Molecular Genetic and Functional Analyses of Surface Molecules of *Theileria annulata* Sporozoites.

By Frank Katzer

For the submission as D.Phil. Thesis

University of York

July 1995

Abstract:

The molecular genetics and potential function of two sporozoite surface antigens (SPAG-1 and SPAG-2) of *Theileria annulata* have been investigated using a variety of molecular and immunochemical techniques. Both of these antigens are candidates for inclusion in a subunit vaccine. In this thesis I will concentrate on the role of SPAG-1, since SPAG-2 only became available during the later stages of this research.

I have been able to demonstrate that SPAG-1 is encoded by a single copy gene. Further, I have identified four SPAG-1 alleles using a PCR-based approach. The sequences of two different full-length SPAG-1 alleles were compared. The comparison of these revealed that the C- and N-termini are highly conserved and that the central region of the antigen gene is highly polymorphic. The implications for vaccine development and functional importance are discussed.

I mapped the 5' end of the SPAG-1 mRNA in an attempt to identify the promoter region involved in the stage specific regulation of the gene. The comparison of the upstream regions of SPAG-1 and p67, the *T. parva* homologue, resulted in the identification of two palindromic, 6 bp long sequence motifs, which are conserved in the 5' region of both genes. They are situated in a highly conserved stretch near the predicted mRNA initiation site and one of these palindromic sequences is repeated. The function of these motifs is unknown. I provide evidence for the existence of a cryptic, 30 bp long, intron in the SPAG-1 gene.

I have been able to show that recombinant SPAG-1 and SPAG-2 bind to a subset of bovine peripheral blood mononuclear cells. Some of these cells are targets for sporozoite invasion. The importance of SPAG-1, SPAG-2 and the elastin receptor for sporozoite invasion is discussed. My results show that SPAG-1 also binds to BL3 cells which are also targets for sporozoite invasion. The SPAG-1 receptor number on BL3 cells is presented.

Contents

Table Legends

Figure Index

Abbreviations

APS	Ammonium persulphate
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BoLA	Bovine leucocyte antigen
bp	Base pairs
BSA	Bovine serum albumin
cDNA	Complementary deoxy-ribonucleic acid
cpm	Counts per minute
CTVM	Centre for Tropical Veterinary Medicine (Edinburgh)
DEPC	Diethyl pyrocarbonate
DMSO	Dimethyl sulfoxide
DNA	Deoxy-ribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra-acetic acid
GST	Glutathione-S-transferase
IAA	Isoamyl alcohol
IPTG	Isopropylthio-β-D-galactoside
kb	Kilobases
kDa	Kilodaltons
LB	Luria-Bertani medium
mAb	Monoclonal antibody
MHC	Major histocompatibility complex
mRNA	Messenger ribo-nucleic acid
NBT	Nitro blue tetrazolium
NMS	Normal mouse serum
OD	Optical density
p67	*T. parva* 67 kDa sporozoite antigen
PBM	Peripheral blood mononuclear cells
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIPES	Piperzaine-NN'-bis-2-ethane sulphonic acid
PMSF	Phenylmethylsulphonyl fluoride
RFLP	Restriction fragment length polymorphism
RNA	Ribo-nucleic acid
RPMI	Roswell Park Memorial Institute medium
RT	Room temperature
SDS	Sodium dodecyl sulphate
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SPAG-1	*T. annulata* sporozoite surface antigen 1
SPAG-2	*T. annulata* sporozoite surface antigen 2
SSC	Sodium citrate
TaA	*Theileria annulata* Ankara
TaH	*Theileria annulata* Hisar
TBL	*Theileria annulata* infected bovine lymphocytes
TE	Tris-EDTA
TNE	Tris-sodium chloride-EDTA
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

Acknowledgements

I have to thank Dr. R. Hall for his helpful discussions, supervision and proof reading. I am indebted to the Department of Biology at the University of York for providing my fees, the Institute for Animal Health in Compton for the CASE Studentship and my parents for providing my maintenance money. I also have to thank Dr. W.I. Morrison for his helpful discussions, funding and use of materials and equipment during my visits to Compton. I also have to thank Dr C.J. Howard, P. Sopp and K. Griffith for their help and advice during the flow cytometry studies. I would like to thank Duncan Brown, Susanna Williamson, Lesley Bell-Sakyi and Gwen Wilkie from the Centre for Tropical Veterinary Medicine, Edinburgh, for providing sporozoite material, the cloned macroschizont cell lines and the 4B11 and 1A7 monoclonal antibodies. I also have to thank Dr. P. Knight for providing the SPAG-2 clones and her sequencing data. I am also indebted to the Nuffield Foundation for providing a travel grant for me to visit the laboratory of Prof. R. Mecham at the Washington University Medical Centre, St. Louis. I would like to thank Prof. R. Mecham for his kind hospitality, teaching me the iodination and binding procedures and for the provision of the two monoclonal antibodies BCZ and BA4. I also have to thank Meg Stark and Pete Crosby for taking the pictures displayed in this thesis.

In Dr. R. Hall's laboratory I would like to thank Daniel Lawson for helping me to get started using the computers, Nicky Boulter for discussing my thesis and making life in the laboratory more enjoyable. I also wish to thank Rob Somerville for being Rob and for keeping me amused, Andy Harrison for helpful discussions and William Starritt for looking after me. Finally but not least I thank Dr. Phil Hunt for providing the ground for my work, his helpful comments and discussion.

Declaration:

The studies reported in this thesis are the work of the author with the help of those people listed in the acknowledgements. This thesis has not been submitted previously for the award of a degree to any university. The following publications include work described in this thesis:

Boulter, N., Knight, P.A., Hunt, P.D., Hennessey, E.S., Katzer, F., Tait, A., Williamson, S., Brown, D., Baylis, H.A. and Hall, R. (1994) *Theileria annulata* sporozoite surface antigen (SPAG-1) contains neutralizing determinants in the C terminus. Parasite Immunology 16, 97-104.

Katzer, F., Carrington, M., Knight, P., Williamson, S., Tait, A., Morrison, I.W. and Hall, R. (1994) Polymorphism of SPAG-1, a candidate antigen for the inclusion in a sub-unit vaccine against *Theileria annulata*. Molecular and Biochemical Parasitology 67, 1-10.

Katzer, F., Knight, P., Williamson, S., Morrison, I.W. and Hall, R. (1994) Sporozoite surface molecules of *Theileria annulata* : Variation and Function. pp 105-109 In: Proceedings of the Third EU Workshop on Tropical Theileriosis. (Antalya, Turkey, 1994).

Knight, P., Musoke, A.J., Gachanja, J.N., Nene, V., Katzer, F., Boulter, N., Hall, R., Brown, C.G.D., Williamson, S., Kirvar, E., Bell-Sakyi, L., Hussain, K. and Tait, A. (submitted) Conservation of neutralising determinants between the sporozoite surface antigens of *Theileria annulata* and *T. parva*. Experimental Parasitology.

Chapter 1

Introduction and literature review.

1.1. The Parasite.

The protozoan parasite *Theileria annulata* was first described by Dschunkowsky and Luhs (1904) and is the causative agent of tropical theileriosis or Mediterranean Coast fever. World-wide this parasite threatens an estimated 250 million cattle and is of major economical importance due to constraints in livestock production in endemic areas due to its disease. The eradication of this parasite is hindered by its ability to infect cells of the hosts' immune system and to infect several host species.

1.1.2. The Phylum.

Theileria annulata belongs to a group of tick-borne apicomplexan parasites which infect a variety of wild and domestic animals throughout the world. The genus *Theileria* contains a number of species, some of which infect cattle as shown in Table 1. The different species and their characteristics have been reviewed by Uilenberg (1981), Dolan (1989) and Morzaria and Nene (1990).

Other apicomplexan parasites which are of major importance are *Plasmodium, Eimeria, Babesia, Sarcocystis* and *Toxoplasma.* All apicomplexan parasites are characterised by possession of an apical complex during at least one of their life cycle stages and by exhibiting at least one intracellular stage within the mammalian host. For example, they invade either erythrocytes, leucocytes or hepatocytes. The closest related parasites to *Theileria* are those of the genus *Babesia.* These two genera can be distinguished by their morphology and their target cells since *Babesia* species only invade erythrocytes while *Theileria* species

Theileria species	Vertebrate host	Invertebrate host	Disease	Distribution	References
T. annulata	*Bos taurus, Bos indicus Bubalus bubalis*	*Hyalomma spp.*	Mediterranean Coast Fever, Tropical Theileriosis	Northern Africa, Egypt to Sudan, Southern Europe Middle East, India and Southern Russia	Dschunkowsky and Luhs, 1904; Hadani et al., 1963; Neitz, 1957
T. camelensis	*camels*	unknown	unknown	Africa and parts of Russia	Dolan, 1989
T. hirci	*sheep & goats*	*Hyalomma spp.*	Malignant ovine and caprine theileriosis	South-east Europe, North Africa, Southern Russia, the Middle East	Dschunkowsky and Urodschevich, 1924; Hooshmand-Rad and Hawa, 1973a, 1973b
T. mutans	*Syncerus caffer, Bos taurus*	*Amblyomma spp.*	Benign African theileriosis I	Africa Caribbean	Theiler, 1906; Uilenberg et al., 1983;Young et al., 1978
T. parva	*Syncerus caffer. Bos taurus*	*Rhipicephalus spp.*	East Coast Fever, Corridor disease	East and Central Africa	Theiler, 1904; Koch, 1906: Cowdry and Ham, 1932
T. sergenti	*Bos taurus*	*Haemaphysalis. spp.*	Oriental Theileriosis	Eastern Russia, Japan, South Korea, Eastern China.	Yakimoff and Dekhtereff, 1930; Warnecke et al., 1979
T. taurotragi	*Taurotragus oryx Bos taurus*	*Rhipicephalus spp.*	Benign African theileriosis II	Africa	Martin and Brocklesby, 1960; Young et al., 1980
T. velifera	*Bos indicus, Syncerus caffer*	*Amblyomma spp.*		Africa	Uilenberg, 1964; Uilenberg et al., 1983

Table 1: ***Theileria*** **species, their tick vectors and their distribution.** Adapted from Irvin and Morrison (1987).

also invade leukocytes. Molecular biological techniques may soon play a more important role in the classification and the development of phylogenetic trees for apicomplexan parasites. The phylogenetic relationship between *Theileria* and other apicomplexan parasites has been established and is based on the sequence comparison of the small sub-unit ribosomal RNA molecules (Barta et al., 1991; Gajadhar et al., 1991). The current classification of the *Theileria* genus is as follows:

Subkingdom	Protozoa
Phylum	Apicomplexa
Class	Sporozoea
Subclass	Piroplasmia
Order	Piroplasmida
Family	Theileriidae
Genus	*Theileria*

Classification according to Irvin (1987) and Irvin and Morrison (1987).

