tryptone, 5.00g bacto-yeast extract, 0.5g NaCl, 0.25mM KCl, 10mM $\text{MgCl}_2$ , 20mM glucose, pH 7.0
SSC	Sodium chloride-sodium citrate buffer 1X SSC is 0.15M NaCl, 0.015M tri-sodium citrate
TAE	Tris-acetate EDTA 1X TAE is 40mM Tris-acetate, 10mM EDTA pH 8.0
TBE	Tris-borate EDTA 1X TBE is 900mM Tris-borate, 2mM EDTA, pH 8.0
TE	Tris-EDTA 10mM Tris.HCl pH 8.0, 1mM EDTA pH 8.0

## 2.2 Preparation of *Plasmodium yoelii* parasites

Donor mice (balb/c strain) infected with *Plasmodium yoelii* YM parasites were bled into a minimal volume of Krebs' saline, 0.2% glucose, and 250 units heparin. Aliquots of 500µl were snap-frozen in liquid nitrogen and stored at -80°C until required. On the day of use, the vial was thawed on ice and 200µl of the stablate injected intra-peritoneally into each of five balb/c mice. Once the parasitaemia of the mice had risen to 30-40% (as determined by performing blood smears), they were bled. One hundred microlitres of the blood was injected intraperitoneally into fifty balb/c mice. When the parasitaemia reached approximately 50% these mice were bled. The volume of blood was made up to 100ml by the addition of ice-cold phosphate buffered saline (PBS). The blood was depleted of leukocytes by passing it twice over a CF-11 (Whatman) column while maintaining the temperature of the sample at less than 4°C. Red blood cells were precipitated by centrifugation at 1000x g for 10 minutes and washed two to three times in ice-cold PBS followed by centrifugation as before. Erythrocytes were lysed by the addition of two volumes of ice-cold 1% acetic acid with incubation for two minutes on ice. The lysis reaction was quenched by the addition of four volumes of ice-cold PBS. Haemoglobin was removed from the sample by repeated washing of the sample with ice-cold PBS and centrifugation at 1000x g. Washed parasites were precipitated by centrifugation, divided into aliquots, snap frozen in liquid nitrogen, and stored at -80°C.



## 2.3 Purification and preparation of nucleic acids

### 2.3.1 Purification of *Plasmodium yoelii* genomic DNA

Frozen parasites prepared as described in section 2.2 were used as the source of *Plasmodium yoelii* YM genomic DNA. The DNA was purified using a Genomic DNA Purification Kit (Qiagen) according to the manufacturer's instructions.

One to two microlitre aliquots of the DNA preparation were run on a 0.5% agarose gel to estimate its concentration and size distribution, and its purity assessed using UV spectrophotometry. DNA was considered to be pure if it had an  $A_{260\text{nm}}/A_{280\text{nm}} > 1.7$ , and of good size if the majority of the DNA ran at  $> 20\text{kb}$  as assessed by agarose gel electrophoresis.

### 2.3.2 Purification of *Plasmodium yoelii* total RNA

100 $\mu\text{l}$  of frozen *P. yoelii* YM parasites prepared as described in section 2.2 were used to prepare total RNA, using a SNAP total RNA isolation kit (Invitrogen) according to the manufacturer's instructions. The concentration of the RNA preparation was determined by measuring the  $A_{260\text{nm}}$ , and using the following formula:

$$\text{RNA concentration} = (A_{260})(0.04\mu\text{g } \mu\text{l}^{-1})D$$

where D is the dilution factor.

### 2.3.3 Synthesis of oligonucleotides

Oligonucleotides were synthesised using an ABI 380B DNA synthesiser by the Sequencing and Synthesis Service at NIMR. They were supplied fully deprotected in 35% ammonia. The oligonucleotides were precipitated by

addition of 0.1 volumes 3M sodium acetate pH 5.2 and 2.5 volumes 100% ethanol, followed by centrifugation at 12000 x g for 30 minutes. The DNA pellet was washed in 70% ethanol and resuspended in distilled, deionised water (DDW) at a final concentration of 10µM.

The concentration of each oligonucleotide was calculated as follows:

$$E = (G \times 11.7) + (A \times 15.4) + (T \times 8.8) + (C \times 7.3)$$

$$c (\mu\text{M}) = \frac{(A_{260\text{nm}} \times 1000)}{E}$$

where E is the molar extinction coefficient, G, A, T and C are the number of G, A, T and C residues in the oligonucleotide, and c is the concentration of the oligonucleotide.

#### 2.3.4 Preparation of plasmid DNA

For small scale preparation of plasmid DNA bacteria were cultured overnight at 37°C in 5ml LB containing 50-100µg ml<sup>-1</sup> ampicillin (Sigma). DNA was prepared from the bacterial suspension using the Wizard Miniprep DNA Plus purification system (Promega) according to the manufacturer's instructions.

Large scale preparations of plasmid DNA were performed by growing bacteria in 250ml LB containing 50-100µg ml<sup>-1</sup> ampicillin overnight at 37°C. Plasmid DNA was purified using the Wizard Maxiprep DNA Plus purification system (Promega) according to the manufacturer's instructions.

The concentration of the DNA was determined by measuring its A<sub>260nm</sub>, assuming an A<sub>260nm</sub> of 1.0 is equivalent to 50µg ml<sup>-1</sup>.

### 2.3.5 Restriction endonuclease digestion of DNA

DNA to be digested with restriction endonucleases was resuspended in 1X buffer specified by the manufacturer of the enzyme (Boehringer or New England Biolabs). In general, 10U restriction enzyme were added per  $\mu\text{g}$  DNA to be digested, and the reaction incubated at the designated temperature for 2-4 hours to ensure complete digestion.

### 2.3.6 Purification of DNA from agarose gels

DNA fragments were purified from low melting temperature agarose using Wizard PCR purification columns (Promega) according to the manufacturer's instructions.

DNA was purified from high melting temperature agarose using the GeneClean kit (BIO 101 Inc) according to the manufacturer's instructions.

### 2.3.7 Preparation of vector and insert DNA for ligation

Vector and insert DNA were digested with the appropriate restriction endonuclease and purified from low melting temperature agarose according to sections 2.3.5 and 2.3.6. The phosphate group was removed from the 5'-termini of the vector DNA by incubation at 30 minutes at 37°C with 0.01U calf intestinal alkaline phosphatase per pmol 5'-ends (Maniatis et al., 1982). The vector was phenol:chloroform extracted, ethanol precipitated, and resuspended in DDW at a concentration of 100 $\mu\text{g ml}^{-1}$ .

### 2.3.8 Ligation of vector and insert DNA

Ligation reactions were carried out in a total volume of 10 $\mu$ l in 1X T4 DNA ligase buffer with 1 Weiss Unit T4 DNA ligase (Boehringer). 100ng vector DNA was ligated to insert DNA in a molar ratio of 1:3. The reaction was incubated overnight at room temperature.

### 2.3.9 Transformation of DNA into competent bacterial cells

20-50ng of ligated DNA was transformed into 100 $\mu$ l competent cells by electroporation. Electroporation-competent *E. coli* SURE cells (Stratagene) were used in an attempt to reduce the chance of unwanted recombination events. The DNA was added to a 50 $\mu$ l aliquot of competent cells and kept on ice for 1 minute. The transformation was transferred to a Genepulser cuvette (Biorad) with a 0.1cm electrode gap. The sample was electroporated using a Genepulser electroporator (Biorad) set at 1.8 kV, 200 $\Omega$  and 25  $\mu$ F. 1ml of SOC medium preheated to 37°C was added, and the transformation was incubated at 37°C with shaking at 225-250 rpm for 60 minutes. Generally, 10-100 $\mu$ l of the transformation was plated onto LB agar plates containing 50-100 $\mu$ g ml<sup>-1</sup> ampicillin, which had been preplated with 35 $\mu$ l 100mM IPTG and 25 $\mu$ l 50mg ml<sup>-1</sup> X-Gal, and incubated overnight at 37°C.

## 2.4 Sequencing of DNA

### 2.4.1 Manual sequencing of double-stranded plasmids

Manual sequencing of recombinant plasmids was performed using the Rapidwell DNA sequencing kit (USB) or Sequenase version 2.0 (USB) according to the manufacturer's instructions. Sequencing reactions were run on 6% denaturing polyacrylamide gels (30 x 40 x 0.04 cm) in 1 X TBE buffer for Sequenase version 2.0 reactions or in 1 X glycerol tolerant buffer (USB) for Rapidwell reactions. The gels were run at 60W as described by Maniatis. Gels were fixed by immersion in 2 litres of a solution of 10% glacial acetic acid, 20% methanol v/v for 15 minutes, dried under vacuum at 80°C, and exposed to Biomax-MR film (Kodak) for 12-24 hours at room temperature.

### 2.4.2 Automated sequencing of DNA

Plasmid templates for DNA sequencing were prepared as described in section 2.3.4. Sequencing reactions were carried out using the ABI PRISM cycle sequencing kit (Applied Biosystems Inc.), which contains fluorescently labelled dideoxynucleotides. Cycle sequencing was carried out in accordance with the manufacturer's instructions, with the exceptions that 500ng DNA template and 5pmol primer were used per reaction, and the reaction volume was scaled down 2-fold, i.e.. the final reaction volume was 10µl rather than 20µl.

PCR products were sequenced directly. 30ng of gel purified PCR product was used per reaction with 1.6pmol primer, and cycle sequencing performed according to the manufacturer's protocol.

Sequencing reactions were loaded onto a 6% denaturing polyacrylamide gel, or a 5% Long Ranger polyacrylamide gel (Flowgen) and the gel run at 30W for 12 hours on an ABI 377 sequencer (Applied Biosystems Inc.). Data were collected using the ABI 377 data collection software (Applied Biosystems Inc.) and analysed using Factura and Autoassembler programs (Applied Biosystems Inc.).

## 2.5 Polymerase Chain Reaction (PCR)

PCR reactions were typically carried out in a total volume of 50 $\mu$ l with 1 $\mu$ M each PCR primer, 200 $\mu$ M dNTPs, and 10ng template DNA in 1X PCR amplification buffer (Promega) using 2.5U AmpliTaq DNA polymerase (Promega). The AmpliTaq enzyme was added in a "hot-start" reaction once the reaction mixture had reached 95°C, to prevent amplification of products caused by annealing of the primers to non-specific sequences at lower temperatures. Amplification of target sequences was achieved using the following cycling procedure:

Denaturation of template DNA:	95°C	1 min
Annealing of primers:	X°C	1 min
Synthesis of target sequence:	72°C	1 min/kb target DNA

The annealing temperature of each primer (X°C) was calculated as follows:

$$T_m = 2(A+T) + 4(G+C)$$

$$X = T_m - 5^\circ\text{C}$$

## Chapter Two

where A represents the number of A residues in the primer, T the number of T residues, G the number of G residues, C the number of C residues, and  $T_m$  denotes the melting temperature of the oligonucleotide.

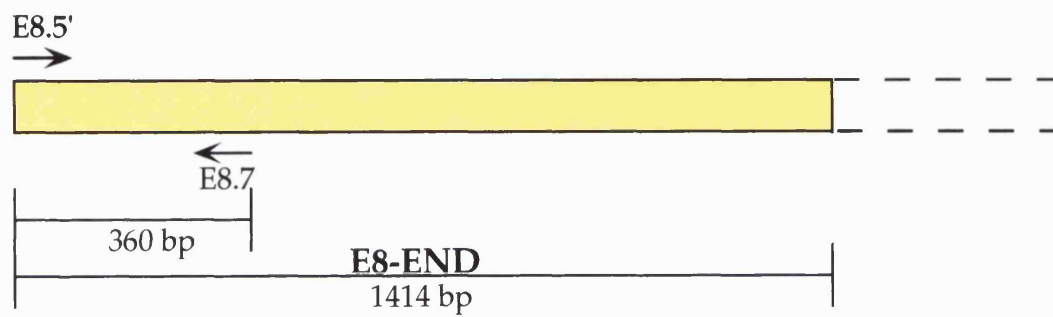
These steps were repeated 30 times using a thermal cycler (Perkin Elmer), and products analysed by running 5-10 $\mu$ l on a 1% agarose gel.

**Figure 2.5.1** Primer sequences and positions in the E8 sequence for PCR to generate a probe

The sequence of primers E8.5' and E8.7 are given with respect to their positions in the clone E8.END. They were used in a PCR reaction to amplify the first 360 bp of clone E8.END for use as a probe in subsequent genomic library screens.

Clone E8.END is an inverse PCR product derived from the E8 gene which has been cloned into the pMAL c2 expression vector, and was supplied by Dr. Jane Keen.





## 2.6 Inverse PCR

The template for inverse PCR was restriction endonuclease-digested *P. yoelii* YM genomic DNA which had been circularised by intramolecular ligation.

### 2.6.1 Digestion of *P. yoelii* YM genomic DNA with restriction endonucleases

Five micrograms of *P. yoelii* YM genomic DNA were incubated at 37°C for two hours with 40U of the selected restriction endonucleases in 1X buffer recommended by the manufacturer of the enzymes. The final volume of the reaction was 100µl. Digested DNA was subsequently purified using GeneClean (BIO 101 Inc.) and the DNA eluted in a total volume of 20µl DDW.

### 2.6.2 Ligation of digested DNA

One hundred nanograms of the restriction endonuclease-digested DNA was incubated in 1X T4 DNA ligase buffer with 2.5U T4 DNA ligase (Boehringer) in a total volume of 25µl. The concentration of DNA was kept low (<4µg ml<sup>-1</sup>) to encourage intramolecular ligation of individual DNA fragments.

The ligation reaction was incubated overnight at room temperature, then drop-dialysed against DDW on 0.2µm nylon filters (Amersham) for 60 minutes at room temperature to remove excess salt.

### 2.6.3 Amplification of target DNA using Touchdown PCR

The entire 25µl ligation reaction was used as a template for PCR with the following additions: 200µM dNTPs, 1.5mM MgCl<sub>2</sub>, 1µM primer E8.11, 1µM primer E8.12, 1X PCR buffer II (Promega), 2.5U AmpliTaq DNA polymerase (Perkin Elmer) in a final volume of 50µl. The polymerase was added in a hot-start reaction as described previously. The mixture was subjected to Touchdown PCR. This is a modification of a conventional PCR program, designed to enhance specific amplification of target DNA sequences and to reduce non-specific amplification products.

Instead of using a constant annealing temperature in each cycle of PCR, the initial annealing temperature used was 5°C above the calculated T<sub>m</sub> of the primers. Every two cycles the annealing temperature was reduced by 2°C, until an annealing temperature 10°C below the T<sub>m</sub> of the primers was reached. Thirty cycles of PCR were performed using this lower annealing temperature, and the reaction held at 4°C.

The logic behind Touchdown PCR is as follows. In the first few cycles of PCR, where the annealing temperature is very close to the T<sub>m</sub> of the primers, only very specific amplification of the target sequence occurs. Therefore the concentration of only the target sequence increases exponentially during these cycles. Once the annealing temperature is lowered to the point where non-specific amplification can occur the concentration of the specifically amplified DNA product is such that it easily out competes non-specific sequences for annealing of the primers. Thus, the specificity of the PCR reaction is enhanced.

5µl of the PCR reactions were run on a 1% agarose gel for analysis.

#### 2.6.4 Purification of inverse PCR products in preparation for sequencing

PCR products were purified using Centricon-100 columns (Amicon). 2ml DDW was added to the sample reservoir and the entire PCR reaction (50µl) added. The column was assembled and centrifuged in a Beckman JA20.1 rotor at 2000x g for 15 minutes. The flow through was discarded, the column inverted, and centrifuged again at 2000x g for 2 minutes to recover the concentrated DNA. The sample was precipitated by the addition of 0.1 volumes 3M sodium acetate pH 5.2 and 2.5 volumes 100% ethanol, followed by centrifugation at 12000x g for 15 minutes in a microfuge. The DNA pellet was resuspended in 10µl DDW for use directly in automated DNA sequencing (section 2.4.2).

**Figure 2.6.1** Primer sequences and positions for inverse PCR experiment

The sequences of the primers used in inverse PCR are shown with respect to their positions in the sequence of clone E8.END. They are in the opposite orientation to primers used in conventional PCR, pointing away from one another.

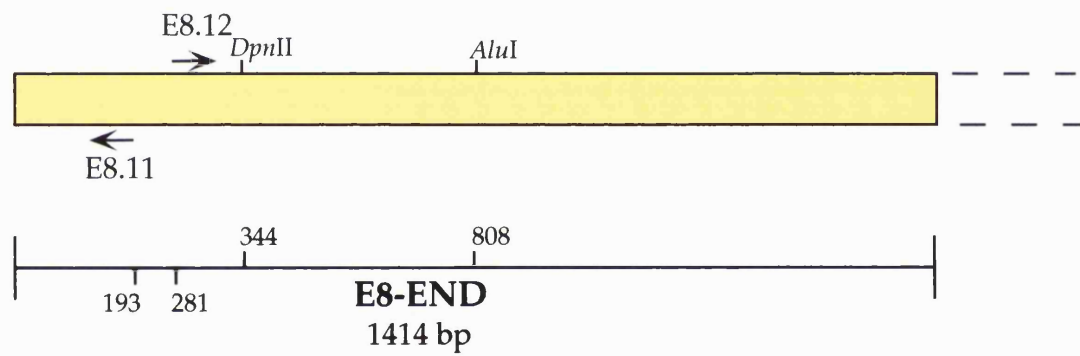
Clone E8.END is an inverse PCR product derived from the E8 gene which has been cloned into the pMAL c2 expression vector, and was supplied by Dr. Jane Keen.

**E8.11:** the reverse complement of nucleotides 170-193 in the sequence of clone E8-END.

5'-GTTTATTAATAATTCCCTAGTG-3'

**E8.12:** nucleotides 281-300 of clone E8-END

5'-GAAGGTGAATTAAATTGGAGT-3'



## 2.7 5'-Rapid Amplification of cDNA Ends (5'-RACE)

### 2.7.1 Synthesis of first strand of cDNA

2.5 pmol of the E8 gene-specific primer, E8.NEW.RC, were added to 1µg of *P. yoelii* YM total RNA and the volume made up to 15.5µl with DEPC-treated water. The reaction was incubated at 70°C for 10 minutes to denature the RNA and subsequently placed on ice for 1 minute. 2.5µl 10X PCR buffer (Gibco BRL), 2.5µl 25mM MgCl<sub>2</sub>, 1µl 10mM dNTPs, and 2.5 µl 0.1M DTT were added to the reaction and incubated at 42°C for 1 minute. 200U of Superscript II Reverse Transcriptase (Gibco BRL) were added, and incubation at 42°C continued for a further 50 minutes. The reaction was terminated by heating the sample at 70°C for 15 minutes. Degradation of RNA was carried out by addition of 1µl DNase-free RNase mix (Gibco BRL) and the reaction incubated at 37°C for 30 minutes. The contents of the tube were collected by brief centrifugation in a microfuge, and placed on ice.

### 2.7.2 Purification of cDNA

The cDNA was purified using the Geneclean kit (BIO 101 Inc) according to the manufacturer's instructions. DNA was eluted from Glassmilk in 20µl DDW.

### 2.7.3 Oligo-dC tailing of cDNA using terminal deoxynucleotidyltransferase

5µl 5X tailing buffer (Gibco BRL) and 2.5µl 2mM dCTP were added to 10µl of the purified cDNA and the volume of the reaction made up to 25µl with DEPC-treated water. The reaction was heated to 95°C for 2 minutes, placed

on ice for 1 minute, and the contents of the tube collected by brief centrifugation. 1µl terminal deoxynucleotidyltransferase (Gibco BRL) was added and the reaction incubated at 37°C for 10 minutes. The terminal deoxynucleotidyltransferase was inactivated by heating at 65°C for 10 minutes, and the reaction was placed on ice.

#### 2.7.4 PCR of oligo-dC tailed cDNA

The following reagents were added to 5µl oligo-dC tailed cDNA:

31.5µl DDW, 5µl 10X PCR buffer (Gibco BRL), 3µl 25mM MgCl<sub>2</sub>, 1µl 10mM dNTPs, 2µl 10µM R1, and 2µl 10µM Universal Anchor Primer (Gibco BRL). The reaction was heated to 94°C for 1 minute followed by addition of 2.5U AmpliTaq DNA polymerase (Promega). Polymerase chain reaction conditions were as follows:

- |                     |       |
|---------------------|-------|
| 1. 94°C             | 1 min |
| 2. 55°C             | 2 min |
| 3. 72°C             | 3 min |
| 4. repeat steps 1-3 |       |
| 35 times            |       |
| 5. 72°C             | 6 min |
| 6. 4°C              | hold  |

10µl of the reaction was analysed by agarose gel electrophoresis on a 1% agarose gel.

Nested PCR was carried out in an identical manner using 5µl of the original PCR reaction as template and substituting R2 for R1, and Abridged



Anchor Primer (Gibco BRL) for UAP. After PCR amplification, 10 $\mu$ l of the nested PCR reactions were run on a 1% agarose gel for analysis.

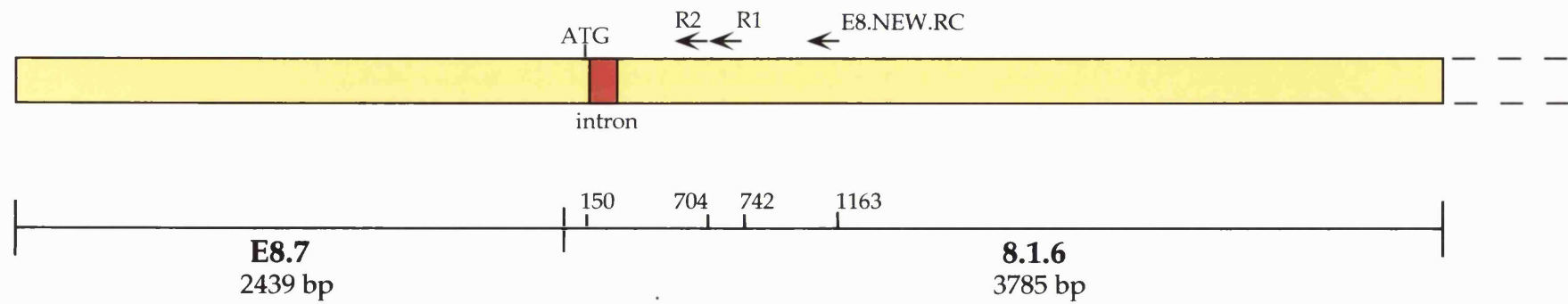
#### 2.7.5 Cloning and sequencing of 5'RACE products

PCR products from 2.7.4 were cloned into the pMOSBlue vector (Amersham) according to the manufacturer's instructions. Purified recombinant plasmids were sequenced using automated DNA sequencing with primers T7 and R2.

**Figure 2.7.1** Sequences and positions of E8-specific primers for 5'RACE

The sequences of the gene specific primers used for 5'RACE are given with respect to their positions in clone 8.1.6.

The putative initiating ATG codon and the position of the intron are also marked.



**E8.NEW.RC:**

5'- GTATTCTTGAGTTTGGAGCTA -3'

**R1:**

5'- CGACCGTATCTAACGTATC -3'

**R2:**

5'- GCGGTTATTTGTAAAAAGGAATT -3'

## 2.8 Reverse Transcriptase PCR (RT-PCR)

### 2.8.1 Synthesis of cDNA using gene-specific primers

1µg total *P. yoelii* YM RNA was added to 2.5pmol E8-specific primer E8.NEW.RC in a total volume of 15.5µl.

The sample was treated exactly as in section 2.7.1 to synthesise a cDNA copy of the RNA. Purification of the cDNA was achieved using the Geneclean kit (BIO 101 Inc.) according to the manufacturer's instructions, and cDNA was eluted in a final volume of 20µl DDW.

### 2.8.2 PCR from cDNA using E8-specific primers

PCR reactions were carried out as described in section 2.5 in a total volume of 50µl, using 5µl purified cDNA as template. The primers that were used in the PCR amplification step were based on the nucleotide sequence either side of the proposed intron of the E8 gene.

Products were analysed by running 5-10µl on a 1% agarose gel and viewed with UV-transillumination.

### 2.8.3 Cloning of RT-PCR products

RT-PCR products were gel purified from high melting temperature agarose (Boehringer MP-agarose) as described in section 2.3.6, and cloned into the pMOSBlue vector (Amersham) according to the manufacturer's instructions.

**Figure 2.8.1** Position and sequences of primers for RT-PCR

The positions of the primers used in RT-PCR are given using the A residue of the putative initiating ATG codon as position 1.



## 2.9 Construction and screening of a sub-genomic DNA library

### 2.9.1 Identification of positive fractions by hybridisation

25µg *P. yoelii* YM genomic DNA was incubated at 37°C overnight with 100U *DpnII* (New England Biolabs). The DNA fragments were separated on a 1% agarose gel. Several thin slices of agarose 1-2mm wide were excised, which contained DNA in the size range of the positive hybridisation signal of the original Southern blot (which had been probed with an ( $\alpha$ -<sup>32</sup>P)dATP-labelled DNA probe corresponding to the first 270 bp of clone E8.END (for details see section 3.1). The DNA in these gel slices was purified from the agarose using the Band-Prep kit (Stratagene) according to the manufacturer's instructions, and eluted in 20µl TE pH 8.0. 5µl of each purified size fraction was run on a 1% agarose gel and photographed with UV-transillumination. A Southern blot was performed on this gel (section 2.11) and the blot hybridised to the same ( $\alpha$ -<sup>32</sup>P)dATP-labelled DNA probe as in the original Southern blot. The blot was washed under stringent conditions and exposed to X-omat AR film (Kodak) overnight.

This procedure enabled identification of the specific DNA fraction in which the target DNA was contained.

### 2.9.2 Preparation of pBluescriptIISK- plasmid

10µg pBluescriptIISK- (Stratagene) was digested with 20U *Bam*HI (Boehringer) in 1X *Bam*HI reaction buffer for 2 hours at 37°C. One unit of calf-intestinal alkaline phosphatase (Boehringer) was added to the digested vector and the reaction incubated at 37°C for a further 30 minutes. The vector was then phenol:chloroform extracted (Maniatis et al., 1982),

precipitated by the addition of 0.1 volume 3M sodium acetate pH 5.2 and 2.5 volumes 100% ethanol with centrifugation for 15 minutes at 12000x g, and the DNA pellet resuspended in DDW at a final vector concentration of 100ng  $\mu\text{l}^{-1}$ .

### 2.9.3 Ligation of purified insert to pBIISK- vector and transformation of *E. coli* SURE cells

Ligation of insert DNA to 100ng vector was carried out at a 3:1 molar ratio of insert:vector. The reaction was carried out in a final volume of 10 $\mu\text{l}$  using 1 Weiss Unit T4 DNA ligase (Boehringer) in 1X T4 DNA ligase buffer, and incubated overnight at room temperature.

Fifty nanograms of the ligation reaction was transformed into *E. coli* SURE electrocompetent cells (Stratagene) as described in section 2.3.9. The percentage of recombinant colonies was assessed by plating an aliquot of the transformation onto an LB agar plate containing 100 $\mu\text{g ml}^{-1}$  ampicillin, and preplated with 35 $\mu\text{l}$  100mM IPTG and 25 $\mu\text{l}$  50mg  $\text{ml}^{-1}$  X-Gal. Recombinant colonies appear white in colour compared to blue non-recombinants owing to the disruption of the *lacZ*  $\alpha$ -complementation factor by insertion of DNA into the vector.

Ten thousand recombinant colonies were plated onto LB plates containing 100 $\mu\text{g ml}^{-1}$  ampicillin and incubated overnight at 37°C.

### 2.9.4 Screening the library

The library was screened by colony hybridisation (section 2.12) using the same 270 bp ( $\alpha$ - $^{32}\text{P}$ )dATP labelled DNA probe as in the original Southern blot experiment (section 3.1).



## Chapter Two

Positive colonies were identified from the plate by realignment of the autoradiograph, filter, and master plate. The appropriate area of the agar plate (5 X 5mm) was excised using a sterile scalpel blade, placed into 1ml LB medium, 1µl of a 1000-fold dilution of this plated onto a fresh LB agar plate containing 100µg ml<sup>-1</sup> ampicillin, and incubated overnight at 37°C.

The second screen was performed in exactly the same way, by colony hybridisation. Single positive colonies were picked into 5ml LB containing 100µg ml<sup>-1</sup> ampicillin, grown overnight at 37°C, and plasmid DNA prepared from the cultures by the method described in section 2.3.4.

The insert DNA was analysed by restriction endonuclease digestion with *Bss*HII, which has sites at either end of the pBluescriptIISK- polylinker. The digests were run on a 1% agarose gel for analysis.

Sequencing of the positive clones was by the dideoxy chain-termination method using the Sequenase Rapidwell Sequencing Kit (USB) according to the manufacturer's instructions.

## 2.10 Construction of a partially *Sau3A* digested, size-fractionated genomic library in pBluescriptII KS+

### 2.10.1 Partial *Sau3A* digestion of *Plasmodium yoelii* genomic DNA

Aliquots of 1µg *Plasmodium yoelii* YM genomic DNA were digested in a final volume of 10µl with the following amounts of *Sau3A* (Boehringer): 1U, 0.5U, 0.25U, 0.125U, 0.06U, 0.03U, 0.015U, and 0.008U.

The digests were incubated at 37°C for 60 minutes, the reaction stopped by the addition of 0.2 volumes of agarose gel loading buffer, and run on a 0.6% agarose gel to determine the degree of digestion in each instance. It was determined that the samples digested with 0.5, 0.25, 0.125, and 0.06U gave a good range of DNA fragments in the 2-5kb size range.

The digests were scaled up by a factor of ten and repeated, i.e.. 10µg of *Plasmodium yoelii* YM genomic DNA was digested in a total volume of 100µl with 5.0U, 2.5U, 1.25U, and 0.6U of *Sau3A*, and incubated for 60 minutes at 37°C. 10µl of agarose gel loading buffer was added to each digest and the samples run on a 0.6% agarose gel.

### 2.10.2 Digestion of pBluescriptII KS+ with *XhoI*

Twenty micrograms pBluescriptIIKS+ (Stratagene) were incubated at 37°C for 12 hours with 100U *XhoI* (Boehringer) in 1X *XhoI* reaction buffer (supplied by the manufacturer) in a total volume of 100µl. The enzyme was inactivated by heating at 60°C for 20 minutes and 100ng of the vector run on a 1% agarose gel to ensure complete digestion.

### 2.10.3 Size-fractionation of insert DNA

DNA in the size range of 2-5kb was excised from the agarose gel while viewed by transillumination under long-wave ultraviolet light. DNA was purified from the agarose using Wizard PCR minicolumns (Promega) according to the manufacturer's instructions, with the following exception. Once DNA had been eluted from the minicolumn with TE, 0.1 volumes of 3M sodium acetate pH 5.2 and 2.5 volumes of 100% ethanol were added to the eluate. The sample was spun at 12000 x g for 30 seconds in a microfuge to precipitate residual agarose and oligosaccharides. The supernatant was removed and centrifuged at 12000 x g for 20 minutes to precipitate the DNA. The DNA pellet was resuspended in 20µl DDW. The DNA was run on a 0.8% agarose gel, and purification of the 2-5kb fraction was repeated in an identical manner as described above. After the final ethanol precipitation the partially *Sau3A* digested DNA was resuspended in 10µl DDW.

### 2.10.4 Partial fill-in of vector and insert ends using the Klenow fragment of DNA polymerase I

10µl of *Plasmodium yoelii* YM genomic DNA, which had been partially digested with *Sau3A* and size-fractionated, was incubated with 1U Klenow (New England Biolabs), 1X Klenow reaction buffer (New England Biolabs), 250µM dGTP, 250µM dATP in a total volume of 20µl. The reaction was incubated at room temperature for 30 minutes and the Klenow enzyme was then heat inactivated by incubation at 70°C for 5 minutes.

The DNA was precipitated by addition of 0.1 volumes of 3M sodium acetate pH 5.2 and 2.5 volumes of 100% ethanol followed by centrifugation at 12000x g for 30 minutes. The supernatant was removed and the DNA

pellet resuspended in 10µl DDW, at a concentration of approximately 50µg ml<sup>-1</sup>.

Similarly, 2µg of *Xho*I-digested pBluescriptII KS+ were incubated with 2U Klenow in 1X Klenow reaction buffer, supplemented with 250µM dCTP and 250µM dTTP. Treatment of the DNA was identical to that of the genomic DNA, except that the vector was resuspended in DDW at a concentration of 100µg ml<sup>-1</sup>.

#### 2.10.5 Ligation and transformation

One hundred nanograms of partially-*Sau*3A digested, partially filled-in, size-fractionated *Plasmodium yoelii* YM DNA was ligated to 100ng of *Xho*I-digested, partially filled-in pBluescriptII KS+. The reaction was carried out in 1X T4 DNA ligase buffer (Boehringer) with 1 Weiss Unit T4 DNA ligase (Boehringer) in a total volume of 10µl. The reaction was incubated for 16 hours at room temperature.

The ligation reaction was drop-dialysed on a 0.2µm nylon filter versus DDW for 60 minutes at room temperature to remove excess salts.

One microlitre of the ligation was transformed into 100µl *E. coli* SURE electrocompetent cells (Stratagene) as described in section 2.3.9. This was repeated nine times.

#### 2.10.6 Assessment of insert size

Ten microlitres of one of the transformations from 2.10.5 was plated onto an 82mm LB agar plate containing 100µg ml<sup>-1</sup> ampicillin, which had been pre-plated with 50µl 100mM IPTG and 25µl 50mg ml<sup>-1</sup> X-Gal. The plate was incubated overnight at 37°C and five white colonies (containing

recombinant plasmid DNA) were picked into 5ml LBamp and grown overnight at 37°C. Plasmid DNA was isolated as described in section 2.3.4. Approximately 1µg of plasmid DNA was digested with 10U each of *Bam*HI and *Kpn*I (Boehringer) to liberate the insert. Digests were run on a 0.8% agarose gel and photographed with UV-transillumination. The average insert size was determined as 2.8 kb.

#### 2.10.7 Plating out the library

Two sheets of Biotrans A transfer membrane (PALL) measuring 222 x 222mm were laid onto each of two LB agar plates (240 x 240mm) containing 100µg ml<sup>-1</sup> ampicillin. 5ml of transformed *E. coli* SURE cells (from 2.10.5) were plated onto each membrane. The filters were placed in a sterile cabinet until no surface moisture was visible, then incubated at 37°C overnight to allow growth of the bacterial colonies.

#### 2.10.8 Preparing replicas of the library

Two replica filters of each master membrane were prepared as follows. The master membrane was removed from the LB amp plate and placed onto a sheet of LB-soaked 3MM Chr paper (Whatman) which was on a firm perspex plate. A fresh Biotrans A transfer membrane (222 x 222mm), which had been pre-wetted on a fresh LB amp agar plate, was placed precisely on top of the master membrane and covered with another piece of LB-soaked 3MM Chr paper. Another perspex plate was placed on top of the assembly. Transfer of the colonies from the master membrane to the replica was achieved by placing the assembly on the floor and requesting a heavy member of the lab staff to stand on it. The perspex plates and 3MM Chr paper were removed and the sandwiched membranes keyed to one

another with a large gauge needle. The membranes were gently separated, placed onto LB amp agar plates, and incubated at 37°C until colonies were clearly visible.

A second replica was prepared in an identical manner, using the first replica as a template.

#### 2.10.9 Storage of the library

Two sheets of 3MM Chr paper (Whatman) (250 x 250mm) were saturated in LB containing 25% glycerol. Each master membrane was placed onto one of these and left at room temperature for 15 minutes. The master membranes were removed and each was placed onto a perspex plate (250 x 250 x 3mm). On top of the membrane was placed an acetate grid (250 x 250mm) onto which the orientation holes of the master membrane were marked. (Another grid was marked in an identical way and stored with the replicas of the master membrane). A second perspex plate was placed on top of the acetate grid and secured with small bulldog clips. The library assemblies were stored at -80°C.

#### 2.10.10 Preparing replicas for screening

The library replicas were incubated at 37°C on LB agar plates containing 100µg ml<sup>-1</sup> ampicillin until colonies became clearly visible. The DNA in the colonies was denatured and neutralised by placing the replicas onto two sheets of 3MM Chr paper (Whatman) soaked in the following solutions:

<u>Treatment</u>	<u>Time</u>
0.5M NaOH	5 min
air dry	5 min
1M Tris.HCl pH 7.5	2 min
air dry	2 min
1M Tris.HCl pH 7.5, 1.5M NaCl	15 min
air dry	15 min

DNA was crosslinked to the membranes using the autocrosslink facility on a UV-Stratalinker (Stratagene). The replicas were stored with their respective alignment grids at 4°C.

#### 2.10.11 Screening the library

Screening of the library was carried out by hybridisation of specific radiolabelled probes to the replica filters. Probes for screening the library were prepared as described in section 2.11.2 and hybridisation as described in section 2.11.3.

Immediately after hybridisation, the replica filters were washed in 1L of 2X SSC, 0.1% SDS at room temperature for 15 minutes, followed by three high stringency washes in 1L of 0.1X SSC, 0.1% SDS at 65°C followed by overnight exposure to X-omat AR film (Kodak) with intensifying screens (Dupont) at -70°C.

Positively reacting colonies were picked by marking their position onto the grid kept with the replicas and subsequent alignment of this grid to the grid kept on the master membrane. An area of 5 x 5mm was excised from the master membrane, using a sterile scalpel blade, and placed in 1ml LB. This

was kept at room temperature for 15 minutes, vortexed vigorously to dislodge the bacterial colonies from the membrane, and 1 $\mu$ l added to a further 1ml of LB. The cells were titrated by plating 1, 10 and 100 $\mu$ l of the 1000-fold dilution onto LB agar plates containing 100 $\mu$ g ml<sup>-1</sup> ampicillin. One thousand colonies were plated onto 132mm LB amp agar plates and screened by colony hybridization with the same probe. A single, well-separated, positively reacting colony could then be isolated.

#### 2.10.12 Analysis of positive clones

Positive clones were grown overnight at 37°C in 250-500ml LB containing 100 $\mu$ g ml<sup>-1</sup> ampicillin, and plasmid DNA isolated. Analysis of insert size was carried out by restriction endonuclease digestion of the plasmids with *Bam*HI and *Kpn*I (Boehringer) followed by agarose gel electrophoresis. Clones of interest were sequenced using automated DNA sequencing, initially using primers based on vector sequences (T3 and M13-20). Further primers were designed based upon new sequence when they were required.

### 2.11 Southern blots

#### 2.11.1 Transfer of nucleic acid to a solid support

After separation of the DNA fragments by agarose gel electrophoresis the gel was treated with 0.2M HCl for 20 minutes to depurinate the DNA to allow transfer of large DNA fragments. This treatment was only carried out if the fragments to be transferred were >10kb. The DNA was denatured and neutralised as follows:



## Chapter Two

1. Washed in 500-1000ml 0.5M NaOH, 1.5M NaCl for 45min
2. Rinsed in DDW
3. Washed in 500-1000ml 1M Tris.HCl pH 7.4, 1.5M NaCl for 45min
4. Rinsed in DDW.

The washes were carried out at room temperature with gentle agitation in a rotary incubator.

Transfer of the DNA fragments to a solid support was carried out exactly as described by Maniatis *et al* (1989), with transfer onto Hybond-N membrane (Amersham) taking place overnight.