1.1.3. The distribution of the *Theileria* genus and their vectors.

The genus *Theileria* is found across all continents but is less common in the Americas (Uilenberg, 1981). The economically more important species which infect cattle can be distinguished by the morphology of their piroplasms and schizonts, by serological differences detected by indirect fluorescent antibody tests (Kimber et al., 1973), and differences in their geographical distribution, pathogenicity and vector species. There are five economically important *Theileria* species and due to differences in treatment and prevention of disease it is important to be able to distinguish between the species. *T. parva* is found in Eastern and Central Africa where it is transmitted by *Rhipicephalus* species ticks and is the cause of East Coast Fever, Corridor Disease and January disease. *T. mutans* is also found in Africa, is transmitted by ticks of the *Amblyomma* genus and causes Benign African theileriosis I. In Africa one can also find *T. taurotragi*, which

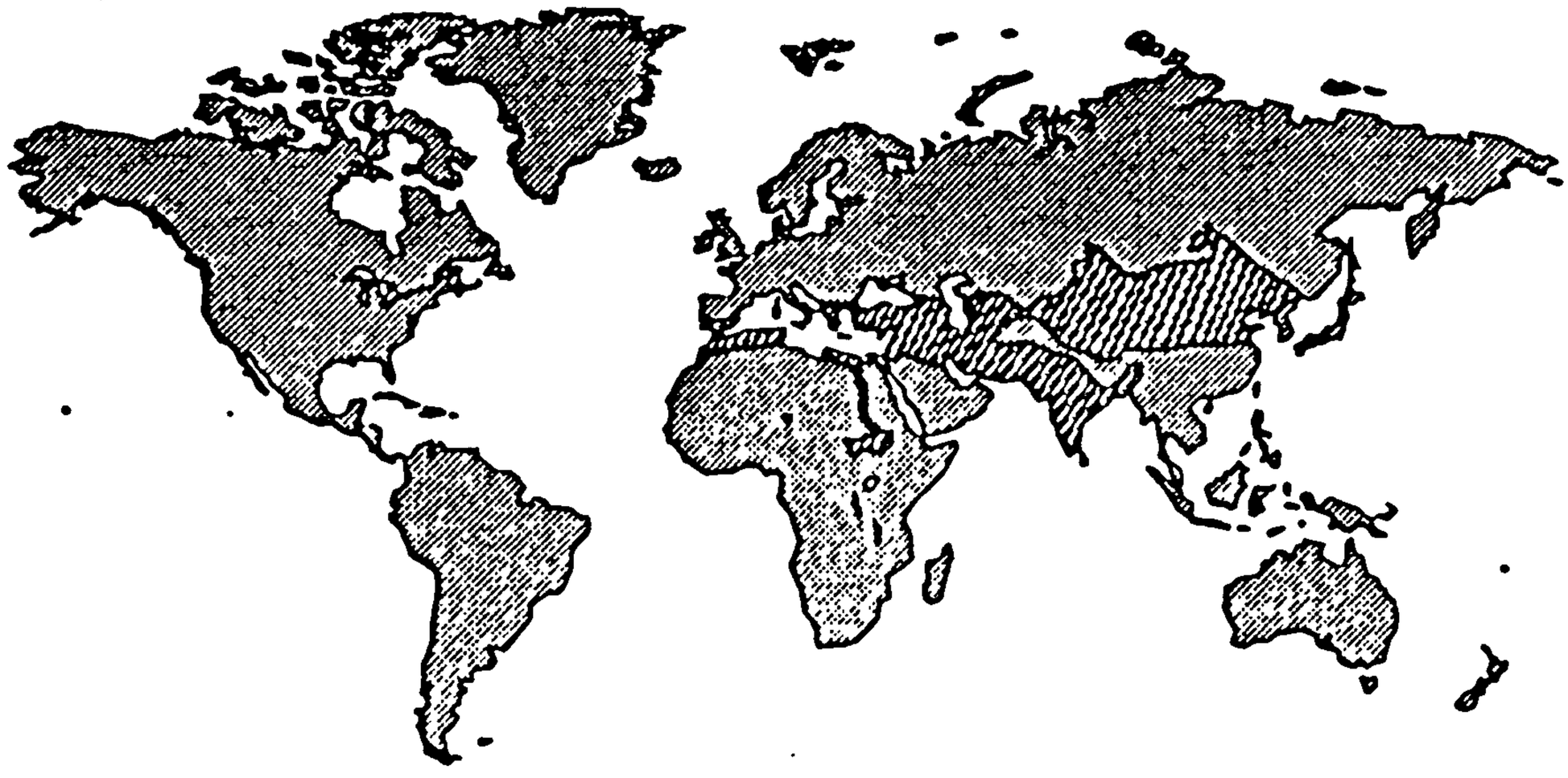


Figure 1: The distribution of *Theileria annulata*. The parasite is endemic in the shaded areas of the map of the world. This figure was adapted from Neitz (1957).

causes Benign African theileriosis II, and like *T. parva*, is transmitted by ticks of the *Rhipicephalus* genus. Diagnosis is made more difficult by another species, *T. velifera*, which is non-pathogenic but also occurs in Africa and is transmitted by ticks of the genus *Amblyomma*. The mammalian host for all these parasites is cattle, although some species also infect a variety of native mammals. In general, the symptoms of the disease are more profound in exotic and cross-bred cattle than in native breeds. This is a major constraint on the improvement of milk production in these countries. *T. annulata* the causative agent of tropical theileriosis has a much wider distribution; it is found in Southern Europe, Northern Africa, Egypt to the Sudan, the Middle East, India, parts of the former Soviet Union and southern China. *T. annulata* is transmitted by ticks of the genus *Hyalomma*, and depending on the geographical location, different species are of importance. For example the most important tick vector in India is *Hyalomma anatolicum anatolicum*, while in the former Soviet Union it is *H. scupense* and in North Africa it is *H. detritum* (Singh et al., 1986). A diagrammatic representation of the world wide distribution of *T. annulata* is shown in Figure 1. Another *Theileria* species which infects cattle and is of economical importance is *T. sergenti*. The disease associated with this parasite is called Oriental theileriosis, and is found in the eastern part of the former Soviet Union, Japan, South Korea and eastern China. It is transmitted by ticks of the genus *Haemaphysalis*. Other *Theileria* species which are also of economical importance but do not infect cattle are *T. hirci* and *T. camelensis*. The former is transmitted by ticks of the genus *Hyalomma* and causes disease in sheep and goats while the latter infects camels but its arthropod host is unknown. These *Theileria* species, their tick vectors, distribution and the disease they cause are listed in Table 1 together with the relevant references.

1.1.4. The life cycle.

T. annulata has a complicated two-host life cycle which is typical of the whole *Theileria* genus. A simplified diagrammatic form is shown in Figure 2. It involves three main phases of multiplication. These are sporogony, after sexual reproduction in the invertebrate host, followed by two asexual phases of multiplication in the vertebrate host:

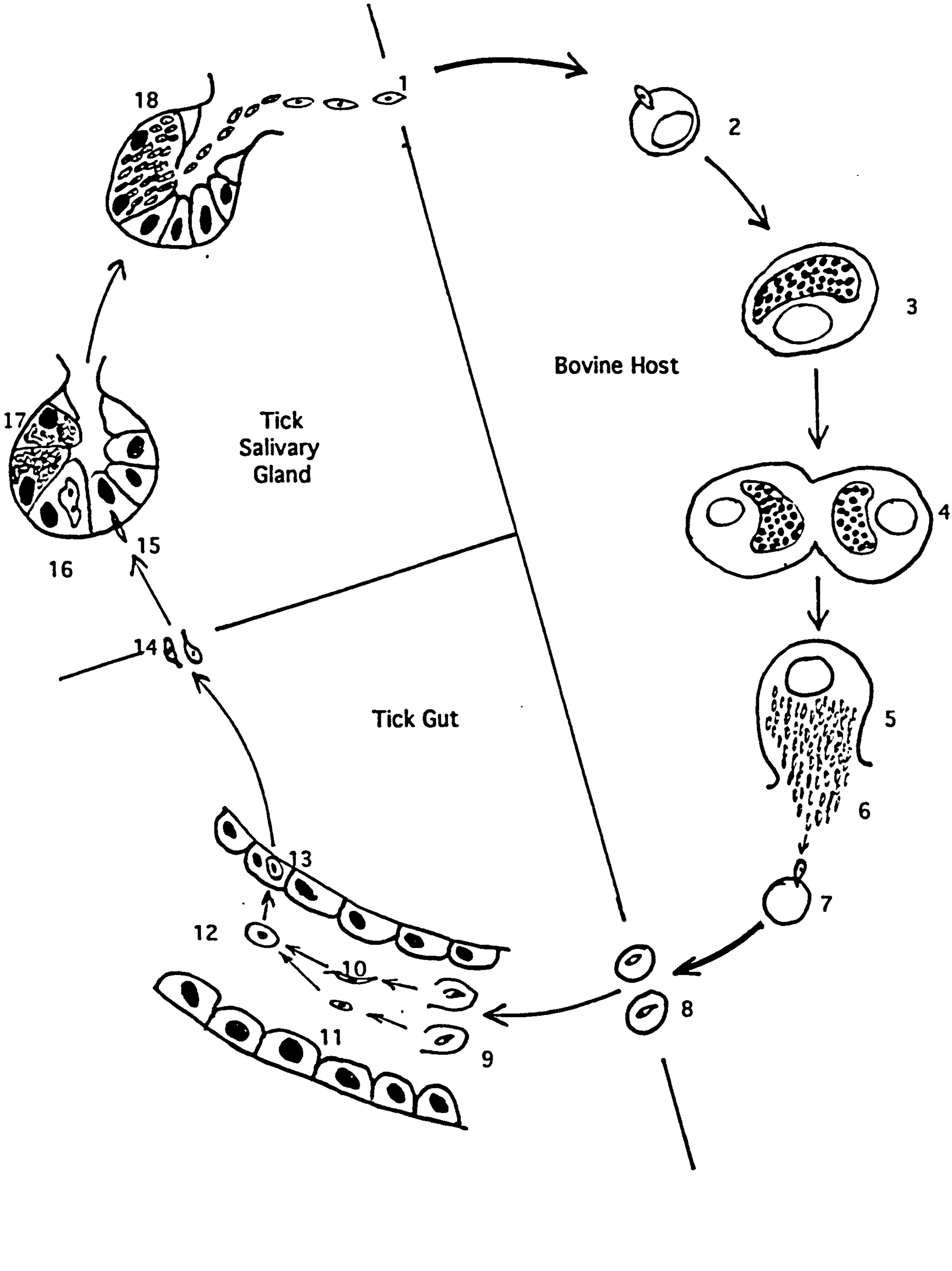


Figure 2: The life cycle of *Theileria annulata*. 1. Sporozoite in saliva of a feeding tick. 2. Sporozoite invading a leukocyte. 3. Macroschizont. 4. Macroschizonts divide in synchrony with host cells. 5. Merogony, release of merozoites. 6. Merozoite. 7. Merozoite invading an erythrocyte. 8. Piroplasms. 9. Release of ovoid stages from blood masses in tick gut. 10. Microgamete. 11. Macrogamete. 12. Zygote. 13. Infected gut epithelial cell. 14. Kinete. 15. Kinete invades salivary gland cell. 16. Asexual reproductionof kinete. 17. Sporoblast. 18. Release of sporozoites. (This figure is adapted from: Mehlhorn and Schein, 1984; Dolan, 1989; Tait and Hall, 1990).

schizogony and merogony. The life cycle of *T. annulata* has been studied by light and electron microscopy, and the following account is based on observations and descriptions by Uilenberg (1981), Fawcett et al. (1982), Mehlhorn and Schein (1984), Tait and Hall (1990) and Morzaria and Nene (1990).

T. annulata is transmitted from an infected tick of the *Hyalomma* genus to the vertebrate host (cattle *Bos taurus* or *Bos indicus* or domestic buffalo, *Bubalis bubalis*), when the tick is feeding. Most vertebrates can only be infected by ticks or by artificial inoculation since no transovarial transmission has been detected. During a blood meal an infected adult tick will inject saliva containing *Theileria* sporozoites into the blood stream of the bovine host. Once in the blood, the sporozoites rapidly invade their host cells. Generally, the target cells for *T. annulata* are leukocytes, such as B cells and macrophages (Glass et al., 1989). The invasion process and the target cells are discussed in more detail in sections 1.7.2 and 1.7.3 respectively.

The invasion process is very rapid, taking around three minutes *in vitro* (Jura, 1984). During the invasion process the sporozoite is enclosed by the host cell membrane. Subsequent to invasion this membrane is broken down, probably by secretions of the sporozoite, and finally the sporozoite comes to lie in the cytoplasm where it develops into the trophozoite. This process is distinct from other apicomplexan parasites such as *Eimeria* and *Plasmodium*, where the sporozoite is retained in a parasitophorous vacuole. The process from beginning of invasion until development of the trophozoite is completed within 30 minutes *in vitro* (Jura et al., 1983).

The trophozoite enlarges by ingesting the cytoplasm of the host cell, and undergoes repeated nuclear division which results in the formation of the macroschizont. The macroschizont then transforms and immortalises its host cell. The macroschizont can first be detected approximately three days after infection (Mehlhorn and Schein, 1984). A typical macroschizont contains six to eight nuclei, but in some cases many more have been observed (Kurtti et al., 1981). The process by which the macroschizont immortalises its host cell is only poorly understood, but elevated levels of some transcription factors (Ap-1, Jun,

NF1 and NF-κB) have been found in infected cells (Baylis et al., 1995; Ivanov and Williams, 1991). Changes in phosphoprotein and protein kinase activities have also been detected (Dyer et al., 1992). Possible mechanisms for transformation of the host cell are discussed by Dyer and Tait (1987). The process of synchronous division of host cell and macroschizont is called schizogony, which is the first phase of asexual multiplication in the bovine host.

The next phase of asexual multiplication for *T. annulata* in the bovine host is called merogony, and this phase follows schizogony. After 8-10 days merozoites start to form from the nuclei at the periphery of the schizont. They appear "rosette-like" and are termed microschizonts; they can be seen under a light microscope. The merozoites develop rhoptries and an apical polar ring and they bud from the surface of the schizont in a synchronous manner (Shaw and Tilney, 1992). Free merozoites in the bloodstream of the bovine host contain a rhoptry complex, a mitochondrion, ribosomes and a nucleus. They are 1-2 μm long and their outer surface consists of a cell membrane and two closely apposed inner membranes (Mehlhorn and Schein, 1984).

Once in the bloodstream, the merozoites rapidly invade erythrocytes. In an infected cow up to 90% of all erythrocytes can be infected. The process of invasion of erythrocytes by merozoites is only poorly understood, but is thought to be mediated by ligand-receptor interactions as demonstrated in malaria (Kawamoto et al., 1990). After the invasion the erythrocyte membrane enveloping the merozoite is disintegrated, probably by secretions from the rhoptries, and the parasite differentiates into the piroplasm form. Two different forms of piroplasms have been observed: a) slender comma shaped forms or b) ovoid or spherical forms (Mehlhorn and Schein, 1984). The frequency with which these two different forms are observed varies for different *Theileria* species, e.g. for *T. annulata* both forms occur in approximately equal numbers whereas in *T. parva* 80% of all piroplasms are comma shaped.

It has been proposed that another cycle of division is linked to the comma-shaped piroplasms. These are thought to divide by binary fission and this nuclear division is associated with cellular division. As a result

no multi-nucleate schizont-like stages can be observed. This process has been described *in vitro* by Conrad et al. (1985), and similar forms were also seen in infected cattle. The merozoites released from erythrocytes are identical to those released from schizont-infected lymphocytes and can re-infect erythrocytes. The relative importance of the two forms for the maintenance of infection is unknown.

When a tick feeds on an infected host, piroplasms are taken up during the blood meal and so transmission between vertebrate and invertebrate hosts occurs. In the gut of the tick, the spherical piroplasms develop into microgametes. These can be observed 2-4 days after the tick stops feeding (Mehlhorn and Schein, 1984). Larger gametes can also be found in the gut of the tick and these gametes are called macrogametes. These two types of gametes fuse to form the zygote, which develops into a kinete; this is a club-like, uninucleate motile stage which in turn invades the gut epithelial cells of the tick. Shortly before the molting of the skin of the tick, kinetes of *T. annulata* can be observed to migrate in the haemolymph on their way to the salivary gland.

Once a kinete reaches the salivary gland it infects salivary gland cells and is thought to lie dormant until the tick starts feeding (Mehlhorn and Schein, 1984). The kinetes are found in the cytoplasm of the gland cells where they undergo a series of differentiation and multiplication steps, after the tick starts to feed. Three to five days after feeding is initiated the sporozoites are formed which, when transmitted to the bovine host, are infective and complete the life cycle of *T. annulata*. In total, about 50,000 sporozoites can be produced from a single infected salivary gland cell.

1.2. Tropical Theileriosis.

1.2.1 The disease.

Worldwide there are 250 million cattle at risk of tropical theileriosis. In endemic areas, indigenous cattle are usually infected as calves and tend to develop only mild symptoms and recover readily. After this recovery they exhibit protective immunity to further infections but are persistent carriers (Brown, 1990a). In recent years, attempts to increase the milk yield in Third World countries have resulted in the introduction of cross-bred and exotic cattle. These were found to be highly susceptible to tropical theileriosis and exhibit a mortality of 40-60 %. Tropical theileriosis is therefore of great economical importance due to the constraints it places on the increase of milk production in endemic areas.

The epidemiology of tropical theileriosis is described by Uilenberg (1981). In subtropical areas the disease has a seasonal character, and most cases are found during the summer months when the ticks by which the disease is transmitted are most active. As would be expected the seasonality of infection is less pronounced in the tropics. An important characteristic of the disease, which could make its eradication virtually impossible, is that exotic and cross-bred cattle, once recovered, remain healthy carriers. These animals thus act as a reservoir for future infections. Another reservoir host is the Asian water buffalo, which only experiences mild disease symptoms and also remains a carrier.

The description below of the pathology of the disease is based on papers by Neitz (1957), Laiblin (1978), Uilenberg (1981), Mehlhorn and Schein (1984) and Tait and Hall (1990). The severity of the disease varies, depending on the susceptibility of the animal, the virulence of the parasite strain and the number of the sporozoites with which the animal is infected. In general, an infected animal develops a fever about 2 weeks after the tick starts to feed. The onset of fever in mechanically-infected animals is usually delayed.

The symptoms of a typical acute infection are as follows. Two days before the onset of the fever, (which is usually above 41^{o}C), schizonts

can be detected in the bloodstream of the infected animal. The fever is accompanied by swelling of the superficial lymph glands, closely followed by swelling of the regional lymph nodes draining the sites of infection. Other symptoms are the cessation of rumination, drooling from the mouth, swelling of the eyelids, accelerated pulse and breathing, general weakness, a reduction in milk production and diarrhoea. After prolonged infection, blood and mucus can be observed in the faeces. The red blood cell count also drops from 7 to 3 million cells per mm^3. The animal becomes markedly emaciated. If the animal continues to feed and the erythrocyte count recovers then the animal has a good chance of recovering completely. If, however, the erythrocyte count does not recover this will lead to severe anaemia (an erythrocyte count of less than 1 million per mm^3) and the animal will die, usually 8-15 days after the onset of the disease.

It is not clear how disease progression is related to parasite development. However, the main pathogenic effects occur during the intra-lymphocytic stage i.e. schizogony, when the infected lymphocytes multiply. It has been suggested that these displace uninfected lymphocytes in the lymph node tissue and thereby induce symptoms similar to leucosis (leucocyte depletion). Haemolytic anaemia, which can be observed during the later stages of infection, occurs during merogony and it has been shown that up to 90 % of all erythrocytes can be infected. The severe anaemia is more likely to be due to phagocytosis of infected erythrocytes than to direct lysis of the infected cells by the parasite. Parasite induced auto-immune responses may be another cause of anaemia, as suggested by Uilenberg (1981). The symptoms for other *Theileria* species vary. For example in *T. mutans*, the pathogenesis is predominantly due to the erythrocytic stage, while in *T. parva* piroplasm replication is not found and haemolytic anaemia is uncommon (Morzaria and Nene, 1990).

The disease is usually diagnosed by evaluation of the clinical symptoms. The diagnosis can be substantiated by detection of macroschizonts and/or piroplasms in Giemsa stained blood or tissue smears. Other techniques are being developed, especially for locations where more than one *Theileria* species is endemic. Such techniques could involve indirect immunofluorescent antibody (IFA) tests and

enzyme-linked-immuno-sorbent-assays (ELISAs) against macroschizonts and piroplasms, or techniques based on Southern blotting or PCR (Allsopp et al., 1989; Katende et al., 1990; Ben Miled et al., 1994)

1.2.2. The immune response of the bovine host.

The immune responses of the bovine host to *T. annulata* infection have been reviewed by Hall (1988), Tait and Hall (1990) and Brown (1990). Protective immune responses against tropical theileriosis have been observed after an animal has overcome the infection following either natural or mechanical infection or immunisation with attenuated macroschizont-infected cell lines (described in section 1.3.4.). The immunity, if not challenged subsequently through natural infection, lasts up to three years. The host's immune responses to the individual life stages of the parasite are discussed below.