The blot was washed in a solution of 5X SSC at room temperature for 5 minutes, air dried, and the DNA crosslinked to the membrane using the autocrosslink facility on a Stratalinker UV-crosslinker (Stratagene).

#### 2.11.2 Preparation of radioactively labelled DNA probes

Double-stranded DNA probes were labelled with ( $\alpha$ - $^{32}\text{P}$ )dATP (Amersham) using random hexamer oligonucleotides and the Klenow fragment of DNA polymerase supplied with the Prime-It II kit (Stratagene) according to the manufacturer's instructions. Unincorporated label was removed using Sephadex G-50 Quickspin columns (Boehringer) according to the manufacturer's instructions.

#### 2.11.3 Hybridisation

Blots were prehybridised for at least one hour at 65°C in a minimal volume of hybridisation solution (5X Denhardt's solution, 5X SSC, 0.1% SDS, 100 $\mu\text{g ml}^{-1}$  denatured salmon sperm DNA (Boehringer)) sufficient to cover the blot. The probe was then heated at 95°C for 5 minutes, placed on

ice, and added to the hybridisation buffer. Between  $5 \times 10^5$  and  $1 \times 10^6$  cpm of the probe were added per ml hybridisation solution. Hybridisation was carried out in an incubator at  $65^\circ\text{C}$  with gentle agitation.

#### 2.11.4 Washing conditions and exposure to X-ray film

All blots were washed at high stringency in solutions of varying salt concentration. A typical washing procedure is as follows:

- |                       |                    |       |
|-----------------------|--------------------|-------|
| 1. 2X SSC, 0.1% SDS   | $65^\circ\text{C}$ | 20min |
| 2. 0.1X SSC, 0.1% SDS | $65^\circ\text{C}$ | 20min |

Step 2 was repeated two more times.

The blot was sandwiched between two pieces of Saran wrap (Dow) and exposed to X-omat AR film (Kodak) at  $-70^\circ\text{C}$  with intensifying screens (Dupont) for 12-24 hours.

### 2.12 Colony hybridisation

Colony hybridisation is a relatively rapid method for screening large numbers of recombinant bacterial colonies for the clone of interest.

LB agar plates containing the colonies of interest were incubated at  $4^\circ\text{C}$  for 1-2 hours prior to transfer, then a sheet of Hybond N (Amersham) of the appropriate dimensions placed directly on top of the colonies. The membrane was keyed to the agar plate by passing a large gauge needle through both the membrane and the plate for subsequent reorientation of the two. The membrane was carefully peeled from the plate and placed,

colony side up, on a piece of 3MM Chr paper (Whatman). The master agar plate was placed at 37°C for several hours to allow regrowth of the colonies. Colonies on the nylon membrane were treated as follows to denature their DNA and allow hybridization to an ( $\alpha$ -<sup>32</sup>P)dATP labelled DNA probe. Two layers of 3MM Chr paper were soaked in the following solutions and the membranes placed colony side up for the appropriate length of time:

10% SDS	5 min
air dry	5 min
0.5M NaOH, 1.5M NaCl	5 min
air dry	5 min
1M Tris.HCl pH 7.4,	10 min
1.5M NaCl	
air dry	60 min

DNA was fixed permanently to the membranes using the autocrosslink facility on a Stratalinker UV crosslinker (Stratagene).

Bacterial debris was removed from the membranes by incubation in a solution of 5X SSC, 0.5% SDS, 1mM EDTA pH 8.0 for 30 minutes at 50°C with gentle agitation. The residual debris was removed by gently rubbing the membranes with gloved fingers.

Hybridisation of the membranes to an ( $\alpha$ -<sup>32</sup>P)dATP labelled DNA probe was carried out as described in section 2.11.3.

### 2.13 Pulsed Field Gradient Gel Electrophoresis (PFGE)

Agarose blocks containing *P. yoelii* YM chromosomes were separated by pulsed field gel electrophoresis using a contour-clamped homogeneous

electric field (CHEF) apparatus. A 1% agarose gel was run at 14°C in 0.25 X TBE over a period of sixty hours. The ramp conditions were as follows:

90 sec to 300 sec : 95 volts for 30 hr

300 sec to 720 sec : 85 volts for 30 hr

## 2.14 Dot blots

One hundred to two hundred nanograms of DNA was heated at 100°C for 10 minutes and the tubes immediately placed on ice. An equal volume of 1M NaOH was added to the solution and the reaction incubated for 20 minutes at room temperature. 0.5 volumes of neutralisation solution were added to the reaction and the tubes placed on ice immediately. The samples were loaded onto Hybond-N nylon membrane (Amersham) in small aliquots of 3µl, allowing the membrane to absorb the first aliquot before adding the next. The membrane was allowed to air dry for 60 minutes and the DNA crosslinked to the membrane using a UV-Stratalinker (Stratagene). The DNA was hybridised to a radioactively labelled DNA probe as described in section 2.11.3.

**CHAPTER THREE**

**CONSTRUCTION OF A SIZE-FRACTIONATED  
SUB-GENOMIC LIBRARY**

### 3.1 Construction of a size-fractionated sub-genomic library

Previous efforts to clone the entire open reading frame for one of the members of the *P. yoelii* YM multigene family encoding the 235 kDa rhoptry proteins had focused on the family member designated Py235.E8. The sequence at the 3'-end of the gene had extended to, and beyond, an in-frame stop codon and putative polyadenylation signal (Keen et al., 1994). The sequence at the 5' end had not extended into non-coding sequence. However, features resembling a signal sequence and a putative initiating methionine residue had been identified in clone E8-END, a clone derived from an inverse PCR product which contained the most 5' sequences of the Py235.E8 gene (Sinha et al., 1996).

The aims of this project were to complete the coding sequence of the Py235.E8 gene, and also to analyse non-coding sequence at the 5'-end of the gene in order to determine the transcription start site and to identify promoter elements in the gene.

To obtain further DNA sequence at the 5'-end of the E8 gene a sub-genomic library was constructed. A size-fractionated subgenomic library was selected to enhance the specificity of the library and to reduce the number of colonies that needed to be screened. A specific size fraction of restriction enzyme digested *Plasmodium yoelii* YM genomic DNA which had been demonstrated to contain the target sequence was cloned into a plasmid and the library screened by colony hybridisation.

Small-scale restriction endonuclease digests were carried out on 1µg *P. yoelii* YM genomic DNA. The enzymes were selected because of the proximity of their cleavage sites to the 5'-end of clone E8-END, but their

sites were far enough downstream to allow hybridisation of a single DNA fragment to a radiolabelled probe corresponding to the first 270bp of clone E8-END. The DNA fragment was generated by digestion of a subclone - E8.5', a *Dra* I restriction fragment of clone E8.END which had been cloned into pUC18 by Dr. J. Keen.

Southern blot analysis showed that *Dpn*II-digested DNA gave a positive hybridisation signal at a size of approximately 950bp (figure 3.1). This enzyme was selected to make the sub-genomic library because DNA digested with it can be cloned directly into *Bam*HI digested pBluescriptSK-.

Twenty-five micrograms of *P. yoelii* YM genomic DNA was digested to completion with *Dpn*II and run on a 1% agarose gel. Several thin slices were excised from the gel at the size indicated by the Southern blot, and the DNA purified from them. One-quarter of each fraction was run on a 1% agarose gel, followed by Southern transfer to a solid support, and hybridisation of the blot to the same 270bp probe as used in the original Southern blot analysis. This demonstrated that the target sequence was present in only one of the fractions, fraction five (figure 3.2).

DNA from the positive size fraction was ligated to pBluescriptIISK- (Stratagene) which had been digested with *Bam*HI and treated with alkaline phosphatase, and the ligation subsequently transformed into *E. coli* SURE cells.

Sambrook *et al* (1989) describe the required number of recombinants to be screened in a genomic library as:

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

where N is the number of recombinants, P is the probability of finding a specific sequence, and f is the fraction of the genome contained in a single recombinant clone. In this case, with a size-fraction of 950bp and a genome size of  $2 \times 10^7$ bp,  $f = 950 / 2 \times 10^7 = 4.75 \times 10^{-5}$  and,

$$N = \frac{\ln(1-0.99)}{\ln(1-4.75 \times 10^{-5})} = 96948$$

The *DpnII*-digested size-fraction purified for the library construction was estimated, by agarose gel electrophoresis and ethidium bromide staining, to contain approximately 5% of the total genomic DNA, indicating a further 20-fold enrichment. Thus, for a probability of 0.99 of detecting a specific sequence in the sub-genomic library, approximately 5000 recombinant colonies had to be screened.

10000 recombinant colonies were screened by colony hybridisation with the most 5' 270bp of the E8 gene. A single positively reacting clone was detected, designated D1, and this was analysed further.

### 3.2 Characterisation of clone D1

The D1 clone was initially characterised by restriction endonuclease digestion and Southern blot analysis. Digestion of the purified plasmid with *Bss*HII to liberate the insert demonstrated an insert size of 2kb, rather than 950bp as anticipated suggesting that the clone contained two tandem inserts. Further digests with *EcoRI* and *XbaI* showed that the clone did contain two inserts, with an *EcoRI* site situated approximately 100bp into the insert of interest.



The clone was sequenced on both strands. Analysis of the sequence of the insert of interest showed that the clone did NOT represent the 5'-end of the E8 gene. It did, however, show an extremely high homology with the 5'-ends of clones E3 and E8 at both the nucleotide and amino acid levels, suggesting that this clone was part of another gene which was a member of the same multigene family.

Analysis of the DNA sequence of the 950 bp of interest of clone D1 demonstrated that it comprised an open reading frame (orf) of 315 amino acids (figure 3.3). This orf, along with the DNA sequence itself, was analysed using BLAST and FASTA homology search programs to find related sequences. The results of such searches showed an extremely high degree of relatedness between D1 and the most 5' sequences of both clones E3 and E8. At the amino acid level, E8 and D1 showed a 45% level of identity, with the identity being particularly marked in the latter half of the D1 sequence (figure 3.4).

The high degree of identity between the three sequences lends weight to the suggestion that clone D1 represents the partial sequence of another member of the multigene family which encodes the 235 kDa rhoptry protein of *Plasmodium yoelii* YM. The high identity of the two sequences is the probable explanation as to why the D1 clone was detected with a probe derived from E8.

The homology between clone D1 and E8 does not extend upstream of the previously identified 5' end of E8. Instead the homology is internal, and extends from position 67 to 382 in the amino acid sequence of E8. This led to the suspicion that the probe which had been used in the original Southern blot (figure 3.1) and in the subsequent library screens was not, in fact, the *Dra*I fragment from the 5' end of clone E8.END.

Analysis of the E8.END sequence shows that there is another *Dra*I fragment of 287bp located further downstream. Additional examination of the positively reacting fragment sizes obtained in the Southern blot (figure 3.1) suggest that the subclones had been incorrectly identified and the clone labelled E8.5' contained sequence from 404 to 691 of clone E8.END. This is summarised in figure 3.6.

Previous reports have suggested that the translation start site of the E3 gene (Keen et al., 1994) and the E8 gene (Sinha et al., 1996) had been identified, along with putative signal sequences. Analysis of the sequence of D1 in the corresponding region showed that the purported initiating methionine residue is absent. This could mean any one of three things:-

1. Translation of the genes does not initiate at the position originally thought,
2. The start site of translation is not identical in the three genes,
3. D1 is not translated at all, i.e.. it is a pseudogene.

Any one of these scenarios is possible, however further work which will be discussed in later chapters suggests that the translation initiation site of the genes had been incorrectly assigned.

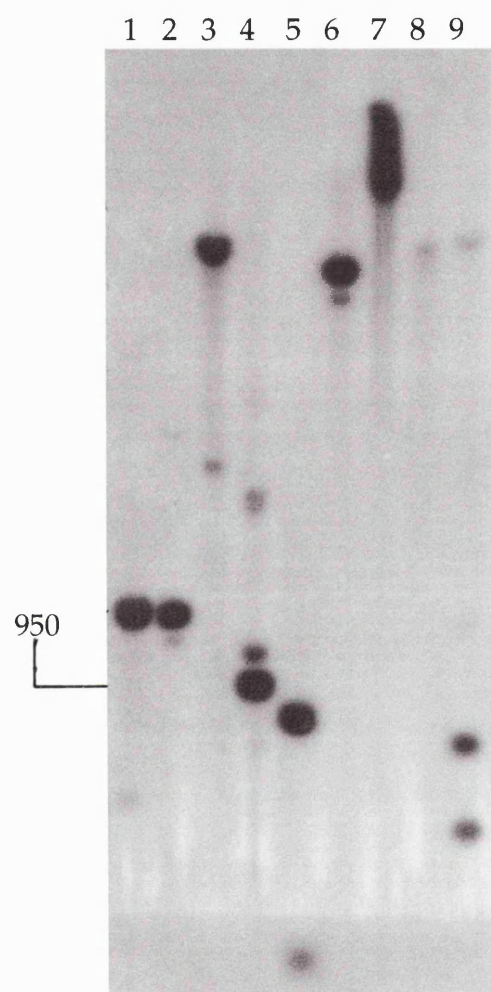
Interestingly, a significant homology with the genes encoding the reticulocyte binding proteins (RBPs) of *Plasmodium vivax* was noted. RBP-1 and 2 are apically localised proteins of *P. vivax* merozoites which have been proposed to be involved in determining host cell specificity for this parasite (Galinski and Barnwell, 1996; Galinski et al., 1992; Sinha et al., 1996). Previous reports have shown a region of high homology between

the proteins in the C-terminal portion, with lower homology throughout the sequences. This homology is present in the D1 clone, again supporting the view that D1 is part of another member of the *P. yoelii* YM multigene family.

**Figure 3.1** Southern blot of restriction enzyme digested  
*Plasmodium yoelii* YM genomic DNA using the an E8  
specific probe

One microgram *P. yoelii* YM genomic DNA was digested to completion with each of the following restriction enzymes. The DNA was run on an agarose gel and Southern blotted onto a nylon membrane. The blot was hybridised to a radioactively labelled probe corresponding to the first 360 bp of clone E8-END, and washed at high stringency (see Materials and Methods).

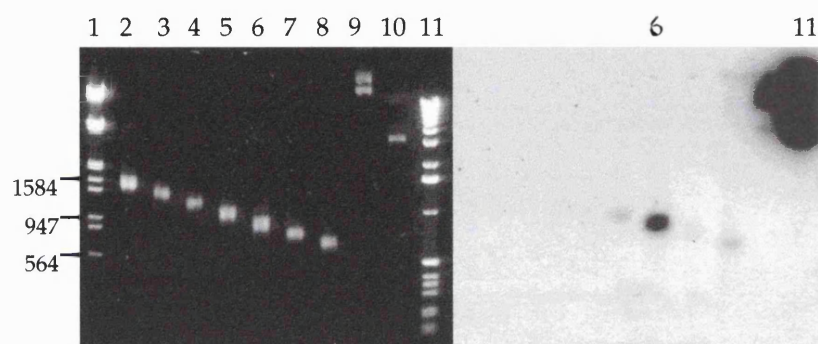
Lane	Enzyme used in digest
1	<i>Alu</i> I
2	<i>Mae</i> III
3	<i>Hph</i> I
4	<i>Dpn</i> II
5	<i>Ase</i> I
6	<i>Eco</i> RI
7	<i>Bam</i> HI
8	<i>Fok</i> I
9	<i>Dde</i> I



**Figure 3.2** Size fractions of *DpnII* digested *Plasmodium yoelii* YM genomic DNA probed with the first 360 bp of clone E8-END

Each size fraction of *DpnII* digested *P. yoelii* YM genomic DNA (lanes 2-8) was run on an agarose gel, Southern blotted, and hybridised to a probe corresponding to the first 360 bp of E8 in order to select the fraction that contained the DNA of interest. The blot was washed at high stringency (see Materials and Methods).

Lane	
1	$\lambda$ DNA digested with <i>EcoRI</i> & <i>HindIII</i>
2-8	Size fractions of <i>DpnII</i> -digested <i>P. yoelii</i> YM genomic DNA
9	clone E3
10	clone E8-END
11	1 kb ladder (Gibco BRL)



**Figure 3.3** DNA and deduced amino acid sequence of D1

Clone D1 was sequenced using Sequenase version 2.0 (USB).

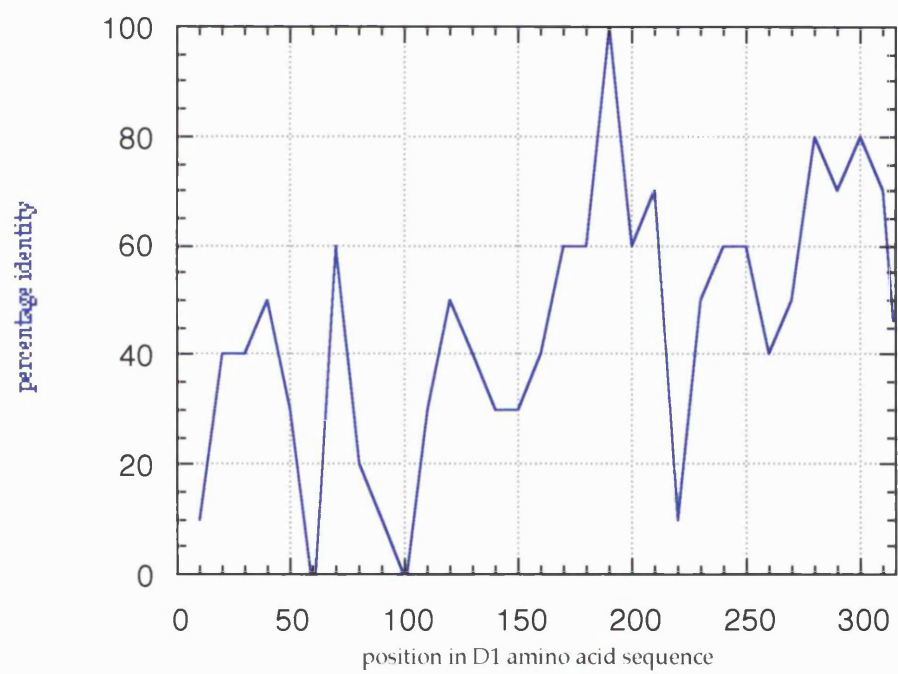
The region of interest is 946 bp long and comprises an open reading frame of 315 amino acids.



GATCCAAATAACTAACGTATTTAAAGAATATATTTATGAATTTAATCTCTGGATACCC 60  
 D P N N T N V F K E Y I Y E F N L W I P  
 AAATCGAAATTAACATTATTAGAAAAAGATTTTTGAAATATTTGAAGAGAAATGGAAT 120  
 K S K L T L L E K R F F E I F E E K W N  
 TCTTATGAAATTAATAAGATATTGACGAAAATAGTAAACAATATAATGTTGTGAAATTA 180  
 S Y E I K K D I D E N S K Q Y N V V K L  
 ATTTTGCAATACATGAAGGAATTAACAGACGTAATTGATTTTATGGAAAATTACCAACCA 240  
 I L Q Y M K E L T D V I D F M E N Y Q P  
 GATGAAGTTCCTACAAGAAGGATAATTGGTTTTGAAGTCGAATGGCGTATAAAAGCATT 300  
 D E V P T R R I I G F E V E W R I K A L  
 CTTTATTCAGAAGTAGAAGCTGGAGTGAAGAAATCTTATGAATCAGTAAAAAATTGGAGA 360  
 L Y S E V E A G V K K S Y E S V K N W R  
 AAATCAATCCTCGAAATAAACAAAAAATTAGAGGAGGAAAATGAAAAAGTTATTAAATTG 420  
 K S I L E I N K K L E E E N E K V I K L  
 GAAACACAAATTAAGATTTATTTAACCAATATTTGAAAATAATTGATGAAAATATATAT 480  
 E T Q I K D L F N Q Y L K I I D E N I Y  
 ATAAACAAGTTAAAATTAGAAATTAAGAAAAAATAAAAAATATATCTGACAAAAATGAA 540  
 I N K L K L E L K E K I K N I S D K N E  
 TATATTAATAAAGCAGTTGACTTAAGAAGATAATAGAAAATAACAATGCATACATTGAT 600  
 Y I K K A V D L K K I I E N N N A Y I D  
 GAATTAGCTAAAACCTCGCCATATCAAGTTCCGGAACATTTAAAGAGCACAGATACAATA 660  
 E L A K T S P Y Q V P E H L K S T D T I  
 TATAGTACAATAAAATCAGAGTTATCCCAAATTTATGAAGATGACATTGATAAACTCTAT 720  
 Y S T I K S E L S Q I Y E D D I D K L Y  
 AATGAATTATCTTCTATAGTTCAAGAAAATGACATTGACAATGTAGAAGATAAAACGAAA 780  
 N E L S S I V Q E N D I D N V E D K T K  
 CTTGACGATTTACAATCTAAATAGATAATGTATATAGTAAATCCAAACATGGAAACT 840  
 L D D L Q S K I D N V Y S K I Q N M E T  
 GCAACAGTTGAATCACATCTAACGAATATCGAACTAACAAAAATAAACTATCAGACACA 900  
 A T V E S H L T N I E T N K N K L S D T  
 ATTGTGGAGATAAAAAAATATATATATGGAGAAATTAGCAAAGATC 946  
 I V E I K K Y I Y G E I S K D

**Figure 3.4** Similarity of clone D1 with E8-END

The sequences of clone D1 and clone E8-END were aligned using the MegAlign program of DNASTar. The overlapping regions were divided into groups of ten amino acids and the identity plotted as a function of their position in the D1 amino acid sequence.



**Figure 3.5** Alignment of D1 and E8

The deduced amino acid sequences of clone D1 and the corresponding region of the E8-END clone were aligned using the MegAlign program of DNASTar. Identical residues are highlighted in green. The numbers indicate the amino acid position in the open reading frame of each clone.

The methionine residue which had previously been suggested to be the site of translation initiation of gene E8 is highlighted in red, and the putative signal sequence which had been assigned is underlined.

[illegible]

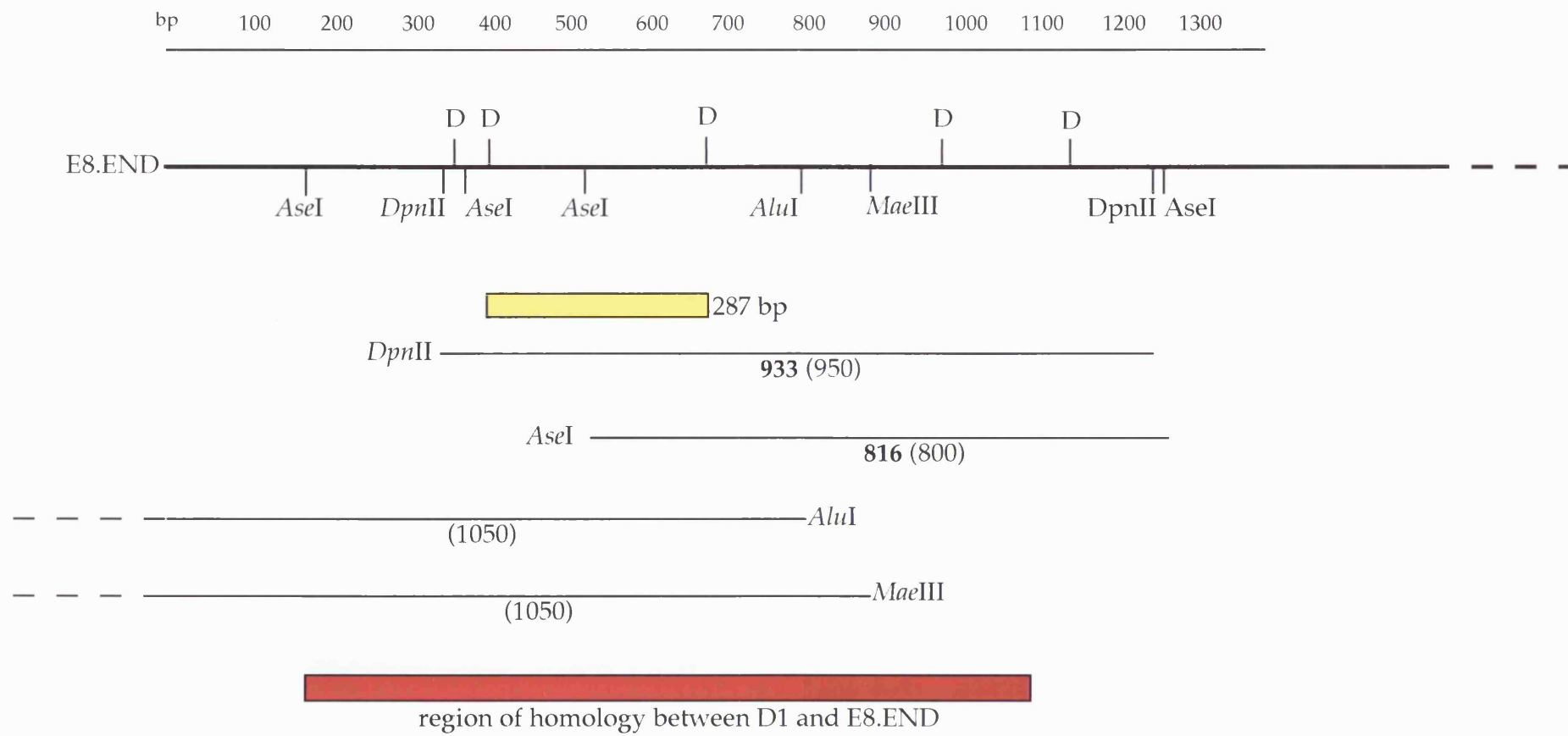
**Figure 3.6** Southern blot analysis

Comparison of the size fragments obtained in the Southern blot (values showed in parentheses) and those which would theoretically be detected if a probe derived from nucleotides 404 to 691 was used in the hybridisation (values in bold type) supports the theory that the clone E8.5' had been incorrectly assigned. Analysis of the additional sequence obtained upstream of clone E8.END described in Chapter Four corroborates this, with restriction sites for *Mae*III and *Alu*I being present at the appropriate positions upstream to give the fragment sizes detected in the Southern blot.

D = *Dra*I



the region of clone E8.END which appears to have been used as a probe in the library screen



## **CHAPTER FOUR**

### **CONSTRUCTION AND SCREENING OF A** ***Plasmodium yoelii* YM DNA GENOMIC LIBRARY**



#### 4.1 Construction of a *Plasmodium yoelii* YM genomic library

The experience with the previous subgenomic library clearly demonstrated the problem of picking out related sequences in a library screen. It was decided to construct a genomic library because the sequence of interest would definitely be represented therein. The method used for construction of the library was a slight variation on the method of Dr. Ekkehard Werner (Werner et al., 1997). *P. yoelii* YM genomic DNA which had been partially digested with the restriction endonuclease *Sau3A* and size-fractionated was cloned into pBluescriptIIS<sup>+</sup>. The 2-5 kb size-fraction was selected because it was thought that these fragments could be stably propagated in plasmids within a bacterial host, where larger A+T-rich sequences have previously been shown to be unstable. In addition *E. coli* SURE cells were used as the host, whose genotype had been altered by deletion of *recA* genes in an attempt to prevent certain recombination events. The library was plated onto Biotrans A transfer membranes for ease of handling and storage. Partial digestion of the genomic DNA also meant that clones within the library would overlap, and could therefore be used to "walk" along genes.

A range of *Sau3A* concentrations was tested in preliminary experiments in order to determine which ones gave a suitable range of molecules in the 2-5 kb size fraction. This range was from 0.008-1U  $\mu\text{g}^{-1}$  *P. yoelii* YM DNA (Figure 4.1.1). On this basis, large scale digests were carried out using 0.06-0.5U *Sau3A*  $\mu\text{g}^{-1}$  DNA.

The 2-5 kb fractions of these digests were gel purified, and the 5'-overhangs partially filled in using the Klenow fragment of DNA polymerase I and dGTP and dATP. The insert DNA was ligated to pBluescriptII KS<sup>+</sup> which had been digested with *XhoI* and the overhang partially filled in using

dCTP and dTTP. The logic behind the partial fill-in reactions was as follows: the partial fill-ins make the overhangs left by the respective restriction endonuclease digests compatible, and it also ensures that only one insert fragment can ligate to one vector fragment as after the partial fill-ins the insert overhangs become incompatible with one another (Figure 4.1.2).

After transformation of the ligated DNA into *E. coli* SURE cells, the library was plated directly onto Biotrans A transfer membranes and stored at -80°C so that it could be used to isolate clones over a long period.

**The library contained approximately 100,000 recombinant clones.**

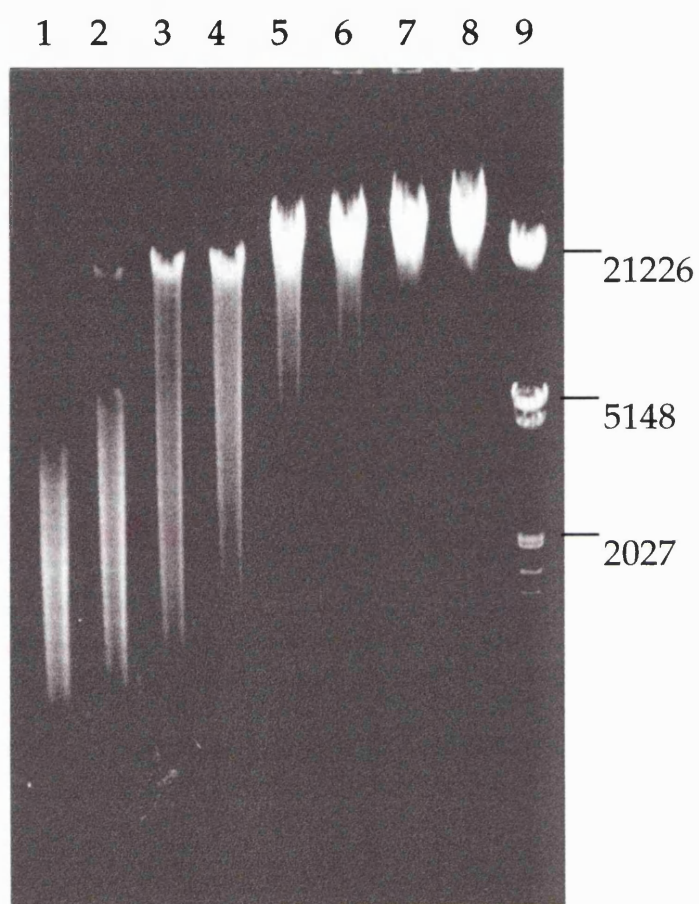
Five clones were picked and the insert sizes determined, which ranged from 2 to 4.8 kb. The average insert size of recombinant clones in the library was determined as 2.8 kb.

**Figure 4.1.1** Partial *Sau*3A digests of *Plasmodium yoelii* YM genomic DNA

Lane	units <i>Sau</i> 3A per microgram DNA
1	1
2	0.5
3	0.25
4	0.125
5	0.063
6	0.031
7	0.016
8	0.008

Each of the digests contained 1µg *Plasmodium yoelii* YM genomic DNA and was incubated at 37°C for one hour.

Lane 9 contains λ DNA digested with *Eco*RI and *Hind*III



**Figure 4.1.2** Construction of a genomic DNA library of  
*Plasmodium yoelii* YM from partially *Sau3A* digested,  
size-fractionated DNA

*Plasmodium yoelii* YM genomic DNA was partially digested with *Sau3A* and the 2-5 kb size fraction purified following agarose gel electrophoresis. The overhangs left by digestion were partially filled in using the Klenow fragment of DNA polymerase I in order to make the ends compatible with *XhoI* digested vector which had also had its ends partially filled in.

Following ligation and transformation into *E. coli* SURE cells, the library was plated directly onto nylon membranes and stored at -80°C.

partially digest *P. yoelii* YM  
genomic DNA with Sau3A

digest pBluescriptII KS<sup>+</sup>  
with XhoI

purify 2-5 kb fraction

partially fill in overhang  
5' ——— GA  
3' ——— CTAG

partially fill in overhang  
TCGAG ——— 3'  
CTC ——— 5'

————— GATCGAG —————  
————— CTAGCTC —————

transform into  
*E. coli* SURE competent cells

plate directly onto  
Biodyne A transfer membranes

store at -80°C

## 4.2 Screening the genomic library with an E8-specific probe to obtain additional 5' sequence

The library was screened with a probe generated from the first 360 bp of E8. The probe was made by PCR, using primers E8-5' and E8-7, from the clone E8-END supplied by Dr. Jane Keen. The primer sequences and positions are shown in figure 4.2.2. This probe was selected as it would hybridise to sequences within the library which contained additional upstream sequence. The first *Sau3A* site within clone E8-END lies at position 344 in the nucleotide sequence. Consequently there would be minimal reactivity of the probe with any clones which did not contain sequence upstream of the 5'-end of clone E8-END.

Four double-positive clones were identified in this screen, and after single colonies had been isolated in a secondary colony hybridisation experiment, plasmid DNA was purified from each one. To ensure that these signals were unambiguous, two methods were employed to screen the clones:

- 1) dot-blot hybridisation to the 360 bp E8 probe
- 2) PCR using primers E8-5' and E8-11. This reaction should give a PCR product of 193 bp based upon previously identified E8 sequence (Dr. Jane Keen, unpublished data). The sequences and positions of the primers are shown in figure 4.2.2.

Dot-blot screening of the clones using the first 360 bp of E8 showed that all four clones gave positive hybridisation signals (figure 4.2.1). PCR analysis suggested that only one of the clones contained E8 sequence - clone 8.2.8 (figure 4.2.3).

This clone was sequenced using T3 and M13-20 primers based on the sequence of pBluescript, and the insert size analysed by restriction enzyme digestion.

Restriction endonuclease digestion of clone 8.2.8 revealed an insert size of 350 bp (data not shown). Considering that the library from which the clone was isolated was constructed in such a way as to give insert sizes of 2-5 kb, there was a high probability that the clone had undergone a rearrangement event in the bacterial host.

Sequence analysis of the clone showed that it was identical to the 5'-end of the sequence previously reported for the E8 gene. There was no cloning site at the 5'-end, and therefore no *Sau3A* site present in *P. yoelii* genomic DNA at this point, again suggesting a recombination event (figure 4.2.4).

To confirm this hypothesis, and to determine the true sequence, a strategy based on inverse PCR was employed.



**Figure 4.2.1** Dot blot analysis of clones isolated from the genomic library with an E8 specific probe

DNA from the clones of interest was denatured and applied to a nylon membrane. The DNA was probed with the first 370 bp of clone E8 and the blot washed at high stringency (0.1 X SSC, 0.1% SDS, 65°C).

	CLONE
1	8.1.6
2	8.2.1
3	8.2.3
4	8.2.4
5	8.2.6
6	8.1.3
7	pBluescript
8	E8-END

The six clones here were obtained in a screen of the genomic library with the probe indicated above. Clone 8.2.8 (not shown in this figure) was obtained in the same screen. Clones 8.1.6, 8.2.1 and 8.2.3 were selected for further analysis owing to the fact that they gave the strongest hybridisation signals in this dot blot. Clone 8.2.8 was sequenced because it gave a positive result in the PCR screen (figure 4.2.3).

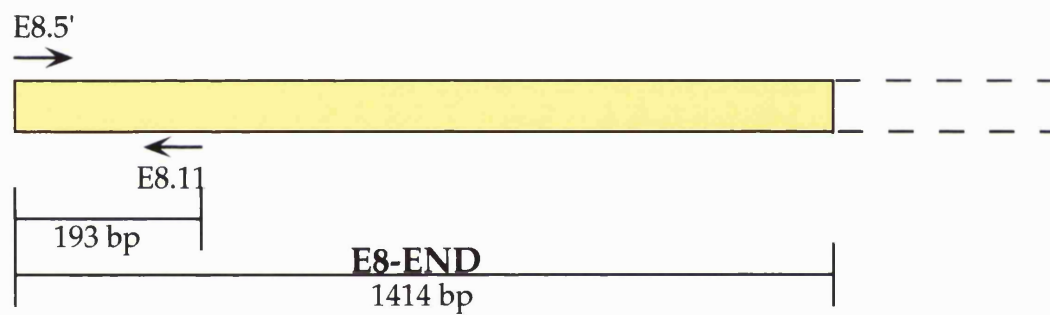


**Figure 4.2.2** Sequences of primers used to screen E8-positive clones

The clones which were positive for the 5' end of E8 by hybridisation were screened by PCR. The sequences of the primers used to screen the clones are given with respect to their position in clone E8.END.

**E8.5'**: nucleotides 1-16 of clone E8-END  
5'-CTTGCATGCCTGCAGG-3'

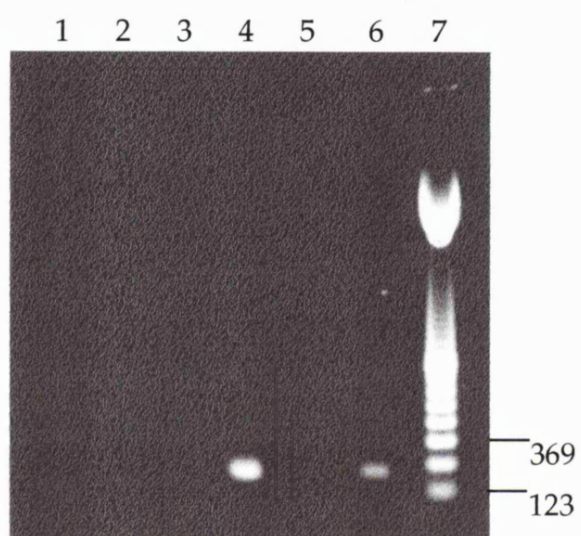
**E8.11**: the reverse complement of nucleotides 340-360 of clone E8-END  
5'-GTATCTTTGAATTGATCAAG-3'



**Figure 4.2.3** PCR of positive clones using primers based on the 5'-end sequence of clone E8-END

PCR reactions were carried out using primers E8-5' and E8-11. Ten microlitres of each reaction was run on a 1.5% agarose gel.

1	8.1.6
2	8.2.1
3	8.2.3
4	8.2.8
5	pBluescript
6	E8
7	123 bp ladder (Gibco BRL)



**Figure 4.2.4** DNA and deduced amino acid sequence of clone 8.2.8

Clone 8.2.8 was sequenced using primers T3 and M13 -20 based on sequences of pBluescriptII flanking the multiple cloning site. The clone is 347 bp long and has an open reading frame of 155 amino acids.

CTTGCATGCCTGCAGGTCGACTCTAGATAATAATAATTGTGATAAACATTTTAATGAAAT 60  
L H A C R S T L D N N N C D K H F N E I  
TAAAAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAATTTTAGAAAATTT 120  
K K I A E S Y S I I F Y N Y T I L E N L  
AGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAGTAGGGAA 180  
E S V N I T Y K E S L D Y F F N S L G K  
ATTATTAATAAACAAAGTAGATTCCGATGGAAATATAATAGAAGAAAATGATATTGATAA 240  
L L I N K V D S D G N I I E E N D I D N  
TTTTGATTTGTCTAAACCCAAAAATAATTTTAAATTATTAGAAGGTGAATTAAATGGAGT 300  
F D L S K P K N N F K L L E G E L N G V  
ATTTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATC 347  
F E N K W N D Y K N N K N L D



### 4.3 Inverse PCR

Inverse PCR was carried out on *P. yoelii* YM genomic DNA in order to determine the true sequence upstream of the 5' end of clone E8-END. The products produced in these experiments would be sequenced directly, without any intervening cloning procedures to prevent possible recombination events. Inverse PCR is a variation on conventional PCR designed to identify sequences which lie outside the bounds of known DNA sequence (Triglia et al., 1988) and therefore can be used to "walk" along genes. Primers in inverse PCR face in the opposite orientation to conventional PCR, i.e.. they face away from one another (figure 4.3.1). *P. yoelii* YM genomic DNA was digested to completion with either *DpnII* or *AluI*, the DNA ligated at very low concentration ( $<3\mu\text{g ml}^{-1}$ ) to encourage intramolecular ligation, and subjected to Touchdown PCR using E8-specific primer pairs E8-11 and E8-12 (figure 4.3.2).

Products were analysed by agarose gel electrophoresis. The *AluI* reaction gave a product of ~900 bp, and the *DpnII* reaction a product of ~500 bp (figure 4.3.3). The *DpnII* reaction product was purified, and sequenced directly using the ABI 377 machine with primers E8-11 and E8-12 (figure 4.3.4). The product was not subcloned as this may have resulted in rearrangement of the insert sequence, as was believed to have happened to clone 8.2.8.

The sequence of the *DpnII* inverse PCR product differed from that of E8 and 8.2.8 in the first 25 bp of the previously reported sequence (figure 4.3.4). As the sequence was obtained by PCR and no cloning steps were involved any possible recombination events liable through passage of the DNA through a prokaryotic host were avoided.

To check the authenticity of the new sequence obtained by inverse PCR, a primer was designed close to the 5'-end - E8.NEW - and used in PCR with primers from the previously reported sequence - E8-7 and E8-11 - using *P. yoelii* YM genomic DNA as a template. A primer based on the "different" region of E8 was designed from the old sequence -E8.5'. The sequences and positions of the primers are shown in figure 4.3.5.

The products obtained with each set of primers were analysed by agarose gel electrophoresis. This confirmed that the new sequence was in fact correct, with the products being of the anticipated sizes. By contrast, PCR using E8.5' and either of the downstream primers gave no products at all (figure 4.3.5).

When the first 25 bp of the old sequence was put through a database search it became obvious that this stretch of nucleotides was derived from the vector polylinker, rather than *P. yoelii* DNA, and appears to occur as a result of rearrangement of insert and vector sequences. The original E8-END sequence was cloned into the pMAL vector which is a derivative of the *E.coli* maltose binding fusion protein vector pCG806, and pBluescript contains an identical sequence (not shown).

Synthetic mRNA leader sequence K.

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

**E8 end:** 1 CTTGCATGCCTGCAGGTCGACTCTA 25

**Subject:** 27 CTTGCATGCCTGCAGGTCGACTCTA 51

Cloning vector pGEM3 polylinker sequence (SP6 polymerase transcription vector)

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Plus / Plus

**E8 end:** 1 CTTGCATGCCTGCAGGTCGACTCTA 25

**Subject:** 27 CTTGCATGCCTGCAGGTCGACTCTA 51

E.coli vector pCG806 encoding modified maltose binding fusion protein

Identities = 25/25 (100%), Positives = 25/25 (100%), Strand = Minus / Plus

**E8 end:** 25 TAGAGTCGACCTGCAGGCATGCAAG 1

**Subject:** 35 TAGAGTCGACCTGCAGGCATGCAAG 59

**Figure 4.3.1 Inverse PCR**

Inverse PCR is a method to amplify regions of DNA which lie outside the bounds of known sequence.

Genomic DNA is first digested to completion with a chosen restriction enzyme. The restriction fragments are incubated with T4 DNA ligase at extremely low DNA concentrations to allow circularisation of individual restriction fragments. Primers, whose sequences are based on the region of known sequence, are designed in such a way so that once circularisation of the template has occurred they face towards one another, whereas in the non-circular DNA fragment they face away from one another.

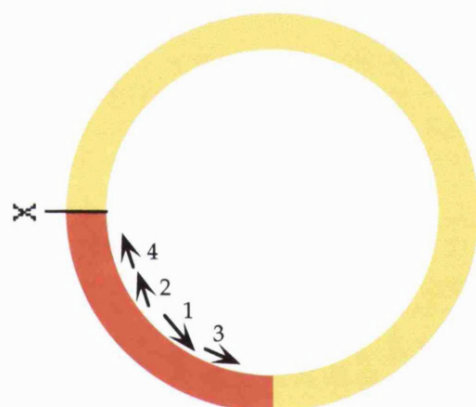
Nested PCR is carried out using two pairs of gene-specific primers. The PCR product contains regions of known sequence at either end and unknown sequence in the middle.



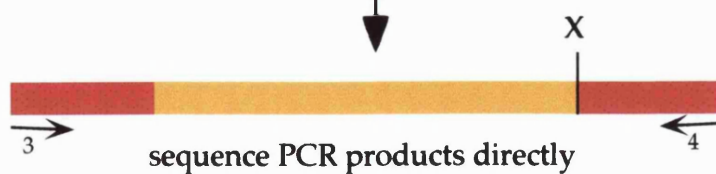
digest DNA to completion  
with restriction enzyme X



ligate at low concentration to  
encourage intramolecular ligation



PCR using primer pair 1 & 2  
Nested PCR using primer pair 3 & 4



**Figure 4.3.2** Inverse PCR primers

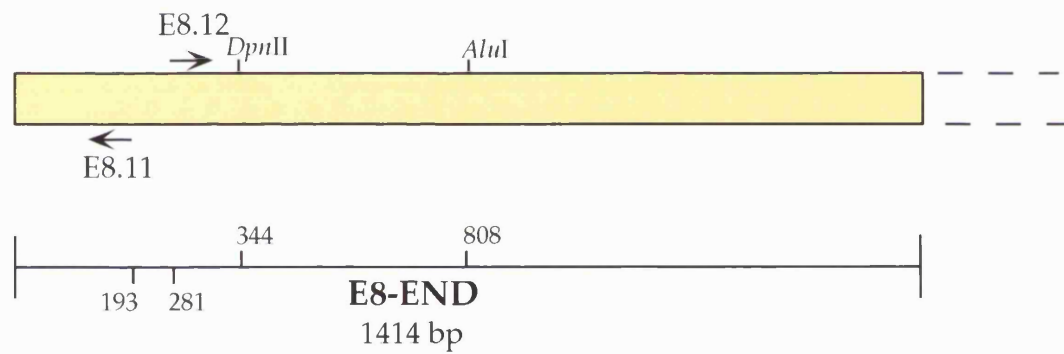
The sequences of the primers used in inverse PCR are shown with respect to their positions in the sequence of clone E8.END. They are in the opposite orientation to primers used in conventional PCR, pointing away from one another.

**E8.11:** the reverse complement of nucleotides 170-193  
in the sequence of clone E8-END.

5'-GTTTATTAATAATTTCCCTAGTG-3'

**E8.12:** nucleotides 281-300 of clone E8-END

5'-GAAGGTGAATTAAATTGGAGT-3'

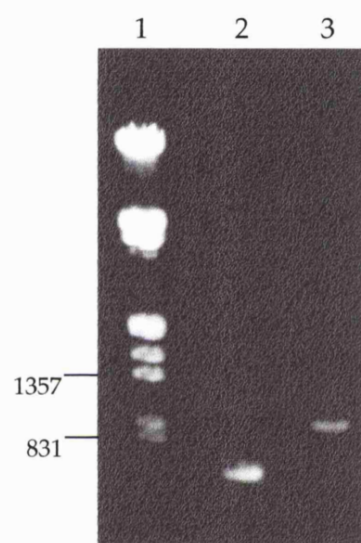


**Figure 4.3.3** Products of inverse PCR

E8-specific primers E8-11 and E8-12 were used in an inverse PCR reaction using restriction enzyme digested *P. yoelii* YM genomic DNA, which had been circularised by ligation, as a template. Ten microlitres of each inverse PCR reaction was run on a 1% agarose gel.

- 1       $\lambda$  DNA digested with *Eco*RI and *Hind*III
- 2      *Dpn*II derived inverse PCR product
- 3      *Alu*I derived inverse PCR product





**Figure 4.3.4** Partial DNA sequence and deduced amino acid of *DpnII* inverse PCR product

Direct sequencing of the PCR product using primers E8-11 and E8-12 gave 463 bp of DNA sequence. This sequence had an open reading frame of 154 amino acids. New sequence is shown in blue, and overlapping sequence with clone E8.END is shown in green. The sequence which differs from the 5'-end of clones E8.END and 8.2.8 is shown in red, and in (i)-blue and (ii)-red.

- (i) 5' sequence of clone E8 and clone 8.2.8
- (ii) Sequence of the corresponding region of *DpnII*-derived inverse PCR product

ACAAATGATATTATTTTCATATGAAGAAAAATTATATACAGATAAATGTAAAAATAGCTCC 60  
 T N D I I S Y E E K L Y T D K C K N S S  
 AAACCTCAAGAATACTACTAAGAAACCTTCTGAATTTTGTAATAATGTGATGAATCTCAAT 120  
 K L K N T T K K P S E F C K N V M N L N  
 GTATCTCGTAAAAATAGTTGGGAAGTTCCCAAAAAATAACGAAATAATTTTCATTTATAGAT 180  
 V S R K N S W E V P K N N E I I S F I D  
 TTTCTAATAGAAAAATTAATCAAAACAATTACCCAATGACTCTTGTAACAAAACCTAGAT 240  
 F L I E K L K S N N Y P M T L V T K L D  
 TTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATATGT 300  
 F I K K Q F E D I K N K H N K H I K I C  
 AAACAAGAAGAGATTGTTGTGAACAAATGTACCAACAACATTGATAATAATAATTGTGAT 360  
 K Q E E I V V N K C T N N I D N N N C D  
 AAACATTTTAATGAAATTAATAATAAGCCGAATCATATAGCATTATATTTTATAATTAT 420  
 K H F N E I K K I A E S Y S I I F Y N Y  
 ACAATTTTAGAAAAATTTAGAATCAGTAAATATTACATATAAGG 463  
 T I L E N L E S V N I T Y K

(i)

L H A C R S T L D N N N C D K H F N E I  
 CTTGCATGCCTGCAGGTCGACTCTAGATAATAATAATTGTGATAAACATTTTAATGAAATT  
 10 20 30 40 50 60

(ii)

V N K C T N N I D N N N C D K H F N E I  
 TGTGAACAAATGTACCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATT  
 10 20 30 40 50 60

**Figure 4.3.5** Confirmation by PCR that the new sequence for E8 is correct

PCR from genomic DNA of *P. yoelii* YM was performed using primers based on the 5' sequence of clone E8-END and the sequence of the *DpnII* inverse PCR product

<u>LANE</u>	<u>PRIMER PAIR</u>
1	E8-NEW & E8-11
2	E8-NEW & E8-7
3	E8-5' & E8-11
4	E8-5' & E8-7

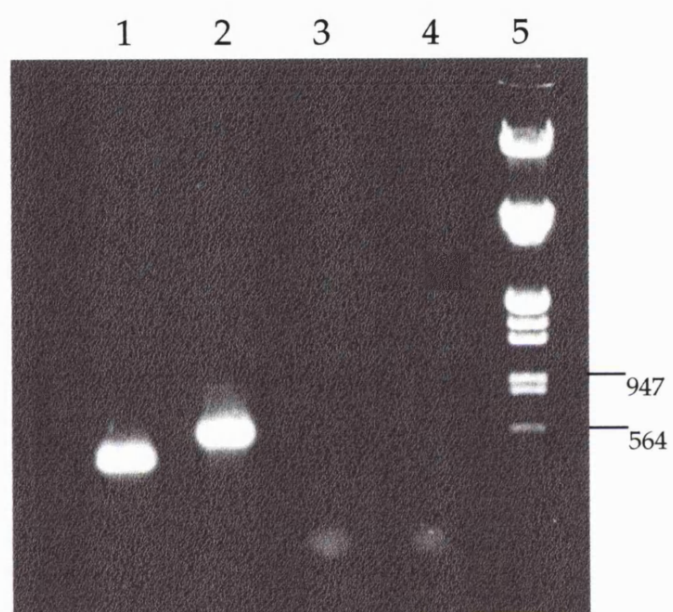
Primer sequences

E8-NEW: 5'- TAGCTCCAAACTCAAGAATAC -3'  
(nucleotides 54-74 of the inverse PCR product sequence shown in figure 4.3.4)

E8-5': 5'- CTTGCATGCCTGCAGG -3'  
(nucleotides 1-16 of E8-END)

E8-7: 5'- GTATCTTTGAATTGATCAAG -3'  
(the reverse complement of nucleotides 340-360 of E8-END)

E8-11: 5'- GTTTATTAATAATTTCCCTAGTG -3'  
(the reverse complement of nucleotides 170-193 of E8-END)



The screen of the genomic library gave four clones which were positive by hybridisation, however only one, 8.2.8, was also positive by PCR. The PCR screen was performed using primers E8-5' and E8-11. The primer E8-5' was designed based on the first sixteen bases of the old E8-END sequence, which has now been demonstrated to be of vector origin. Clone 8.2.8 was the only clone to be positive by PCR because it was the only one to have undergone an identical rearrangement event as the original E8-END clone. Further sequence analysis of clones 8.1.6, 8.2.1, and 8.2.3 demonstrated that these did, in fact, contain additional sequence upstream of the original E8-END clone.

It may be of interest to note that of all the positive clones, the orientation of the insert was the same, except for clones E8-END and 8.2.8 whose 5' ends were proximal to the T7 promoter of the plasmid. It may be possible that such rearrangements are orientation specific. This is merely speculation but is an observation which could, perhaps, warrant further investigation.

#### 4.4 Analysis of E8-positive clones

As has been mentioned, clones 8.1.6, 8.2.1, and 8.2.3 all contained additional sequence upstream of that previously identified for the E8 gene. The insert sizes for each clone were analysed by restriction endonuclease digestion with *Apa*I and *Sma*I. This showed that clone 8.1.6 had an insert of ~3.5 kb, 8.2.1 an insert of ~1.2 kb, and 8.2.3 an insert of ~2.8 kb (figure 4.4.1). As the insert of clone 8.2.1 was smaller than the size range of the original 2-5 kb fraction used for construction of the library, the clone was not analysed further as it was possible that it had undergone some form of rearrangement or deletion in the bacterial host.

Clones 8.1.6 and 8.2.3 were sequenced in their entirety using automated sequencing. The largest of the two clones, 8.1.6, overlapped with the previously known E8 sequence by 2388 bp and gave 1423 bp of new sequence. Clone 8.2.3 had the same 5'-end as clone 8.1.6 but had a shorter overlap with E8 of 1366 bp.

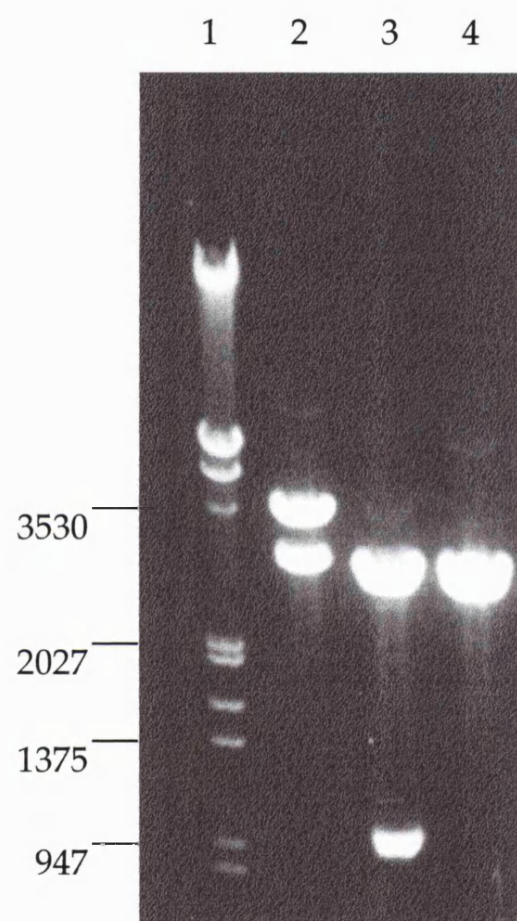
Analysis of the new sequence showed an open reading frame of 361 amino acids continuing upstream of the old sequence (figure 4.4.2). This was preceded by several in-frame stop codons. It was observed that further upstream was a methionine followed by two charged residues then a stretch of hydrophobic residues, consistent with a signal sequence. This sequence will be analysed in more detail in section 5.3.

**Figure 4.4.1** Restriction enzyme digest analysis of the insert sizes of clones 8.1.6, 8.2.1, and 8.2.3

One microgram of each clone was digested with restriction enzymes ApaI and SmaI. The digests were run on a 1% agarose gel.

- |   |  |
|---|--|
| 1 | $\lambda$ DNA digested with <i>Eco</i> RI &<br><i>Hind</i> III |
| 2 | 8.1.6  |
| 3 | 8.2.1  |
| 4 | 8.2.3  |





**Figure 4.4.2** DNA and deduced amino acid sequence of clone 8.1.6

Sequencing of clone 8.1.6 indicated an insert size of 3785 bp.

An open reading frame was apparent from position 340 of the sequence which continues until the end of the clone. The sequence which overlaps with previously identified sequence in clone E8-END and lambda clone E8 is shown in green. A putative intron, with conserved 5' and 3' splice sites (shown in red), is present at position 203 to 340. A putative initiating methionine residue is encoded by an ATG codon at position 150-153, followed by sequence characteristic of a signal sequence, shown in orange. These features will be discussed in further detail in chapter six.

GATCGACAAGTAGAAAATTAAAGCGATTAATGTAGAACAAAATTAATATTAAGAAATAAA 60

ATAAAAATAAAACAAAACATCAAAAATATAGAAGAAAAAACATATATATAATATATTTTT 120

TAAACTCAAAAATATATATTATATTTAACCATGAAAAAATATATTTATATTATTAGTTTG 180

M K K Y I Y I I S L

GCTGCATTTTTTATTTCTTCGTAATCATAGATATCTTTAACAAATTATTTCAATTCTT 240

A A F F I S F .....

GGGTTTACATATTTTATTTAATATGTTTATGTGTATTAATTATGCAAATGTGTAAATTCC 300

.....

TATACTAATATTACTTAATTTGTTTCATCATTAAATATGCAGATATAATTTGTGCAACCAA 360

..... D I I C A T K

AATTGAGGAAAAATAAAACAAGAATGCCACAAAGTTAAACAATTATAATCCATACCATAA 420

I E E N K N K N A T K L N N Y N P Y H N

TTTAGAAGAAAGTGATTTTGATAATTCAAACTTTTTAAATGAAAAAAATATGATAATAA 480

L E E S D F D N S N F L N E K K Y D N K

AAATAATTCCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCT 540

N N S I N N E K Y I Q P H S I N N N F L

TCAAAATAAAAAAGATATAAACAATACAAAAATCACATTGTATAATGAATATAATAAGAC 600

Q N K K D I N N T K I T L Y N E Y N K T

AAATAATATTGATAATTTTAGAAATTTTAATCATGAACATGAAAAACAAATAACATATT 660

N N I D N F R N F N H E H E K T N N I L

AATTGAAAATAATAATTCCTTTTTACAAATAACCGCTATATCTGAAATAAATAAGATAC 720

I E N N N S F L Q I T A I S E I N K D T

GTTAGATACGGTCGATACTATATATGGTGTAATGATAATGAAAGCGTATTGTGGTTTTT 780

L D T V D T I Y G V N D N E S V L W F F

TTATAAATTCTATATTCTGTTGAATATATTAATAAATGTTAAGTAACCTAAATATAGG 840

Y K L L Y S V E Y I K K M L S N L N I G

TCTATCCAGTGAGCAAACTAAAATAAAGACAGATCAAAATGTTGTTATTAATGACGTCAG 900

L S S E Q T K I K T D Q N V V I N D V R

ACAAACAAACGCGTCATGTTATTCTGAAAAAATGAATTAATAAATAAATTGGGCAATAT 960

Q T N A S C Y S E K N E L I N K L G N I

ACACAATCCTTTTTATAATTTTCATAATGATCGTTTCGATAGATGATTATAATTTATATAC 1020

H N P F Y N F H N D R S I D D Y N L Y T

ATCATCAAAAAATAATTTTGTAATTTGTTTAAAAAAGGATTTGAAAATATAAAAAA 1080

S S K N N F V N C L K K G F E N I K K K

AACAAATGATATTATTTTCATATGAAGAAAAATTATATACAGATAAATGTAAAAATAGCTC 1140

T N D I I S Y E E K L Y T D K C K N S S

CAAACCTCAAGAATACTACTAAGAAACCTTCTGAATTTTGTA AAAATGTGATGAATCTCAA 1200

K L K N T T K K P S E F C K N V M N L N

TGTATCTCGTAAAAATAGTTGGGAAGTTCCCAAAAATAACGAAATAATTTCAATTTATAGA 1260

V S R K N S W E V P K N N E I I S F I D

TTTTCTAATAGAAAAATTA AAATCAAACAATTACCCAATGACTCTGTAAACAAACTAGA 1320

F L I E K L K S N N Y P M T L V T K L D

TTTTATAAAAAACAATTCGAGGATATTA AAAATAAACACAATAAACACATAAAAAATATG 1380  
 F I K K Q F E D I K N K H N K H I K I C  
 TAAACAAGAAGAGATTGTTGTGAACAAATGTACCAACAACATTGATAATAATAATTGTGA 1440  
 K Q E E I V V N K C T N N I D N N N C D  
 TAAACATTTTAATGAAATTA AAAAAATAGCCGAATCATATAGCATTATATTTTATAATTA 1500  
 K H F N E I K K I A E S Y S I I F Y N Y  
 TACAATTTTAGAAAAATTTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTT 1560  
 T I L E N L E S V N I T Y K E S L D Y F  
 TTTTAATTCAGTAGGGAAATTATTAATAAACAAAGTAGATTCCGATGGAAATATAATAGA 1620  
 F N S L G K L L I N K V D S D G N I I E  
 AGAAAATGATATTGATAATTTTGATTTGTCTAAACCCAAAAATAATTTTAAATTATTAGA 1680  
 E N D I D N F D L S K P K N N F K L L E  
 AGGTGAATTAAATGGAGTATTTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCT 1740  
 G E L N G V F E N K W N D Y K N N K N L  
 TGATCAATTCAAAGATACAATGAAAAAATAATATTATTAATTATACAATATATGAACGA 1800  
 D Q F K D T M K K I I L L I I Q Y M N E  
 ATTTAAAGGCTTGAATGACGCTATGACAAAATAAAAAATGAAGGTATTTACAAAAAATT 1860  
 F K G L N D A M T K L K N E G I S Q K F  
 TGTGATTAATAATCAAATAAAACAAAAATTCGATAAATCAACTTATGATGAAAAAAAAGA 1920  
 V I N N Q I K Q K F D K S T Y D E K K E  
 AGGGTTTGAAAGTTCTCTTGAATTAGCAAAAAATTTGGGAAAAAAAAAAAACTCGAAATAAT 1980  
 G F E S S L E L A K N W E K K K L E I I  
 AACGGAATTAAAAAATAAATGAAGAACTGTTCAATTGGATATAAAAAATTAGAGAACT 2040  
 T E L K K K N E E T V Q L D I K I R E L  
 AATTAACAAATCAAAGATATTATAGAGGAACAAAAAATAGTGAACGATTTAAAATTAGA 2100  
 I K Q I K D I I E E Q K I V N D L K L E  
 ATTAACAAAAAATAAAGAAATAACTGAGAAAATTGAATATATTA AAAAAAGCAGTTGA 2160  
 L N K K I K E I T E K I E Y I K K A V D  
 CTTAAGAAAGAAATAGAAAAAGACAACGTATATATTGATGAATTAGCTAAGGAACCAACC 2220  
 L K K E I E K D N V Y I D E L A K E P P  
 ATATCAAATTACAAAATATATAGAAAAAATAAATGAAATATATAATACAATAAAATCAGA 2280  
 Y Q I T K Y I E K K N E I Y N T I K S D  
 TTTTGACAAAATTTATGTAGGCGATATTGAACAACCTTTACAATGAAATGTTTTCTGTAGT 2340  
 F D K I Y V G D I E Q L Y N E M F S V V  
 TCAAGAAAGTAACATTGAGCATATAGAAAATAAACAGAAATTCTAACTTTAAAAACGAA 2400  
 Q E S N I E H I E N K T E I L T L K T K  
 AATAGATAATGTATATAATAATATCCAAAACATGGAAACTGAAACAGTTAAATCACATCT 2460  
 I D N V Y N N I Q N M E T E T V K S H L  
 AAAGAATATAGAACTAACAACTATCAGAAACAATTTTGGATATAATAAAATATAT 2520  
 K N I E T N N K L S E T I L D I I K Y I  
 ATATGGAGAGATTACCAACGAGCTAAATAAACATTAGAAGACTTTAAAAATAAAGAAAA 2580  
 Y G E I T N E L N K T L E D F K N K E K  
 AGGACTATCCAATAAAATAGATGAATACGCTAAGGAAATGTCCAATTAATGTTTATAA 2640  
 G L S N K I D E Y A K E N V Q L N V Y K

ATCTAACATATTAGAAATCAGAAAGCATTATAATGATCAAATTAATATAGACAATATTAA 2700  
S N I L E I R K H Y N D Q I N I D N I K  
GGAAAAAGAAGCCAAACAAACTATGATCAATTCAAAGAACATATGAAAACAATACCACC 2760  
E K E A K Q N Y D Q F K E H M K T I P P  
TAACGAAAATGAAATATCAAAAACCATCAATGAGATAAAAATTATGAAAGACGAATTCTT 2820  
N E N E I S K T I N E I K I M K D E F L  
GAGTAAAGTAAATAAATATAATGATTTTGACAAAGTTTATAAAGAAAAGGTTGAATCAGA 2880  
S K V N K Y N D F D K V Y K E K V E S E  
ACACAATAAATTTACTGAATTAACAAATAAAATAAAAACAGAAGTTTCGGATGAAGAAAT 2940  
H N K F T E L T N K I K T E V S D E E I  
AAAAAATATGAAAACAAATTTAATGATAGCAAATCTTTAATTAATGAAACAAAAAATC 3000  
K K Y E N K F N D S K S L I N E T K K S  
CATTGAAGAGGAATATCAAAACATTAATACACTTAAAAAGGTAGATGATTATATAAAAGT 3060  
I E E E Y Q N I N T L K K V D D Y I K V  
ATGTCTAAATCAAAATGAATTAATAACAAATTGTCATAATAAACAACTACATTAAGA 3120  
C L N T N E L I T N C H N K Q T T L K D  
CAAATTAATCAAAATATTAATAACCATAAAAGAACTAATTCAATAGATAAAATTTATAC 3180  
K L N Q N I K T I K E T N S I D K I Y T  
AGACAAGTTTGAAAATATATTAACAGATAAAAAGACAGAATTAGAAACAAAATTCACAGG 3240  
D K F E N I L T D K K T E L E T K F T G  
ATTGTCTTTAAATAATCACGAATCGAATAACAAAGAATTGCTAACATATTTTTATGATT 3300  
L S L N N H E S N N K E L L T Y F Y D L  
AAAGGCAAATTTAGGAAAAAATAAGAAAACATGTTATATAAACAATTCAATGAAAAAGA 3360  
K A N L G K N K E N M L Y K Q F N E K E  
AAAAGCTGTTGAAGATATTAAGAAAAAAATGTAGATATAAATAAAATTGTTTCAAATAT 3420  
K A V E D I K K K N V D I N K I V S N I  
TGAAATAACAATTTATACATCCATTTATAATATTAATGAAGATACAGAAAATGAAATTGG 3480  
E I T I Y T S I Y N I N E D T E N E I G  
AAAAAGTATAGAATTACTAAATACCAAAGTACTTGAAAAGGTAAAAGCAAACGTAACCAA 3540  
K S I E L L N T K V L E K V K A N V T N  
TTTGAATGAAATAAAGGAAAAATTAAGATTATGATTTTCAGGATTTTGAAAAGAGAA 3600  
L N E I K E K L K D Y D F Q D F G K E K  
AAATATTAAATATCCTGATGAAAACAAATTAATAATGATATTGATACATTAAACCAAAA 3660  
N I K Y P D E N K I K N D I D T L N Q K  
AATCGATAAAAGTATAGAGACATTAACGGAATAAAAAATTCAGAGAACCATATTGA 3720  
I D K S I E T L T E I K K N S E N H I D  
TGAAATAAAGGACAAATAGATAAATTAAAAAAAGTACCAATAAAACAATGTTTAACGA 3780  
E I K G Q I D K L K K V P N K T M F N E  
GGATC 3785

D

#### 4.5 Screen of library with the 5' end of clone 8.1.6

Clone 8.1.6 appeared to contain the initiating methionine and the N-terminal coding portion of the Py235-E8 gene. It also contained 150 bp of sequence upstream of the putative translation start site of the gene. One of the aims of the project was to analyse sequences upstream of the transcription start site of one or more of the genes encoding the 235 kDa protein.

In order to get more 5' sequence, the genomic library was screened with a probe corresponding to the first 950 bp of clone 8.1.6. This probe was generated by PCR from *P. yoelii* YM genomic DNA using primers E8-NEW and 8.1.6-5'. The PCR product was separated by agarose gel electrophoresis and purified.

After three rounds of hybridisation to isolate single colonies, two clones were isolated which were positive by hybridisation, clone 5 and clone 7.

##### 4.5.1 Analysis of clone 5 and clone 7

Restriction enzyme digestion of the plasmids with *Bam*HI and *Kpn*I to liberate the inserts showed that clone 5 contained an insert of ~2.2 kb, and clone 7 an insert of ~2.4 kb with an internal *Bam*HI or *Kpn*I site (figure 4.5.1). The clones were sequenced using the ABI 377 system. (figures 4.5.2 and 4.5.3).

Sequencing of the two clones demonstrated that clone 5 had significant homology with the 5'-end of the E8 gene, but appeared to represent part of another member of the gene family. An alignment of the two open reading frames is shown in figure 4.5.4. Clone 5 also contains what appears

to be the 3' portion of an intron in an identical position as the proposed intron of the E8 gene. The 3' consensus dinucleotide splice site for a eukaryotic intron - AG - is present. The intron contains 80% A+T nucleotides. This is not significantly higher than the 77% A+T content of the rest of the clone. Often in *Plasmodium* the non-coding intron sequence contains a higher A+T content than the coding sequence. A comparison of the nucleotide sequences of the two proposed introns was carried out and they were shown to be 91% identical.

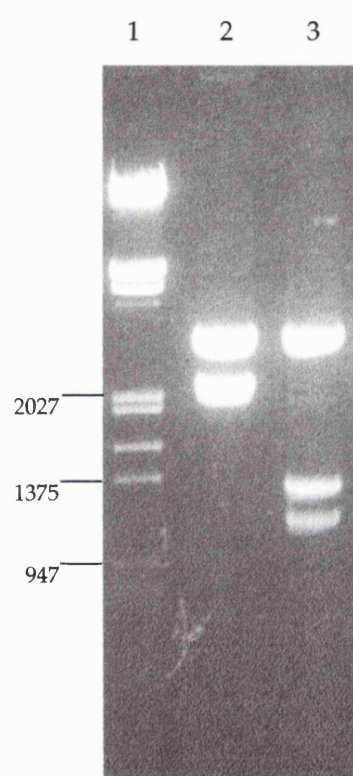
Clone 7 contained more sequence upstream of the 5'-end of clone 8.1.6. The clone is 2439 bp and has an overlap of 823 bp with the 5'-end of clone 8.1.6, and therefore contains 1616 bp of additional sequence. Analysis of potential promoter elements contained within this sequence will be carried out in further chapters.

**Figure 4.5.1** Restriction enzyme digest analysis of the insert sizes of clones 5 and 7

One microgram of each clone was digested with restriction enzymes *Bam*HI and *Kpn*I. The digests were run on a 1% agarose gel.

- 1       $\lambda$  DNA digested with *Eco*RI and *Hind*III
- 2      clone 5
- 3      clone 7





**Figure 4.5.2** DNA and deduced amino acid sequence of clone 7

Clone 7 is 2439 bp long and has an overlap with clone 8.1.6 of 823 bp, which begins at position 1616 and continues to the 3' end of the clone. There is no more coding sequence in the additional 1616 bp. The intron splice acceptor and donor dinucleotides are highlighted in red, and the putative signal sequence is shown in blue.

GATCAAGTAATAGTTTATTCGCCAATTACACACCCATGAAAATTATGTTTATTTATTAC 60  
TTATTATTACTACTATTTTATTATTGTATACTACAGCGGTCTAATTAATGAATACAATAG 120  
AATACTTTAAACCAATCTATAACATTCCATTTATATTTTTTTTAAAATAATTTATTATA 180  
ATATATATATTCTAATAAAAAACGATATATTTAAAACATATATTACGAATTTTATATCT 240  
ATAATACCATAAAATATAAATTA AAAACAAATCATTTGCTATACAAATTTTCTATAAAATC 300  
TACATTCATATATAAATATATTAATTTGCCTGAATCAAATAATATTCAGTATTTAATCAT 360  
CAATATAACATTAAAAAATGTATATAGTAGGATAAATAATTCCAATTTTATATCCCATT 420  
TTTATTCATTATTACAAACATTCTTATATTATACCCATTTTAAAATATATTAATTTTATA 480  
CAAATTTACACTTCCCAAACCTATATATGTGTAGCTATTAAAAATAAAAAATAATTTTAAT 540  
AATATAGTGTTAACCTTTATGAAAAATAAATTAACCATATAAACCATTTCTGCTTTATT 600  
ATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTGTTAATATAATC 660  
CAAATTTCCATTATTTATTATTTCATATATTTATAATAAATTTTATTATTAAGATATATA 720  
ATTAAGTATGGGCAAAAATATAATAATTACCAATTTGGGGATAATTTATAAAAGTTGTTA 780  
TATATGATTAACCAAACTTACCATTAAAATATAATATACTAATATTATACATTAATAAT 840  
GCTTAATCTATATAAACTCATCATATTTTAAGACTAAATTTGATTATAAATTTTTTATTT 900  
ACGATGTTACTTTTATGTTAATTCCTTATATATCTCTCCTGTGTCCAAAATTTTGTCTTA 960  
TCTAGAGGTTAAAAATTCAGGTTGTAGAGATTTTTTTTATAATTTTTACTTATATTTAAA 1020  
CAATTTATTATTTATATATATTGTTTAATAAATATTATCAAATTCTACATATTTGATTAA 1080  
AACAAAACCTCTACACACGGAACCTGGGTACCATCTTAATTTTGGTTCTATATACATGACA 1140  
CTAATTATTTATTCTAAAACCGAAACATGGAGAATTTTAAATTAATTTCCATTGATTT 1200  
TCTGCATGCATAAAAAGTAAATAAAAAGGTCTTTATTTTGAATCACTATTTCCCTTTAT 1260  
TTGGAAATTATTATAATATTTTTTTAATTAATATGTGTAAGTGTTTTGGTGTTTTCCAAT 1320

TTTTGTCTGACACATTTATCTTAAAAAAGTTATAATGCAATTATTAATTTTTTTATT 1380  
 TATATAAATTAAAAAAATGAGAGCATAATTATTAATTCCTTTTTTTATCTACACACC 1440  
 AATAAATAATTTGTTAATTTTTAATTTAAAAATTTTATTTAATGGAAAATTATTTGTTTT 1500  
 ATATACTTACACAAATATTTTTTATTTATATATAAAAAATGAAGAATATTATCTTTAAATA 1560  
 AAATGATCGACAAGTAGAAAATTAAGCGATTAATGTAGAACAAAATTAATATTAAGAAA 1620  
 TAAAAATAAAATAAAACAAAAACATCAAAAATATAGAAGAAAAACATATATATAATATAT 1680  
 TTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAAATATATTTATATTATTAG 1740  
 M K K Y I Y I I S  
 TTTGGCTGCATTTTTTATTTCTTCGTAATCATAGATATCTTTAACAAATTATTTCAAT 1800  
 L A A F F I S F .....  
 TCTTGGGTTTACATATTTTATTTAATATGTTTATGTGTATTAATTATGCAAATGTGTAAA 1860  
 .....  
 TTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATATGCAGATATAATTTGTGCAA 1920  
 ..... D I I C A  
 CCAAAATTGAGGAAAAATAAAAAACAAGAATGCCACAAAGTTAAACAATTATAATCCATACC 1980  
 T K I E E N K N K N A T K L N N Y N P Y  
 ATAATTTAGAAGAAAGTGATTTTGATAATTCAAACCTTTTTAAATGAAAAAATATGATA 2040  
 H N L E E S D F D N S N F L N E K K Y D  
 ATAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATT 2100  
 N K N N S I N N E K Y I Q P H S I N N N  
 TCCTTCAAAATAAAAAAGATATAAAACAATACAAAAATCACATTGTATAATGAATATAATA 2160  
 F L Q N K K D I N N T K I T L Y N E Y N  
 AGACAAATAATATTGATAATTTTAGAAAATTTTAATCATGAACATGAAAAACAAATAACA 2220  
 K T N N I D N F R N F N H E H E K T N N  
 TATTAATTGAAAATAATAATTCCTTTTTACAAATAACCGCTATATCTGAAATAAATAAAG 2280  
 I L I E N N N S F L Q I T A I S E I N K  
 ATACGTTAGATACGGTCGATACTATATATGGTGTAATGATAATGAAAGCGTATTGTGGT 2340  
 D T L D T V D T I Y G V N D N E S V L W  
 TTTTTTATAAACTTCTATATTCTGTTGAATATATTAAAAAATGTTAAGTAACTTAAATA 2400  
 F F Y K L L Y S V E Y I K K M L S N L N  
 TAGGTCTATCCAGTGAGCAAACATAAAATAAAGACAGATC 2439  
 I G L S S E Q T K I K T D

**Figure 4.5.3** DNA and deduced amino acid sequence of clone 5

Clone 5 is 2153 bp long and has an open reading frame of 675 amino acids which begins at position 128 in the DNA sequence. The sequence from position 1 to 128 does not contain coding information in any frame. At positions 127 and 128 is an AG dinucleotide (highlighted in red). This is consistent with the consensus 3' splice site in eukaryotes, and as will be shown in figure 4.5.4 and 4.5.5, aligns exactly with the splice site present in Py235.E8.