The sporozoite.

Immunity against sporozoites is not essential for a protective immune response, since cattle infected with attenuated macroschizont-infected cell-lines develop a protective immune response which does not recognise the sporozoite stage (Brown et al., 1990). However, humoral immune responses against the sporozoite stage have been detected (Gray and Brown, 1981). They show that serum from immune cattle is capable of neutralising sporozoite infectivity of lymphocytes *in vitro*. This was confirmed by Preston and Brown (1985). Subsequently, monoclonal antibodies were raised against surface molecules from sporozoites and some of these were found to prevent sporozoite infection of lymphocytes *in vitro* (Williamson, 1988; Williamson et al., 1989). Two of these antibodies, 1A7 and 4B11, and the antigens they bind to will be discussed in section 1.4. These findings imply that if sufficiently high antibody titres were present in cattle they might prevent infection. Data to support this theory were obtained from an immunisation trial using a part of a recombinant sporozoite antigen expressed in the e1 loop of Hepatitis B core antigen (Boulter et al., 1995).

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

The macroschizont.

The macroschizont stage seems to be the most important target for protective immune responses, as immunisations of naive cattle with attenuated macroschizont-infected cell lines induce protective immunity (Hall, 1988; Brown et al., 1990). This immunity seems to be predominantly a cellular response employing a variety of lymphocyte sub-populations, while the humoral response, though observed, was not found to be of importance (Pipano, 1977). It was shown that cytotoxic cells directed against the macroschizont can be found in animals which recover, or have recovered, from tropical theileriosis whereas none were found in animals which died of the disease (Preston et al., 1983). Both cytotoxic T cells as well as NK cells were postulated to be involved in this immune response. Subsequently it has been shown that the immunological memory contains cytotoxic T cells which are directed against surface antigens of the macroschizont stage (Preston and Brown, 1988). They also found macrophage-mediated cytostasis to be of importance and this response could be observed consistently after both immunisation and challenge.

Three origins of the antigens recognised by the immune system have been postulated (Hall, 1988): a) they could be parasite antigens which are expressed on the surface of the parasitised cell; b) they could be surface antigens of the host which have been altered by the parasite;

or c) they could be host antigens which are not normally expressed. Attempts have been made to isolate antigens from the surface of macroschizont-infected cell lines. The method chosen was to raise monoclonal antibodies against surface molecules of macroschizont infected cell lines (Shiels et al., 1986). One of these monoclonal antibodies, 4H5, binds to a 95-120 kDa surface molecule of macroschizont infected lymphocytes (Shiels et al., 1989). 4H5 lyses macroschizont infected cells in a complement-dependent manner and suppresses infected cell proliferation (Preston et al., 1986). Unfortunately no data is available to indicate whether this antigen induces cell-mediated immune responses to the macroschizont stage, but it is worth studying this antigen further with regard to inclusion in a recombinant vaccine.

In *T. parva,* it has been shown that the most important immune response against East Coast Fever is cellular (Morrison et al., 1989). Some of the antigens recognised by the bovine immune system have been isolated and they are reviewed by Morrison et al. (1989). Both cytotoxic and T helper clones, which react with macroschizont-infected lymphocytes, have been derived from *T. parva*-infected animals (Brown et al., 1990). When these cytotoxic T cell clones were analysed they were found to be either parasite strain specific or cross-reactive. This indicates that the T cell epitopes seen by the cytotoxic T cell clones can be conserved but can also be polymorphic (Brown et al., 1990)

The merozoite/piroplasm.

Serum from immune cattle does not appear to react with infected erythrocytes but a humoral response to merozoites and piroplasms was found (Shiels et al., 1989). The antigens recognised by the immune serum can be immuno-precipitated and seem to originate from the merozoite stage. The response of the bovine immune system to these antigens has not yet been elucidated. If the immune system can decrease the number of merozoites it might be able to reduce the anaemia associated with the disease and also reduce the piroplasm transmission to the tick.

Until recently it was difficult to isolate sufficient numbers of merozoites to either study their invasion of erythrocytes or the effect of

the bovine immune system on their survival. However, merozoites have now been produced *in vitro* by culturing macroschizont-infected lymphocytes at 41°C (Glascodine et al., 1990). Subsequently, a surface antigen of the merozoite and piroplasm stage was isolated (Dickson and Shiels, 1993). The role of this antigen, and its interaction with the bovine immune system, however, still remain to be investigated.

The tick.

Little is known about the bovine immune response to the tick vector of *T. annulata*. However, it may be important to investigate the immune responses to the tick; generally, ticks remain attached to their host for at least a week, which is sufficient time to allow the immune system to raise an adequate response. Innate and acquired immune responses to ixodid ticks, to which the *Hyalomma* genus belongs, have been described (Wakelin, 1984).

It has been shown that tick bites induce an inflammatory reaction and in immune hosts elicit a rapid immune response which might prevent the tick from feeding and may even result in its death. The immune response of guinea pigs to the ixodid tick *Amblyomma americanum* was studied by Brown and Askenase (1983). They found an inflammatory response at the site of penetration of the mouth parts of the tick. Initially the bite site was infiltrated predominantly by neutrophils and after 3-5 days by basophils and eosinophils. The recruitment of basophils and eosinophils was induced by sensitised T cells and IgG1 antibodies produced in response to *A. americanum* antigens. This immune response seemed to be typical for a number of ixodid ticks. If an immune response to ticks of the *Hyalomma* genus could be induced in cattle, which would prevent ticks from feeding, it could stop the transmission of *T. annulata*.

1.3. Treatment and Control.

There are three main types of control for tropical theileriosis and East Coast fever: chemotherapy, vector control and vaccination. These control measures are reviewed by Uilenberg (1981), Irvin and Morrison (1987), Dolan (1989), Brown (1990), Morzaria and Nene (1990), Musisi (1990) and Tait and Hall (1990). The following account of present control measures for tropical theileriosis is based on these reviews.

1.3.1. Treatment of Tropical Theileriosis.

Chemotherapy has been used on a large scale to treat *T. parva* infection. It has also been used as a component in the immunisation of cattle using the "Infection and Treatment" method (see section 1.3.4.). In *T. annulata*, on the other hand, chemotherapy has hardly been used, although it is very effective. This is due to high costs and the availability of attenuated cell line vaccines (see section 1.3.3.) The earliest drug used was chlorotetracycline. During the last 15 years many new drugs have been tested, some of which were found to be promising. These drugs were tested by screening for anti-theilerial activity in *in vitro* cultures (Brown, 1989). These tests were followed by *in vivo* trials. Drugs which were found to be successful are: the anticoccidial drug halofuginone, a napthaquinone menoctone and the menoctone analogues parvaquone and buparvaquone. Menoctone and buparvaquone were found to be active against *T. annulata* and *T. parva* macroschizonts in culture (McHardy, 1978; McHardy et al., 1985). Subsequently, field trials have shown that buparvaquone is a highly effective therapeutic agent against both *T. annulata* and *T. parva* infection in cattle (McHardy, 1991). Interestingly, drugs which are effective in the treatment of *Babesia*, *Eimeria* and *Plasmodium* were ineffective against *Theileria* (McHardy, 1978). It was postulated that these drugs, proguanil, diaveridine and chloroquine, fail to penetrate the infected lymphocyte.

1.3.2. Vector control.

One of the methods used for controlling tropical theileriosis and East Coast fever is by minimising tick infestation. The current method for reducing the tick burden is by either spraying or dipping cattle in acaricides such as butocarb or amitraz. The advantage of this method is that it disposes of all ticks regardless of which parasitic infection they carry. Therefore, one can reduce the incidence of *Theileria* infection as well as other tick-borne diseases, e.g. heartwater and babesiosis by dipping cattle with acaricides. Dipping or spraying of cattle is usually done once a week since the acaricide residues are sufficient for about four days and sporozoite transmission only occurs three days after tick attachment (Urquhart et al., 1987). This method is more commonly used for the prevention of East Coast fever rather than tropical theileriosis.

There are some disadvantages associated with this method. For example one will not be able to eradicate *Theileria* by spraying or dipping cattle in acaricides since there are other reservoir hosts. Therefore this method will only permit the disease to be kept under control. Further, it is a relatively expensive method for Third World countries if it is used on a large scale. For this technique to be effective, farms need to be highly organised and rigorous as any lapse in treatment could lead directly to outbreaks of infection. Another problem is the development of resistance of ticks to the acaricide over prolonged use.

Other methods of controlling infection are the separation of infected cattle from uninfected susceptible cattle, and prevention of infected cattle moving into areas where the disease is otherwise not found. Cattle have been kept in zero grazing environments and thereby the cattle were isolated from infected stock, but this is a very expensive method and has been used only in very rare circumstances. In Morocco, it was found that some ticks behave as barn ticks and these hibernate in cracks in the walls of sheds in which the animals are kept. In these circumstances, spraying the sheds with acaricides proved effective.

1.3.3. Vaccination with attenuated macroschizont-infected cell lines.

The use of attenuated macroschizont infected cell lines as vaccines is reviewed by Pipano (1981), Hall (1988) and Brown (1990). This method is very efficient and is widely used in the prevention of tropical theileriosis. The development of this method was possible after continuous cultures of macroschizont-infected cell lines had been established *in vitro* (Hulliger et al., 1964). These infected lymphocyte cell lines became attenuated during their prolonged culture *in vitro*. This means that if these cells are injected into cattle they are less virulent, producing milder clinical symptoms and a lower parasitaemia after increasing time in culture. A cell-line is fully attenuated when it no longer produces any piroplasms when injected into cattle. The attenuation of *T. annulata* infected leucocytes is achieved in between 20-300 passages (2 months to over 2 years) in culture depending on the *T. annulata* isolate. The mechanisms involved in attenuation are not yet understood but generally, when attenuated, the parasites lose their ability to differentiate into piroplasms (Pipano, 1989) and in some cases changes in the expression of proteases occur (Baylis et al., 1992a). Once attenuation has been achieved, 10^6 - 10^7 infected cells are usually used per inoculation (Hall, 1988). Normally one immunisation is adequate to induce cross-protective immunity since the immunity is reinforced by subsequent tick challenges in the field. However, Friesian cattle usually require a second immunisation from a heterologous schizont stock of a lower passage culture to provide adequate protection (Pipano, 1981)

The attenuated macroschizont vaccine was first developed in Israel over 20 years ago (Pipano, 1981) and has drastically reduced tropical theileriosis there (Brown, 1990). Subsequently, this vaccine has become established in India where it has also proved very effective (Singh, 1990). It is currently on trial in Iran, Russia, Morocco and Turkey (Hall, 1988). So far, it has not been possible to develop an attenuated vaccine for *T. parva*. It has been postulated that the reason for this is failure of transfer of schizonts into lymphocytes of the host itself (Dolan, 1989). Musisi (1990) suggested that the BoLA mis-match inhibits the development of immunity in *T. parva* but not in *T. annulata*.