GATCTTTAACAAATTATTTCAATTCGTGGGGTTACATATTTTATTTAATATATGTTTATA 60  
.....  
TATTAATTATGCAAATTTGTATATTCCTATACTAATATTACTTAATTTTTTTCACCATTA 120  
.....  
ATATGCA<sup>AG</sup>ATATAATTTGTGCAACCAAAATTGAGGAAAATAACAACAAACATGCCCCAAA 180  
..... D I I C A T K I E E N N N K H A P N  
CTCAAACAATTATATTCATACCATAATTTAGAAGAAAGTGATTTTAAATGATTCAAACCT 240  
S N N Y I P Y H N L E E S D F N D S N F  
TTCAAATGAAAAACAATATAATAATGAAAATAATTCCATTAACAATGAAAAATATATACA 300  
S N E K Q Y N N E N N S I N N E K Y I Q  
ACCACATTCAATAATTAACAACCTCCCTTCTAAATAAAAAAGACATAAATAATACAAAACCT 360  
P H S I I N N S L L N K K D I N N T K L  
AACATTATATAATGAATATAATAAGACAAATAATCTTGATGATTTTAGAAAAATTTAATCA 420  
T L Y N E Y N K T N N L D D F R K F N H  
TGAACATGAAAAACAATAACAGCTTAATTGAAATTAATAAGTCTTTTTTACAAATGAG 480  
E H E K T N N S L I E I N K S F L Q M S  
CATCGTATCTCAAGTAAGTCATGAAGTATTCGATATGATCGAAGTTCTATATGATGTAAA 540  
I V S Q V S H E V F D M I E V L Y D V N  
TGATAATGACAATGTGTTATGGCCTTTTTATAAACCTATATATTCTTCTAAATATATTAT 600  
D N D N V L W P F Y K P I Y S S K Y I I  
AAAATTATTCGAGAAATTCAGTATACCCTCAACTTCTATTAATAAAAAAATAAAATGAAGAA 660  
K L F E K F S I P S T S I K K N K M K K  
AGACCATGACCAGTTTTTAATAAAATCGCAATAACAAACCGTTATGCATAACTGAAAG 720  
D H D Q V F N K I A I T N R L C I T E R  
AAATGAATTAATTAATCAAAATGGGGAATATACACAATCCCTTATATAATTTTTATAATAA 780  
N E L I N Q M G N I H N P L Y N F Y N N  
CCCTCCGATGGATGACTATAACTTATATAAATCATCAAAAACCAAGTTTGTAATTTGTTT 840  
P P M D D Y N L Y K S S K T K F V N C L  
AAAAAATAATTTTGAAAACATAAAAAATGAAATAAATAATGTTATTTCAAGTGAAGAAAC 900  
K N N F E N I K N E I N N V I S S E E T  
ATTGTATAAAACGGAATGTAATAATAGCAACAATCTCAAGAGTTCTACTCAGAAAGATTC 960  
L Y K T E C K N S N N L K S S T Q K D S  
TGGATTTTGTAATAATGTGATAAGTATCGGTATAACTTATAAAAAAATCGGACATTCC 1020  
G F C K N V I S I G I T Y K K K S D I P  
CAACGATAATGATATTATTCCATTTATAGATTTTCTAATAGAAAAATTAAACCAAAAGA 1080  
N D N D I I P F I D F L I E K L K P K D  
TGATTTAAAGAGTCTTATAATCAAACCTAAATTTTATAAAGAACAATTCAAGGATATGAA 1140  
D L K S L I I K L N F I K E Q F K D M K  
AAATAAACATGACAAATACACAAAAATATGTCAAAAAGAAGAGGCTTCTGTGAAAAAATG 1200  
N K H D K Y T K I C Q K E E A S V K K C  
TAATGAACCCGATTTAAACAATAATAATTGTGGTAATTATTTTAATGATTTTAAAAAGAA 1260  
N E P D L N N N N C G N Y F N D F K K K  
AGCCGATTCGTATAGTATTTTAGATTATAATGATAAAATTATACAAAATTTAGAAGCAAT 1320  
A D S Y S I L D Y N D K I I Q N L E A I

ATATTATACTTATAAAGAATCCTTAGAATACTTTTTTAAATCAATAGGGGAATTATTAAT 1380  
Y Y T Y K E S L E Y F F K S I G E L L I  
AGACAAACCAGATTCAAATGGAAATATAATAGACAAAGATGGTATTAATGATTTTCGATT 1440  
D K P D S N G N I I D K D G I N D F D L  
GTCTACACCCAAATCTAGATTAAAATCTTTAGAAACCGAATTCCATACAATATTTGAAAA 1500  
S T P K S R L K S L E T E F H T I F E N  
TACATTGGATTTTTATAAAAAATAAACAAAATCTTGATAATTCCAATGATACAAAAAAAAA 1560  
T L D F Y K N K Q N L D N S N D T K K N  
TATAATATCATTAAATTTTACCACTTATGGATAAATTTACAGATTTAAATAAGTCTATGTT 1620  
I I S L I L P L M D K F T D L N K S M L  
AAATTTAATAAATGATGGCGTTTTACAAAAACATATGACTATTACTCAAATAAAACAAAA 1680  
N L I N D G V L Q K H M T I T Q I K Q K  
ATTAAATCAATCAACTTATGAAGAAAGAAAAGAAGGGTTTACATCTGCCCTCGAATTAGC 1740  
L N Q S T Y E E R K E G F T S A L E L A  
AGAACGTTGGGAAAAGAAAAAACTCGAAACAATAACGAAGTTAAACAAAGAAAACGAAGA 1800  
E R W E K K K L E T I T K L N K E N E E  
AACTGTAAATTGGAAAAAGAAATTAGAGAATTATTTAAAAAATACTCAGATGAGGTGGC 1860  
T V K L E K E I R E L F K K Y S D E V A  
TGAAAAGAAACATGTAGAAGAATTAATAATTAATAATTAAGAAACAATAAAAGACATATA 1920  
E K K H V E E L K L K L K E T I K D I Y  
TAACAAAAAGGAATATATTAATAAAGCAATTGACTTAAAGAAAGCAATAGAAAATAATAA 1980  
N K K E Y I K K A I D L K K A I E N N N  
TAAATACATTGATGAATTAGGTAAAAATACGCCATTTCAAATTGAGGAATATGTAAAGAA 2040  
K Y I D E L G K N T P F Q I E E Y V K K  
AAAAGATACAATATATAGTACAATAAAATCAGAGTTATCCGAAATTTATAAAGGTAACAT 2100  
K D T I Y S T I K S E L S E I Y K G N I  
TGTTGAGCTTTACAATGAATTGTCTTCTGTAGTTCAAGAAAATACCATTGATC 2153  
V E L Y N E L S S V V Q E N T I D

**Figure 4.5.4** Alignment of clone 5 with E8

The overlapping amino acid sequences of clone 8.1.6 and clone 5 were aligned using the MegAlign program of DNASTar. Identical residues are shaded in green. There is considerable identity between the two sequences and this, along with the conservation of the 5'-end intron, suggests that the two genes have evolved from a common ancestral gene and are members of the same multigene family of *P. yoelii*. There are eight conserved cysteine residues, shaded in red.



[illegible]

**Figure 4.5.5** Alignment of the introns of clone 5 and E8

The intron of E8 and the proposed intron of clone 5, of which only partial sequence was available, were aligned using the MegAlign program of DNASTar. Identical nucleotides are shaded in green. The identity of the introns is 91%.

202	G G T A A T C A T A G A T	A T C T T T A A C A A A T T A T T	E8 intron
1	- - - - - - - - - - - G	A T C T T T A A C A A A T T A T T	E8.5 intron
233	T C A A T T C T T G G G T	T T A C A T A T T T T A T T T A A	E8 intron
19	T C A A T T C G T G G G G	T T A C A T A T T T T A T T T A A	E8.5 intron
263	T A T G T T T A T G T G	T A T T A A T T A T G C A A A T G T	E8 intron
49	T A T A T G T T T A T A	T A T T A A T T A T G C A A A T T T	E8.5 intron
293	G T A A A T T C C T A T A C T A A T A T T A C T T A A T T T		E8 intron
79	G T A T A T T C C T A T A C T A A T A T T A C T T A A T T T		E8.5 intron
323	G T T T C A T C A T T A A T A T G C A G		E8 intron
109	T T T T C A C C A T T A A T A T G C A G		E8.5 intron

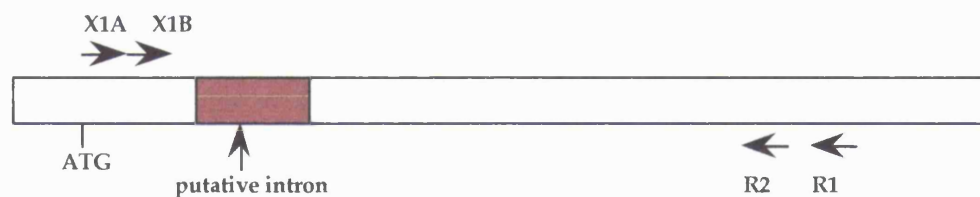
**CHAPTER FIVE**

**STRUCTURE OF THE Py235-E8 GENE**

To test the hypothesis that the gene contains a short intron at its 5'-end reverse transcriptase PCR (RT-PCR) was carried out on total *P. yoelii* YM RNA.

### 5.1 RT-PCR

Primers were designed to amplify the region either side of, and across, the proposed intron sequence. PCR product amplified from genomic DNA will contain the intron, whereas PCR product derived from messenger RNA, from which the intron will have been removed by the parasite's splicing machinery, will not. Thus, if the hypothesis that the gene contains an intron is correct, a smaller product should be seen in the RT-PCR reaction as compared to the genomic DNA control reaction.



Primer positions and sequences:

X1A: 5'-ACCATGAAAAAATATATTTATATTATTA-3'

1

25

X1B: 5'-TTTGGCTGCATTTTTTATTTC-3'

27

48

R1: 5'-CGACCGTATCTAACGTATC-3'

592

572

R2: 5'-GCGGTTATTTGTAAAAAGGAATT-3'

553

531

(all sequence positions are given using the A residue of the putative initiating ATG codon as position 1).

cDNA was synthesised from 1µg total *P. yoelii* YM RNA using the primer E8.NEW.RC. Purified cDNA was used as a template for PCR using primers X1A and R1, and a subsequent nested PCR reaction was performed on a 100-fold dilution of the first PCR using primers X1B and R2. One would anticipate a PCR product of 520 bp from genomic DNA or unspliced RNA, and a reduction in size of approximately 140 bp from the cDNA derived from spliced mRNA if the intron boundaries had been correctly assigned.

From agarose gel analysis of the PCR products, the sizes were consistent with the presence of a 139 bp intron being present in the gene at the anticipated location (figure 5.1.1). It should be noted that both the larger and smaller bands are present in the reaction with RNA as the template. This is due to the presence of contaminating genomic DNA in the RNA preparation. The control reaction which did not include RT shows only the larger band, indicating that the smaller band is derived from RNA. The RNA derived product was cloned into the vector pMOSBlue and sequenced (figure 5.1.2).

This analysis showed that the smaller, RNA-derived product did, indeed, have 139 bp missing when compared to the corresponding genomic DNA sequence. This was indicative of there being an intron at the 5'-end of the gene which is removed by the parasite's RNA-processing machinery after transcription. The intron is marked by consensus intron junctions which have been described for *Plasmodium* genes (Weber, 1988) and other eukaryotic genes, with the dinucleotide -GT- at the 5' boundary and the dinucleotide -AG- at the 3' intron boundary.

It is very common for *Plasmodium* genes to contain a short intron just downstream of the translation start site and signal sequence. Such an arrangement is seen in a number of *Plasmodium* genes, for example reticulocyte binding protein-1 (RBP-1) of *P. vivax* (Galinski et al., 1992), knob associated histidine rich protein (KAHRP) of *P. falciparum* (Triglia et al., 1987), and RhopH3 of *P. falciparum* (Brown and Coppel, 1991).

In addition to the RT-PCR data establishing the presence of an intron in the E8 gene, they also prove that the E8 gene is indeed transcribed. This is the first evidence that a specific member of the 235 kDa rhoptry protein multigene family is expressed.

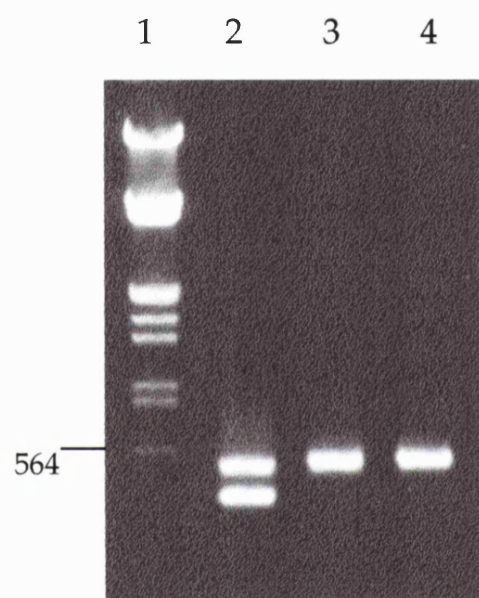
**Figure 5.1.1** RT-PCR of *P. yoelii* YM RNA

Ten microlitres of each RT-PCR reaction using primers X1B and R2 was run on a 1% agarose gel.

- 1       $\lambda$  DNA digested with *EcoRI* & *HindIII*
- 2      RNA as a template + reverse transcriptase
- 3      RNA as a template - reverse transcriptase
- 4      DNA as a template

In lane 2, two bands are present. The upper PCR product is DNA-derived owing to the presence of contaminating genomic DNA in the RNA preparation. This band is also present in the control reaction performed without the addition of reverse transcriptase (lane 3) and the genomic DNA control reaction (lane 4). The lower band in lane 2 is RNA derived.





**Figure 5.1.2** DNA and deduced amino acid sequence of the RNA  
derived RT-PCR product

The RNA derived RT-PCR product was cloned into the TA cloning vector and sequenced. The sequencing confirmed the presence of a 139 bp intron at the 5'-end of the Py235.E8 gene. The splice site is denoted by an arrow.

splice site



```
TTTGGCTGCATTTTTTATTTTCCTTCGATATAATTTGTGCAACCAAAATTGAGGAAAATAA 60
  L A A F F I S F D I I C A T K I E E N K
AAACAAGAATGCCACAAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGA 120
  N K N A T K L N N Y N P Y H N L E E S D
TTTTGATAATTCAAACTTTTTAAATGAAAAAAATATGATAATAAAAAATAATTCCATTAA 180
  F D N S N F L N E K K Y D N K N N S I N
TAATGAAAAATATATACAGCCACATTCAATAAATAACAATTCCTTCAAAATAAGAAAGA 240
  N E K Y I Q P H S I N N N F L Q N K K D
TATAACAATACAAAAATCACATTGTATAATGAATATAATAAGACAAATAATATTGATAA 300
  I N N T K I T L Y N E Y N K T N N I D N
TTTTAGAAATTTTAATCATGAACATGAAAAAACAAATAACATATTAATTGAAAATAATAA 360
  F R N F N H E H E K T N N I L I E N N N
TTCCTTTTTACAAATAACCG 380
  S F L Q I T
```

## 5.2 Analysis of the putative translation start site

The nucleotide sequence surrounding, and more importantly preceding, the initiating methionine of all genes has been shown to be important in determining the start site of translation of that gene (Kozak, 1986). According to the scanning model of translation initiation, the 40S ribosome subunit binds to the m<sup>7</sup>G cap at the 5'-end of the messenger RNA, stopping only to initiate translation when it encounters an AUG codon which lies in a context optimal for translation initiation. Analysing the effects of single base substitutions around the AUG initiator codon has shown that the sequence surrounding the initiating methionine codon which leads to optimal translation is

-3 -2 -1 1 2 3 4  
**A C C A U G G**

The dominant effect of a purine residue, and particularly an adenine nucleotide, in position -3 has been established (Kozak, 1986).

The nucleotide sequence surrounding the initiating AUG codons of *Plasmodium falciparum* genes has been analysed and a consensus sequence revealed (Saul and Battistutta, 1990).

-3 -2 -1 1 2 3 4  
**A A A A T G A**

The putative initiating methionine codon of the Py235-E8 gene shows a surrounding nucleotide sequence of

-3 -2 -1 1 2 3 4  
**A C C A T G A**

This is extremely similar to the optimal eukaryotic sequence, but varies from the *P. falciparum* consensus sequence. The sequence identified for *P. yoelii* genes would not necessarily be expected to be so closely related to *P. falciparum* sequences as the rodent parasite does not have the extremely A+T biased genome of the human parasite.

A search of DNA databases to analyse the nucleotide sequences surrounding the initiating AUG codon of other *P. yoelii* genes was undertaken (references shown in footnote). The sequences from position -4 to +4 were compared. A consensus was calculated for *P. yoelii* sequences, with the numbers in subscript denoting the percentage occurrence of the nucleotide in that position.

	-4	-3	-2	-1	1	2	3	4
Merozoite surface protein-1 <sup>1</sup>	G	A	A	A	A	T	G	A
Sporozoite surface protein-2 <sup>2</sup>	A	A	A	C	A	T	G	A
cAMP-dependent protein kinase <sup>3</sup>	T	A	A	A	A	T	G	A
ATPase <sup>4</sup>	T	A	T	A	A	T	G	G
Cytochrome oxidase b <sup>5</sup>	T	T	A	T	A	T	G	A
235 kDa rhoptry protein	A	A	C	C	A	T	G	A
Majority	T <sub>50</sub>	A <sub>83</sub>	A <sub>66</sub>	A <sub>50</sub>	A	T	G	A <sub>83</sub>
<i>P. falciparum</i> consensus <sup>6</sup>	A	A	A	A	A	T	G	A
Eukaryotic optimal <sup>7</sup>		A	C	C	A	T	G	G

<sup>1</sup>. References:

- 1 Lewis A.P. (1989) Molecular & Biochemical Parasitology **36**, 271-282
- 2 Rogers W.O. *et al* (1992) Molecular & Biochemical Parasitology **53**, 45-52
- 3 Saito-Ito A. *et al* (1995) Biochimica Biophysica Acta **1269**, 1-5
- 4 Murakami K. *et al* (1990) Journal of Cellular Science **97** (3), 487-495
- 5 Vaidya A.B. *et al*; (1989) Molecular & Biochemical Parasitology **35**, 97-107
- 6 Saul A. & Battistutta D. (1990) Molecular & Biochemical Parasitology **42**, 55-62
- 7 Kozak M. (1986) Cell **44**, 283-292

### 5.3 Analysis of the proposed signal sequence of Py235-E8

The proposed signal sequence of Py235-E8 fits into the rather loose rules defining which sequences can act as signal sequences (vonHeijne, 1985). The two lysine residues following the initial methionine are positively charged, followed by a stretch of largely hydrophobic residues. More significantly, the proposed signal sequence shows a high degree of homology to the putative signal sequence of another *Plasmodium* gene - the acidic basic repeat antigen (ABRA) of *Plasmodium falciparum*, an exoantigen which is secreted into the parasitophorous vacuolar space, but has also been shown to be loosely associated with the merozoite surface (Weber, 1988).



- identical residues
- similar residues

Comparison of the first twenty-five residues of the two proteins shows that they share 9 of 25 residues (36%), lending support to the hypothesis that the sequence acts as a signal peptide for the 235 kDa protein of *P. yoelii*.

From a detailed analysis of thirty-nine signal sequences and their cleavage sites, Perlman and Halvorson have proposed certain rules which allow the assignment of signal sequences and cleavage sites for proteins for which the amino acid sequence is deduced from cloned DNA sequences (Perlman and Halvorson, 1983). The proposed signal sequence for the E8 gene encoding the 235 kDa rhoptry protein of *P. yoelii* YM is

1	33
<b>MKKYIYIISIAAFFISFDIICATKIEENKNKNA</b>	
<i>N-terminal</i>	<i>C-terminal</i>

There are several features of this sequence which are suggestive of its role as a signal peptide for the 235 kDa protein, consistent with the guidelines for the assignment of such sequences (Perlman and Halvorson, 1983):

1. The sequence contains a core of largely hydrophobic residues. In the case of the 235 kDa protein's sequence, this consists of 14 residues. In general, the hydrophobic core contains 12 residues on average, with the observed lengths ranging from 8 to 15 residues.

It has been proposed that the hydrophobic core of the signal sequence participates in the transmembrane orientation, binding and alignment of the signal sequence for cleavage by a signal peptidase enzyme.

2. Immediately preceding the hydrophobic core there are two lysine residues. Again, it is a common observation that the majority of signal

sequences possess such an arrangement, that is, the presence of one or more positively charged amino acids preceding the hydrophobic core of the signal sequence. From the analysis of Perlman and Halvorson, who studied thirty-nine signal sequences of proteins from a wide range of organisms, 87% of the sequences were found to contain one or more positively charged residues in the three positions preceding the hydrophobic core, with lysines at these positions being considerably favoured compared to other residues (Perlman and Halvorson, 1983).

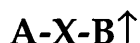
3. Repeating pairs of residues are commonly observed in signal sequences. In the case of the signal sequence of E8, there are three such repeated pairs: YI at positions 4 & 5 and 6 & 7, II at positions 7 & 8 and 19 & 20, and IS at positions 8 & 9 and 15 & 16. The function of such pairs of residues in the signal sequence is unknown, but they are present in a significant numbers of such sequences. It has been suggested that a symmetrical subunit complex in the membrane may recognise such signal sequence symmetry units (Perlman and Halvorson, 1983).

4. Immediately after the hydrophobic core, there is a negatively charged aspartic acid residue at position 18. Again this is commonly observed in signals sequences.

5. The sequences of signal peptides tend to favour the formation of  $\beta$ -strand secondary structure. The presence of positive residues at the N-terminus of the sequence and a negatively charged residue at the C-terminus of the signal sequence negatively correlates with the tendency of a sequence to form  $\alpha$ -helices. The high isoleucine content (5 out of 14 residues : 36%) of the hydrophobic core gives it a bias for the formation of  $\beta$ -strands (Perlman and Halvorson, 1983).



6. The signal peptide cleavage sites have been analysed and a consensus sequence suggested. This is



where A is any aliphatic or hydroxy amino acid (and cysteine), B is a subset of A but excludes the larger aliphatic amino acids such as leucine, valine and isoleucine, X is any amino acid, and  $\uparrow$  denotes the site of cleavage by signal peptidase. The position in the amino acid sequence which is most favoured for cleavage by the signal peptidase has been observed to be located between 4 and 6 residues C-terminal of the end of the hydrophobic core (Perlman and Halvorson, 1983). In the proposed signal sequence for the 235 kDa protein of *P. yoelii*, there are three possible cleavage sites located at positions from 3 to 6 residues C-terminal to the end of the hydrophobic core:

---- Ile Ile Cys ----

---- Ile Cys Ala ----

---- Cys Ala Thr ----

The most likely cleavage site of the three is Ile-Cys-Ala, when compared to other known cleavage sites, where one of forty proteins studied had a cleavage site of Ile-X-Cys $\uparrow$ , none had a cleavage site of Cys-X-Thr $\uparrow$ , and three of forty had a cleavage site of Ile-X-Ala $\uparrow$  (Perlman and Halvorson, 1983).

7. It is common for the secondary structure of the sequence after the hydrophobic core and the signal peptidase cleavage site to be capable of the formation of a  $\beta$ -turn. It has been speculated that the turn could be important for allowing peptidase access to the cleavage site, and that the signal sequence is inserted into the endoplasmic reticulum membrane as a hairpin structure (Perlman and Halvorson, 1983). Analysis of the sequence of the proposed signal peptide of the 235 kDa rhoptry protein using the Chou-Fasman structure prediction demonstrated that a region of the sequence from position 29 to 33 was likely to form a  $\beta$ -turn structure. This region is 6 residues from the proposed signal peptidase cleavage site and 10 residues from the end of the hydrophobic core.

The comments above are summarised in figure 5.3.1.

Such an analysis is purely theoretical, but is based on the studies of known signal sequences and their cleavage sites.

The overall structure and features of the proposed signal sequence for the 235 kDa rhoptry protein are consistent with those described for other *Plasmodium* proteins. RhopH3 of *P. falciparum* has a signal sequence which consists of a predominantly hydrophobic core of nineteen residues, flanked by positively charged lysine residues (Brown and Coppel, 1991). If the criteria of Perlman and Halvorson are applied to this sequence, the hydrophobic core begins after four residues with two positively charged residues preceding it. The core extends for sixteen residues before another charged residue is encountered, signalling the end of the hydrophobic core. Cleavage of the signal sequence has been demonstrated to occur between a glycine and lysine residue positioned four residues after the end of the hydrophobic core (Cooper et al., 1989), with the A-X-B $\uparrow$  sequence in this

case being Valine-Tryptophan-Glycine, again consistent with the rules suggested by Perlman and Halvorson.

N-terminal amino acid sequence of *P. falciparum* MSP-1 has shown that the cleavage of this protein's signal sequence occurs between a cysteine and valine residue (Holder et al., 1985). The signal sequence of this protein comprises a seventeen residue hydrophobic sequence flanked on the N-terminal side by a lysine residue.

Previous analysis of the sequence of clone E8-END had led to the assignment of a group of residues in this clone as the putative signal sequence. These residues lie at position 587-622 in the amino acid sequence of clone 8.1.6.

5' - MKKIILLIIQYMNEFKGLNDAMTKLKNEGISQKFV - 3'

587

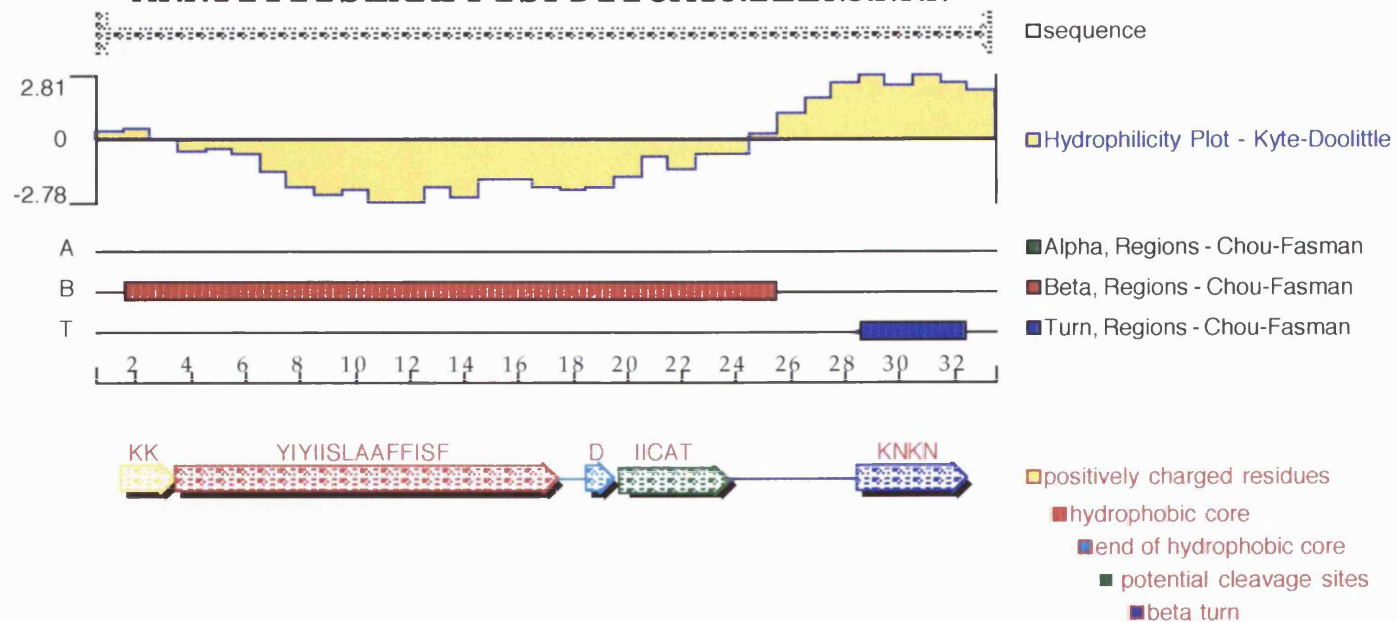
622

Applying the same criteria to this sequence as the signal sequence at the 5' end of the gene, it is apparent that this sequence fulfils many of the rules laid out for assignment of signal sequences. It contains two positively charged lysine residues after the methionine, has a hydrophobic core of ten residues immediately after this, possesses pairs of residues in this case KKIILLII, possesses a negatively charged glutamate residue following the hydrophobic core, and has sequence capable of forming a  $\beta$ -turn structure after the hydrophobic core. It does not possess any obvious candidate sequences for cleavage by a signal peptidase, however the assignment of these sites is not a particularly precise art and the "rules" are rather vague.

**Figure 5.3.1** Analysis of the putative signal sequence of the Py235-E8 gene

The first 32 amino acids of the Py235.E8 protein were analysed using the Protean<sup>TM</sup> secondary structure prediction program of DNASTar<sup>TM</sup>. This shows the presence of a hydrophobic core, which is predicted to take up a  $\beta$ -strand structure, followed by a region capable of forming a  $\beta$ -turn. Consistent with other signal sequences, there is no  $\alpha$ -helical structure predicted.

MKKYIYIISLAFFISFDIICATKIEENKNKN



## 5.4 5' RACE

By definition, the core promoter of any given gene lies upstream of the transcription start site of that gene. In order to determine the start site of transcription for the Py235-E8 gene, 5'-RACE (Rapid Amplification of cDNA Ends) analysis was performed.

5'-RACE is a procedure which allows amplification of nucleic acid sequences situated between an internal known sequence, and the unknown 5' end of the mRNA. The specificity of the amplification is single-sided, as cDNA is synthesised from mRNA using a gene-specific antisense oligonucleotide - GSP-1. Terminal deoxynucleotidyl transferase (TdT) is used to add a homopolymeric dC tail to the 3' end of the purified single stranded cDNA. Amplification of the sequence is achieved by using another gene specific primer (GSP2) upstream of GSP1 and a primer which hybridises to the homopolymeric dC tail at the 3' end of the cDNA (figure 5.4.1). Nested PCR amplification using a third gene specific primer (GSP3) can also be carried out.

Using *P. yoelii* YM RNA as a template in 5'RACE with E8.RC, R1, and R2 as the three gene specific primers (figure 5.4.2), and the abridged anchor primer (supplied by Gibco BRL), a 5'RACE product of approximately 250 bp was produced (figure 5.4.3). This product was cloned into the pMOSBlue vector and sequenced (figure 5.4.4). Such analysis demonstrated that the product was specific for the E8 gene, but suggested that the transcription start site for the gene lay downstream of the proposed initiating methionine.

Such a result appears to be inconsistent with the previous RT-PCR work described in section 5.1 which suggested that the transcript is much longer, extending at least 300 bp beyond this point.

The product generated in the 5'RACE experiment is almost certainly RNA-derived, as double stranded genomic DNA cannot be used as a template by TdT, would not possess an oligo-dC tail and therefore would not be a viable template for PCR amplification as it would possess an annealing site for only one of the primers. The RNA in the sample may have degraded, but one would anticipate observing a smear of different length 5'RACE products if this were the case.

A possible explanation for the observations could be the presence of alternative transcription start sites in the E8 gene. The transcription start site identified by 5'RACE lies 300 bp downstream of the proposed initiating methionine. 1330 bp downstream of the transcription start site lies another methionine residue followed by a sequence characteristic of a signal sequence, which had previously been proposed as the translation start site. It is possible that the gene has two transcription start sites, or two forms of transcript, one a truncated version derived from the other.

If there are two forms of mRNA for the E8 gene, why was only the shorter one detected in 5'RACE? It is possible that the shorter form of the message is much more abundant than the full length transcript, or that PCR amplification in such circumstances will always favour amplification of smaller products.

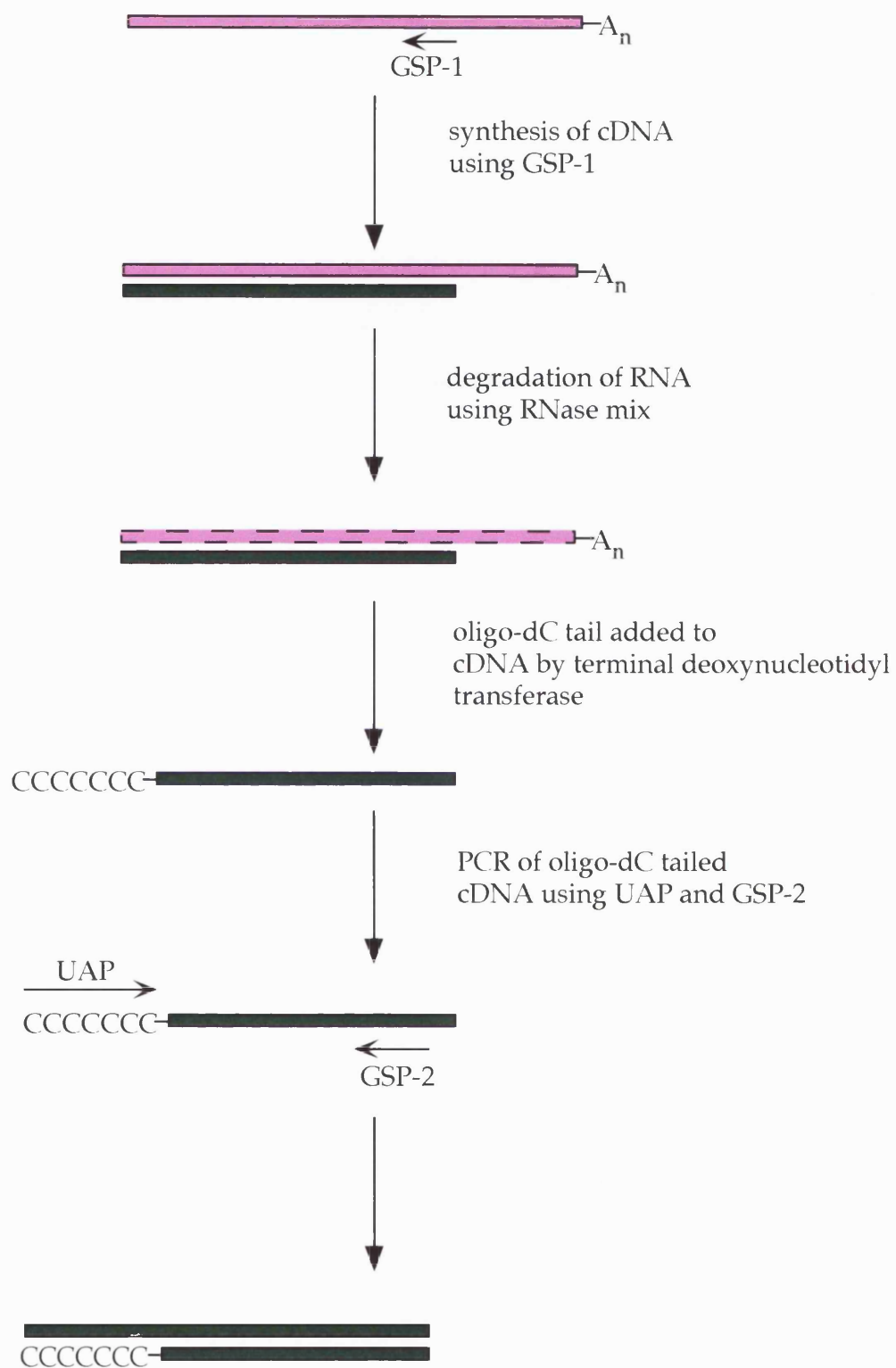
Although the scenarios described would explain the observations, there is not really enough evidence at present to say that this is certainly the case. Further work is required in order to determine if there are indeed two

forms of mRNA for the Py235-E8 gene, and if this is shown to be the case, whether there are two independent transcription start sites or the shorter transcript is derived from the longer mRNA by post-transcriptional processing events such as splicing.



**Figure 5.4.1 5' RACE**

The template for 5' RACE is messenger RNA (mRNA), here shown in pink. A DNA copy of the mRNA (cDNA), shown in green, is made using a gene-specific primer (GSP-1). The mRNA is degraded and an oligo-dC tail is added to the 3'-end of the cDNA using terminal deoxynucleotidyl transferase. Amplification of the cDNA is achieved by using a second gene-specific primer (GSP-2) which is internal to the first, and a universal anchor primer (UAP) supplied by the manufacturer which hybridises to the oligo-dC tail. In order to sequence the 5' RACE product, it was cloned into the pMOS Blue vector.

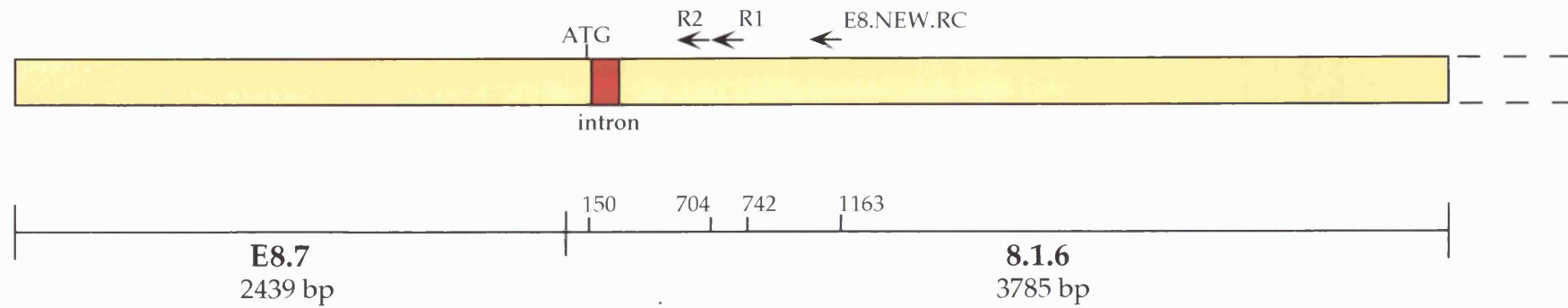


clone 5'RACE PCR products into  
pMOS Blue vector and sequence

**Figure 5.4.2 Primers used in 5' RACE**

The sequences of the gene specific primers used for 5'RACE are given with respect to their positions in clone 8.1.6.

The putative initiating ATG codon and the position of the intron are also marked.



**E8.NEW.RC:**

5'- GTATTCTTGAGTTTGGAGCTA -3'

**R1:**

5'- CGACCGTATCTAACGTATC -3'

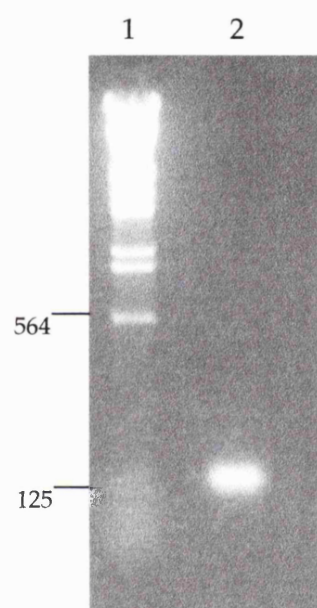
**R2:**

5'- GCGGTTATTTGTAAAAAGGAATT -3'

**Figure 5.4.3** 5' RACE product from *P. yoelii* YM RNA using E8 specific primers

Ten microlitres of the 5'RACE PCR reaction was run on a 1% agarose gel. This showed a product of approximately 250 bp. DNA sizes of the markers are shown in base pairs.

- 1       $\lambda$  DNA digested with *Eco*RI and *Hind*III
- 2      5'-RACE product



**Figure 5.4.4** DNA and deduced amino acid sequence of 5' RACE product

The 5'RACE product was sequenced using the ABI 377 automated sequencing set-up. The product was 271 bp long which included a 13 residue long C tail at the 5' end (here shown in red). The aspartate residue (D) codon at the beginning of the 5'RACE product is 288 bp downstream of the putative initiating methionine and 99 bp downstream of the 3' splice site of the intron identified in section 5.1.

GGGGGGGGGGGGTGATAATTCAAACTTTTTAAATGAAAAAAATATGATAATAAAAAATA 60  
D N S N F L N E K K Y D N K N  
ATTCCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTCCTTCAAA 120  
N S I N N E K Y I Q P H S I N N N F L Q  
ATAAAAAAGATATAAACAATACAAAAATCACATTGTATAATGAATATAATAAGACAAATA 180  
N K K D I N N T K I T L Y N E Y N K T N  
ATATTGATAATTTTAGAAATTTTAATCATGAACATGAAAAACAAATAACATATTAATTG 240  
N I D N F R N F N H E H E K T N N I L I  
AAAATAATAATTCCTTTTTACAAATAACCGCAAT 274  
E N N N S F L Q I T A M



## 5.5 Promoter motifs

If the transcription start site for the Py235-E8 gene is genuinely the one indicated by the 5'-RACE experiment described in the previous section, it should be preceded by promoter sequences that control the start site of transcription, such as the TATA box, and others that control the frequency of transcription initiation such as the CAAT box. The function of such motifs has been described in more detail in section 1.10.1.

Analysis of the sequence upstream of the putative transcription start site of the E8 gene showed that there is a putative TATA box at position 1966-1972 in the gene sequence. This is 36 bp upstream of the transcription start site, entirely consistent with the spacing described for eukaryotic genes. The sequence of the putative TATA box is

**5'-TATAATC-3'**

which agrees with the consensus sequence for eukaryotic genes described in section 1.10.1.

At position 1925-1928 of the Py235-E8 gene sequence is a sequence which is consistent with an inverted CAAT box. The CAAT box is 77 bp upstream of the putative transcription start site as determined by 5'-RACE, a position where a CAAT box is commonly found in eukaryotic promoters. The sequence of the putative inverted CAAT box is

**5'-CCTCAATTT-3'**

which is in agreement with the consensus described for eukaryotic genes.

The base numbering refers to that in figure 4.5.2.

If initiation of transcription of the E8 gene does indeed occur at two sites, that is at the site suggested by the 5'-RACE experiment and also at a site upstream of the putative initiating methionine suggested by RT-PCR experiments, it would be expected that there would be promoter elements upstream of the initiating methionine regulating expression of the gene from this site. However, without knowing where transcription initiates, it is extremely difficult to assign putative promoter motifs with any degree of accuracy. Obviously, promoter elements will lie upstream of the initiating ATG codon, so a fairly simplistic search of this region was undertaken to look for elements which are consistently found in promoters of *Plasmodium* genes, GC-rich elements. The results are summarised in figure 5.5.1.

An element with the consensus sequence

**5'- (T/A) T (A/T) C C C (A/T) (T/A) -3'**

was found on nine occasions in the non-coding sequence of 1715 bp. The sequence was detected only once in the subsequent 725 bp of clone E8.7. The role of such GC-rich sequences is not known, but some have been implicated in interactions with parasite nuclear proteins (Lanzer et al., 1993).

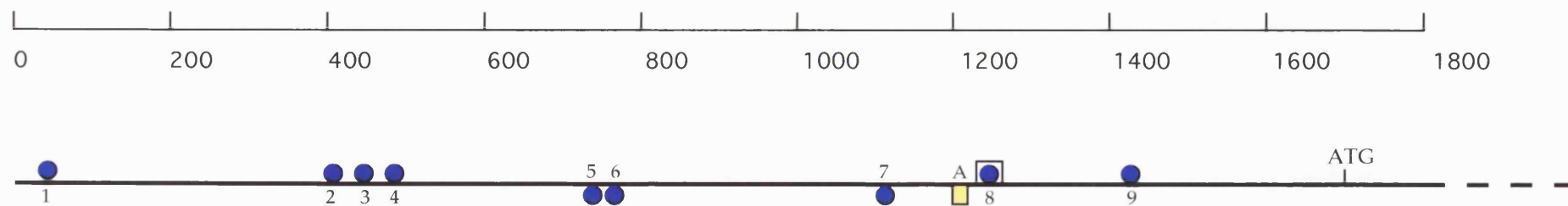
**Figure 5.5.1** Identification of GC-rich elements in the non-coding region of clone E8.7

The region of clone E8.7 upstream of the putative initiating ATG codon was searched for GC-rich elements. Nine which fit the consensus sequence 5'- (T/A) T (A/T) C C C (A/T) (T/A) -3' were identified, and are shown here by blue circles. The circle which is enclosed in a rectangle is identical to the sequence of a repetitive GC-rich element in the circumsporozoite protein of *P. knowlesi* (Ruiz-Altamira et al., 1987)

The yellow rectangle shows the location of a region identical to a repetitive GC-rich element found in the promoter of the GBP-130 gene of *P. falciparum* (Ravetch et al., 1985).

Sequences

1. 5'- ACACCCAT -3'
  2. 5'- TATCCCAT -3'
  3. 5'- ATACCCAT -3'
  4. 5'- CTTCCCAA -3'
  5. 5'- TTGCCCAT -3'
  6. 5'- ATCCCCTT -3'
  7. 5'- GTACCCAG -3'
  8. 5'- TTTCCCTT -3'
  9. 5'- ATTCCCTT -3'
- A. 5'- TGCATGCA -3'



## **CHAPTER SIX**

### **SCREEN OF GENOMIC LIBRARY WITH D1**

## 6.1 Screen of genomic library with D1

In order to obtain more sequence upstream of the clone D1, the genomic library was screened with the entire D1 insert. This gave several positively reacting colonies on the initial library screen, but on secondary screening, only two clones were identified, designated D1.2 and DX. The insert sizes of the clones were estimated by restriction enzyme digestion with *Bam*HI and *Kpn*I followed by agarose gel electrophoresis. This showed that clone D1.2 had an insert of approximately 1.25 kb and clone DX had an insert of approximately 2.3 kb. Sequencing of the clones was completed using the ABI 377 system.

## 6.2 Analysis of clone D1.2

Clone D1.2 did contain additional sequence for gene D1. The entire sequence of clone D1 is contained in clone D1.2, with 375 bp of sequence upstream of the 5' end of clone D1 which, when translated, extended the open reading frame by 125 amino acids (figure 6.2.1).

D1.2 has an open reading frame of 439 amino acids. When the amino acid sequence was compared to that of Py235-E8 the overall identity was found to be 40%, with a similarity of 56%. The identity increases markedly in the latter half of the sequence, which is shown graphically in figure 6.2.2. The overlap of the sequences extends from position 314 to 751 in the amino acid sequence of Py235-E8.

Searches of protein databases using the FASTA search program of the DNA Data bank of Japan showed that the open reading frame of clone D1.2

has homologies with both of the *P. vivax* reticulocyte binding proteins and the yeast integrin homologue *uso-1*. The significance of such homologies has been discussed with respect to Py235-E8, but the same considerations also apply to clone D1.2.

**Figure 6.2.1 DNA and deduced amino acid sequence of clone D1.2**

The nucleotide and deduced amino acid sequence of clone D1.2 are shown. Clone D1.2 is 1319 bp long and has an open reading frame of 439 amino acids. The translated sequence which overlaps with clone D1 (315 amino acids) is shown in green, while novel sequence (124 amino acids) is shown in blue.



GATCTATGATATGGGACTTCGTTACATGGAAACAGACTAATTA AAAATTTTAGAGTCTGA 60  
I Y D M G L R S H G N R L I K I L E S E  
GGGGGAACAAGGGAATATTATAAGTAAACTAAAATTTTATCAAACGAAATTGAATATAC 120  
G E Q G N I I S K L K F L S N E I E Y T  
TTTAAACAAATCTGATAATTACAAAGACGAATGTACGTCGAATTATACATTTATGGTCAA 180  
L N K S D N Y K D E C T S N Y T F M V N  
TTATGTACAACGTGATACATATACGCAATATTACTATTATTACTTGAGGGAAATTA AAAA 240  
Y V Q R D T Y T Q Y Y Y Y Y L R E I K K  
AATAGCACAACAAAGAGCTAATTTCAATTATAATTTTAATAAACTAAAACATTTGAAATT 300  
I A Q Q R A N F N Y N F N K L K H L K L  
ATTATATAAACAGCAAGAAGACATAATGAGAAATTTTCATAAAACGATAGAAAGACTAAT 360  
L Y K Q Q E D I M R N F H K T I E R L I  
ATTTGGGAAACAGATCCAAATAATACTAACGTATTTAAAGAATATATTTATGAATTTAA 420  
F G K P D P N N T N V F K E Y I Y E F N  
TCTCTGGATACCCAAATCGAAATTAACATTATTAGAAAAAAGATTTTGTGAAATATTTGA 480  
L W I P K S K L T L L E K R F F E I F E  
AGAGAAATGGAATTCCTATGAAATTA AAAAAGATATTGACGAAAATAGTAAACAATATAA 540  
E K W N S Y E I K K D I D E N S K Q Y N  
TGTTGTGAAATTAATTTTGAATACATGAAGGAATTAACAGACGTAATTGATTTTATGGA 600  
V V K L I L Q Y M K E L T D V I D F M E  
AAATTACCAACCAGATGAAGTTCCTACAAGAAGGATAATTGGTTTTGAAGTCGAATGGCG 660  
N Y Q P D E V P T R R I I G F E V E W R  
TATAAAAGCATTACTTTATTCAGAAGTAGAAGCTGGAGTGAAGAAATCTTATGAATCAGT 720  
I K A L L Y S E V E A G V K K S Y E S V  
AAAAAATTGGAGAAAATCAATCCTCGAAATAAACAAAAAATTAGAGGAGGAAAATGAAAA 780  
K N W R K S I L E I N K K L E E E N E K  
AGTTATTA AATTGGAAACACAAATTAAGATTTATTTAACCAATATTTGAAAAATAATTGA 840  
V I K L E T Q I K D L F N Q Y L K I I D  
TGAAAAATATATATAAAACAAGTTAA AATTAGAATTA AAAGAAAAAATAAAAAATATATC 900  
E N I Y I N K L K L E L K E K I K N I S  
TGACAAAAATGAATATATTA AAAAAGCAGTTGACTTAAAGAAGATAATAGAAAAATAACAA 960  
D K N E Y I K K A V D L K K I I E N N N  
TGCATACATTGATGAATTAGCTAAACCTCGCCATATCAAGTTCGGAACATTTAAAGAG 1020  
A Y I D E L A K T S P Y Q V P E H L K S  
CACAGATACAATATATAGTACAATAAAATCAGAGTTATCCCAAATTTATGAAGATGACAT 1080  
T D T I Y S T I K S E L S Q I Y E D D I  
TGATAAACTCTATAATGAATTATCTTCTATAGTTCAAGAAAATGACATTGACAATGTAGA 1140  
D K L Y N E L S S I V Q E N D I D N V E  
AGATAAAACGAACTTGACGATTTACAATCTAAAATAGATAATGTATATAGTAAAATCCA 1200  
D K T K L D D L Q S K I D N V Y S K I Q  
AAACATGGAACTGCAACAGTTGAATCACATCTAACGAATATCGAACTAACAAAAATAA 1260  
N M E T A T V E S H L T N I E T N K N K  
ACTATCAGACACAATTGTGGAGATAAAAAAATATATATATGGAGAAATTAGCAAAGATC 1319  
L S D T I V E I K K Y I Y G E I S K D

**Figure 6.2.2** Alignment of clone D1.2 with Py235.E8

The homology between clone D1.2 and Py235.E8 is shown in this alignment. Identical residues are shaded in green. The homology extends from position 313 to 751 in the open reading frame of Py235.E8.

The alignment was performed using the MegAlign™ program of DNASTar™.

[illegible]

### 6.3 Analysis of clone DX

Clone DX has an open reading frame of 691 amino acids (figure 6.3.1). When the amino acid sequence was compared to that of Py235-E8 the overall identity was found to be 58%, with a similarity of 69%. The region of homology of the two sequences extends from position 440 to 1111 in the open reading frame of Py235-E8 (figure 6.3.2). The identity of the sequences increases towards the latter half of the alignment.

Comparison of the sequence of clone DX with that of clone D1.2 reveals a striking degree of similarity. The overall identity is 92%, with a similarity of 95% (see figure 6.3.3). The homology extends from position 124 to the end (position 439) of clone D1.2, and position 1 to 315 of clone DX.

Once again, clone DX also shows homology with *P. vivax* reticulocyte binding proteins 1 and 2, and the yeast integrin homologue *uso-1*.

**Figure 6.3.1** DNA and deduced amino acid sequence of clone DX

The nucleotide and deduced amino acid sequence of clone DX are shown. Clone DX is 2074 bp long and has an open reading frame of 691 amino acids.

GATCCAAATAATACTAACGTATTTAAAGAATATATTTATGAATTTAATATCTCGATACCC 60  
D P N N T N V F K E Y I Y E F N I S I P  
AAATCGAAATTAACAGCATTAGAAAAAGATTTTTGAAATATTTGAAGAGAAATGGAAT 120  
K S K L T A L E K R F F E I F E E K W N  
TCTTATGAAATTAAGATATTGACGAAAATAGTAAACAATATAATGTTGTGAAATTA 180  
S Y E I K K D I D E N S K Q Y N V V K L  
ATTTTGAATACATGAAGGAATTAACAGACTTAATTGATTTTATGGACAATTACCAACCA 240  
I L Q Y M K E L T D L I D F M D N Y Q P  
GATGAAGTTCCTACAAGAAGTATAATTGGTTTTGAAGTCGAATGGCGTATAAATCAAACA 300  
D E V P T R S I I G F E V E W R I N Q T  
CTTTATACAGAAGTAGAAGCTGGAGTGAAGAAATCTTATGAATCAGTAAAAAATTGGAGA 360  
L Y T E V E A G V K K S Y E S V K N W R  
AAATCAATGGTCGAAATAAACAAAAAATTAGAGGAAGAAAATGAAAAAGTTATTAATTG 420  
K S M V E I N K K L E E E N E K V I K L  
GAAACACAAATTAACGATTTATTTAACCAATATTTGAAAATAAATGATGAAAATATATAT 480  
E T Q I N D L F N Q Y L K I N D E N I Y  
CTAAACAAGTTAAATTAGAATTAAGAAAAAATAAAAAATATATCTGACAAAAATGAA 540  
L N K L K L E L K E K I K N I S D K N E  
TATGTTAAAAAAGCAGTTGACTTAAAGAAGACAATAGAAAATAACAATATATACATTGAT 600  
Y V K K A V D L K K T I E N N N I Y I D  
GAATTAATAAAACCTCGCCATATCAAGTTCGGAACATTTAAAGAACACAGATACAATA 660  
E L T K T S P Y Q V P E H L K N T D T I  
TATAATACAATAAAATTAGAGTTATCCCAAATTTATGAAGATGACATTGATAAACTCTAT 720  
Y N T I K L E L S Q I Y E D D I D K L Y  
AATGAATTATCTTCTATAGTTCAAGAAAATGACATTGACAATGTAGAAGATAAAACGAAA 780  
N E L S S I V Q E N D I D N V E D K T K  
CTTGACGATTTACAATCTAAAATAGATAATGTATATAGTAAATCCAAAACATGGAAAAT 840  
L D D L Q S K I D N V Y S K I Q N M E N  
GAAACAGTTGAATCACATCTAACGAATATCGAAACTAACAAAAATAAACTATCAGACACA 900  
E T V E S H L T N I E T N K N K L S D T  
ATTTTGGCGATAAAAAAATATATATATGGAGAAATTAGCAAAGATCTAAATAAAACGTTA 960  
I L A I K K Y I Y G E I S K D L N K T L  
GAAGATTTTAAAAATAAAGAAAAAGAATTATCAAATAAAATAAATGATTACGCTAAGGAA 1020  
E D F K N K E K E L S N K I N D Y A K E  
AATGACCAATTAATGTATATAAATCTAAAATATCAGAAATCAGAAATCATTATAATAGT 1080  
N D Q L N V Y K S K I S E I R N H Y N S  
CAAATTAATATAGACAATACAAAAGAAGGAGAAGCCAAGCAAAACTATGATAAATCCAAT 1140  
Q I N I D N T K E G E A K Q N Y D K S N  
GAGCATATGACAAAAATATCAACTAACGAAAATGAAATATCAAAATTCATAAATGAGGTA 1200  
E H M T K I S T N E N E I S K F I N E V  
AAAAGTATGAAAGACGCATTTTAAAGTAAAGTAGATAAATATATTAATTTTGACAATAAT 1260  
K S M K D A F L S K V D K Y I N F D N N  
TATAAAGAAAATGTTAATTTAGAGCACACCCAATTTACTGAATTAACAGATAAAATAAAA 1320  
Y K E N V N L E H T Q F T E L T D K I K



GCAGAAGTTTCGGATGAAAAATTAAGCAAACATGAAAAAAGTTTAAATGATAGTAAATCT 1380  
A E V S D E K L S K H E K S F N D S K S  
TTAATTAATGAAACAAAAAACTCCATTGAAAAAGAATACCAAAACATTAATACCCTTAAA 1440  
L I N E T K N S I E K E Y Q N I N T L K  
AAGGTAGATGAGTATATAAAAGTATGCGAAATCACCAAAGAGTCAATAACAAAATTTAGT 1500  
K V D E Y I K V C E I T K E S I T K F S  
AGTAAACAAAATACATTAATAATGCATGTTAAATCAAAATATTAACCGTAAAGGAAACT 1560  
S K Q N T L K C M L N Q N I K T V K E T  
AATTCAATAGAAAACCTTTTATAAGACAAGTTTGAAAATACATTGACCAATAAAATTAAT 1620  
N S I E N F Y K D K F E N T L T N K I N  
GAGTTAGATAAAACATTTAAAGATGCATCTTTAAATGATTACGAATTAAATAACAATGAA 1680  
E L D K T F K D A S L N D Y E L N N N E  
TTAATGCAATATTTCAATAATTTAAAAGAAAATTTAGGAAAAGATAAAGAAAACATGTTA 1740  
L M Q Y F N N L K E N L G K D K E N M L  
TATAATCAATTAGCTGAAAAAGAAAAAACTTTTAAATGATATTAAGAAAAATAATACGCAT 1800  
Y N Q L A E K E K T F N D I K K N N T H  
ATAAATGAAGAAAATTTCAAATATCGAAATAGCAATTTATGCATCAATTTATAATATTAGT 1860  
I N E E I S N I E I A I Y A S I Y N I S  
GAAGAAATAAAAAATGAAATTGGAAAAAATATAGAATTACTAAATGCCAAAGTAGTTGAA 1920  
E E I K N E I G K N I E L L N A K V V E  
AAGGTAAAAGCAAACGTAACCAATTTGAATGAAATAAAGGAAAAATTAACAAATTATAAT 1980  
K V K A N V T N L N E I K E K L K H Y N  
TTTCACGATTTTGGGAAAGACGAAAATATAAAATATGCTAATGAAGTTAACAAAATTAAA 2040  
F H D F G K D E N I K Y A N E V N K I K  
GATGACATTAAGACTGTAAGCCAACAAATCGATC 2074  
D D I K T V S Q Q I D

**Figure 6.3.2** Alignment of clone DX with Py235.E8

An alignment of clone DX with Py235.E8 shows that the homology extends from position 420 to 1111 in the open reading frame of Py235.E8. Identical residues are shaded in green.



420 D S D G N I I E E N D I D N F D L S K P K N N F K L L E G E L N G V F E N K W N E8  
 1 D P N N T N V F K E Y I Y E F N I S I P K S K L T A L E K R F F E I F E E K W N DX  
 461 D Y K N N K N L D Q F K D T M K K I I L L I I Q Y M N E F K G L N D A M T K L K E8  
 41 S Y E I K K D I D E - N S K Q Y N V V K L I L Q Y M K E L T D L I D F M D N Y Q DX  
 501 N E G I S Q K F V I N N Q I K Q K F D K S T Y D E K K E G F E S S L E L A K N W E8  
 80 P D E V P T R S I I G F E V E W R I N Q T L Y T E V E A G V K K S Y E S V K N W DX  
 541 E K K K L E I I T E L K K K N E E T V Q L D I K I R E L I K Q I K D I I E E Q K E8  
 120 R K S M V E I N K K L E E E N E K V I K L E T Q I N D L F N Q Y L K I N D E N I DX  
 581 I V N D L K L E L N K K I K E I T E K I E Y I K K A V D L K K E I E K D N V Y I E8  
 160 Y L N K L K L E L K E K I K N I S D K N E Y V K K A V D L K K T I E N N N I Y I DX  
 621 D E L A K E P P Y Q I T K Y I E K K N E I Y N T I K S D F D K I Y V G D I E Q L E8  
 200 D E L T K T S P Y Q V P E H L K N T D T I Y N T I K L E L S Q I Y E D D I D K L DX  
 661 Y N E M F S V V Q E S N I E H I E N K T E I L T L K T K I D N V Y N N I Q N M E E8  
 240 Y N E L S S I V Q E N D I D N V E D K T K L D D L Q S K I D N V Y S K I Q N M E DX  
 701 T E T V K S H L K N I E T N - N K L S E T I L D I I K Y I Y G E I T N E L N K T E8  
 280 N E T V E S H L T N I E T N K N K L S D T I L A I K K Y I Y G E I S K D L N K T DX  
 740 L E D F K N K E K G L S N K I D E Y A K E N V Q L N V Y K S N I L E I R K H Y N E8  
 320 L E D F K N K E K E L S N K I N D Y A K E N D Q L N V Y K S K I S E I R N H Y N DX  
 780 D Q I N I D N I K E K E A K Q N Y D Q F K E H M K T I P P N E N E I S K T I N E E8  
 360 S Q I N I D N T K E G E A K Q N Y D K S N E H M T K I S T N E N E I S K F I N E DX  
 820 I K I M K D E F L S K V N K Y N D F D K V Y K E K V E S E H N K F T E L T N K I E8  
 400 V K S M K D A F L S K V D K Y I N F D N N Y K E N V N L E H T Q F T E L T D K I DX  
 860 K T E V S D E E I K K Y E N K F N D S K S L I N E T K K S I E E E Y Q N I N T L E8  
 440 K A E V S D E K L S K H E K S F N D S K S L I N E T K N S I E K E Y Q N I N T L DX  
 900 K K V D D Y I K V C L N T N E L I T N C H N K Q T T L K D K L N Q N I K T I K E E8  
 480 K K V D E Y I K V C E I T K E S I T K F S S K Q N T L K C M L N Q N I K T V K E DX  
 940 T N S I D K I Y T D K F E N I L T D K K T E L E T K F T G L S L N N H E S N N K E8  
 520 T N S I E N F Y K D K F E N T L T N K I N E L D K T F K D A S L N D Y E L N N N DX  
 980 E L L T Y F Y D L K A N L G K N K E N M L Y K Q F N E K E K A V E D I K K K N V E8  
 560 E L M Q Y F N N L K E N L G K D K E N M L Y N Q L A E K E K T F N D I K K N N T DX  
 1020 D I N K I V S N I E I T I Y T S I Y N I N E D T E N E I G K S I E L L N T K V L E8  
 600 H I N E E I S N I E I A I Y A S I Y N I S E E I K N E I G K N I E L L N A K V V DX  
 1060 E K V K A N V T N L N E I K E K L K D Y D F Q D F G K E K N I K Y P D E - N K I E8  
 640 E K V K A N V T N L N E I K E K L K H Y N F H D F G K D E N I K Y A N E V N K I DX  
 1099 K N D I D T L N Q K I D K S I E T L T E I K K N S E N H I D E I K G Q I D K L K E8  
 680 K D D I K T V S Q Q I D DX  
 1139 K V P N K T M F N E D E8  
 691 DX

**Figure 6.3.3** Alignment of clone DX with clone D1.2

An alignment of the deduced amino acid sequences of clones D1.2 and DX, with identities shaded in green, demonstrates that the homology extends from position 125 to 439 in the open reading frame of clone D1.2.

The alignment was performed using the MegAlign program of DNASTar.

81	I A Q Q R A N F N Y N F N K L K H L K L L Y K Q Q E D I M R N F H K T I E R L I	D1.2
	- - - - -	DX
121	F G K P D P N N T N V F K E Y I Y E F N L W I P K S K L T L L E K R F F E I F E	D1.2
1	- - - - D P N N T N V F K E Y I Y E F N I S I P K S K L T A L E K R F F E I F E	DX
161	E K W N S Y E I K K D I D E N S K Q Y N V V K L I L Q Y M K E L T D V I D F M E	D1.2
37	E K W N S Y E I K K D I D E N S K Q Y N V V K L I L Q Y M K E L T D L I D F M D	DX
201	N Y Q P D E V P T R R I I G F E V E W R I K A L L Y S E V E A G V K K S Y E S V	D1.2
77	N Y Q P D E V P T R S I I G F E V E W R I N Q T L Y T E V E A G V K K S Y E S V	DX
241	K N W R K S I L E I N K K L E E E N E K V I K L E T Q I K D L F N Q Y L K I I D	D1.2
117	K N W R K S M V E I N K K L E E E N E K V I K L E T Q I N D L F N Q Y L K I N D	DX
281	E N I Y I N K L K L E L K E K I K N I S D K N E Y I K K A V D L K K I I E N N N	D1.2
157	E N I Y L N K L K L E L K E K I K N I S D K N E Y V K K A V D L K K T I E N N N	DX
321	A Y I D E L A K T S P Y Q V P E H L K S T D T I Y S T I K S E L S Q I Y E D D I	D1.2
197	I Y I D E L T K T S P Y Q V P E H L K N T D T I Y N T I K L E L S Q I Y E D D I	DX
361	D K L Y N E L S S I V Q E N D I D N V E D K T K L D D L Q S K I D N V Y S K I Q	D1.2
237	D K L Y N E L S S I V Q E N D I D N V E D K T K L D D L Q S K I D N V Y S K I Q	DX
401	N M E T A T V E S H L T N I E T N K N K L S D T I V E I K K Y I Y G E I S K D -	D1.2
277	N M E N E T V E S H L T N I E T N K N K L S D T I L A I K K Y I Y G E I S K D L	DX
	- - - - -	D1.2
317	N K T L E D F K N K E K E L S N K I N D Y A K E N D Q L N V Y K S K I S E I R N	DX

**CHAPTER SEVEN**

**CHROMOSOMAL LOCATIONS OF Py235 GENES**

## 7.1 Chromosome blots

The chromosomal location of each of the new genes belonging to the 235 kDa family was determined by hybridisation of gene-specific probes to a Southern blot of *Plasmodium yoelii* YM chromosomes. The chromosomes were separated using pulsed-field gel electrophoresis on a 1% agarose gel, and transferred to a nylon membrane following depurination and denaturation of the DNA as described in sections 2.13 and 2.11 of Chapter Two.

In each case, the probe for subsequent hybridization was generated by restriction enzyme digestion of the appropriate clone with *Bam*HI and *Kpn*I, the digest run on a 1% agarose gel, and the liberated insert purified.

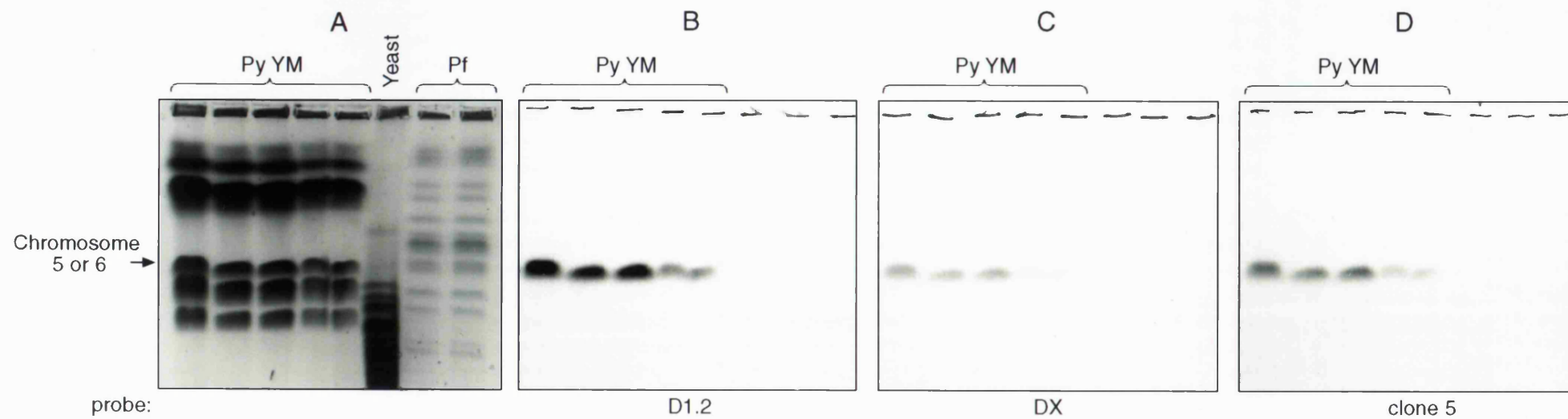
Hybridization of the chromosome blot to the inserts of clone D1.2, DX and clone 5 demonstrated that the sequences are all present on the same chromosome, chromosome 5 or 6. There is no cross-reactivity with *Plasmodium falciparum* chromosomes (figure 7.1)

The fact that all three of the newly isolated Py235 sequences are present on the same chromosome strongly suggests that they arose through gene duplication of a common ancestral gene. The sequences were isolated through hybridization with a probe generated from the E8 gene sequence, which has been localised to chromosome 1 (C. Owen, unpublished). However, the other member of the Py235 multigene family which has been extensively sequenced, E3, has been localised to chromosome 5. A comparison of the sequence homology of the new sequences with the corresponding regions of E3 and E8 may be of use when examining the

evolutionary relationships of the gene family members. This will be discussed in detail in Chapter Eight.

**Figure 7.1**                      Chromosome blot

- a)     Ethidium bromide stained CHEF gel separating the chromosomes of *P. yoelii* YM.
- b)     Southern blot performed using the insert of clone D1.2 as a probe. Hybridisation of the probe is seen to only one *P. yoelii* YM chromosome, chromosome 5 or 6. There is no cross-reactivity with *P. falciparum* chromosomes.
- c)     Southern blot performed using the insert of clone DX as a probe. Hybridisation of the probe is seen to only one *P. yoelii* YM chromosome, chromosome 5 or 6. There is no cross-reactivity with *P. falciparum* chromosomes.
- d)     Southern blot performed using the insert of clone 5 as a probe. Again, hybridisation of the probe is seen to only one *P. yoelii* YM chromosome, chromosome 5 or 6. There is no cross-reactivity with *P. falciparum* chromosomes.





**CHAPTER EIGHT**

**GENERAL DISCUSSION**

## 8.1 Homologies within the Py235 multigene family

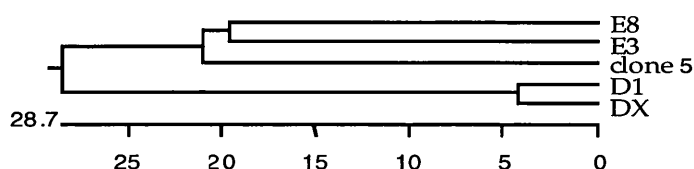
The sequences identified in the preceding chapters represent regions of different members of the same multigene family. It is of interest to examine the relatedness of the sequences, which may give some insight into the evolutionary origin of such a large multigene family. For example, a member of the family may have more sequence homology with another specific member of the family than to any others. This may mean that these family members have a common ancestral gene which other members do not share, as the gene duplication event which gave rise to these members may have occurred relatively recently.

An alignment of the sequences is shown in figure 8.1.1

The table below shows a comparison of the deduced amino acid sequences of the genes. The figures shown represent the percentage amino acid identity of the pairs of sequences in the region where all of the sequences overlap, which is shown in figure 8.1.1.

	E8	E3	D1	DX	clone 5
E8	***	61	40	40	51
E3	61	***	41	40	59
D1	40	41	***	94	44
DX	40	40	94	***	44
clone 5	51	59	44	44	***

From these analyses, it appears that the sequences can be split into two fairly distinct types. DX and D1 fall into one group, and E3, E8, and clone 5 fall into another, on the basis of sequence identity. An assessment of relatedness was performed using the phylogenetic analysis in the MegAlign™ program of DNASTar.



From this examination, it would seem that D1 and DX are the most closely related, followed by E3 and E8. Clone 5 is more closely related to E3 and E8 than D1 and DX. This concurs with the hypothesis that there appear to be two types of Py235 gene.

The two types of sequence may be equivalent to reticulocyte binding protein-1 (RBP-1) and reticulocyte binding protein-2 (RBP-2) in *P. vivax* (Galinski and Barnwell, 1996; Galinski et al., 1992). RBP-1 and RBP-2 are significantly related to one another, and form heterodimers which bind to the surface of reticulocytes. These proteins are discussed in more detail in the following section.

**Figure 8.1.1**      Alignment of overlapping regions of E8, E3, D1, DX, and clone5

The overlapping regions of the Py235 clones' open reading frames were aligned using the Pileup program of GCG.

The homologous regions of each clone were as follows:

E8:            residues 437-691

E3:            residues 54-308

clone 5:      residues 424-678

D1.2:        residues 125-378

DX:           residues 1-254

■ residues which are identical in all five sequences

■ residues which match sequence E8, but not D1

■ residues which match sequence D1, but not E8

■ residues which match both E8 and D1

437	D	S	D	G	N	I	I	E	E	N	D	I	D	N	F	D	L	S	K	P	K	N	N	F	K	L	L	E	G	E	E8
54	D	S	D	G	N	I	I	E	K	D	E	I	S	D	F	N	L	L	K	P	Q	S	T	F	K	S	L	E	A	E	E3
424	D	S	N	G	N	T	N	V	F	K	E	Y	I	Y	E	F	N	L	W	I	P	K	S	K	L	T	L	E	K	R	clone 5
125	D	P	N	N	T	N	V	F	K	E	Y	I	Y	E	F	N	I	S	I	P	K	S	K	L	T	A	L	E	K	R	D1.2
1	D	P	N	N	T	N	V	F	K	E	Y	I	Y	E	F	N	I	S	I	P	K	S	K	L	T	A	L	E	K	R	DX

467	L	N	G	V	F	E	N	K	W	N	D	Y	K	N	N	K	N	L	D	Q	F	K	D	T	M	K	K	I	I	L	E8
84	F	L	K	I	F	E	N	K	W	D	F	Y	K	N	T	Q	N	L	N	G	T	K	D	P	M	K	N	T	I	S	E3
454	F	H	T	I	F	E	N	T	L	D	F	Y	K	N	K	Q	N	L	D	N	S	N	D	T	K	K	N	T	I	S	clone 5
155	F	F	E	I	F	E	E	K	W	N	S	Y	E	I	K	K	D	I	D	E	-	N	S	K	Q	Y	N	V	V	K	D1.2
31	F	F	E	I	F	E	E	K	W	N	S	Y	E	I	K	K	D	I	D	E	-	N	S	K	Q	Y	N	V	V	K	DX

497	L	I	I	Q	Y	M	N	E	F	K	G	L	N	E	D	A	M	T	K	L	K	N	E	G	I	S	Q	K	F	V	I	E8
114	L	I	L	P	L	M	N	Q	F	K	L	N	K	S	M	I	L	N	L	I	N	D	G	V	L	Q	K	H	M	T	E3	
484	L	I	L	P	L	M	N	Q	F	K	L	N	K	S	M	I	L	N	L	I	N	D	G	V	L	Q	K	H	M	T	clone 5	
184	L	I	L	Q	Y	M	K	E	L	T	D	V	I	D	F	M	E	N	Y	Q	P	D	E	V	P	T	R	R	I	I	D1.2	
60	L	I	L	Q	Y	M	K	E	L	T	D	V	I	D	F	M	E	N	Y	Q	P	D	E	V	P	T	R	S	I	I	DX	

527	N	N	Q	I	K	Q	K	F	D	K	V	S	T	Y	D	E	K	K	E	G	F	E	S	S	L	E	L	A	K	N	W	E8	
144	S	N	Q	I	K	Q	K	F	L	N	V	S	T	Y	P	E	E	G	R	E	G	F	T	S	S	L	E	L	A	K	S	R	E3
514	I	T	Q	I	K	Q	K	F	L	N	Q	S	T	Y	E	E	R	K	E	G	F	T	S	A	L	E	L	A	E	R	W	clone 5	
214	G	F	E	V	E	W	R	I	K	A	L	L	Y	S	T	E	V	E	A	G	V	K	K	S	Y	F	S	V	K	N	W	D1.2	
90	G	F	E	V	E	W	R	I	N	Q	T	L	Y	S	T	E	V	E	A	G	V	K	K	S	Y	F	S	V	K	N	W	DX	

557	E	K	K	L	E	I	I	T	E	L	K	K	S	N	E	E	T	V	Q	L	D	I	K	I	R	E	L	I	K	E8	
174	E	K	K	L	E	I	I	T	E	L	K	K	S	N	E	E	T	V	R	L	E	K	E	I	R	E	L	I	K	E3	
544	E	K	K	L	E	I	I	T	E	L	K	K	S	N	E	E	T	V	K	L	E	K	E	I	R	E	L	I	K	clone 5	
244	R	K	S	I	L	E	I	N	K	K	L	E	E	E	N	E	K	V	I	K	L	E	T	Q	T	K	D	I	F	N	D1.2
120	R	K	S	M	V	F	I	N	K	K	L	E	E	E	N	E	K	V	I	K	L	E	T	Q	T	K	D	I	F	N	DX

587	Q	I	K	D	I	I	E	Q	K	I	V	N	D	L	K	L	E	L	N	K	K	I	K	E	I	T	E	K	I	E8	
204	K	Y	L	D	E	E	A	R	K	Y	L	E	G	L	K	L	E	L	N	K	K	I	K	D	I	I	A	K	I	E3	
574	K	Y	S	D	E	V	A	E	K	K	H	V	E	L	K	L	E	L	N	K	K	I	K	E	T	I	Y	N	K	clone 5	
274	Q	Y	L	K	I	I	D	E	N	I	Y	I	N	K	L	K	L	E	L	K	E	K	I	K	N	I	S	D	K	N	D1.2
150	Q	Y	L	K	I	I	D	E	N	I	Y	I	N	K	L	K	L	E	L	K	E	K	I	K	N	I	S	D	K	N	DX

617	E	Y	I	K	A	V	D	L	K	K	E	I	E	K	D	N	V	Y	I	D	E	L	A	K	E	P	P	Y	Q	E8	
234	E	Y	V	K	N	T	V	E	L	K	K	E	I	E	K	N	N	A	Y	I	D	E	L	A	N	Q	S	P	Y	K	E3
604	E	Y	I	K	A	I	D	L	K	K	A	I	E	N	N	N	K	Y	I	D	E	L	G	K	N	T	P	F	Q	clone 5	
304	E	Y	I	K	A	V	D	L	K	K	I	E	N	N	N	N	A	Y	I	D	E	L	A	K	T	S	P	Y	Q	D1.2	
180	E	Y	V	K	A	V	D	L	K	K	T	I	E	N	N	N	I	Y	I	D	E	L	T	K	T	S	P	Y	Q	DX	

647	I	T	K	Y	I	E	K	K	N	E	I	Y	N	T	I	K	S	D	F	D	K	I	Y	V	G	D	I	E	Q	L	E8	
264	V	T	G	Y	I	E	N	K	N	T	I	Y	N	T	I	K	S	S	E	L	S	E	I	Y	K	G	N	I	V	E	L	E3
634	I	E	E	Y	V	K	K	D	T	I	Y	N	T	I	K	S	S	E	L	S	E	I	Y	K	G	N	I	V	E	L	clone 5	
334	V	P	E	H	L	K	S	T	D	T	I	Y	N	T	I	K	S	S	E	L	S	Q	I	Y	E	D	D	I	D	K	L	D1.2
210	V	P	E	H	L	K	N	T	D	T	I	Y	N	T	I	K	S	S	E	L	S	Q	I	Y	E	D	D	I	D	K	L	DX

677	Y	N	E	M	F	S	V	V	Q	E	S	N	I	E	E8
294	Y	N	E	L	S	S	V	V	K	E	D	P	I	D	E3
664	Y	N	E	L	S	S	V	V	Q	E	N	T	I	D	clone 5
364	Y	N	E	L	S	S	V	V	Q	E	N	D	I	D	D1.2
240	Y	N	E	L	S	S	V	V	Q	E	N	D	I	D	DX

## 8.2 Homologies with other proteins

Analysis of the entire open reading frame of the E8 gene showed that the protein has significant homology to both *P. vivax* reticulocyte binding protein-1 (RBP-1) and *P. vivax* reticulocyte binding protein-2 (RBP-2). These are high molecular mass proteins found at the anterior pole of *P. vivax* merozoites which have been demonstrated to bind to reticulocytes in an erythrocyte binding assay, and are implicated in controlling the specificity of red blood cell invasion by *P. vivax* merozoites (Galinski and Barnwell, 1996; Galinski et al., 1992). The homology is more pronounced in the comparison with PvRBP-2, but as the entire sequence for the RBP-2 gene is not available it is not known whether the homology extends throughout the entire gene.