Unfortunately, this method is not without its disadvantages. For example, a "cold chain" is required. This means that the macroschizont-infected cells need to be kept at 4°C or frozen between the laboratory, where they were generated, and the site of the immunisation. This might pose a problem for immunisations in areas which are not easily accessible (Tait and Hall, 1990). Another potential problem might be the infection of the immunised cattle with other pathogens. As the macroschizont-infected cell lines were established from other animals, these could have been infected with other pathogens, or subsequently the cell cultures could be contaminated, i.e. in the laboratory. Another risk of vaccination with attenuated macroschizont-infected cell lines is that the parasite could become virulent again. This has not yet been detected but if it did happen it could result in the infection of susceptible cattle. Another drawback is that the immunisation of cattle with attenuated cell lines does not protect the cattle from becoming healthy carriers following subsequent tick challenges in the field (Tait and Hall, 1990). This technique will therefore never eradicate tropical theileriosis.

1.3.4. The infection and treatment method.

The infection and treatment method is reviewed by Brown (1990) and Morzaria and Nene (1990). In this immunisation method, cattle are infected using either virulent sporozoites, from a cryo-preserved stabilate, or infected ticks and subsequently are treated with chemotherapeutic agents during the latent period. The drugs used tend to be tetracyclines and more recently buparvaquone. Simultaneous administration of drugs and parasites have also been tested. This resulted in reduced symptoms and cross-protective immunity was usually still found. This method is used on a large scale with *T. parva* in Burundi, Kenya, Malawi, Rwanda and Zambia since no attenuated macroschizont vaccines are available. This method is rarely used in *T. annulata* since it tends to be more expensive than attenuated cell vaccines.

The main disadvantage of this method is the high cost of the chemotherapeutic drugs for Third World countries. There are other practical disadvantages, as this method needs to be administered by a qualified veterinarian and the cattle become carriers. There is also a

danger that the parasite may be passed on to uninfected cattle (Musisi, 1990; Mozaria and Nene, 1990). This method also shares two disadvantages with the attenuated macroschizont vaccine, i.e. the "cold chain" and infection with other pathogens. Another problem is that the infection and treatment method fails to induce cross protective immunity. Therefore a "cocktail" of several parasite strains are included but as a result one might introduce parasite strains to regions where they were not originally found (Musisi, 1990). It is also expensive to titre a dose of stabilate as it has to be tested on cattle and there is a large error range attached to dose evaluation (Musisi, 1990).

1.3.5. Recombinant sub-unit vaccines.

There is no sub-unit vaccine available for the prevention of either tropical theileriosis or East Coast fever, but a large effort has been made to develop such vaccines. The target for such a vaccine is to induce cross-protective immunity, to be inexpensive so that people in the Third World can afford it, to be easy to administer (ideally in a single dose) and to have a long shelf life. It would have advantages over the other preventative measures, since a cold chain would not be necessary, the product would not be contaminated with other pathogens and it could not revert to virulence.

So far a number of antigens of both *T. annulata* and *T. parva* have been isolated from several life cycle stages but small vaccination trials have only been conducted with three sporozoite antigens. p67, a sporozoite surface antigen of *T. parva*, was the first candidate to give some positive results. Most cattle immunised with p67 were protected against homologous challenge (Musoke et al., 1992; Musoke 1993). Similarly, positive results were obtained by Boulter et al. (1995) using a part of the cloned gene encoding the surface antigen SPAG-1 of *T. annulata* (this antigen is discussed in more detail in section 1.4.1.). Another trial using part of the SPAG-2 antigen (a different surface antigen of *T. annulata* sporozoite is discussed in section 1.4.2.) did not result in any detectable protection (Knight personal communication). Some of these trials indicate that sporozoite antigens might induce protective immunity, but it seems that a sub-unit vaccine containing

antigens from all parasite life-cycle stages might be more effective. If the vaccine is solely based on the sporozoite stage, some sporozoites might escape the immune responses due to their rapid invasion of host cells. If a sporozoite manages to escape the immune response it has the potential to develop into a macroschizont and subsequently into thousands of merozoites. The inclusion of antigens from different life cycle stages in the multi-stage vaccine might help to overcome the parasites at later stages of the life cycle. This might also reduce the probability of selecting polymorphic parasite strains since variation must arise in more than one of the vaccine components.

1.4. Sporozoite Antigens.

So far only one sporozoite antigen of *T. annulata* has been cloned and sequenced completely, while another has been only partly cloned and sequenced. These two antigens have been studied in this thesis. The relevant background of both antigens is described below.

1.4.1. The sporozoite surface antigen SPAG-1.

A panel of mouse monoclonal antibodies was raised against the sporozoite stage of *T. annulata* (Williamson, 1988). One of these antibodies, 1A7, was found to block invasion of sporozoites into purified bovine peripheral blood mononuclear cells *in vitro* by 66% (Williamson, 1988 and Williamson et al., 1989). 1A7 was shown to react with the surface of sporozoites using an indirect fluorescent antibody test, and binds to a protein (called SPAG-1) which is only expressed at the sporozoite stage. This protein is completely absent during subsequent life cycle stages of the parasite (Williamson et al., 1989). The protein recognized by 1A7 has a molecular weight of 104 kDa and is processed into forms on the mature sporozoite with molecular masses of 85 kDa, 70 kDa, 63 kDa and 54 kDa, which are recognised by 1A7 on Western blots (Williamson et al., 1989). 1A7 was used to screen a λgt11 expression library containing genomic *T. annulata* Hisar DNA. A clone, called λgt11-SR1, containing a 330 bp insert was isolated and sequenced (Williamson, 1989). SR1 was used to

screen a cDNA library containing *T. annulata* Hisar DNA and a clone of 2.8 kb was isolated and sequenced. This clone contained the full length SPAG-1 gene (Hall et al., 1992) whose sequence is shown in Figure 3. The SR1 region maps to the C-terminus of SPAG-1. The 1A7 epitope has been mapped to 16 amino acids within the first 45 amino acids of SR1 (Boulter et al., 1994). SPAG-1 has been shown to be expressed on the surface of the sporozoite via immuno-gold electron microscopy (Knight 1993). Sequence comparison of the predicted SPAG-1 amino acid sequence to p67, a sporozoite surface antigen of *T. parva*, revealed an overall identity of 47% (Nene et al., 1992 and Katzer et al., 1994). Another interesting finding was that the predicted SPAG-1 amino acid sequence contains two blocks of repeats as shown diagrammatically in Figure 4a. SPAG-1 contains a total of 17 repeats of the PGVGV motif, which is identical to the repeats found in bovine elastin (Hall et al., 1992 and Hall, 1994). Bovine elastin has 11 tandem repeats of this motif. An alignment of the first block of the *Theileria* repeats and the bovine sequence is shown in Figure 4b. It has been speculated that the importance of these repeats lies in the evasion of the hosts immune system.

The blocking data of the 1A7 monoclonal antibody gave the first indication that SPAG-1 is a candidate antigen for the inclusion is a sub-unit vaccine. Further support of the importance of this antigen as a sub-unit vaccine candidate was provided by data from a vaccination trial using the SR1 region of SPAG-1, expressed in the e1 loop of the hepatitis B core antigen (Boulter et al., 1995). This trial indicated that SPAG-1 induces protective immune responses against the parasite.

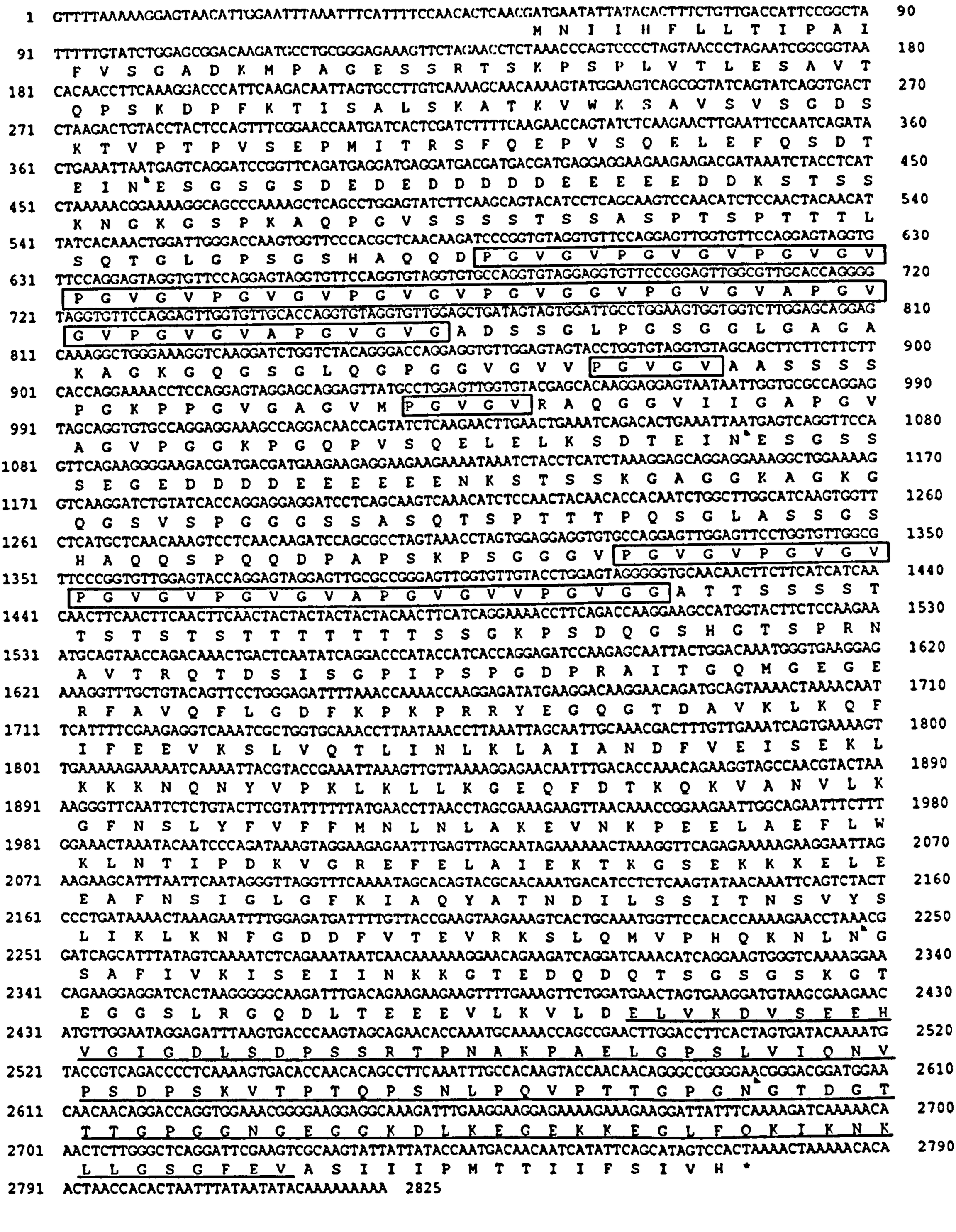


Figure 3: cDNA sequence of *T. annulata* surface antigen (SPAG-1). Reproduced from Hall et al. (1992). The predicted amino acid sequence is shown below. The PGVGV repeats showing homology to elastin are boxed. The C-terminal region encoded by the insert of λgt11-SR1, used to isolate this clone, is underlined. Putative N-linked glycosylation sites are marked with arrowheads.