The overall identity of Py235-E8 with PvRBP-1 is 18%, with a similarity of 38%. This extends over the whole protein, with regions of higher identity/similarity occurring. The alignment of the two sequences is represented graphically in figure 8.2.1.

The identity of Py235-E8 with PvRBP-2 is 22%, with a similarity of 42%. This extends from amino acid position 1160 in the sequence of Py235-E8, and therefore does not include any of the new sequence identified in the studies described in this thesis.

The significance of the homology of the rhoptry protein with *P. vivax* reticulocyte binding proteins is considerable. Previous work has shown that there is a functional homology between the three proteins, in that they appear to be responsible for selection of the subset of red blood cells which a parasite is capable of invading (Galinski and Barnwell, 1996; Galinski et al., 1992). Both of the reticulocyte binding proteins are predicted to have a secondary structure consisting almost entirely of  $\alpha$ -helices, with a large external N-terminal domain, membrane-spanning domains, and a short

cytoplasmic C-terminal tail (Galinski and Barnwell, 1996; Galinski et al., 1992). The predicted structure of the full length Py235-E8 protein is strikingly similar to this, both in the structural domains and the overall secondary structure predictions. This is shown in figure 8.2.2.

PvRBP-1 has two RGD amino acid motifs which could serve as adhesive domains in binding reticulocyte ligands. Such motifs have been shown to be integrin binding sites in a number of (Hynes, 1987; Hynes, 1992; Kuhn and Eble, 1994). Integrins are a family of related proteins which are usually expressed on the surface of cells and are involved in intracellular adhesion. It is possible that PvRBP-1 binds to an integrin-like molecule on the surface of the reticulocyte *via* its RGD motif, or that the interaction which holds RBP-1 and RBP-2 in their complex is mediated by the interaction of a domain of RBP-2 with the RGD motifs.

Interestingly, Py235-E8 also shows homology to a *Saccharomyces cerevisiae* integrin homologue, *uso1* (Hostetter et al., 1995). The identity between the two amino acid sequences is 23%, and the similarity is 56% extending over a stretch of 1015 amino acids (including gaps) from position 230 to position 1200 in the deduced amino acid sequence of Py235-E8. This is represented graphically in figure 8.2.3. The significance of such a relationship is not clear, as the identity of the ligand for Py235 on the red blood cell surface is not yet known, nor has the region of Py235 which is responsible for the binding interaction. Integrins are cell surface proteins which are responsible for a wide range of cell-cell interactions, and subsequent signal transduction within those cells. It may be that within the region of homology between *uso1* and Py235.E8 lies a binding domain which is conserved between species. However, no other integrins or integrin homologues were present when a database search for similar proteins to Py235.E8 was carried out. This suggests that the similarity between the two

proteins, although relatively high, may not be significant. However, this does perhaps warrant a further, more detailed, investigation.

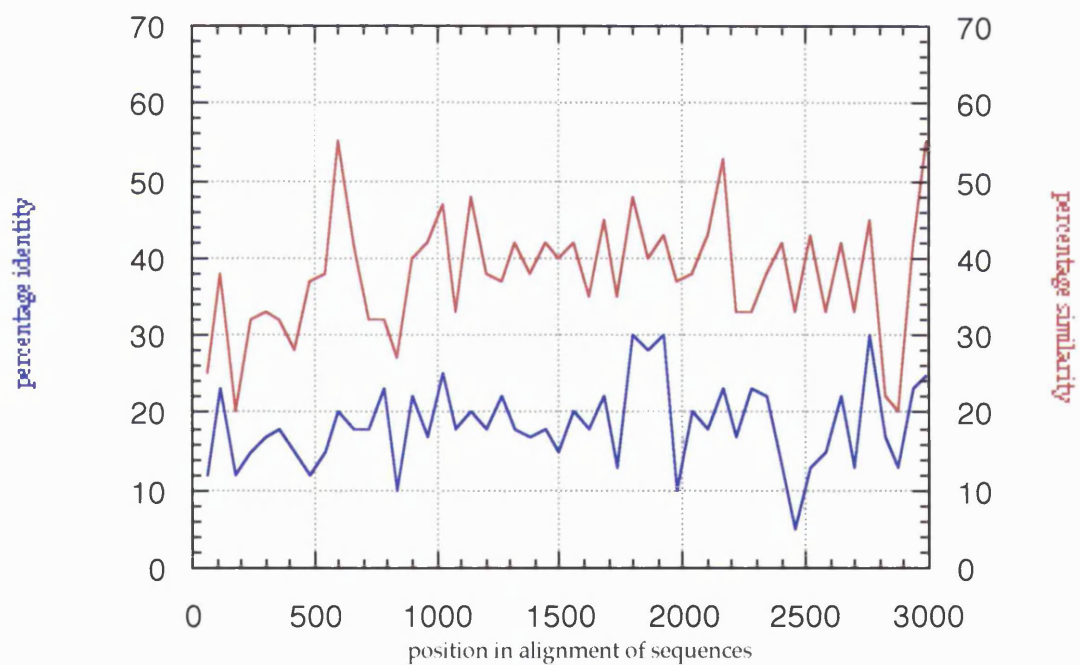
In the N-terminal sequence of Py235.E8, there are eight cysteine residues. These are conserved in position in clone 5 (figure 4.5.4). Cysteine residues are often important determinants of the three dimensional structure which a folded polypeptide assumes. They form covalently associated disulphide bridges with other cysteine residues to form a stable structure. The structure of certain *Plasmodium* antigens has been suggested from analysis of the relative positions of their cysteine residues. This has been performed for three proteins of *P. falciparum*, Pfs230, Pfs48/45, and Pf12 and the results have suggested that the proteins are uniquely related through structure (Carter et al., 1995). The disulphide bonds which are present in *P. chabaudi* AMA-1 have been determined and demonstrated to be crucial to the molecule's ability to induce an immune response (Hodder et al., 1996). Examples of *Plasmodium* proteins in which cysteine residues and disulphide bridges have been shown to play a crucial role in structure and function are the Duffy binding protein and merozoite surface protein-1. In the case of the latter, two epidermal growth factor (EGF)-like modules are present at the C terminus of the protein. EGF-like modules have also been demonstrated to be present in Pfs25, a sexual stage antigen of *P. falciparum* that is present on the ookinete surface and contains four predicted EGF-like modules (Kaslow et al., 1989), merozoite surface protein-4 which contains one EGF-like module (Marshall et al., 1997) The conformation of these structures is maintained by disulphide bridges between cysteine residues. It is therefore possible that the cysteine residues seen in the Py235 sequences are important in maintaining its structure and may play a role in determining the protein's ligand binding specificity.



**Figure 8.2.1 Homology of Py235.E8 with PvRBP-1**

The aligned sequences (not shown) were divided sequentially into blocks of 60 amino acids and the identity and homology of these individual blocks of sequence were expressed as percentages.

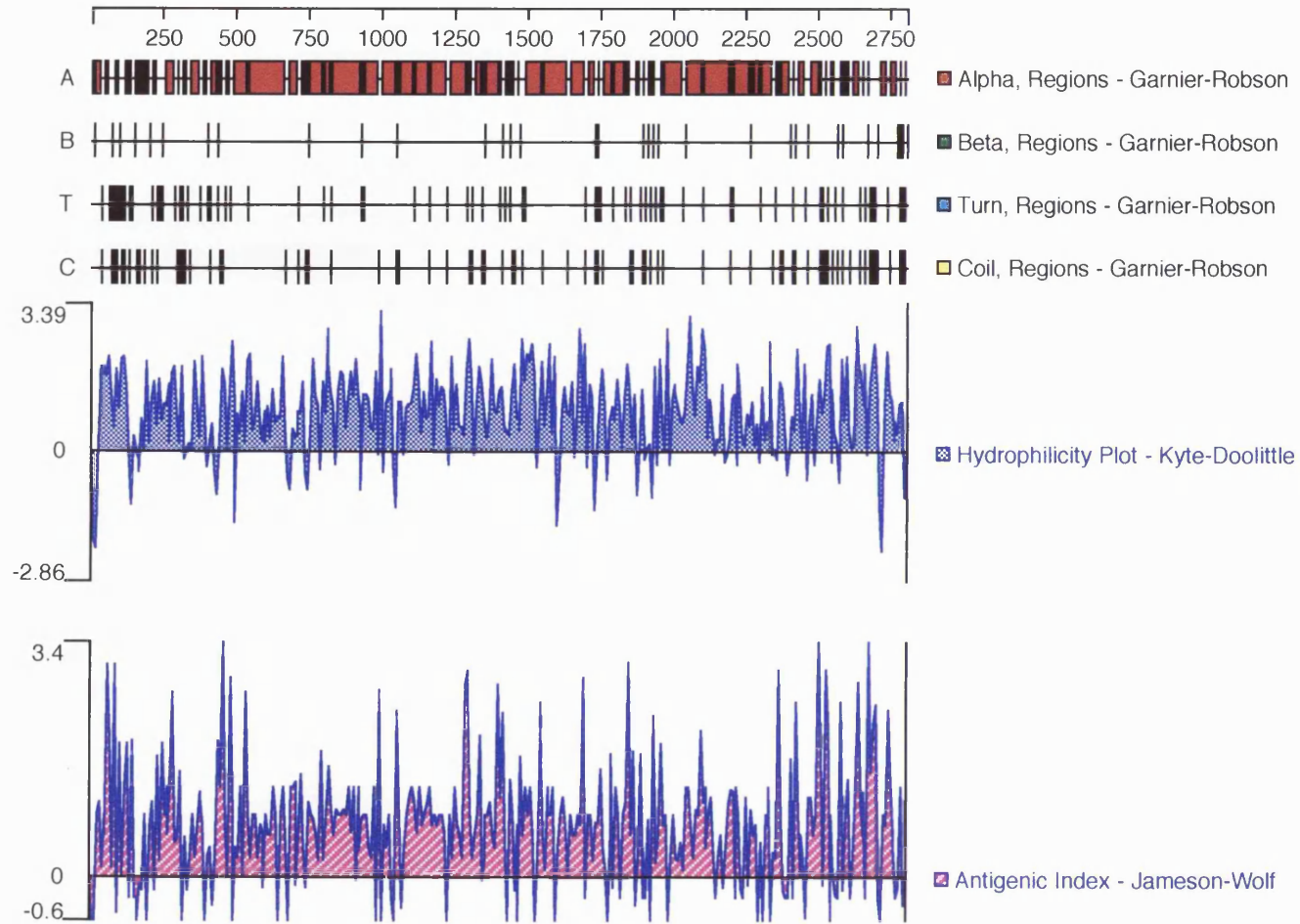
### Comparison of E8 and PvRBP-1 sequences



**Figure 8.2.2** Predicted secondary structure of Py235.E8

The secondary structure of Py235.E8 was analysed using the Protean™ program of DNASTar™.

The results of such an analysis show that the protein is predicted to be largely  $\alpha$ -helical.

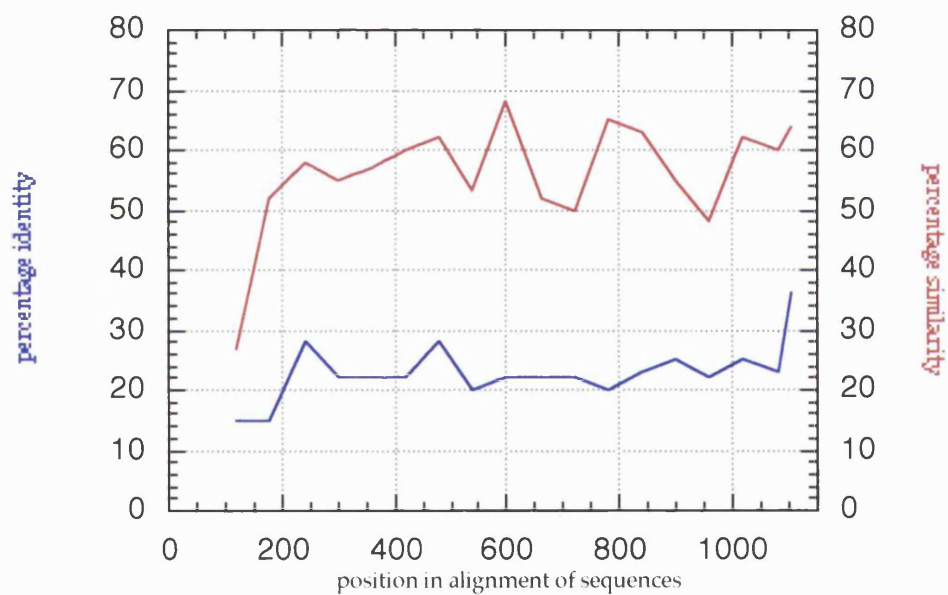


**Figure 8.2.3** Homology of Py235.E8 with the *S. cerevisiae* integrin homologue

The identity between the two amino acid sequences is 23%, and the similarity is 56%, extending over a stretch of 1015 amino acids (including gaps) from position 230 to position 1200 in the deduced amino acid sequence of Py235-E8.

The aligned sequences (not shown) were divided sequentially into blocks of 60 amino acids and the identity and homology of these individual blocks of sequence were expressed as percentages.

### Comparison of E8 and yeast integrin homolog sequences



### 8.3 Future work

Although one of the initial aims of this project was to identify promoter elements for members of the Py235 multigene family, this was not achieved. The main reason for this was the inability to identify, unequivocally, the transcription start site of the E8 gene. Clone E8.7, described in section 4.5, contains 1710 bp of DNA upstream of the putative initiating ATG codon. Although some of this will be untranslated leader sequence, one would anticipate that within this sequence are contained promoter elements. As has been demonstrated in section 5.5, the region contains many GC-rich elements, some identical to those seen in other *Plasmodium* genes. The majority of promoter elements identified in *Plasmodium* genes to date are within 1000 bp of the transcription start site of that gene and the untranslated regions of *Plasmodium* genes tend to be relatively short, generally no more than a few hundred base pairs (Lanzer et al., 1993).

Identifying promoter regions for the different members of the multigene family may give a clue to how expression of the genes is co-ordinated. Recently, it has been demonstrated that different parasite lines express a different pattern of transcription of Py235 genes (Dr. P. Preiser, Experimental Parasitology, in press). This has been discussed in detail in the introduction. Although it is not clear that the process of screening identified transcripts from every member of the gene family, it does show that the non-lethal 17X line appears to express a larger subset of genes than the lethal YM line. It would be interesting to see whether this expression pattern is due to differences in the genes' promoters. Perhaps a translocation event occurs which brings the gene to be expressed into an environment which results in its expression.

With the availability of sequence for five distinct gene family members now available, it would be simple to analyse the expression of these family members in the two parasite lines. Although this analysis would not cover such a large number of genes as the use of a probe from the 3' end of the gene, the 5' ends of the genes are much less conserved than the 3' end and therefore the study would be more specific.

The eight conserved cysteine residues in the N-terminus of Py235 genes E8 and clone 5 may be involved in determining the structure and ligand binding specificity, as discussed in the previous section. Binding studies with regions of the protein containing the cysteine residues would be informative in determining the proteins' ligand(s) on the surface of the red blood cell.

Further investigations are warranted into whether there are, indeed, two "sub-families" within the Py235 multigene family, the first represented by E8, E3, and clone 5 and the second by DX and D1. It is possible that the two types of gene have distinct expression patterns, as discussed above, or that they form a heterodimeric complex in an analogous fashion to the reticulocyte binding proteins of *P. vivax* (Galinski et al., 1996).



**CHAPTER NINE**

**REFERENCES**

**References**

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

Adams, J. H., Sim, B. K. L., Dolan, S. A., Fang, X., Kaslow, D. C., and Miller, L. H. (1992). A family of erythrocyte binding proteins of malaria parasites. *Proc Natl Acad Sci USA* 89, 7085-7089.

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

Alt, F. W., Blackwell, T. K., and Yancopoulos, G. D. (1987). Development of the primary antibody repertoire. *Science* 238, 1079-1087.

Bannister, L. H., and Mitchell, G. H. (1995). The role of the cytoskeleton in *Plasmodium falciparum* merozoite biology: an electron-microscopic view. *Annals of Tropical Medicine and Parasitology* 89, 105-111.

Baruch, D. I., Pasloske, B. L., Singh, H. B., Bi, X., Ma, X. C., Feldman, M., Taraschi, T. F., and Howard, R. J. (1995). Cloning the *P. falciparum* gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. *Cell* 82, 77-87.

Berendt, A. R., Ferguson, D. J. P., and Newbold, C. I. (1993). Sequestration in *Plasmodium falciparum* malaria: Sticky cells and sticky problems. *Parasitology Today* 6, 247-254.

Biggs, B. A., Anders, R. F., Dillon, H. E., Davern, K. M., Martin, M., Peterson, C., and Brown, G. V. (1992). Adherence of infected erythrocytes to venular endothelium selects for antigenic variants of *Plasmodium falciparum*. *Journal of Immunology* 149, 2047-2054.

Blackman, M. J., Whittle, H., and Holder, A. A. (1991). Processing of the *Plasmodium falciparum* merozoite surface protein-1: identification of a 33 kilodalton secondary processing product which is shed prior to erythrocyte invasion. *Molecular and Biochemical Parasitology* 49, 35-44.

Blackman, M. J., Scott-Finnigan, T. J., Shai, S. and Holder, A. A. (1994). Antibodies inhibit the protease-mediated processing of a malaria merozoite surface protein. *Journal of Experimental Medicine* 180, 389-393.

Borre, M. B., Owen, C. A., Keen, J. K., Sinha, K. A., and Holder, A. A. (1995). Multiple genes code for high-molecular-mass rhoptry proteins of *Plasmodium yoelii*. *Molecular and Biochemical Parasitology* 70, 149-155.

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

Borst, P., and Rudenko, G. (1994). Antigenic variation in African trypanosomes. *Science* 264, 1872-1873.

Braun-Breton, C., and Silva, L. P. d. (1988). Activation of a *Plasmodium falciparum* protease correlated with merozoite maturation and erythrocyte invasion. *Biol. Cell.* 64, 223-231.

Breathnach, and Chambon (1981). Organization and expression of eucaryotic split genes coding for proteins. *Annual Review of Biochemistry* 50, 349-383.

Brown, H. J., and Coppel, R. L. (1991). Primary structure of a *Plasmodium falciparum* rhoptry antigen. *Molecular and Biochemical Parasitology* 49, 99-110.

Bushell, G. R., Ingram, L. T., Fardoulis, C. A., and Cooper, J. A. (1988). An antigenic complex in the rhoptries of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 28, 105-112.

Carter, R., Coulson, A., Bhatti, S., Taylor, B. J., and Elliot, J. F. (1995). Predicted disulfide bonded structures for three uniquely related proteins of *Plasmodium falciparum*, Pfs230, Pfs48/45 and Pf12. *Molecular and Biochemical Parasitology* 71, 203-210

Chappel, J. A., and Holder, A. A. (1993). Monoclonal antibodies that inhibit *Plasmodium falciparum* invasion *in vitro* recognise the first growth factor-like domain of merozoite surface protein-1. *Molecular and Biochemical Parasitology* 60, 303-312.

Chitnis, C. E., and Miller, L. H. (1994). Identification of the erythrocyte binding domains of *Plasmodium vivax* and *Plasmodium knowlesi*

proteins involved in erythrocyte invasion. *The Journal of Experimental Medicine* 180, 497-506.

Clark, J. T., Anand, R., Akoglu, T., and McBride, J. S. (1987). Identification of proteins associated with the rhoptry organelles of *Plasmodium falciparum*. *Parasitology Research* 73, 425-434.

Clark, J. T., Donachie, S., Anand, R., Wilson, C. F., Heidrich, H. G., and McBride, J. S. (1989). 46-53 kDa glycoprotein from the surface of *Plasmodium falciparum* merozoites. *Molecular and Biochemical Parasitology* 32, 15-24.

Cooper, J. A., Atkins, A., and Saul, A. J. (1989). N-terminal sequencing of the 105 kilodalton rhoptry antigen of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 33, 203-204.

Cooper, J. A., Ingram, L. T., Bushell, G. R., Fardoulis, G. A., Stenzel, D., Schofield, L., and Saul, A. J. (1988). The 140/130/105 kilodalton soluble complex in the rhoptries of *Plasmodium falciparum* consists of discrete polypeptides. *Molecular and Biochemical Parasitology* 29, 251-260.

Coppel, R. S., Bianco, A. E., Culvenor, J. G., Crewther, P. E., Brown, G. V., and Kemp, D. J. (1987). A cDNA clone expressing a rhoptry protein of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 25, 73-81.

Crabb, B. S., and Cowman, A. F. (1996). Characterization of promoters and stable transfection by homologous and non-homologous recombination in *Plasmodium falciparum*. *Proc Natl Acad Sci USA* 93, 7289-7294.

Crewther, P. E., Culvenor, J. G., Silva, A., Cooper, J. A., and Anders, R. F. (1990). *Plasmodium falciparum*: Two antigens of similar size are located in different compartments of the rhoptry. *Experimental Parasitology* 70, 193-206.

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

Das, A., Elmendorf, H. G., Li, W. I., and Haldar, K. (1994). Biosynthesis, export and processing of a 45 kDa protein detected in membrane clefts of erythrocytes infected with *Plasmodium falciparum*. *Biochemical Journal* 302, 487-496.

Deans, J. A., Thomas, A. W., and Cohen, S. (1983). Stage-specific protein synthesis by asexual blood stage parasites of *Plasmodium knowlesi*. *Molecular and Biochemical Parasitology* 8, 31-44.

Dluzewski, A. R., Mitchell, G. H., Fryer, P. R., Griffiths, S., Wilson, R. J. M., and Gratzer, W. B. (1992). Origins of the parasitophorous vacuole membrane of the malaria parasite, *Plasmodium falciparum*, in human red blood cells. *Journal of Cell Science* 102, 527-532.

Dvorak, J. A., Miller, L. H., Whitehouse, W. C., and Shiroishi, T. (1975). Invasion of erythrocytes by malaria merozoites. *Science* 187, 748-750.

Dynan, W. S., and Tjian, R. (1985). Control of eukaryotic mRNA synthesis by sequence-specific DNA-binding proteins. *Nature* 316, 774-778.

Dynan, W. S., and Tjian, R. (1983). Isolation of transcription factors that discriminate between different promoters recognized by RNA polymerase II. *Cell* 32, 669-680.

Eisenberg, D., Schwarz, E., Komaromy, M., and Wall, R. (1984). Analysis of membrane and surface protein sequences with the hydrophobic plot. *The Journal of Molecular Biology* 179, 125-142.

Etzion, Z., Murray, M. C., and Perkins, M. E. (1991). Isolation and characterization of rhoptries of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 47, 51-62.

Facer, C. A. (1995). Erythrocytes carrying mutations in spectrin and protein 4.1 show differing sensitivities to invasion by *Plasmodium falciparum*. *Parasitol Res* 81, 52-57.

Falanga, P. B., DaSilveira, J. F., and DaSilva, L. P. (1982). Proteins synthesized during specific stages of the schizogonic cycle and conserved in the merozoites of *Plasmodium chabaudi*. *Molecular and Biochemical Parasitology* 6, 55-65.

Freeman, R. R., Trejdosiewicz, A. J., and Cross, G. A. M. (1980). Protective monoclonal antibodies recognising stage-specific merozoite antigens of a rodent malaria parasite. *Nature* 284, 366-368.

Galinski, M. R., and Barnwell, J. W. (1996). *Plasmodium vivax*: Merozoites, invasion of reticulocytes and considerations for malaria vaccine development. *Parasitology Today* 12, 20-29.

Galinski, M. R., Medina, C. C., Ingravallo, P., and Barnwell, J. W. (1992). A reticulocyte-binding protein complex of *Plasmodium vivax* merozoites. *Cell* 69, 1213-1226.

Grau, G. E. (1989). Tumour necrosis factor and disease severity in children with falciparum malaria. *New England Journal of Medicine* 320, 1586-1589.

Harnyuttanakorn, P., McBride, J. S., Donachie, S., Heidrich, H. G., and Ridley, R. G. (1992). Inhibitory monoclonal antibodies recognise epitopes adjacent to a proteolytic cleavage site on the RAP-1 protein of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 55, 177-186.

Hodder, A. N., Crewther, P. E., Matthew, M. L., Reid, G. E., Moritz, R. J., and Anders, R. F. (1996). The disulfide bond structure of *Plasmodium* apical membrane antigen-1. *Journal of Biological Chemistry* 271, 29446-29452.

Holder, A. A., and Freeman, R. R. (1981). Immunisation against blood-stage rodent malaria using purified parasite antigens. *Nature* 294, 361-364.

Holder, A. A., Lockyer, M. J., Odink, K. G., Sandhu, J. S., Riveros, M. V., Davey, L. S., Tizard, M. L. V., Schwarz, R. T., and Freeman, R. R. (1985). Primary structure of the precursor to the three major surface antigens of *Plasmodium falciparum* merozoites. *Nature* 317, 270-273.

Holder, A. A., Sandhu, J. S., Hillman, Y., Davey, L. S., Nicholls, S. C., Cooper, H., and Lockyer, M. J. (1987). Processing of the precursor to the major merozoite surface antigens of *Plasmodium falciparum*. *Parasitology* 94, 199-208.



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

Hostetter, M. K., Tao, N. J., Gale, C., Herman, D. J., McClellan, M., Sharp, R. L., and Kendrick, K. E. (1995). Antigenic and functional conservation of an integrin I-domain in *Saccharomyces cerevisiae*. *Biochemical and Molecular Medicine* 55, 122-130.

Howard, R. F., Stanley, H. A., Campbell, G. H., and Reese, R. T. (1984). Proteins responsible for a punctate fluorescence pattern in *Plasmodium falciparum* merozoites. *American Journal of Tropical Medicine and Hygiene* 33, 1055-1059.

Howard, R. J., and Pasloske, B. L. (1993). Target antigens for asexual malaria development. *Parasitology Today* 9, 369-372.

Howard, R. F., and Schmidt C. M. (1995). The secretory pathway of *Plasmodium falciparum* regulates transport of p82/RAP-1 to the rhoptries. *Molecular and Biochemical Parasitology* 74, 43-54.

Hynes, R. O. (1987). Integrins: a family of cell surface receptors. *Cell* 48, 549-554.

Hynes, R. O. (1992). Integrins: versatility, modulation, and signalling in cell adhesion. *Cell* 69, 11-25.

Ishii, S., Xu, Y. H., Stratton, R. H., Roe, B. A., Merlino, G. T., and Pastan, I. (1985). Characterization and sequence of the promoter region of the human epidermal growth factor (EGF) receptor gene. *PNAS USA* 82, 4920-4924.

Iyer, V., and Struhl, K. (1995). Poly(dA:dT), a ubiquitous promoter element that stimulates transcription via its intrinsic DNA structure. *EMBO J* 14, 2570-2579.

Jaikaria, N. S., Rozario, C., Ridley, R. G., and Perkins, M. E. (1993). Biogenesis of rhoptry organelles in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 57, 269-280.

Kaslow, D. C., Syin, C., McCutchan, T. F., and Miller, L. H. (1989). Comparison of the primary structure of the 25kDa ookinete surface antigens of *Plasmodium falciparum* and *Plasmodium gallinaceum* reveal six conserved regions. *Molecular and Biochemical Parasitology* 33, 283-287.

Keen, J., Holder, A., Playfair, J., Lockyer, M., and Lewis, A. (1990). Identification of the gene for a *Plasmodium yoelii* rhoptry protein. Multiple copies in the parasite genome. *Molecular and Biochemical Parasitology* 42, 241-246.

Keen, J. K., Sinha, K. A., Brown, K. N., and Holder, A. A. (1994). A gene coding for a high-molecular mass rhoptry protein of *Plasmodium yoelii*. *Molecular and Biochemical Parasitology* 65, 171-177.

Klausner, R. D., Donaldson, J. G., and Lippincott-Schwartz, J. (1992). Brefeldin A: insights into the control of membrane traffic and organelle structure. *Jornal of Cell Biology* 116, 1071-1080.

Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283-292.

Kuhn, K., and Eble, J. (1994). The structural basis of integrin-ligand interactions. *Trends in Cell Biology* 4, 256-261.

Kwiatkowski, D. (1993). Anti-tumour necrosis factor therapy inhibits fever in cerebral malaria. *Q.J. Med.* 86, 91-98.

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

Lanzer, M., DeBruin, D., and Ravetch, J. V. (1992). Transcription mapping of a 100 kb locus of *Plasmodium falciparum* identifies an intergenic region in which transcription terminates and reinitiates. *The EMBO Journal* 11, 1949-1955.

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

Lewis, A. P. (1990). Sequence analysis upstream of the gene encoding the precursor to the major merozoite surface antigens of *Plasmodium yoelii*. *Molecular and Biochemical Parasitology* 39, 285-288.

Ling, I. T., Ogun, S. A., and Holder, A. A. (1994). Immunisation against malaria with a recombinant protein. *Parasite Immunology* 16, 63-67.

Locher, C. P., Vandekerckhove, F., and Tam, L. Q. (1992). Three matrix metalloproteinases form a non-covalent association with the rhoptry-associated protein-1 of *Plasmodium falciparum*. *Biochimica et Biophysica Acta* 1160, 275-280.

Lustigman, S., Anders, R. F., Brown, G. V., and Coppel, R. L. (1988). A component of an antigenic rhoptry complex of *Plasmodium falciparum* is modified after invasion. *Molecular and Biochemical Parasitology* 30, 217-224.

Magowan, C., Wollish, W., Anderson, L., and Leech, J. (1988). Cytoadherence by *Plasmodium falciparum*-infected erythrocytes is correlated with the expression of a family of variable proteins on infected erythrocytes. *Journal of Experimental Medicine* 168, 1307-1320.

Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). *Molecular Cloning - a laboratory handbook*: Cold Spring Harbor Laboratory Publications).

Marshall, V. M., Silva, A., Foley, M., Cranmer, S., Wang, L., McColl, D. J., Kemp, D. J., and Coppel, R. L. (1997). A second merozoite surface protein (MSP-4) of *Plasmodium falciparum* that contains an epidermal growth factor-like domain. *Infection and Immunity* 65, 4460-4467.

McAndrew, M. B., Read, M., Sims, P. F. G., and 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.

McBride, J. S., and Heidrich, H. G. (1987). Fragments of the polymorphic Mr 185000 glycoprotein from the surface of isolated *Plasmodium falciparum* merozoites form an antigenic complex. *Molecular and Biochemical Parasitology* 23, 71-84.

Mendis, K. N., and Carter, R. (1995). Clinical disease and pathogenesis in malaria. *Parasitology Today* 11, PTI2-PTI15.

Myler, P., Saul, A., and Kidson, C. (1983). The synthesis and fate of stage-specific proteins in *Plasmodium falciparum* cultures. *Molecular and Biochemical Parasitology* 9, 37-45.

Myler, P. J. (1989). Nucleotide and deduced amino acid sequence of the gp195 (MSA-1) gene from *Plasmodium falciparum* Palo Alto PLF-3/B11. *Nucleic Acids Research* 17, 5401.

Newbold, C. I., Boyle, D. B., Smith, C. C., and Brown, K. N. (1982). Stage specific protein and nucleic acid synthesis during the asexual cycle of the rodent malaria *Plasmodium chabaudi*. *Molecular and Biochemical Parasitology* 5, 33-44.

Newbold, C. I., and Marsh, K. (1990). Antigens on the *Plasmodium falciparum* infected erythrocyte surface are parasite derived: A reply. *Parasitology Today* 6, 320-322.

Nichols, B. A., Chiappino, M. L., and O'Connor, G. R. (1983). Secretion from the rhoptries of *Toxoplasma gondii* during host-cell invasion. *J. Ultrastruct. Res* 83, 85-98.

Nickoloff, J. A., and Reynolds, R. J. (1990). Transcription stimulates homologous recombination in mammalian cells. *Molecular Cell Biology* 10, 4837-4845.

Nordheim, A., and Rich, A. (1983). Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. *Nature* 303, 674-679.

Ogun, S. A., and Holder, A. A. (1996). A high molecular mass *Plasmodium yoelii* rhoptry protein binds to erythrocytes. *Molecular and Biochemical Parasitology* 76, 321-324.

Ogun, S. A., and Holder, A. A. (1994). *Plasmodium yoelii*: Brefeldin A-sensitive processing of proteins targeted to the rhoptries. *Experimental Parasitology* 79, 270-278.

Oka, M., Aikawa, M., Freeman, R. R., and Holder, A. A. (1984). Ultrastructural localization of protective antigens of *Plasmodium yoelii* merozoites by the use of monoclonal antibodies and ultrathin cryomicrotomy. *American Journal of Tropical Medicine and Hygiene* 33, 342-346.

Okada, A., and Alt, F. W. (1994). Mechanisms that control antigen receptor variable region gene assembly. *Semin. Immunol.* 6, 185-196.

Perkins, M., and Rocco, L. J. (1988). Sialic acid-dependent binding of *Plasmodium falciparum* merozoite surface antigen, Pf200, to human erythrocytes. *Journal of Immunology* 141, 3190-3196.

Perkins, M. E., and Zieffer, A. (1994). Preferential binding of *Plasmodium falciparum* SERA and rhoptry proteins to erythrocyte membrane inner leaflet phospholipids. *Infection and Immunity* 62, 1207-1212.

Perlman, D., and Halvorson, H. O. (1983). A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. *Journal of Molecular Biology* 167, 391-409.

Peterson, M. G., Marshall, V. M., Smythe, J. A., Crewther, P. E., Lew, A., Silva, A., Anders, R. F., and Kemp, D. J. (1989). Integral membrane protein located in the apical complex of *Plasmodium falciparum*. *Molecular and Cellular Biology* 9, 3151-3154.

Ravetch, J. V., Kochan, J., and Perkins, M. (1985). Isolation of the gene for a glycophorin-binding protein implicated in erythrocyte invasion by a malaria parasite. *Science* 227, 1593-1597.

Ridley, R. G., Lahm, H. W., Takacs, B., and Scaife, J. G. (1991). Genetic and structural relationships between components of a protective rhoptry antigen complex from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 47, 245-246.

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

Ridley, R. G., Takacs, B., Lahm, H. W., Delves, C. J., Goman, M., Certa, U., Matile, H., Woollett, G. R., and Scaife, J. G. (1990). Characterisation and sequence of a protective rhoptry antigen from *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 41, 125-134.

Roberts, D. J. (1993). Protection, pathogenesis and phenotypic plasticity in *Plasmodium falciparum* malaria. *Parasitology Today* 9, 281-286.

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

Rockett, K. A. (1994). Tumour necrosis factor and interleukin-1 synergy in the context of malaria pathology. *American Journal of Tropical Medicine and Hygiene* 50, 735-742.

Roger, N., Dubremetz, J. F., Delplace, P., Fortier, B., Tronchin, G., and Vernes, A. (1988). Characterisation of a 225 kilodalton rhoptry protein of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 27, 135-142.

Ruiz-i-Altaba, A., Ozaki, L. S., Gwadz, R. W., and Godson, G. N. (1987). Organization and expression of the *Plasmodium knowlesi* circumsporozoite antigen gene. *Molecular and Biochemical Parasitology* 23, 233-245.

Sam-Yellowe, T. Y., and Perkins, M. E. (1991). Interaction of the 140/130/110 kDa rhoptry protein complex of *Plasmodium falciparum* with the



erythrocyte membrane and liposomes. *Experimental Parasitology* 73, 161-171.

Sam-Yellowe, T. Y., Shio, H., and Perkins, M. E. (1988). Secretion of *Plasmodium falciparum* rhoptry protein into the plasma membrane of host erythrocytes. *Journal of Cell Biology* 106, 1507-1512.

Saul, A., and Battistutta, D. (1990). Analysis of the sequences flanking the translational start sites of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 42, 55-62.

Saul, A., Cooper, J., Hauquitz, D., Irving, D., Cheng, Q., Stowers, A., and Limpaiboon, T. (1992). The 42-kilodalton rhoptry-associated protein of *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 50, 139-150.

Sinha, K. A., Keen, J. A., Ogun, S. A., and Holder, A. A. (1996). Comparison of two members of a multigene family coding for high-molecular mass rhoptry proteins of *Plasmodium yoelii*. *Molecular and Biochemical Parasitology* 76, 329-332.

Smith, J. D., Chitnis, C. E., Craig, A. G., Roberts, D. J., Hudson-Taylor, D. E., Peterson, D. S., Pinches, R., Newbold, C. I., and Miller, L. H. (1995). Switches in expression of *Plasmodium falciparum* var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. *Cell* 82, 101-110.

Staudt, L. M., Clerc, R. G., Singh, H., Lebowitz, J. H., Sharp, P. A., and Baltimore, D. (1988). Cloning of a lymphoid-specific cDNA encoding a protein binding the regulatory octamer DNA motif. *Science* 241, 577-580.

Staudt, L. M., Singh, H., Sen, R., Wirth, T., Sharp, P. A., and Baltimore, D. (1986). A lymphoid-specific protein binding to the octamer motif of immunoglobulin genes. *Nature* 323, 640-643.

Stern, A., and Meyer, T. F. (1987). Common mechanism in controlling phase and antigenic variation in pathogenic neisseriae. *Molecular Microbiology* 1, 5-12.

Su, X., Heatwole, V. M., Wertheimer, S. P., Guinet, F., Herrfeldt, J. A., Peterson, D. S., Ravetch, J. V., and Wellems, T. E. (1995). The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. *Cell* 82, 89-100.

Thomas, B. J., and Rothstein, R. (1989). Elevated recombination rates in transcriptionally active DNA. *Cell* 56, 619-630.

Tjian, R. (1995). Molecular machines that control genes. *Scientific American*.

Topol, J., Ruden, D. M., and Parker, C. S. (1985). Sequences required for *in vitro* transcriptional activation of a *Drosophila* hsp70 gene. *Cell* 42, 1527-1537.

Triglia, T., Peterson, M. G., and Kemp, D. J. (1988). A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Research* 16, 8186.