a)

: [PGVGV]n : amino acids 179-236 and 424-456

: SR1 : amino acids 778-891

b)

```
S  179 PGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGGVPGVGVAPGVGVPGVGVAPGVGV
       :::::::::::::::::::::::::::::::::::: :::::: :::::::::: :::::
E  335 PGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVG.VPGVGV.PGVGVPGVGV.PGVGV
```

Figure 4: Mimicry of bovine elastin by *T. annulata* surface antigen SPAG-1. a) Schematic representation of the structure of the SPAG-1 polypeptide emphasizing the location of the elastin homologous regions and the SR1 region. b) Sequence comparison between the most N terminal block of the SPAG-1 PGVGV repeats and bovine elastin. The VGVAPG regions are underlined. S denotes the SPAG-1 sequence and E denotes the bovine elastin sequence.

1.4.2. The sporozoite surface antigen SPAG-2.

4B11, another monoclonal antibody raised against *T. annulata* sporozoites, blocks the invasion of sporozoites into host cells by about 100% (Williamson, 1988; Knight, 1993). The antigen, to which 4B11 binds has a molecular weight of 17 kDa, as detected by Western blot analysis of mature sporozoite material. The antigen identified by 4B11 was named SPAG-2 and like SPAG-1 it is also processed. The processed forms of SPAG-2 have molecular weights of 150, 67 and 17-20 kDa as seen on Western blots (Knight, 1993). The 4B11 antibody was used to screen a λgt11 expression library containing genomic *T. annulata* DNA and a λgt11 clone containing a 980 bp insert was isolated. This clone was called λgt11-KP8. KP8 was partially sequenced and cloned into pGEX-1λT and called pGEX-KP8 (Knight, 1993). 4B11 reacts with the GST-KP8 fusion protein (expressed by pGEX-KP8) on Western blots (Knight, 1993) confirming that the sequence cloned codes for the parasite protein recognised by 4B11. Unlike SPAG-1, SPAG-2 is not exclusively expressed during the sporozoite stage, but also during the macroschizont stage as shown by Northern blotting (Knight, 1993). No sequence homology between KP8 and any known gene could be detected when the gene banks were searched (Knight, personal communication). The remaining part of the SPAG-2 gene is currently being cloned and sequenced (Knight, personal communication).

1.5. Antigenic Polymorphism.

Parasites use four main techniques to evade the immune system of their vertebrate host (Borst, 1991): (1) invading cells or hiding in sites in the body where the immune system is less effective, (2) mimicry of host proteins, (3) suppressing the immune system of the host and (4) antigenic polymorphism. In *Theileria* all these mechanisms have been observed during some of the life-cycle stages in the vertebrate host and the importance of antigenic polymorphism is discussed in more detail below. When a molecule is considered for inclusion in a sub-unit vaccine it is essential to understand the degree and nature of variation in its structure. Ideally the aim is to identify an immunologically important

region which is not polymorphic and will induce immune responses cross-protective against different strains of the parasite.

1.5.1. Antigenic polymorphism in *Plasmodium* with relevance to vaccine development.

People who overcome malaria infection are usually protected against homologous challenge but not against different parasite isolates. This observation has subsequently been linked to the existence of antigenic polymorphism among malaria parasites in natural populations (Langsley and Roth, 1987; Kemp et al., 1990). Antigenic polymorphism has been well documented, for example in two of the major malarial vaccine candidates, merozoite surface protein-1 (MSP-1) and cirumsporozoite protein (CSP) (Miller et al., 1993). Attempts to map the host's immune responses to locations on known antigens have shown that the immune system tends to interact with polymorphic regions of these antigens (McCutchan et al., 1992; Mendis et al., 1991). Conserved and cross-protective epitopes prove difficult to find in malaria parasites probably due to immunological selection. For example, B-cell epitopes have been located in variant parts of the CSP antigen of *Plasmodium vivax* (Rosenberg et al., 1989), the MSP-1 (Peterson et al., 1990; Tanabe et al., 1987) and MSA2 of *P. falciparum* (Peterson et al., 1990; Fenton et al., 1990), the erythrocyte surface antigens of both *P. falciparum* (Marsh and Howard, 1986; Mendis et al., 1983) and *P. vivax* (Mendis et al., 1988) and the sexual stage antigens of *P. vivax* (Premawansa et al., 1990) and, to a lesser extent, of *P. falciparum* (Petterson et al., 1987). Variation has also been observed in regions where T-cell epitopes are located in some of these antigens (Good et al., 1988a and 1988b; Lockyer et al., 1989; Guttinger et al., 1988; Suss et al., 1993). As a result it is very difficult, if not impossible, to find non-polymorphic T-cell or B-cell epitopes. Therefore, it will be necessary to devise a sub-unit vaccine which includes a cocktail of variants of polymorphic epitopes from as many strains as possible in order to induce protective immune response against a range of parasite strains.

1.5.2. The detection of polymorphism in *Theileria*.

Polymorphism in *Theileria* has been well documented and studied. The first documented evidence for strain variation of *T. annulata* is based on the results observed during the vaccination of calves by the infection and treatment method. It was shown that calves vaccinated by this method were protected against homologous challenges but were not protected against a heterologous challenge (Gill et al., 1980). This indicates the existence of different parasite strains from distinct geographical areas. In *T. parva* the variation of geographically distinct stocks was characterised by raising monoclonal antibodies against macroschizonts from different stocks. Some monoclonal antibodies raised this way were specific for certain *T. parva* strains (Minami et al., 1983). This shows that different *T. parva* strains exist and these antibodies can be used to characterise parasite stocks found in the field. Subsequently, monoclonal antibodies were raised which demonstrate variation between *T. annulata* strains thus allowing strain typing (Shiels et al., 1986). Glucose phosphate isomerase polymorphisms have also been demonstrated in both *T. annulata* and *T. parva* (Melrose et al., 1984 and Wilkie et al., 1986). This is another method which can be employed in strain typing.

More recent advances in molecular biology have allowed further characterisation of the polymorphism at the DNA level. These were first discovered by studying the melting point of DNA from different *T. parva* strains (Allsopp et al., 1988) and later by studying restriction fragment length polymorphism (RFLP) patterns (Allsopp et al., 1988, Morzaria et al., 1990 and Bishop et al., 1993). RFLPs were found to exist for *T. annulata* DNA digested with *Eco* RI and probed with the SR1 fragment of SPAG-1 (Williamson, 1988, Williamson et al., 1989, Knight, 1993 and Katzer et al., 1994). PCR, is another technique which has also been used extensively in an attempt to detect and characterise polymorphic strains of *T. annulata* (Ben Miled et al., 1994).

Further indications of polymorphism at the protein level were obtained by SDS-PAGE. The *T. parva* polymorphic immunodominant macroschizont antigen (PIM) was shown to have a size polymorphism (Toye et al., 1991). Subsequently, similar size polymorphisms were

observed in the *T. parva* immunodominant schizont surface antigen from different stocks using 2 dimensional SDS-PAGE followed by Western blotting (Sugimoto et al., 1992). This antigen is probably be the same as PIM. A protein size polymorphism was also detected for the major merozoite surface molecule in *T. annulata*. This protein was shown to have a molecular mass of 30-kDa in isolates from Ankara (Turkey) and a molecular mass of 32-kDa in isolates from Gharb (Morocco) (Dickson and Shiels 1993). This size difference on Western blots is probably due to differential glycosylation. The detection of size polymorphism is yet another technique which can be used for parasite strain typing but yields little information with relevance to vaccine development and none about the function of a protein.

1.5.3. The study of polymorphism for treatment and vaccine purposes.

So far the main emphasis of studying polymorphism was on the identification of all polymorphic parasite strains in a given geographical location. This would permit the selection of appropriate strains for the infection and treatment method for *T. parva* and also the selection of cross protective attenuated cell lines for vaccinations against *T. annulata*. So far no single technique has been able to detect all polymorphic parasite strains known and it is very unlikely that any one technique will ever be able to do so. It might be more successful to develop an infection and treatment method for *T. parva* which incorporates all parasite strains found in a given geographical location. A better alternative to either the infection and treatment method for *T. parva* and vaccination with attenuated cell lines for *T. annulata* is the development of subunit vaccines which consist of cross-protective epitopes from various life cycle stages of all known parasite strains for either *T. annulata* or *T. parva* . One might also be able to identify protective T-cell or B-cell epitopes which are identical for all *Theileria* species.

1.6. Stage-specific gene regulation.

Stage specific gene regulation has been well documented and studied in protozoan parasites. Some of these stage-regulated genes encode molecules which are involved in cell adhesion and penetration as well as those which are involved in the evasion of the host defence mechanism (Parsons, 1990). Although many stage-regulated parasite genes have been isolated, we are just beginning to understand which mechanisms of gene regulation are of importance for protozoan parasites. The stages at which the expression of genes can be controlled are at the pre-transcriptional, transcriptional and post-transcriptional levels. Transcriptional regulation can either be constitutive, suppressible or inducible. Post-transcriptional regulation can be controlled by mRNA stability, mRNA splicing or by regulating translation. Transcriptional regulation is predicted to be the most economical mode of regulation (Latchman, 1990) and the expression of many genes has been found to involve this process; some examples will be described in the following sections.

1.6.1. Stage specific gene regulation in trypanosomes.

The best studied parasites with regard to gene regulation are the trypanosomes. One of the transcriptionally-regulated genes in these organisms encodes PARP (procyclic acidic repetitive protein)(Clayton et al., 1989). It's mRNA is α-amanitin resistant, indicating that it is transcribed by RNA polymerase I (Rudenko et al., 1989). A single stranded DNA binding site has been identified in its predicted promoter region (Brown and van der Ploeg, 1994).

The VSG (variant surface glycoprotein) genes of trypanosomes are a well studied gene family and their transcriptional control involves gene rearrangements (Pays et al., 1989; Latchman, 1990). About 1000 VSG genes are known but only one VSG transcription initiation site has been identified. During the promastigote stage, this transcription initiation site is operating constitutively, but only one VSG gene is transcribed at a time and the others lie dormant. The switching of expression from one gene to another happens by duplication of one of the silent genes which

replaces the previously transcribed gene at the expression site (Pays et al., 1989; Latchman, 1990).

Post-transcriptional gene regulation has also been demonstrated in trypanosomes. For example the expression of the phosphoglycerate kinase isozyme is regulated by mRNA stability (Gibson et al., 1988). The fructose biphosphate aldolase gene is also post-transcriptionally regulated. This was suggested since the enzyme is stage-specifically expressed but the mRNA is not stage-specifically transcribed; the factors involved still need to be identified (Vijayasarathy et al., 1990; Parsons, 1990).

In general, the regulation of trypanosome gene expression is well studied, whereas for a lot of other important parasites very little is known. The reason why so much more is known about gene regulation in trypanosomes is linked to the transfection systems which are available for these organisms. Once trypanosomes could be transformed, it became possible to study and isolate promoters, possible enhancers and other factors which might be involved in stage-specific gene regulation, such as regions responsible for mRNA stability. For example, transformation experiments have shed some light on the function of the PARP promoter (Rudenko et al., 1990; Lee and Ploeg, 1990) and the VSG promoter (Lee, 1995).

1.6.2. Stage-specific gene regulation in *Plasmodium*.

Stage-specific gene regulation is also well documented in *Plasmodium* but very little is known about the mechanisms involved in this process since a stable transformation system has only just been achieved (Wu et al., 1995). A whole array of *Plasmodium* genes have been cloned and some of them are known to be stage-specifically regulated. The most common stage for gene regulation is at the level of transcription. Table 2 contains ten *Plasmodium* genes all of which are stage-specifically regulated and involve some level of transcriptional control. More detailed studies might reveal that post-transcriptional levels of control are also effective at later stages after the initial

Name of gene	Isolation reference	Stage where the gene product is found	Stage specificity refernece	Type of gene regulation	Gene regulation reference	Multiple or unique transcription initiation sites	Reference for mapping of transcription initiation
CS	Ellis et al., 1983	sporozoite	Ellis et al., 1983	transcriptional	Ruiz i Altaba et al., 1987	multiple	Ruiz i Altaba et al., 1987
GBP130	Ravetch et al., 1985	trophozoite	Perkins, 1988	transcriptional	Lanzer et al., 1990	unique	Lanzer et al., 1990
PfHRP-II	Wellems and Howard, 1986	erythrocytic stages	Wellems and Howard, 1986	transcriptional	Lanzer et al., 1992	not known	not known
KAHRP	Pologe and Ravetch, 1986	trophozoites	Pologe and Ravtech, 1986	transcriptional	Pologe and Ravtech, 1986	unique	Lanzer et al., 1992
pfMSA-1	Hall et al., 1984	merozoite	Hall et al., 1984	transcriptional and post-transcritional	Myler, 1989 and Lanzer et al., 1992	multiple	Myler, 1989
py230	Lewis, 1989	merozoite	Lewis, 1989	transcriptional	Lewis, 1990	multiple	Lewis, 1990
Pf actin II	Wesseling et al., 1989	gametes/ zygotes	Wesseling et al., 1989	transcriptional	Wesseling et al., 1989	not known	not known
α-tubulin II	Holloway et al., 1990	male gametes	Rawlings et al., 1992	transcriptional suggested	Holloway et al., 1990	not known	not known
A-gene rRNA	Gunderson et al., 1987	asexual stages	Gunderson et al., 1987	transcriptional	Waters et al., 1989	not known	not known
C-gene rRNA	Gunderson et al., 1987	late oocysts/ sporozoites	Gunderson et al., 1987	transcriptional	Waters et al., 1989	not known	not known

Table 2: *Plasmodium* genes and their stage specific gene regulation.

regulation of transcription. One known case is pfMSA-1, where the gene is transcribed during the erythrocytic stages (Myler, 1989) but the mRNA only accumulates during the late erythrocytic stages, as demonstrated in a comparison of Northern blot and nuclear run-on data (Lanzer et al., 1992). The py230 gene is the *Plasmodium yoelii* equivalent to the pfMSA-1 gene of *Plasmodium falciparum*. Interestingly it has been shown that the pfMSA-1 gene is post-transcriptionally as well as transcriptionally regulated (Myler, 1989: Lanzer et al., 1992), while in the case of py230, transcriptional regulation is the only detected mode of regulation so far (Lewis, 1990). This shows that further levels of gene regulation might act at later stages which have not yet been identified.

The process of stage-specific gene regulation in *Plasmodium* is being studied in order to gain a better understanding of parasite genetics. In the future this might lead to a form of treatment which is based entirely on a genetic approach by disrupting parasite specific gene expression. This might act by blocking developmental stages of the parasite by inhibiting unique, parasite-specific transcription factors. The first step towards an understanding of transcriptional regulation in *Plasmodium* was achieved by mapping the site for transcription initiation of the CS gene (Ruiz i Altaba et al., 1987). This gene is transcribed only during the sporozoite stage. Subsequently the site for transcription initiation has been mapped for another four *Plasmodium* genes: pfMSA-1 (Myler, 1989), py230 (Lewis, 1990), GBP 130 (Lanzer et al., 1990) and KAHRP (Lanzer et al., 1992). Interestingly, it has been noted that both KAHRP and GBP 130 have unique transcription initiation sites (Lanzer et al., 1990; Lanzer et al., 1992) while CS, HRP II and pfMSA-1 have multiple sites (Ruiz i Altaba et al., 1987; Myler, 1989; Lewis, 1990). For the actin genes, pf-actin I and pf-actin II, of *Plasmodium falciparum* it has been shown that the sites for mRNA initiation are a maximum of 1100 bp and 450 bp, respectively, 5' from the ATG start codon (Wesseling et al., 1989). The beginning of the mRNA has not yet been mapped to specific sequences since only 250 bp and 331 bp, respectively, of the 5' untranslated region has been cloned and sequenced for these genes.

Once the transcription initiation sites for the five genes were mapped and the sequences of the 5' regions of these initiation sites were available, it became possible to search for putative transcription factor

binding sites. In eukaryotes it has been shown that promoter binding proteins usually bind within 100 bp of the transcription initiation site (Dynman & Tjian, 1985; Maniatis et al., 1987). In the 5' region of the multiple transcription initiation site of the py230 gene TATA boxes were found. The consensus sequence of a TATA box is T A T A A/T A A/T (Corden et al., 1980) and has been shown to be the site responsible for binding by the transcription factor TFIID (Horikoshi et al., 1989). TFIID is responsible for transcription of genes by RNA polymerase II (Buratowski et al., 1989; Pugh & Tjian 1992; Flores et al., 1992) but is also needed for the transcription of genes which do not contain a TATA box in their promoter region (Pugh & Tjian 1991). The TATA box binding protein (TBP) of *Plasmodium falciparum* has been cloned and sequenced (McAndrew et al., 1993). Although TATA boxes were found in the 5' regions of stage-specifically regulated genes in *Plasmodium* their role in stage-specific gene regulation is questionable as the genome of these parasites is highly A and T rich and the intergenic regions consists of 90% A's and T's (Hyde and Sims, 1987). The TATA box motif has also been found in abundance in both constitutive and developmentally regulated genes, for example the α-tubulin I, the α-tubulin II and β-tubulin genes (Holloway et al., 1989; Holloway et al., 1990).

When the 5' region of the transcription sites of the pfMSA-1 gene were searched for other motifs, recognized by other transcription factors, two immunoglobulin octamer (Oct-1) sequences were found but no CCAAT box (Lewis, 1990). The consensus sequences for Oct-1 binding and the CCAAT box is shown in Table 3. Since two Oct-1 sites were found in the region 5' of transcription initiation it has been suggested that Oct-1 might be involved in transcriptional regulation of pfMSA-1 (Lewis, 1990) but more evidence is needed to substantiate this theory.

The 5' region of the GBP130 gene was also analysed and it was shown that the transcription initiation site maps to 980 bp upstream of the ATG codon (Lanzer et al., 1990). The GBP 130 gene is developmentally regulated as shown in Table 2 and transcripts of the gene are only found during the trophozoite stage (Lanzer et al., 1990). The region 5' to transcription initiation contains a region which is homologous to the SV40 enhancer sequence (Lanzer et al., 1992 and Weiher 1983). Subsequently it has been shown that nuclear extracts from the

Name of Sequence	Name of DNA-Binding Protein	DNA Sequence	Relevant references
	ANTP	TCAATTAAAT	Treisman et al., 1990
	API	TGAGTCG	Lee et al., 1987
	AP2	CCCCAGGC	Imagawa et al., 1987
	bicoid	TCTAATCCC	Hanes and Brent, 1989
CRE	CREB	t/g t/a CGTCA	Yamamoto et al., 1988
CCAAT BOX	CTF, NF1	TGTGGCTNNNAGCCAA	Santoro et al., 1988; Mermod et al., 1989
	C/EBP	ATTGCGCAAT	Landschultz et al., 1989
	ets	CACTTCCT	Wasylyk et al., 1990
	FOS	same sequence as API	Neuberg et al., 1989
	Ftz	TCAATTAAATGA	Jaynes and O'Farrell, 1988
GRE Promoter	Glucocorticoid	GGTACANNNTGTTCT	Hollenberg and Evans, 1988
GRE Supressor		ATYACNNNNTGATCW	Godowski et al., 1988
	GAL 4	TGTGGATATATG	Carey et al., 1990
	GCN4	TGA g/c TCAT	Hope and Struhl, 1986
HSE		CTNGAATNTTCTAGA	Bienz and Pelham, 1986
	HSU2	GCGTTCCGGGA	Gaston and Fried, 1992
	H2TF1	TGGGATTCCCCA	Singh et al., 1988
	JUN	same sequence as API	Neuberg et al., 1989
	KROX-20	GCGG c/g GGCG	Nardelli et al., 1991
	NF-κB	TGGGGATTCCCCA	Lenardo and Baltimore, 1989
	myb	c/t AAC c/t G	Biedenkapp et al., 1988
	myc	CACGTG	Murre et al., 1989
	Oct1, Oct2	ATGCAAAT or ATTTGCAT	Falkner et al., 1986; Latchman, 1990
	Oct1+VP16	TAATG a/g AT	O'Hare and Goding, 1988; Preston et al., 1988
	Oestrogen	AGGTCANNNTGACCT	Kumar and Chambon, 1988
	Pit-1	ATGAATA t/a	Ingraham et al., 1988
	rel	GTGGAGATGGGGAATCCCCA	Gilmore, 1990
Serum Response Element		GATGTCCATATTAGGACATC	Norman et al., 1988
	spi-1	GAGGAA	Goebl, 1990
Sp1 Box		GGGGCGG	Courey and Tjian, 1988; Kadonga et al., 1987
TATA Box	TFIID	TATA t/a A t/a	Corden et al., 1980
Thyroid Response Element	Thyroid hormone/ Retinoic acid	TCAGGTCATGACCTGA	Evans, 1988; Umesono and Evans, 1989

Table 3: Transcription factors and the DNA sequence motifs they bind to.

erythrocytic stages of the parasite bind to this region (Lanzer et al., 1993), indicating that it is very likely to be involved in the regulation of the gene. The 5' region prior to the transcriptional initiation site of the CS gene also has homology to the SV40 enhancer sequence (Lanzer et al., 1992; Lanzer et al., 1993). A transformation system needs to be operational to determine whether these sequences or any other sequences, which are speculated to be promoter binding sites, are involved in stage specific gene regulation. The stable transformation of *Plasmodium* blood stages has only just been achieved (Wu et al., 1995) and this system will hopefully provide the required method to study the involvement of these sequence motifs in stage-specific gene regulation in *Plasmodium* .