Triglia, T., Stahl, H. D., Crewther, P. E., Scanlon, D., Brown, G. V., Anders, R. F., and Kemp, D. J. (1987). The complete sequence of the gene for the knob-associated histidine-rich protein from *P. falciparum*. *EMBO J* 6, 1413-1419.

van-der-Wel, A. M., Tomas, A. M., Kocken, C. H., Malhotra, P., Janse, C. J., Waters, A. P., and Thomas, A. W. (1997). Transfection of the primate malaria parasite *Plasmodium knowlesi* using entirely heterologous constructs. *Journal of Experimental Medicine* 185, 1499-1503.

vonHeijne, G. (1985). Signal sequences: the limits of variation. *The Journal of Molecular Biology* 184, 99-105.

Ward, G. E., Fujioka, H., Aikawa, M., and Miller, L. H. (1994). Staurosporine inhibits invasion of erythrocytes by malarial merozoites. *Experimental Parasitology* 79, 480-487.

Waters, A. P., Syin, C., and 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.

Weber, J. L., Lyon, J. A., Wolff, R. H., Hall, T., Lowell, G. H., and Chulay, J. D. (1988). Primary structure of a *Plasmodium falciparum* malaria antigen located at the merozoite surface and within the parasitophorous vacuole. *The Journal of Biological Chemistry* 263, 11421-11425.

Weiher, H., Konig, M., and Gruss, P. (1983). Multiple point mutations affecting the simian virus 40 enhancer. *Science* 219, 626-631.

Werner, E., Patel, K., and Holder, A. A. (1997). Construction of a library for sequencing long regions of malarial genomic DNA. *Biotechniques* 23, 20.

Wertheimer, S. P., and Barnwell, J. W. (1989). *Plasmodium vivax* interaction with the human Duffy blood group glycoprotein: identification of a parasite receptor-like protein. *Experimental Parasitology* 69, 340-350.

Wesseling, J. G., deRee, J. M., Ponnudurai, T., Smits, M. A., and Schoenmakers, J. G. (1988). Nucleotide sequence and deduced amino acid sequence of a *Plasmodium falciparum* actin gene. *Molecular and Biochemical Parasitology* 27, 313-320.

Wesseling, J. G., Sjniders, P. J., vanSomeren, P., Jansen, J., Smits, M. A., and Schoenmakers, J. G. (1989). Stage-specific expression and genomic organization of the actin genes of the malaria parasite *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 35, 167-176.

Wesseling, J. G., Smits, M. A., and Schoenmakers, J. G. (1988). Extremely diverged actin proteins in *Plasmodium falciparum*. *Molecular and Biochemical Parasitology* 30, 143-153.

Wilson, R. J. M., Denny, P. W., Preiser, P. R., Rangachari, K., Roberts, K., Roy, A., Whyte, A., Strath, M., Moore, D. J., Moore, P. W., and Williamson, D. H. (1996). Complete gene map of the plastid-like DNA of the malaria parasite *Plasmodium falciparum*. *Journal of Molecular Biology* 261, 155-172.

Wu, Y., Kirkman, L. A., and Wellems, T. E. (1996). Transformation of *Plasmodium falciparum* malaria parasites by homologous integration of plasmids that confer resistance to pyrimethamine. *PNAS USA* 93, 1130-1134.

Wu, Y., Sifri, C. D., Lei, H. H., Su, X. Z., and Wellems, T. E. (1995). Transfection of *Plasmodium falciparum* within human red blood cells. *Proc Natl Acad Sci USA* 92, 973-977.

Yoeli, M., Hargreaves, B., Carter, R., and Walliker, D. (1975). Sudden increase in virulence in a strain of *Plasmodium berghei yoelii*. *Ann. Trop. Med. Parasitol.* 69, 173-178.

Mitchell GH and Bannister LH (1988). Malaria parasite invasion: interactions with the red cell membrane. *Crit Rev Oncol Hematol* 8: 225-310.

van Dijk M., R. Vinkenoog, J. Ramesar, R. Vervenne, A. Waters and C. Janse (1997). Replication, expression and segregation of plasmid-borne DNA in genetically transformed malaria parasites. *Molecular and Biochemical Parasitology* 86: 155-162.

van Dijk M., A. Waters and C. Janse (1995). Stable transfection of malaria parasite blood stages. *Science* 268: 1358-1362.

**APPENDIX ONE**  
SEQUENCE OF Py235.E8

## Appendix One

During the course of these investigations, the sequencing of the entire coding region gene Py235.E8 has been completed. The final sequence, and its deduced amino acid sequence is as follows, with the putative signal sequence underlined and the splice acceptor and donor dinucleotides highlighted in red:

ATGAAAAAATATATTATATTATTAGTTTGGCTGCATTTTTTATTTCTTTCGTAATCAT 60  
M K K Y I Y I I S L A A F F I S F . . .

AGATATCTTTAACAAATTATTTCAATTCTTGGGTTTACATATTTTATTTAATATGTTTAT 120  
. . . . .

GTGTATTAATTATGCAAATGTGTAAATTCCTATACTAATATTACTTAATTTGTTTCATCA 180  
. . . . .

TTAATATGCAAGATATAATTTGTGCAACCAAAATTGAGGAAAAATAAAAAACAAGAATGCCAC 240  
. . . D I I C A T K I E E N K N K N A T

AAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAA 300  
K L N N Y N P Y H N L E E S D F D N S N

CTTTTTAAATGAAAAAAATATGATAATAAAAATAATTCCATTAATAATGAAAAATATAT 360  
F L N E K K Y D N K N N S I N N E K Y I

ACAGCCACATTCAATAAATAACAATTTCTTCAAAAATAAAAAAGATATAAACAATACAAA 420  
Q P H S I N N N F L Q N K K D I N N T K

AATCACATTGTATAATGAATATAATAAGACAAATAATATTGATAATTTTAGAAATTTTAA 480  
I T L Y N E Y N K T N N I D N F R N F N

TCATGAACATGAAAAACAAATAACATATTAATTGAAAATAATAATTCCTTTTTACAAAT 540  
H E H E K T N N I L I E N N N S F L Q I

AACCGCTATATCTGAAATAAATAAAGATACGTTAGATACGGTCGATACTATATATGGTGT 600  
T A I S E I N K D T L D T V D T I Y G V

AAATGATAATGAAAGCGTATTGTGGTTTTTTTATAAACTTCTATATTCTGTTGAATATAT 660  
N D N E S V L W F F Y K L L Y S V E Y I

TAAAAAAATGTTAAGTAACTTAAATATAGGTCTATCCAGTGAGCAAACCTAAAATAAAGAC 720  
K K M L S N L N I G L S S E Q T K I K T

AGATCAAAATGTTGTTATTAATGACGTCAGACAAACAAACGCGTCATGTTATTCTGAAAA 780  
D Q N V V I N D V R Q T N A S C Y S E K



AAATGAATTAATAAATAAATTGGGCAATATACACAATCCTTTTTATAATTTTCATAATGA 840  
N E L I N K L G N I H N P F Y N F H N D

TCGTTGATAGATGATTATAATTTATATACATCATCAAAAAATAATTTTGTAATTGTTT 900  
R S I D D Y N L Y T S S K N N F V N C L

AAAAAAGGATTTGAAAATATAAAAAAAAAAACAATGATATTATTTTCATATGAAGAAAA 960  
K K G F E N I K K K T N D I I S Y E E K

ATTATATACAGATAAATGTAAAAATAGCTCCAACTCAAGAATACTACTAAGAAACCTTC 1020  
L Y T D K C K N S S K L K N T T K K P S

TGAATTTTGTA AAAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTTGGGAAGTTCC 1080  
E F C K N V M N L N V S R K N S W E V P

CAAAAATAACGAAATAATTTCAATTTATAGATTTTCTAATAGAAAAATTAAATCAAACAA 1140  
K N N E I I S F I D F L I E K L K S N N

TTACCCAATGACTCTTGTAACAAACTAGATTTTATAAAAAACAATTCGAGGATATTAA 1200  
Y P M T L V T K L D F I K K Q F E D I K

AAATAAACACAATAAACACATAAAAAATATGTAAACAAGAAGAGATTGTTGTGAACAAATG 1260  
N K H N K H I K I C K Q E E I V V N K C

TACCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAAAAAATAGC 1320  
T N N I D N N N C D K H F N E I K K I A

CGAATCATATAGCATTATATTTTATAATTATACAATTTTAGAAAAATTAGAAATCAGTAAA 1380  
E S Y S I I F Y N Y T I L E N L E S V N

TATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAGGGAAATTATTAATAAAA 1440  
I T Y K E S L D Y F F N S L G K L L I N

CAAAGTAGATTCCGATGGAAATATAATAGAAGAAAATGATATTGATAATTTTGATTGTC 1500  
K V D S D G N I I E E N D I D N F D L S

TAAACCCAAAAATAATTTTAAATTATTAGAAGGTGAATTAATGGAGTATTTGAAAATAA 1560  
K P K N N F K L L E G E L N G V F E N K

ATGGAATGATTATAAAAAATAATAAAATCTTGATCAATTCAAAGATACAATGAAAAAAT 1620  
W N D Y K N N K N L D Q F K D T M K K I

AATATTATTAATTATACAATATATGAACGAATTTAAAGGCTTGAATGACGCTATGACAAA 1680  
I L L I I Q Y M N E F K G L N D A M T K

ACTAAAAATGAAGGTATTTACAAAAATTTGTGATTAATAATCAAATAAAACAAAAATT 1740  
L K N E G I S Q K F V I N N Q I K Q K F

CGATAATCAACTTATGATGAAAAAAGAAGGGTTTGAAAGTTCTCTTGAATTAGCAAA 1800  
D K S T Y D E K K E G F E S S L E L A K

AAATTGGGAAAAAAAAAACTCGAAATAATAACGGAATTAATAAAAAAAAAAATGAAGAAAC 1860  
N W E K K K L E I I T E L K K K N E E T

TGTTCAATTGGATATAAAAAATTAGAGAACTAATTAACAAATCAAAGATATTATAGAGGA 1920  
V Q L D I K I R E L I K Q I K D I I E E

ACAAAAATAGTGAACGATTTAAAATTAGAATTAACAAAAAATAAAAGAAATAACTGA 1980  
Q K I V N D L K L E L N K K I K E I T E

GAAATTTGAATATATTAATAAGCAGTTGACTTAAAGAAAGAAATAGAAAAAGACAACGT 2040  
K I E Y I K K A V D L K K E I E K D N V

ATATATTGATGAATTAGCTAAGGAACCATATCAAATTACAAAATATATAGAAAAAAA 2100  
Y I D E L A K E P P Y Q I T K Y I E K K

AAATGAAATATATAATACAATAAAATCAGATTTTGACAAAATTTATGTAGGCGATATTGA 2160  
N E I Y N T I K S D F D K I Y V G D I E

ACAACTTTACAATGAAATGTTTTCTGTAGTTCAAGAAAGTAACATTGAGCATATAGAAAA 2220  
Q L Y N E M F S V V Q E S N I E H I E N

TAAACAGAAATTTCTAACTTTAAAAACGAAAATAGATAATGTATATAATAATATCCAAAA 2280  
K T E I L T L K T K I D N V Y N N I Q N

CATGGAACTGAAACAGTTAAATCACATCTAAAGAATATAGAACTAACAACAACTATC 2340  
M E T E T V K S H L K N I E T N N K L S

AGAAACAATTTTGGATATAATAAAATATATATATGGAGAGATTACCAACGAGCTAAATAA 2400  
E T I L D I I K Y I Y G E I T N E L N K

AACATTAGAAGACTTTAAAAATAAAGAAAAAGGACTATCAAATAAAATAGATGAATACGC 2460  
T L E D F K N K E K G L S N K I D E Y A

TAAGGAAATGTCCAATTAAATGTATATAAATCTAACATATTAGAAATCAAAAAGCATTAA 2520  
K E N V Q L N V Y K S N I L E I K K H Y

TAATGATCAAATTAATATAGACAATATAAAGGAAAAAGAAGCAAAACAAAACCTATGATCA 2580  
N D Q I N I D N I K E K E A K Q N Y D Q

ATTCAAAGAACATATGAAAACAATACCACCTAACGAAATGAAATATCAAAAACCATCAAT 2640  
F K E H M K T I P P N E M K Y Q K P S I

CGAGATAAAAATTATGAAAGACGAATTCTTGAGTAAAGTAAATAAATATAATGATTTTGA 2700  
E I K I M K D E F L S K V N K Y N D F D

CAAAGTTTATAAAGAAAAGTTGAATCAGAACACAATAAATTTACTGAATTAACAAATAA 2760  
K V Y K E K V E S E H N K F T E L T N K

AATAAAAACAGAAGTTTCGGATGAAGAAATAAAAAAATATGAAAACAAATTTAATGATAG 2820  
I K T E V S D E E I K K Y E N K F N D S

CAAATCTTTAATTAATGAAACAAAAAAATCCATTGAAGAGGAATATCAAAACATTAATAC 2880  
K S L I N E T K K S I E E E Y Q N I N T

ACTTAAAAAGGTAGATGATTATATAAAAGTATGTCTAAATACAAATGAATTAATAACAAA 2940  
L K K V D D Y I K V C L N T N E L I T N

TTGTCATAATAAACAAACTACATTAAAAGACAAATTAATCAAAATATTAACCATAAAA 3000  
C H N K Q T T L K D K L N Q N I K T I K

AGAAACTAATTCAATAGATAAAATTTATACAGACAAGTTTGAAAATATATTAACAGATAA 3060  
E T N S I D K I Y T D K F E N I L T D K

AAAGACAGAATTAGAAACAAAATTCACAGGATTGTCTTTAAATAATCACGAATCGAATAA 3120  
K T E L E T K F T G L S L N N H E S N N

CAAAGAATTGCTAACATATTTTTATGATTAAAGGCAAATTTAGGAAAAAATAAGAAAA 3180  
K E L L T Y F Y D L K A N L G K N K E N

CATGTTATATAACAATTCAATGAAAAAGAAAAAGCTGTTGAAGATATTAAGAAAAAAA 3240  
M L Y K Q F N E K E K A V E D I K K K N

TGTAGATATAAATAAAATTGTTTCAAATATTGAAATAACAATTTATACATCCATTTATAA 3300  
V D I N K I V S N I E I T I Y T S I Y N

TATTAATGAAGATACAGAAAATGAAATTGGAAAAAGTATAGAATTACTAAATACCAAAGT 3360  
I N E D T E N E I G K S I E L L N T K V

ACTTGAAAAGGTAAAAGCAAACGTAACCAATTTGAATGAAATAAAGGAAAAATTAAAAAGA 3420  
L E K V K A N V T N L N E I K E K L K D

TTATGATTTTCAGGATTTTGGAAAAGAGAAAAATATTAATATCCTGATGAAAACAAAAT 3480  
Y D F Q D F G K E K N I K Y P D E N K I

TAAAAATGATATTGATACATTAAACCAAAAAATCGATAAAAGTATAGAGACATTAACGGA 3540  
K N D I D T L N Q K I D K S I E T L T E

AATAAAAAAAATTCAGAGAACCATATTGATGAAATAAAAGGACAAATAGATAAAATTA 3600  
I K K N S E N H I D E I K G Q I D K L K

AAAAGTACCAAATAAAACAATGTTTAACGAGGATCCAAAGGAAATTGAAAAAAAATAGA 3660  
K V P N K T M F N E D P K E I E K K I E

AAATATAGTAGAAAAAATAGACAAAAAATAATATATATAAAGAAATAGATAAATTATT 3720  
N I V E K I D K K K N I Y K E I D K L L

AAATGAAATATCAAAAATAGAAAATGATAAACTTCGTTAGAAAAATTAAAAATATAAA 3780  
N E I S K I E N D K T S L E K L K N I N

TTTATCATATGGAAAAAGTTTAGGCAACTTATTTTTGCAACAAATTGATGAAGAAAAAAA 3840  
L S Y G K S L G N L F L Q Q I D E E K K

AAAGGCTGAACATACGATAAAAGCAATGGAAGCATATATTGATGATCTTGATAATATAAA 3900  
K A E H T I K A M E A Y I D D L D N I K

AAAAAATCACAAGAAATAGAAAAAGAAATGAACATAAACATGGACATAAAAAATGGACAT 3960  
K K S Q E I E K E M N I N M D I K M D I

ACATAAGGAAATGAAAGCGCTTAATATATCACATGACGACTACAAAATATATCATACTAC 4020  
H K E M K A L N I S H D D Y K I Y H T T

AAGTAAGAATCATGAAGAAAAAATTTCTGACATTCGTAAAAATTCTTTAAAAATAATACA 4080  
S K N H E E K I S D I R K N S L K I I Q

AGATTTTTCTGAGGAATCATATATAAATGATATTAAGGAATTAGAGAAAAATGTTTT 4140  
D F S E E S Y I N D I K K E L E K N V L

AGAATCCCAAATAATAATACTGATATTAATCAATATTTAAGCAAAATTGAAAACATATA 4200  
E S Q N N N T D I N Q Y L S K I E N I Y

TAATATCTTAAATTAATAAAATTAAGATTATTGATAAAGTAAAGAATATACTGA 4260  
N I L K L N K I K K I I D K V K E Y T D

TGAAATCGAAAAGAATAATAAGATAAATGCTGAATTAAGTAATTCAGAAAAATAAT 4320  
E I E K N N K K I N A E L S N S E K I I

TACACAACATAAAAGAAAATTCAAGTTTAAAGAATGCCAATCAAAAATAAAATCAACTAT 4380  
T Q L K E N S S L K E C Q S K I K S T I

AGATGATAATTATGTTAGTGAATGTATAAAGAATATTACAAATTTAAAACTTATATTGT 4440  
D D N Y V S E C I K N I T N L K T Y I V

AAACGAAAAAATAACATCAACACTTATTTTAAAAATGCCGAAGAGTATAATCAAAATGT 4500  
N E K N N I N T Y F K N A E E Y N Q N V

ATCATTAACCTTTAATAATATAGAAATGGCGGATACTAAATCACAATATATATTAACAT 4560  
S L N F N N I E M A D T K S Q Y I L N I

TAAAAAAAATAATGGCACTAATAACACTGATTATAATATCAAAGAATTGAAAGAACACAA 4620  
K K N N G T N N T D Y N I K E L K E H K

GAAGAAGTCTAATGTTTATAAAGATGAGGCTGGTAAAAATACACAAGAAATCAAAAAAAA 4680  
K K S N V Y K D E A G K N T Q E I K K N

TAAGGAATTATTTGAAAAATATGAACAAGAAGTAACTGTACTTTTAAATAAATATTATGC 4740  
K E L F E K Y E Q E V T V L L N K Y Y A

GGTAGAATTAATAAATAAATTTGATAAAACAAAAATTATTTCAGAACAAATCATAAAGGA 4800  
V E L K N K F D K T K N Y S E Q I I K E

AATAAAAGACGCACACAACACTTTTACATCCCAAGCAGACAAATCTGAAAAAAAAATGAA 4860  
I K D A H N T F T S Q A D K S E K K M N

TGAAATAAAAAACGAACAGATTCGTATTGAAGACGAAGTCGCTAAAAATAATAAATCTAA 4920  
E I K N E Q I R I E D E V A K N N K S N

TAAAGCAATACTAGATATTCAACTATCCGTAGAGCCATTCAAAATAAAATTCTTAAAAAT 4980  
K A I L D I Q L S V E P F K I K F L K I

AAAGGATCTAAGAACAAATCAGATGATTGTTTAAAAGAGACCAAAGACATAGAGACCAA 5040  
K D L R T K S D D C L K E T K D I E T K

AATATCAAATTTATCTATAGATACTCAAGAAACAAAATAATAGAGAACAAGAACATATT 5100  
I S N L S I D T Q E T K L I E N K N I L

AAATACCCTTGAGAACTTTTTAGAATCTCTCAAAAACCAAAAAAAAAAATATTGAAGACCA 5160  
N T L E K L L E S L K N Q K K N I E D Q

AAAAAAGAATTAGATGAAGTTAATTCCAAAATTAAAAATATAGAAAGCAATGTAAACCA 5220  
K K E L D E V N S K I K N I E S N V N Q

GCATAAAAAAATTACGAAATTGGAATTGTAGAAAAAATAAATGAAATCGCCAAAGCAAA 5280  
H K K N Y E I G I V E K I N E I A K A N

TAAAGACCAAATTGAATCAACACAAAAATTAATAATACCAACAATAAAAAATTTAATATC 5340  
K D Q I E S T Q K L I I P T I K N L I S

TCCTTTTAAAGCTAATGATTTAGAAGGTATTGACACTAATAAAAACTTGGGAAAATATAA 5400  
P F K A N D L E G I D T N K N L G K Y N

TACAGAAATGAATAATATATATGAAGAATTTATTAAATCATACGATCTAATAACACATTA 5460  
T E M N N I Y E E F I K S Y D L I T H Y



TTTAGAAACGGTTTCAAAGAACCATAACATATGAACAAATTAATAAGCGAATCAC 5520  
L E T V S K E P I T Y E Q I K N K R I T

CGCACAAATGAACTCTTAACAAATATAAAAAATGTAAATAAGCCAAATCCTATTTAGA 5580  
A Q N E L L T N I K N V N K A K S Y L D

TGATATAGAAGCAAATGAATTTGATAGAATAGTCACACATTTTAAAAACAAATTAATGA 5640  
D I E A N E F D R I V T H F K N K L N D

TGTGAATGATAAATTTACAAACGAATATTCAAAGTTAAACAAAGGGTTTGATAATATTC 5700  
V N D K F T N E Y S K V N K G F D N I S

AAACTCTATTAATAATGTAAAAAATCAACTGATGAAAATTTATTATTAAATATACTAAA 5760  
N S I N N V K K S T D E N L L L N I L N

CCAAACAAAAGAAATGTATGCAAATATTGTCAGTAAAAAATATTATAGTTATAAATATGA 5820  
Q T K E M Y A N I V S K K Y Y S Y K Y E

GGCAGAAAACATATTTATAAATATTCCGAAATTAGCAAATTCCTTTAAATATTCAAATAAA 5880  
A E N I F I N I P K L A N S L N I Q I K

AAGCAGTTCAGGAATAGATTTATTTAAGAACATTAATATAGCTATATTACCTTACTTGGA 5940  
S S S G I D L F K N I N I A I L P Y L D

TTCCCAAAAAAAGATACGCTAACCTTTATTCCATCTCCAGAAAAAACATCAGAAACATA 6000  
S Q K K D T L T F I P S P E K T S E T Y

TACAAAAATAAGCGATTCTTACAATACTCTTCTTGATATATTAAGAAGTCAAGAATT 6060  
T K I S D S Y N T L L D I L K R S Q E L

GCAGAAAAAAGAACAACAAGCATTAAATCTTATATTGAAAACCGACTTTTACATGACAA 6120  
Q K K E Q Q A L N L I F E N R L L H D K

AGTCCAAGCAACCAACGAATTAAGACACATTAAGTGATTAAAAAATAAAAAAGAACA 6180  
V Q A T N E L K D T L S D L K N K K E Q

AATATTAATAAAGTTAACTACTTTTACATAAATCTAATGAATTAACAAATTATCATG 6240  
I L N K V K L L L H K S N E L N K L S C

CAATTCTCAAATTATGATACCATTTTAGAATCATCAAAGTATGATAAAATAAAAGAAAA 6300  
N S Q N Y D T I L E S S K Y D K I K E K

AAGCAATAATTATGAAAAGGAAAAAGAAAACCTTGGGATAAATTTTGATGTAAAAGCTAT 6360  
S N N Y E K E K E N L G I N F D V K A M

GGAAGAACAATTTAATAATGATATTAAAGATATAGAAAAATTAGAAAATAATTACAAACA 6420  
E E Q F N N D I K D I E K L E N N Y K H

TTCAGAGAAAGATAATTACAATTTTTCAGAGGAAAATAATAATATTTTACAATCTAAAAA 6480  
S E K D N Y N F S E E N N N I L Q S K K

AAAACTAAAAGAACTAACTAACGCATTTAATGCTGAAATAAAAAAAATTGAGGATAAAAT 6540  
K L K E L T N A F N A E I K K I E D K I

AATAGAAAAAATGGTTTAATTAATAAATTAATAGAAACGAGAAAAGATTGCATGCTTTT 6600  
I E K N G L I N K L I E T R K D C M L F

TACATATAAAACATTAGTCGAGACTCTTAAATAAAAAACAACCTGATTACACGAAATTCAT 6660  
T Y K T L V E T L K I K T T D Y T K F I

AACGTCTGCAACTAAATTTTCAAAGAATTTTAAAAATACATTGATGCTACTTCCAATTC 6720  
T S A T K F S K E F L K Y I D A T S N S

TTTAAATGATGACATCAACACGTTGCAAACAAAATATGATTTAAATCAAATAAACAAGCA 6780  
L N D D I N T L Q T K Y D L N Q I N K H

TGTAGCAAGTATGGTTGCAGATGCAACTAATGATAATAATAATTTAATAGAAAAGGAAAA 6840  
V A S M V A D A T N D N N N L I E K E K

GGAAGCAACTAAGACAATCAACAATTTGACCGAGCTATTTACAATAGATTCAAATAAGAT 6900  
E A T K T I N N L T E L F T I D S N K I

CGATGCCGATGGATTACATAATAATAAAATACAAATTATTTATTTCAATTCTGAACTTCA 6960  
D A D G L H N N K I Q I I Y F N S E L H

TAAATCAATCGACTCCATAAAACAACCTTTATAAAAAAATGCATGCCTTTAAATTATTA 7020  
K S I D S I K Q L Y K K M H A F K L L N



TATAGGTCACATTAATAAAAAATATTTTGATATATCCAAAGAATTTGATAATATTTTACA 7080  
I G H I N K K Y F D I S K E F D N I L Q

GCTGCAGGAAAGTGAATTAACAGCAAATTTAAATGATTTAAAGGAAATTGGTCAAAAGAT 7140  
L Q E S E L T A N L N D L K E I G Q K I

TTCTGATAAAAAAAGCAATTCCTTCATGCACTAAGTGAAACTCCAATTCCTCAACTTCAA 7200  
S D K K K Q F L H A L S E T P I P N F N

TACACTTAAAGAGATATATCATGATATTGTTAAGTATAAAAAATCAGATAGATGAAATAGA 7260  
T L K E I Y H D I V K Y K N Q I D E I E

AAATATTACTAACGAAGAAAACGAAAATATAACTTTATATATGGATATAATTACTAAATT 7320  
N I T N E E N E N I T L Y M D I I T K L

AATGAAAAAAGTACAAAGTATTTTAAATTTTGTACAACTTATGAAAATGATAGTAATAT 7380  
M K K V Q S I L N F V T T Y E N D S N I

AATCAAGCAACATATTCAAGACAATAATGAAAATGATGTATCAAAAATTAAGAATCTTT 7440  
I K Q H I Q D N N E N D V S K I K E S L

AGAAACCACAATTCAATCATTTTCAGAAAATTCTAAATAAACTAAATGAAATCAAAGCTCA 7500  
E T T I Q S F Q K I L N K L N E I K A Q

ATTTTATGATAATAACAATATAAATAATGTTATATCTACCATATCACAAGATGTAAATGA 7560  
F Y D N N N I N N V I S T I S Q D V N D

TGTTAAAAAACATATTTCTAAAGATTTAACTATAGAAAACGAACTTATCCAAATACAAAA 7620  
V K K H I S K D L T I E N E L I Q I Q K

GAGTTTAGAAGATATTAAAAAATCTACTTATGATATCAGAAGCGAACAAATAACTAAATA 7680  
S L E D I K K S T Y D I R S E Q I T K Y

TGTCAATCCTATACACGATTATGTTGAGCAACAACTAAAAAAATTCAAAATAATCCAAA 7740  
V N P I H D Y V E Q Q T K K I Q N N P N

TAAAGACGAAATAGACGATCTAATACAAGAAATCGTCAATTATAATAAAGAATCAGAACT 7800  
K D E I D D L I Q E I V N Y N K E S E L

AAAATTACCCACCATTATAAATAATAAAGATAACGTTACACCAATAATCTCTCGTATTGA 7860  
K L P T I I N N K D N V T P I I S R I D

CAAAGTCATTAATTTAATAAAATCAGAATATAATAACAATGATAATGTATCATATAATGT 7920  
K V I N L I K S E Y N N N D N V S Y N V

TGCCAAAAAAGCTTGAAGAAGATGCCAATAACATAATTCGTGATTTAGATACGAGTCATAA 7980  
A K K L E E D A N N I I R D L D T S H N

CATGCTTAATGATTTAATACAGAAAAATTTTAAAATTATAGACGATTTAAAAAATAAAAA 8040  
M L N D L I Q K N F K I I D D L K N K K

ACAGGAGATAGAAAATCGTAATAATTTACAGACTATTAATAGGGAACAAGAAATAACGCA 8100  
Q E I E N R N N L Q T I N R E Q E I T Q

AACAGAACATGTTAACAATACATATCATCATGATTAATGACATTAATGATGTTATTGA 8160  
T E H V N N T Y H H D I N D I N D V I D

TATTAATGATATTAATGATACAAACGATATAAATCAAAATCACCAAACTCAAGCTCAGA 8220  
I N D I N D T N D I N Q N H Q N S S S D

TAAAAAGGACTCTTCCAAAACAAGAAATACAGGGAATACGATTAGATATGCCGGAGCAAT 8280  
K K D S S K T R N T G N T I R Y A G A I

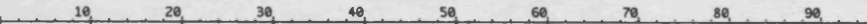
TGCATTTGGTTTAGTAACATTTTATGTAATTATAAGAATAAAAGAAAAAAGGATAAAGA 8340  
A F G L V T F Y V I I R I K E K K D K D


TGAAATGGAATTCGATAAATCTAAAGGTTTTTATGATGGCGGTGAAAGTACGCTTTTTGA 8400  
E M E F D K S K G F Y D G G E S T L F E

AAGAGAAGATGAAGTTATAGAAATAGATATGAACGAGGATTTATCATTTAATTA~~AAAAAAC~~ 8460  
R E D E V I E I D M N E D L S F N . K T

ATTTGTTTATAAAAGTTAATATTATAGGGAATTTACAAC~~TACAAATAAC~~AAAAAGTATTT 8520  
F V Y K S . Y Y R E F T T T N N K K Y F

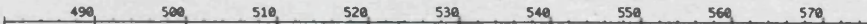
TATATGGATATTTACCATATTAATGCAACAATACTATGTT 8560  
I W I F T I L M Q Q Y Y V

8.1.6T7II.Assembla.  GATCAAGTAAAGTTTATTCGCCAATTACACACCCATGAAATATGTTTATTTATTACTTATTATTACTACTATTTTATTATTGTATACACAG  
7.M13 -> GATCAAGTAAAGTTTATTCGCCAATTACACACCCATGAAATATGTTTATTTATTACTTATTATTACTACTATTTTATTATTGTATACACAG

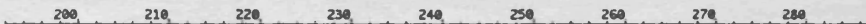
8.1.6T7II.Assembla.  CGGTCTAATTAAATGAATACAATAGAACTTTAAACCAATCTATAACATCCATTATATTTTTTTAAAAATAATTATTTATAATATATATATTC  
7.M13 -> CGGTCTAATTAAATGAATACAATAGAACTTTAAACCAATCTATAACATCCATTATATTTTTTTAAAAATAATTATTTATAATATATATATTC


8.1.6T7II.Assembla.  TAGTAGGATAAAATATCCAATTTTATATCCCATTTTTATTCATTATTACAACATCTTATATTATACCCATTTAAAAATATATTAATTTTATA


7.M13 -> TagTAGGATAAAATATCCAATTTTATATCCCATTTTTATTCATTATTACAACATCTTATATTATACCCATTTAAAAATATATTAATTTTATA  
7.m13.1 -> TAGTAGGATAAAATATCCAATTTTATATCCCATTTTTATTCATTATTACAACATCTTATATTATACCCATTTAAAAATATATTAATTTTATA  
7.m3 -> TAGTAGGATAAAATATCCAATTTTATATCCCATTTTTATTCATTATTACAACATCTTATATTATACCCATTTAAAAATATATTAATTTTATA  
7.m2 -> TAGTAGGATAAAATATCCAATTTTATATCCCATTTTTATTCATTATTACAACATCTTATATTATACCCATTTAAAAATATATTAATTTTATA  
7.4.1 -> TTTAAAAATATATTAATTTTATA

8.1.6T7II.Assembla.  CAAATTTACACTTCCCAACTATATATGTGTAGCTATTAAAAATAAAAAATAATTTTAAATAATATAGTGTTAAACCTTTATGAAAAATAAATTAAC

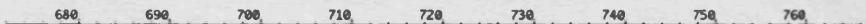
7.m13.1 -> CAAATTTACACTTCCCAACTATATATGTGTAGCTATTAAAAATAAAAAATAATTTTAAATAATATAGTGTTAAACCTTTATGAAAAATAAATTAAC  
7.m3 -> CAAATTTACACTTCCCAACTATATATGTGTAGCTATTAAAAATAAAAAATAATTTTAAATAATATAGTGTTAAACCTTTATGAAAAATAAATTAAC  
7.m2 -> CAAATTTACACTTCCCAACTATATATGTGTAGCTATTAAAAATAAAAAATAATTTTAAATAATATAGTGTTAAACCTTTATGAAAAATAAATTAAC  
7.4.1 -> MAAATTTAACTNCCCAACTATATATGTGTAGCTATTAAAAATAAAAAATAATTTTAAATAATATAGTGTTAAACCTTTATGAAAAATAAATTAAC  
7.5'ii -> AAATAATTTTAAATAATATAGTGTTAAACCTTTATGAAAAATAAATTAAC

8.1.6T7II.Assembla.  TAATAAAAACGATATATTTAAACATATATTTACGAATTTTATATCTATAATACCATAAAATATAAATTAACAAATCATTTCGTATACAAATT  
7.M13 -> TAATAAAAACGATATATTTAAACATATATTTACGAATTTTATATCTATAATACCATAAAATATAAATTAACAAATCATTTCGTATACAAATT  
7.m13.1 -> TCATTTCGTATACAAATT  
7.m3 -> AATT

8.1.6T7II.Assembla.  TTCTATAAAATCTACATTCATATATAAATATATTAATTTGCCTGAATCAAATAATATTCAGTATTTAATCATCAATATAACATTAAAAATGTATA  
7.M13 -> TtCTATAAAATCTACATTCATATATAAATATATTAATTTGCCTGAATCAAATAATATTCAGTATTTAATCATCAATATAACATTAAAAATGTATA  
7.m13.1 -> TTCTATAAAATCTACATTCATATATAAATATATTAATTTGCCTGAATCAAATAATATTCAGTATTTAATCATCAATATAACATTAAAAATGTATA  
7.m3 -> TTCTATAAAATCTACATTCATATATAAATATATTAATTTGCCTGAATCAAATAATATTCAGTATTTAATCATCAATATAACATTAAAAATGTATA  
7.m2 -> TCATCAATATAACATTAAAAATGTATA

8.1.6T7II.Assembla.  CATATAAACCATTTCTGCTTTTATTATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTTTAAATAATCCAAATTTCCATT

7.m13.1 -> CATATAAACCATTTCTGCTTTTATTATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTTTAAATAATCCAAATTTCCATT  
7.m3 -> CATATAAACCATTTCTGCTTTTATTATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTTTAAATAATCCAAATTTCCATT  
7.m2 -> CATATAAACCATTTCTGCTTTTATTATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTTTAAATAATCCAAATTTCCATT  
7.4.1 -> CATATAAACCATTTCTGCTTTTATTATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTTTAAATAATCCAAATTTCCATT  
7.5'ii -> CATATAAACCATTTCTGCTTTTATTATTCGAAAATTAACAAAAAATTTAAATAATTAATTCAAAAAATGTTTAAATAATCCAAATTTCCATT

8.1.6T7II.Assembla.  ATTTATTATTTCATATATTTATAATAAATTTTATTATTAAAGATATATAAATTAAGTATGGGCAAAAATATAAATAATACCAATTTGGGGATAATTTA

7.m13.1 -> ATTTATTATTTCATATATTTATAATAAATTTTATTATTAAAGATATATAAATTAAGTATGGGCAAAAATATAAATAATACCAATTTGGGGAT-ATTTA  
7.m3 -> ATTTATTATTTCATATATTTATAATAAATTTTATTATTAAAGATATATAAATTAAGTATGGGCAAAAATATAAATAATACCAATTTGGGGATAATTTA  
7.m2 -> ATTTATTATTTCATATATTTATAATAAATTTTATTATTAAAGATATATAAATTAAGTATGGGCAAAAATATAAATAATACCAATTTGGGGATAATTTA  
7.4.1 -> ATTTATTATTTCATATATTTATAATAAATTTTATTATTAAAGATATATAAATTAAGTATGGGCAAAAATATAAATAATACCAATTTGGGGATAATTTA  
7.5'ii -> ATTTATTATTTCATATATTTATAATAAATTTTATTATTAAAGATATATAAATTAAGTATGGGCAAAAATATAAATAATACCAATTTGGGGATAATTTA

770 780 790 800 810 820 830 840 850 860  
8.1.6T7II.Assembly. TAAAAGTTGTTATATATGATTAACCAAACCTTACCATTAAATATAATATACTAATATTATACATTAATAATGCTTAATCTATATAAACTCATCAT

7.m13.1 -> TAAAAGTTGTTATATATGAATAACC-AAAMTTACCATTNAAATATAATATACTAATATTATACCTTTATT-ATGCTT-ATCTATATAAACNC-TCAT  
7.m3 -> TAAAAGTTGTTATATATGATTAACCAAACCTTACCATTAAATATAATATACTAATATTATACATTAATAATGCTTAATCTATATAAACTCATCAT  
7.m2 -> TAAAAGTTGTTATATATGATTAACCAAACCTTACCATTAAATATAATATACTAATATTATACATTAATAATGCTTAATCTATATAAACTCATCAT  
7.4.1 -> TAAAAGTTGTTATATATGATTAACCAAACCTTACCATTAAATATAATATACTAATATTATACATTAATAATGCTTAATCTATATAAACTCATCAT  
7.5'ii -> TAAAAGTTGTTATATATGATTAACCAAACCTTACCATTAAATATAATATACTAATATTATACATTAATAATGCTTAATCTATATAAACTCATCAT

870 880 890 900 910 920 930 940 950 960  
8.1.6T7II.Assembly. ATTTTAAGACTAAATTTGATTATAAAATTTTATTACGATGTTACTTTTATGTTAATTCCTATATATCTCTCTGTGCCAAAATTTTGTCTTTA

7.m13.1 -> ATTTTANANTAAA-TTGATATAAAATTTTATTTA  
7.m3 -> ATTTTAAGACTAAATTTGATTATAAAATTTTATTACGATGTTACTTTTATGTTAATTCCTATATATCTCTCTGTGCCAAAATTTTGTCTTTA  
7.m2 -> ATTTTAAGACTAAATTTGATAT-ATTTTATTACGATGTTACTTTTAKGTAAATCTATATATCNCNCCTGTGCCAAAATTTTG  
7.4.1 -> ATTTTAAGACTAAATTTGATTATAAAATTTTATTACGATGTTACTTTTATGTTAATTCCTATATATCTCTCTGTGCCAAAATTTTGTCTTTA  
7.5'ii -> ATTTTAAGACTAAATTTGATTATAAAATTTTATTACGATGTTACTTTTATGTTAATTCCTATATATCTCTCTGTGCCAAAATTTTGTCTTTA

1160 1170 1180 1190 1200 1210 1220 1230 1240  
8.1.6T7II.Assembly. TCTAAAACCGAAACATGGAGAAATTTTAAATAAAATTTCCATTGATTTTCTGCATGCATAAAAAGTAAATAAAAAGGCTTTTATTTTGAATCACT

7.4 -> TCTAAAACCGAAACATGGAGAAATTTTAAATAAAATTTCCATTGATTTTCTGCATGCATAAAAAGTAAATAAAAAGGCTTTTATTTGAATCACT  
7.2 -> TCTAAAACCGAAACATGGAGAAATTTTAAATAAAATTTCCATGATGTTTCTGCATGCATAAAAAGTAAATAAAAAGGCTTTTATTTTGAATCACT  
7.4.1 -> TCTAAAACCGAAACATGGAGAAATTTTAAATAAAATTTCCATTGATTTTCTGCATGCATAAAAAGTAAATAAAAAGGCTTTTATTTTGAATCACT  
7.5'ii -> TCTAAAACCGAAACATGGAGAAATTTTAAATAAAATTTCCATTGATTTTCTGCATGCATAAAAAGTAAATAAAAAGGCTTTTATTTTGAATCACT

1250 1260 1270 1280 1290 1300 1310 1320 1330 1340  
8.1.6T7II.Assembly. ATTTCCCTTTATTTGGAAATTTTATAATATTTTAAATAATATGTTAAGTGTTTTGGTGTTTTCCAATTTTGTGCGGTACACATTTATCTTA

7.4 -> ATTTCCCTTTATTTGGAAATTTTATAATATTTTAAATAATATGTTAAGTGTTTTGGTGTTTTCCAATTTTGTGCGGTACACATTTATCTTA  
7.2 -> ATTTCCCTTTATTTGGAAATTTTATAATATTTTAAATAATATGTTAAGTGTTTTGGTGTTTTCCAATTTTGTGCGGTACACATTTATCTTA  
7.tII -> CCTTTATTTGGAAATTTTATAATATTTTAAATAATATGTTAAGTGTTTTGGTGTTTTCCAATTTTGTGCGGTACACATTTATCTTA

970 980 990 1000 1010 1020 1030 1040 1050  
8.1.6T7II.Assembly. TCTAGAGGTTAAAAATTCAGGTTGTAGAGATTTTTTATAATTTTACTTATATTTAAACAATTTATTATTATATATATGTTTAAATAATATT

7.m3 -> TCTAGAGGTTAAAAATTC

7.4.1 -> TCTAGAGGTTAAAAATTCAGGTTGTAGAGATTTTTTATAATTTTACTTATATTTAAACAATTTATTATTATATATATGTTTAAATAATATT  
7.5'ii -> TCTAGAGGTTAAAAATTCAGGTTGTAGAGATTTTTTATAATTTTACTTATATTTAAACAATTTATTATTATATATATGTTTAAATAATATT

1060 1070 1080 1090 1100 1110 1120 1130 1140 1150  
8.1.6T7II.Assembly. ATCAAAATTCACATATTTGATTAACCAAACCTTACACACGGAACCTGGGTACCCTTAATTTTGGTCTATATACATGACACTAATTTTAT

7.4 -> ATTTGATTAACCAAACCTTACACACGGAACCTGGGTACCCTTAATTTTGGTCTATATACATGACACTAATTTTAT  
7.2 -> ATTTGATTAACCAAACCTTACACACGGAACCTGGGTACCCTTAATTTTGGTCTATATACATGACACTAATTTTAT

7.4.1 -> ATCAAAATTCACATATTTGATTAACCAAACCTTACACACGGAACCTGGGTACCCTTAATTTTGGTCTATATACATGACACTAATTTTAT  
7.5'ii -> ATCAAAATTCACATATTTGATTAACCAAACCTTACACACGGAACCTGGGTACCCTTAATTTTGGTCTATATACATGACACTAATTTTAT

1350 1360 1370 1380 1390 1400 1410 1420 1430 1440  
8.1.6T7II.Assembly. AAAAAAGTTATAATGCAATTTATTAATTTTATTTATATAAATTAATAAATGAGAGCATAATTTATTAATTCCTTTTATCTACACACC

7.4 -> AAAAAAGTTATAATGCAATTTATTAATTTTATTTATATAAATTAATAAATGAGAGCATAATTTATTAATTCCTTTTATCTACACACC  
7.2 -> AAAAAAGTTATAATGCAATTTATTAATTTTATTTATATAAATTAATAAATGAGAGCATAATTTATTAATTCCTTTTATCTACACACC  
7.tII -> AAAAAAGTTATAATGCAATTTATTAATTTTATTTATATAAATTAATAAATGAGAGCATAATTTATTAATTCCTTTTATCTACACACC

1450 1460 1470 1480 1490 1500 1510 1520 1530  
8.1.6T7II.Assembly. AATAAATAATTTGTTAATTTTAAATTTAAATTTTATTTAATGGAATTTATTTGTTTATATACTTACACAAATTTTATTTATATATAAA

7.4 -> AATAAATAATTTGTTAATTTTAAATTTAAATTTTATTTAATGGAATTTATTTGTTTATATACTTACACAAATTTTATTTATATATAAA  
7.2 -> AATAAATAATTTGTTAATTTTAAATTTAAATTTTATTTAATGGAATTTATTTGTTTATATACTTACACAAATTTTATTTATATATAAA  
7.tII -> AATAAATAATTTGTTAATTTTAAATTTAAATTTTATTTAATGGAATTTATTTGTTTATATACTTACACAAATTTTATTTATATATAAA



1540 1550 1560 1570 1580 1590 1600 1610 1620 1630

8.1.6T7II.Assembla AATGAAGAATATTATCTTTAAATAAAATGATCGACAAGTAGAAAAATAAGCGATTAAATGTAGACAAAATTAATATTAGAAAATAAAATAAAATA

7.4 <- AATGAAGAATATTATCTTTAAATAAAATGATCGACAAGTAGAAAAATAAGCGATTAAATGTAGACAAAATTAATATTAGAAAAT

7.2 <- AATGAAGAATATTATCTTTAAATAAAATGATCGACAAGTAGAAAAATAAGCGATTAAATGTAGACAAAATTAATATTAGAAAATAAAATAAAATA

7.tII <- AATGAAGAATATTATCTTTAAATAAAATGATCGACAAGTAGAAAAATAAGCGATTAAATGTAGACAAAATTAATATTAGAAAATAAAATAAAATA

8.2.3.5,R2.RC <- GATCGACAAGTAGAAAAATAAGNGTTTAAKTAGACAAAATAATAToTtoGGAATAAAATAAATTA

8.1.6,R2.RC <- GATCGACAAGTAGAAAAATAAGCGATTAAATGTAGACAAAATAATAToTtoGGAATAAAATAAATTA

8.1.6T3 -> GATCGACAAGTAGARAoYTAAGCGATTAAATGTAGACAAAATAATATTAAAGAAAATAAAATA

8.2.3.5,R1.RC <- TCGAMAAGTAGAAAAATAAGCGATTAAATGTAGACAAAATAATAToTtoGGAATAAAATAAATTA

8.1.6,R1.RC <- AAGTAGAAAAATAAGCGATTAAATGTAGAAAAATAATAToTtoGGAATAAAATAAATTA

1640 1650 1660 1670 1680 1690 1700 1710 1720

8.1.6T7II.Assembla AAACAAAAACATCAAAAATATAGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAATATAT

7.2 <- AAACAAAAACATCAAAAATATAGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATAT

7.tII <- AAACAAAAACATCAAAAATATAGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAATATAT

8.2.3.5,R2.RC <- AAACAAAAACATCAAAAATGKaGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAATATAT

8.1.6,R2.RC <- AAACAAAAACATCAAAAATATAGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAATATAT

8.1.6T3 -> AAACAAAAACATCAAAAATATAGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAATATAT

8.2.3.5,R1.RC <- AAACAAAAACATCAAAAATTAToGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTNMACCoGAAAAATATAT

8.1.6,R1.RC <- AAACAAAAACATCAAAAATAToGAAGAAAAACATATATATAATATATTTTTTAACTCAAAAATATATATTATATTTAACCATGAAAAATATAT

1930 1940 1950 1960 1970 1980 1990 2000 2010

8.1.6T7II.Assembla CCAAAATTGAGGAAAAATAAAACAAGAAATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAACT

7.T3 <- CCAAAATGGAGG-AAATAAAACCGGATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGGAGGAAGTGATTTTGATAATTCAAACT

7.tII <- CCAAAATTGAGGAAAAATAAAACAAGAAATGCC

8.2.3.5,R2.RC <- CCAAAATTGAGGAAAAATAAAACAAGAAATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAAYT

8.1.6,R2.RC <- CCAAAATTGAGGAAAAATAAAACAAGAAATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAACT

8.1.6T3 -> CCAAAATTGAGGAAAAATAAAACAAGAAATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAACT

8.2.3.5,R1.RC <- CCAAAATTGAGGAAAAATAAAACAAGAAATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAACT

8.1.6,R1.RC <- CCAAAATTGAGGAAAAATAAAACAAGAAATGCCACAAGTTAAACAATTATAATCCATACCATAATTTAGAAGAAAGTGATTTTGATAATTCAAACT

2020 2030 2040 2050 2060 2070 2080 2090 2100 2110

8.1.6T7II.Assembla TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCCCTTCAAAATA

7.T3 <- TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCCCTTCAAAATA

8.1.6E8.NEWRc+ <- GAAAAAAATATGoTaAtaGAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCCCTTCAAAATA

8.2.3.5,R2.RC <- TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAAKGAAAAATATATACAGCCACATTCAATAAATAACAATTTCTTCAAAATA

8.1.6,R2.RC <- TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCCCTTCAAAATA

8.1.6T3 -> TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCCCTTCAAAATA

8.2.3.5,R1.RC <- TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCCCTTCAAAATA

8.1.6,R1.RC <- TTTTAAATGAAAAAAATATGATAATAAAAAATAATTCATTAATAATGAAAAATATATACAGCCACATTCAATAAATAACAATTTCTTCAAAATA

1730 1740 1750 1760 1770 1780 1790 1800 1810 1820

8.1.6T7II.Assembla TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

7.tII <- TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

8.2.3.5,R2.RC <- TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

8.1.6,R2.RC <- TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

8.1.6T3 -> TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

8.2.3.5,R1.RC <- TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

8.1.6,R1.RC <- TTATATTATTAGTTGGCTGCATTTTTTATTTCTTCGGTAATCATAGATATCTTTAACAAATTAATTTCAATCTTGGGTTTACATATTTTATTTA

1830 1840 1850 1860 1870 1880 1890 1900 1910 1920

8.1.6T7II.Assembla ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

7.T3 <- AHAATTTGYGC-A

7.tII <- ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

8.2.3.5,R2.RC <- ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

8.1.6,R2.RC <- ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

8.1.6T3 -> ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

8.2.3.5,R1.RC <- ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

8.1.6,R1.RC <- ATATGTTTATGTGTTAATTATGCAAAATGTGAAATTCCTATACTAATATTACTTAATTTGTTTCATCATTAAATGCAGATATAATTTGTGCAA

2120 2130 2140 2150 2160 2170 2180 2190 2200

8.1.6T7II.Assembla AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

7.T3 <- AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

8.1.6E8.NEWRc+ <- AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

8.2.3.5,R2.RC <- AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

8.1.6,R2.RC <- AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

8.1.6T3 -> AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

8.2.3.5,R1.RC <- AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

8.1.6,R1.RC <- AAAAAGATATAACAATACAAAAATCACATTGTATAATGAATATAAAGACAAATAATATTGATAAATTTAGAAATTTTAAATCATGAACATGAAA

2210 2220 2230 2240 2250 2260 2270 2280 2290 2300

8.1.6T7II.Assembla AAACAAATAACATATTAATTGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

7.T3 <- AAACAAATAACATATTAATTGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

8.1.6E8.NEWRc+ <- AAACAAATAACATATTAATTGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

8.2.3.5,R2.RC <- AAACAAATAACATATTAATGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

8.1.6,R2.RC <- AAACAAATAACATATTAATGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

8.1.6T3 -> AaotAAToAaotGTTAATTTGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

8.2.3.5,R1.RC <- AAACAAATAACATATTAATTGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

8.1.6,R1.RC <- AAACAAATAACATATTAATTGAAAAATAAATTCCTTTTACAAATAACCGCTATATCTGAAATAAATAAGATACGTTAGATACGGTCGATACTA

Applied Biosystems AutoAssembler™ 1.4.0 E8 contig-1 Page 13 of 28

2310 2320 2330 2340 2350 2360 2370 2380 2390 2400

8.1.6T7II.Assembla. TATATGGTGAATGATAATGAAAGCGTATTGTGGTTTTTTTATAAACCCTATATCTCTGTTGAATATATTAATAAATGTTAAGTAACCTAAATA

7.T3 <- TATATGGTGAATGATAATGAAAGCGTATTGTGGTTTTTTTATAAACCCTATATCTCTGTTGAATATATTAATAAATGTTAAGTAACCTAAATA

8.1.6E8.NEWRc+ <- TATATGGTGAATGATAATGAAAGCGTATTGTGGTTTTTTTATAAACCCTATATCTCTGTTGAATATATTAATAAATGTTAAGTAACCTAAATA

2410 2420 2430 2440 2450 2460 2470 2480 2490

8.1.6T7II.Assembla. TAGGCTATCCAGTGAGCAAACTAAAATAAGACAGATCAAAATGTTGTTATTAATGACGTCAGACAAACAAACGCGCTCATGTTATCTGAAAAA

7.T3 <- TAGGCTATCCAGTGAGCAAACTAAAATAAGACAGATCAAAATGTTGTTATTAATGACGTCAGACAAACAAACGCGCTCATGTTATCTGAAAAA

8.1.6E8.NEWRc+ <- TAGGCTATCCAGTGAGCAAACTAAAATAAGACAGATCAAAATGTTGTTATTAATGACGTCAGACAAACAAACGCGCTCATGTTATCTGAAAAA

Applied Biosystems AutoAssembler™ 1.4.0 E8 contig-1 Page 15 of 28

2690 2700 2710 2720 2730 2740 2750 2760 2770 2780

8.1.6T7II.Assembla. AATGTAATAATAGCTCCAAACTCAAGAACTACTAAGAAACCTCTGAATTTTGTAATAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTT

8.1.6IP <- AATGTAATAATAGCTCCAAANTCAAGAACTANTANGAAGAACCTCTGAATTTTGTAATAATGTG-KGAATCTCAATGTATNTYGTAAAAATAGTK

8.1.6E8.NEWRc+ <- AATGTAATAATAGCTCCAAACTCAAGAACTACTAAGAAACCTCTGAATTTTGTAATAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTT

8.1.62E <- AATGTAATAATAGCTCCAAACTCAAGAACTACTAAGAAACCTCTGAATTTTGTAATAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTT

8.2.1T3 <- AATGTAATAATAGCTCCAAACTCAAGAACTACTAAGAAACCTCTGAATTTTGTAATAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTT

8.2.1REV <- AATGTAATAATAGCTCCAAACTCAAGAACTACTAAGAAACCTCTGAATTTTGTAATAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTT

8.1.6E8.7.1jgl <- AATGTAATAATAGCTCCAAACTCAAGAACTACTAAGAAACCTCTGAATTTTGTAATAATGTGATGAATCTCAATGTATCTCGTAAAAATAGTT

2790 2800 2810 2820 2830 2840 2850 2860 2870 2880

8.1.6T7II.Assembla. GGGAGTTCCTCCAAAAATACGAAATAATTTTCATTTATAGATTTTCTAATAGAAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

8.1.6IP <- GGGAGTTCCTCCAAAAATACGAAATAATTTTCATTTATAGATTTTCTAATAGAAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

8.1.6IP2 <- TTTATAGATTTTCTAATAG-AAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

8.1.62E <- GGGAGTTCCTCCAAAAATACGAAATAATTTTCATTTATAGATTTTCTAATAGAAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

8.2.1T3 <- GGGAGTTCCTCCAAAAATACGAAATAATTTCTTTATAGATTTTCTAATAGAAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

8.2.1REV <- GGGAGTTCCTCCAAAAATACGAAATAATTTCTTTATAGATTTTCTAATAGAAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

8.1.6E8.7.1jgl <- GGGAGTTCCTCCAAAAATACGAAATAATTTTCATTTATAGATTTTCTAATAGAAAAATTAATAACAAACATACCCAATGACTCTTGTAAACAAAAC

Applied Biosystems AutoAssembler™ 1.4.0 E8 contig-1 Page 14 of 28

2500 2510 2520 2530 2540 2550 2560 2570 2580 2590

8.1.6T7II.Assembla. ATGAATTAATAATAAATTTGGGCAATATACACAATCCTTTTATAATTTTCATAATGATCGTTCGATAGATGATTATAATTTATATACATCATCAA

8.1.6E8.NEWRc+ <- ATGAATTAATAATAAATTTGGGCAATATACACAATCCTTTTATAATTTTCATAATGATCGTTCGATAGATGATTATAATTTATATACATCATCAA

8.1.62E <- CATAATGATCGTTCGATAGATGATTATAATTTATATACATCATCAA

8.2.1T3 <- GATCGTTCGATAGATGATTATAATTTATATACATCATCAA

8.2.1REV <- GATCGTTCGATAGATGATTATAATTTATATACATCATCAA

2600 2610 2620 2630 2640 2650 2660 2670 2680

8.1.6T7II.Assembla. AAAATAATTTTGTAATTTGTTAAAAAAGGATTGAAAATATAAAAAAACAATGATATTATTTTCATATGAAGAAAAATATATACAGATA

8.1.6IP <- AAAATAATTTTGTAATTTGTTAAAAAAGGATTGAAAATATAAAAAAACAATGATATTATTTTCATATGAAGAAAAATATATACAGATA

8.1.6E8.NEWRc+ <- AAAATAATTTTGTAATTTGTTAAAAAAGGATTGAAAATATAAAAAAACAATGATATTATTTTCATATGAAGAAAAATATATACAGATA

8.1.62E <- AAAATAATTTTGTAATTTGTTAAAAAAGGATTGAAAATATAAAAAAACAATGATATTATTTTCATATGAAGAAAAATATATACAGATA

8.2.1T3 <- AAAATAATTTTGTAATTTGTTAAAAAAGGATTGAAAATATAAAAAAACAATGATATTATTTTCATATGAAGAAAAATATATACAGATA

8.2.1REV <- AAAATAATTTTGTAATTTGTTAAAAAAGGATTGAAAATATAAAAAAACAATGATATTATTTTCATATGAAGAAAAATATATACAGATA

Applied Biosystems AutoAssembler™ 1.4.0 E8 contig-1 Page 16 of 28

2890 2900 2910 2920 2930 2940 2950 2960 2970

8.1.6T7II.Assembla. TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.1.6IP <- TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.1.6IP2 <- TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.1.62E <- TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.2.1T3 <- TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.2.1REV <- TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.1.6E8.7.1jgl <- TAGATTTTATAAAAAACAATTCGAGGATATTAATAATAACACAATAAACACATAAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

8.1.6.9 <- ATTTTATAAAAAACAATTCGGGTTTTTAAATTAAC-CANTAAACCTTAAAAATGTAAACAAGAGAGATTGTTGTGAACAAATGTA

2980 2990 3000 3010 3020 3030 3040 3050 3060 3070

8.1.6T7II.Assembla. CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.1.6IP <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.1.6IP2 <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.1.62E <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.2.1T3 <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.2.1REV <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.1.6E8.7.1jgl <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.1.6.9 <- CCAACAACATTGATAATAATAATTGTGATAAACATTTTAATGAAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.2.1T7ii <- TTTAATG-AATT-AAAAATAG-CGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.2.1T7 <- ATGAATTAATAAATAGCCGAATCATATAGCATTATATTTTATAATTATACAAATTT

8.2.1M13 <- ATTTATA-TTTATAATTATACAAATTT



3080 3090 3100 3110 3120 3130 3140 3150 3160

8.1.6T7II.Assembly TAGAAAATTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAC TAGGGAAATTTAATAAACAAAGTAGATTCCG

8.1.6IP <- TAGAAAATTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAC TAGGGAAATTTAATAAACAAAGTAG-ITCCG

8.1.6IP2 <- TAGAAAATTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAC TAGGGAAATTTAATAAACAAAGTAGATTCCG

8.1.62E <- TAGAAAATTAGAATCAGTAAATA

8.2.1T3 <- TAGAAAATTAAGa

8.2.1REV <- TAGAAAATTAAGAA-C

8.1.6E8.7.1jg1 <- TAGAAAATTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAC TAGGGAAATTTAATAAACAAAGTAGATTCCG

8.1.6.9 <- TAGAAAATTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAC TAGGGAAATTTAATAAACAAAGTAGATTCCG

8.2.1T711 <- TAGAAAATTAGAATCAGTAAATATTACATATAAGG-ITCCTTAGATTACTTTTTTAATTCAC TAGGG- AATTATTAATAAACCAAGTAGATTCCG

8.2.1T7 <- TAGAAAATTAGAATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAC TAGGGAAATTTAATAAACAAAGTAGATTCCG

8.2.1M13 <- TAGAAAATTAGaATCAGTAAATATTACATATAAGGAGTCCTTAGATTACTTTTTTAATTCAS TAGGGAAATTTAATAAACAAAGTAGATTCCG

3170 3180 3190 3200 3210 3220 3230 3240 3250 3260

8.1.6T7II.Assembly ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

8.1.6IP <- ATGGAATATAATAGAAG

8.1.6IP2 <- ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTTAAATGGAGTAT

8.1.6E8.7.1jg1 <- ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

8.1.6.9 <- ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

8.2.1T711 <- ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

8.2.1T7 <- ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

8.2.1M13 <- ATGGAATATAATAGAAGAAATGATATTGATAATTTTGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

8.1.6.1E <- ATGNTATTGTTAATTTKGATTGTCTAAACCCAAAAATAATTTAAATTATTAGAAGGTGAATTAATGGAGTAT

3460 3470 3480 3490 3500 3510 3520 3530 3540 3550

8.1.6T7II.Assembly ATAAATCAACTTATGATGAAAAAAGAAGGGTTTGAAGTTCTTGAATTAGCAAAAAATTTGGAAAAAAGAACTCGAAATATAACGGAAT

8.1.63E <- ATAAATCAACTTATGATGAAAAAAGAAGGGTTTGAAGTTCTTGAATTAGCAAAAAATTTGGAAAAAAGAACTCGAAATATAACGGAAT

8.1.61E <- ATAAATCAACTTATGATGAAAAAAGAAGGGTTTGAAGTTCTTGAATTAGCAAAAAATTTGGAAAAAAGAACTCGAAATATAACGGAAT

8.1.6.9 <- ATAAATCAACTTATGATGAAA

8.2.1T711 <- ATAAATCAA

8.2.1T7 <- ATAAATCAA

8.2.1M13 <- ATAAATCAA

8.1.6.1E <- ATAAATCAACTTATGATGAAAAAAGAAGGGTTTGAAGTTCTTGAATTAGCAAAAAATTTGGAAAAAAGAACTCGAAATATAACGGAAT

3560 3570 3580 3590 3600 3610 3620 3630 3640

8.1.6T7II.Assembly TAAAAAAGAAATGAAGAACTGTTCAATTGGATATAAAATTAGAGAACTAATTAACAAATCAAGATATTATAGAGGAACAAAAATAGTGA

8.1.63E <- TAAAAAAGAAATGAAGAACTGTTCAATTGGATATAAAATTAGAGAACTAATTAACAAATCAAGATATTATAGAGGAACAAAAATAGTGA

8.1.6.5 <- TAAAAAAGAAATGAAGAACTGTTCAATTGGATATAAAATTAAGAACTAATTAACAAATCAAGATATTATAGAGGAACAAAAATAGTGA

8.1.61E <- TAAAAAAGAAATGAAGAACTGTTCAATTGGATATAAAATTAGAGAACTAATTAACAAATCAAGATATTATAGAGGAACAAAAATAGTGA

8.1.6.1E <- TAAAAAAGAAATGAAGAACTGTTCAATTGGATATAAAATTAGAGAACTAATTAACAAATCAAGATATTATAGAGGAACAAAAATAGTGA

3270 3280 3290 3300 3310 3320 3330 3340 3350 3360

8.1.6T7II.Assembly TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATACAAATATATGA

8.1.63E <- AAAGATACAATGAAAAAATAATATTATTAAATTATACCATATATGA

8.1.6IP2 <- TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATAC

8.1.61E <- TACAATATATGA

8.1.6E8.7.1jg1 <- TTGAAAATAAATGGRRGTGATTAT-AAAAATA

8.1.6.9 <- TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATACAAATATATGA

8.2.1T711 <- TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATACAAATATATGA

8.2.1T7 <- TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATACAAATATATGA

8.2.1M13 <- TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATACAAATATATGA

8.1.6.1E <- TTGAAAATAAATGGAATGATTATAAAAAATAAAAAATCTTGATCAATTCAAAGATACAATGAAAAAATAATATTATTAAATTATACAAATATATGA

3370 3380 3390 3400 3410 3420 3430 3440 3450

8.1.6T7II.Assembly ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATGAAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.1.63E <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATgaAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.1.61E <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATgaAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.1.6.9 <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAT TAAAAATgaAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.2.1T711 <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATGAAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.2.1T7 <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATGAAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.2.1M13 <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATgaAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

8.1.6.1E <- ACGAATTTAAAGGCTTGAATGACGCTATGACAAAAC TAAAAATgaAGGTATTTACAAAAATTTTGATTAAATCAAAATAAAACAAAAATTCG

3650 3660 3670 3680 3690 3700 3710 3720 3730 3740

8.1.6T7II.Assembly ACGATTTAAATTAGAATTAACAAAAAATAAAGAAATAAC TGAGAAAATGAATATATTAAAAAGCAGTTGACTTAAAGAAAGAAATAGAAA

8.1.63E <- ACGATTTAAATTAGAATTAACAAAAAATAAAGAAATAAC TGAGAAAATGAATATATTAAAAAGCAGTTGACTTAAAGAAAGAAATAGAAA

8.1.6.5 <- ACGATTTAAATTAGAATTAACAAAAAATAAAGAAATAAC TGAGAAAATGAATATATTAAAAAGCAGTTGACTTAAAGAAAGAAATAGAAA

8.1.61E <- ACGATTTAAATTAGAATTAACAAAAAATAAAGAAATAAC TGAGAAAATGAATATATTAAAAAGCAGTTGACTTAAAGAAAGAAATAGAAA

8.1.6.1E <- ACGATTTAAATTAGAATTAACAAAAAATAAAGAAATAAC TGAGAAAATGAATATATTAAAAAGCAGTTGACTTAAAGAAAGAAATAGAAA

3750 3760 3770 3780 3790 3800 3810 3820 3830 3840

8.1.6T7II.Assembly AAGACAACGTATATATTGATGAATTAGCTAAGGAACCAACATACAAATTACAAAATATATAGAAAAAAGAAATGAAATATATATACAAATAAAAT

8.1.63E <- AAGACAACGTATATATTGATGAATTAGCTAAGGAACCAACATACAAATTACAAAATATATAGAAAAAAGAAATGAAATATATATACAAATAAAAT

8.1.6.5 <- ARGACAACGTATATATTGATGAATTAGCTAAGGAACCAACATACAAATTACAAAATATATAGAAAAAAGAAATGAAATATATATACAAATAAAAT

8.1.61E <- AAGACAACGTATATATTGATGAATTAGCTAAGGAACCAACATACAAATTACAAAATATATAGAAAAAAGAAATGAAATATATATACAAATAAAAT

8.1.6.1E <- AAGACAACGTATATATTGATGAATTAGCTAAGGAACCAACATACAAATTACAAAATATATAGAAAAAAGAAATGAAATATATATACAAATAAAAT

3850 3860 3870 3880 3890 3900 3910 3920 3930  
8.1.6T7II.Assembla. CAGATTTTGACAAAATTTATGTAGCGATTTGAACAACCTTACAATGAAATGTTTTCTGTAGTTCAAGAAAGTAACATTGAGCATATAGAAAATA  
8.1.63E <- CAGATTTTGACAAAATTTATKKAG  
8.1.6.5 <- CRGATTTTGACAAAATTTATGTNGCGATTTGAACAACCTTACAATGAAATGTTTTCTGTAGTTCAAGAAAGTAACATTGAGCATATAGAAAATA  
8.1.61E <- CAGATTTTGACAAAATTTATGTAGCGATTTGAACAACCTTACAATGAAATGTTTTCTGTAGTTCAAGAAAGTAACATTGAGCATATAGAAAATA

8.1.6.1E <- CAGATTTTGACAAAATTTATGTAGCGGA

3940 3950 3960 3970 3980 3990 4000 4010 4020 4030  
8.1.6T7II.Assembla. AAACAGAAATTCCTAACTTTAAAAACGAAATAGATAATGTATATAATAAATCCAAAACATGGAACGTAAATCACATCTAAGGAATA  
8.1.6.5 <- AAACAGAAATTCCTAACTTTAAAAACGAAATAGATAATGTATATAATAAATCCAAAACATGGAACGTAAATCACATCTAAGGAATA

4230 4240 4250 4260 4270 4280 4290 4300 4310 4320  
8.1.6T7II.Assembla. GAAAGCATTATAATGATCAAAATTAATATAGACAATATTAAGGAAAAAGAGCCAAACAAACTATGATCAATTCAAAGAACATATGAAACCAATAC  
8.1.6.4 -> GAAAGCATTATAATGATCAAAATTAATATAGACAATATTAAGGAAAAAGAGCCAAACAAACTATGATCAATTCAAAGAACATATGAAACCAATAC

4330 4340 4350 4360 4370 4380 4390 4400 4410  
8.1.6T7II.Assembla. CACCTAACGAAATGAAATATCAAAAACCATCAATGAGATAAAATTTATGAAAGACGAATCTTGAGTAAAGTAAATAATATAATGATTTGACA  
8.1.6.4 -> CACCTAACGAAATGAAATATCAAAAACCATCAATGAGATAAAATTTATGAAAGACGAATCTTGAGTAAAGTAAATAATATAATGATTTGACA  
8.1.6T1a <- ACAGAAATGAAATATCAAAAACCATCAATGAGATAAAATTTATGAAAGACGAATCTTGAGTAAAGTAAATAATATAATGATTTGACA

4040 4050 4060 4070 4080 4090 4100 4110 4120  
8.1.6T7II.Assembla. TAGAACTAACACAAACCTTTCAGAAACAATTTTGGATATAATAAATATATATATGGAGAGATTACCAACGAGCTAAATAAACATTAGAAGACT  
8.1.6.4 -> GAAACAATTTTGGATATAATAAATATATATATGGAGAGATTACCAACGAGCTAAATAAACATTAGAAGACT  
8.1.6.5 <- TAGAACTAACACAAACCTTTCAGAAACAATTTTGGATATAATAAATATATATATGGAGAGATTACCAACGAGCTAAATAAACATTAGAAGACT

4130 4140 4150 4160 4170 4180 4190 4200 4210 4220  
8.1.6T7II.Assembla. TTAAAAATAAGAAAAAGGACTATCCATAAATAGATGAATACGCTAAGGAAAAATGCCAATTAAATGTTTATAAATCTAACATATTAGAAATCA  
8.1.6.4 -> TTAAAAATAAGAAAAAGGACTATCCATAAATAGATGAATACGCTAAGGAAAAATGCCAATTAAATGTTTATAAATCTAACATATTAGAAATCA  
8.1.6.5 <- TTAAAAATAA

4420 4430 4440 4450 4460 4470 4480 4490 4500 4510  
8.1.6T7II.Assembla. AAGTTTATAAAGAAAAGTTGAATCAGAACACAATAAATTTACTGAATTAACAAATAAAATAAAACAGAGTTTCGGATGAAGAAATAAAAAAT  
8.1.6.4 -> AAGTTTATAAAGAAAAGTTGAATCAGAACACAATAAATTTACTGAATTAACAAATAAAATAAAACAGAGTTTCGGATGAAGAAATAAAAAAT  
8.1.6T1a <- AAGTTTATAAAGAAAAGTTGAATCAGAACACAATAAATTTACTGAATTAACAAATAAAATAAAACAGAGTTTCGGATGAAGAAATAAAAAAT

4520 4530 4540 4550 4560 4570 4580 4590 4600  
8.1.6T7II.Assembla. ATGAAAACAATTTAATGATAGCAATCTTTAATTAATGAAACAAAAAATCCATTGAAGAGGAATATCAAAACATTAAATACACTTAAAAAGGTAG  
8.1.6.4 -> ATGAAAACAATTTAATGATAGCAATCTTTAATTAATGAAACAAAAAATCCATTGAAGAGGAATATCAAAACATTAAATACACTTAAAAAGGTAG  
8.1.6T1a <- ATGAAAACAATTTAATGATAGCAATCTTTAATTAATGAAACAAAAAATCCATTGAAGAGGAATATCAAAACATTAAATACACTTAAAAAGGTAG



4610 4620 4630 4640 4650 4660 4670 4680 4690 4700

8.1.6T7II.Assembla\_ ATGATTATATAAAAGTATGCTAAATACAAATGAATTAATAACAAATTGTCATAATAACAACTACATTAAGACAAATTAAATCAAATATTA

8.1.6.4 -> AtgATTATATAAAAGTATgT

8.1.6T1a <- ATGATTATATAAAAGTATGCTAAATACAAATGAATTAATAACAAATTGTCATAATAACAACTACATTAAGACAAATTAAATCAAATATTA

4710 4720 4730 4740 4750 4760 4770 4780 4790 4800

8.1.6T7II.Assembla\_ AAACCATAAAAGAACTAATTCAATAGATAAAATTTATACAGACAAGTTTGAAAATATATTAAACAGATAAAAAGACAGAATTAGAAACAAATTC

8.1.6T1a <- AAACCATAAAAGAACTAATTCAATAGATAAAATTTATACAGACAAGTTTGAAAATATATTAAACAGATAAAAAGACAGAATTAGAAACAAATTC

5000 5010 5020 5030 5040 5050 5060 5070 5080

8.1.6T7II.Assembla\_ CAATTTATACATCCATTTATAATATTAATGAAGATACAGAAAATGAAATTGGAAAAAGTATAGAATTACATAATACCAAGTACTTGAAAAGGTAA

8.1.6T7ii <- CAATTTATACATCCATTTATAATATTAATGAAGATACAGAAAATGAAATTGGAAAAAGTATAGAATTACATAATACCAAGTACTTGAAAAGGTAA

8.1.6T1a <- CAATTTATACATCCATTTATAATMTTAAATGaaG

8.1.6M13 <- CAATTTATACATCCATTTATAATATTAATGAAGATACAGAAAATGAAATTGGAAAAAGTATAGAATTACATAATACCAAGTACTTGAAAAGGTAA

8.1.6T7 <- CAATTTATACATCCATTTATAATATTAATGAAGATACAGAAAATGAAATTGGAAAAAGTATAGAATTACATAATACCAAGTACTTGAAAAGGTAA

5090 5100 5110 5120 5130 5140 5150 5160 5170 5180

8.1.6T7II.Assembla\_ AAGCAAACGTAACCAATTTGAATGAAATAAAGGAAAAATTAAGAGATTATGATTTTCAGGATTTTGAAAAAGAGAAAAATTAATATATCCTGATG

8.1.6T7ii <- AAGCAAACGTAACCAATTTGAATGAAATAAAGGAAAAATTAAGAGATTATGATTTTCAGGATTTTGAAAAAGAGAAAAATTAATATATCCTGATG

8.1.6M13 <- AAGCAAACGTAACCAATTTGAATGAAATAAAGGAAAAATTAAGAGATTATGATTTTCAGGATTTTGAAAAAGAGAAAAATTAATATATCCTGATG

8.1.6T7 <- AAGCAAACGTAACCAATTTGAATGAAATAAAGGAAAAATTAAGAGATTATGATTTTCAGGATTTTGAAAAAGAGAAAAATTAATATATCCTGATG

4810 4820 4830 4840 4850 4860 4870 4880 4890

8.1.6T7II.Assembla\_ CAGGATTGCTTTTAAATAATCACGAATCGAATAACAAGAATTGCTAACATATTTTATGATTTAAAGGCAAAATTTAGGAAAAATAAGAAAAACA

8.1.6T7ii <- GCTAACATATTTTATGATTTAAAGGCAAAATTTAGGAAAAATAAGAAAAACA

8.1.6T1a <- CAGGATTGCTTTTAAATAATCACGAATCGAATAACAAGAATTGCTAACATATTTTATGATTTAAAGGCAAAATTTAGGAAAAATAAGAAAAACA

8.1.6M13 <- GGCAAAATTTAGGAAAAATAAGAAAAACA

4900 4910 4920 4930 4940 4950 4960 4970 4980 4990

8.1.6T7II.Assembla\_ TGTATATAAACAAATTCATGAAAAAGAAAAAGCTGTTGAAGATATTAAGAAAAAAATGTAGATATAAATAAAATGTTTCAAAATATTGAAATAA

8.1.6T7ii <- TGTATATAAACAAATTCATGAAAAAGAAAAAGCTGTTGAAGATATTAAGAAAAAAATGTAGATATAAATAAAATGTTTCAAAATATTGAAATAA

8.1.6T1a <- TGTATATAAACAAATTCATGAAAAAGAAAAAGCTGTTGAAGATATTAAGAAAAAAATGTAGATATAAATAAAATGTTTCAAAATATTGAAATAA

8.1.6M13 <- TGTATATAAACAAATTCATGAAAAAGAAAAAGCTGTTGAAGATATTAAGAAAAAAATGTAGATATAAATAAAATGTTTCAAAATATTGAAATAA

8.1.6T7 <- TGTATATAAACAAATTCATGAAAAAGAAAAAGCTGTTGAAGATATTAAGAAAAAAATGTAGATATAAATAAAATGTTTCAAAATATTGAAATAA

5190 5200 5210 5220 5230 5240 5250 5260 5270 5280

8.1.6T7II.Assembla\_ AAAACAAAATTAATAATGATATTGATACATTAACCAAAAAATCGATAAAAGTATAGAGACATTAACGGAAATAAAAAAAATTCAGAGAACCAT

8.1.6T7ii <- AAAACAAAATTAATAATGATATTGATACATTAACCAAAAAATCGATAAAAGTATAGAGACATTAACGGAAATAAAAAAAATTCAGAGAACCAT

8.1.6M13 <- AAAACAAAATTAATAATGATATTGATACATTAACCAAAAAATCGATAAAAGTATAGAGACATTAACGGAAATAAAAAAAATTCAGAGAACCAT

8.1.6T7 <- AAAACAAAATTAATAATGATATTGATACATTAACCAAAAAATCGATAAAAGTATAGAGACATTAACGGAAATAAAAAAAATTCAGAGAACCAT

5290 5300 5310 5320 5330 5340 5350

8.1.6T7II.Assembla\_ TTGATGAAATAAAGGACAAATAGATAAATTAATAAAAGTACCAATAAAACAATGTTTAAACGAGGATC

8.1.6T7ii <- TTGATGAAATAAAGGACAAATAGATAAATTAATAAAAGTACCAATAAAACAATGTTTAAACGAGGATC

8.1.6M13 <- TTGATGAAATAAAGGACAAATAGATAAATTAATAAAAGTACCAATAAAACAATGTTTAAACGAGGATC

8.1.6T7 <- TTGATGAAATAAAGGACAAATAGATAAATTAATAAAAGTACCAATAAAACAATGTTTAAACGAGGATC