1.6.3. Stage-specific gene regulation in *Theileria*.

Little is known about the process of stage specific gene regulation in *Theileria*. The first indication that some level of control of expression must occur in *T. annulata* is the existence of life-cycle stage specific antigens. The existence of stage specific antigens was first recognized by raising monoclonal antibodies against surface molecules of different parasite stages for vaccine development. Shiels et al. (1986) raised a panel of monoclonal antibodies against macroschizonts. It was shown that some of these mAbs reacted exclusively with schizonts, some reacted with schizonts, piroplasms and sporozoites, while others just reacted with schizonts and piroplasms or schizonts and sporozoites. It was therefore concluded that *Theileria* expresses life-cycle stage specific antigens. Subsequently, another four monoclonal antibodies were raised which were specific for macroschizonts and these did not react with piroplasms, kinetes or sporozoites (Shapiro et al., 1987). These immunological experiments showed that stage specific gene expression occurred, but they yielded no information as to which level of regulation is of importance.

More information about gene regulation in *Theileria* was obtained when the first genes were isolated and molecular biology techniques could be employed to study at which stage their gene regulation was controlled. So far three genes have been isolated which are transcribed

and expressed constitutively. Firstly the apocytochrome B gene of *T. annulata* (Megson et al., 1991) has been characterised. This gene is located on the 6.3 kb extrachromosomal DNA element (Hall et al., 1990). It is thought that this gene product might be a suitable target for the development of a treatment for tropical theileriosis, as its expression is thought to be essential for the parasite (Megson et al., 1991). The second constitutively expressed gene is the cysteine protease of *T. annulata* (Baylis et al., 1992) and the third is the hsp 70.1 gene of *T. annulata* (Mason et al., 1989). The latter has been shown to be transcribed during the sporozoite, schizont and piroplasm stages (Mason et al., 1989) but the level of expression is still further inducible. Although these genes are constitutively transcribed during the life-cycle stages in the bovine host very little is known about gene expression during the early life-cycle stages in the tick. If these stages were studied, it might become apparent that some genes which are constitutively expressed in the bovine host might not be transcribed at all during the stages in the tick, or at least not for all stages in the tick.

So far, most *Theileria* genes isolated are stage-specifically expressed and all these genes are regulated to some extent at the transcriptional level. A list of these genes is shown in Table 4. Some of these genes are transcribed only during a single stage of the life cycle of the parasite, for example p67 (Nene et al., 1992). It is only transcribed during the sporozoite stage and ORF-1, an open reading frame which is located 5' to the p67 gene, is only transcribed during the schizont stage and not during the sporozoite nor the piroplasm stages (Nene et al., 1992). The function of this gene is as yet unknown (Nene et al., 1992). 3' to the p67 gene is another open reading frame, ORF-2, which is transcribed during the sporozoite, schizont and piroplasm stages. Its function is unknown but it is another potential constitutively expressed gene (Nene et al., 1990). Interestingly the 117 kDa rhoptry protein of *T. annulata* is only found in merozoite rhoptries and not in those of the sporozoite, and the gene is transcribed 2 to 3 days prior to rhoptry formation. Afterwards, transcription is down regulated (McDonald, personal communication). These examples indicate that transcriptional control is very important for the development of *Theileria* but unfortunately very little is known about how these genes are regulated.

Transcription initiation sites have been mapped for only one *Theileria* gene, the hsp 70.1 gene (Mason et al., 1989). The beginning of the mRNA was mapped to 215 bp 5' of the ATG start codon using the S1 mapping technique. Two sequences which show high homology to the heat-shock element binding site consensus sequence, shown in Table 3, were found in the 5' region upstream of the transcription initiation site (Mason et al., 1989). Furthermore, a putative TATA box was found in the 5' region and polyadenylation signals were found in the 3' untranslated region. Since the hsp 70.1 gene is heat inducible it was speculated that these conserved promoter binding sequences are functional in *Theileria*. To date, no promoter binding proteins have been isolated in *Theileria*.

Name of gene	Stage where the gene product is found.	References
QP protein	schizont, piroplasm	Baylis et al., 1992
PIM	sporozoite, schizont	Toye et al., 1991
ORF-1	schizont	Nene et al., 1992
SPAG-1	sporozoite	Williamson et al., 1989
SPAG-2	sporozoite, schizont	Knight, 1993
p67	sporozoite	Nene et al., 1992
117 kDa rhoptry protein	merozoite	McDonald, pers. comm.
30-32 kDa merozoite antigen	merozoite, piroplasm	Glascodine et al., 1990 Dickson and Shiels, 1993

Table 4: A list of *Theileria* genes which are stage specifically regulated and the level of regulation involves some transcriptional regulation.

1.7. Host cell recognition and invasion.

In order to evade the immune responses of their host, a very successful strategy has been developed by a number of protozoan parasites; they become intracellular for one or more of their life-cycle stages (Bloom, 1979; Borst, 1991). The parasites have develop adaptive processes that allow host cell penetration and intracellular survival. Ideally, vaccine and treatment development should target the interception or destruction of the parasite before it becomes intracellular and thereby establishes an infection. Therefore, one needs to understand how the parasite interacts with host cells and how defence mechanisms function against them (Vanderberg and Stewart, 1990).

1.7.1. Host cell recognition and invasion by malaria parasites.

Plasmodium exhibits three invasive stages: a) the sporozoite invading hepatocytes in the vertebrate host, b) the merozoite whose target cell is the erythrocyte and c) the ookinete which infects the midgut endothelial cell of the mosquito vector (Sinden, 1985). The first two stages are of importance for the development of therapeutic and preventative treatments. The specificity of the recognition event and the speed of the invasion process implies that the invasion process is receptor-ligand mediated (Hadley et al., 1986). Further, it has been suggested that the invasion of *Plasmodium* is a multi-step process requiring a series of interactions involving several different host and parasite molecules (Mitchell et al., 1986). The current knowledge of the mechanisms of host recognition and invasion during hepatocyte and erythrocyte invasion are described below.

Recognition of hepatocytes by Plasmodium *sporozoites.*

When sporozoites are transmitted by their mosquito vector into their mammalian host they circulate through the blood stream until they reach the liver. Once there they have to penetrate or move between either Kupffer cells or endothelial cells, which line the lumen of the

hepatic sinusoid, in order to reach their target cells, the hepatocytes (Vanderberg and Stewart, 1990). The sporozoites are well adapted and as many as 95% of all injected sporozoites will reach and invade their target cells (Vanderberg, 1968). The CS protein has been proposed as a candidate ligand for the putative hepatic cell receptor (Nussenzweig and Nussenzweig, 1989). This antigen was proposed for three reasons: a) it is expressed on the surface of the infective (salivary gland) sporozoite (Yoshida et al., 1981) but is only present in small numbers on non-infective sporozoites from oocysts (Aikawa et al., 1981); b) the CS protein contains a region, region II, which bears a striking homology to a cell adhesion domain of thrombospondin (Prater et al., 1991; Tuszynski et al., 1989) and to similar regions of several other proteins (Clarke et al., 1990; Hedstrom et al., 1990; Robson et al., 1988; Goundis and Reid, 1988); and c) sporozoite invasion of hepatocytes can be blocked by monoclonal antibodies which react with the CS protein (Potocnjak et al., 1980). Two other characteristics of the CS protein have been suggested as assisting in host cell invasion. The first is that the CS protein contains a Ca^{2+} binding region which interacts with the phospholipid membrane of host cells. It is thereby of critical function during attachment, invasion and the development of the malaria parasites in hepatic cells (Verdini et al., 1991). The second observation is that the CS protein possess both positively and negatively charged sites which are accessible on the surface of the sporozoite, and are thought to be components of surface ligands during hepatocyte invasion (Mathews and Vanderberg, 1994). Conclusive evidence showing that the CS protein is involved in hepatocyte invasion was provided by Cerami and co-workers (1992). They showed that recombinant CSP of *P. falciparum* binds specifically to the membrane surface of hepatocytes and not to other cells. They also showed that the binding can be inhibited by synthetic peptides representing the evolutionarily-conserved region II of the CS protein. Further, this region has been shown to be responsible for binding to the heperan sulfate proteoglycans on the surface of the hepatocytes (Cerami et al., 1992; Panacake et al., 1992; Frevert et al., 1993; Frevert, 1994; Cerami et al., 1994). The binding domain has been found in CS proteins from all *Plasmodium* species as well as another sporozoite protein SSP2/TRAP (Rogers et al., 1992a; Rogers et al., 1992b; Wizel et al., 1994; Robson et al., 1988). The SSP2/TRAP protein is located in the micronemes and the surface of the sporozoite stage and has been shown to bind to sulfated

glycoconjugates on the surface hepatocytes (Miller et al., 1993). It has been proposed that the recognition and initiation of hepatocyte invasion is achieved by a cascade of receptor-ligand interaction. It is therefore likely that the CS protein and SSP2/TRAP act together with other sporozoite proteins during the invasion process (Hedstorm et al., 1990; Moelans, et al., 1991; Fidock et al., 1994a; Fidock et al., 1994b). Interestingly, the immunodominant microneme protein of *Eimeria tenella* has the same thrombospondin-related region as the CS protein and it has also been speculated that it is involved in host cell invasion (Clarke et al., 1990; Tomley et al., 1991). Another surface molecule of the sporozoite stage, CSP-2 has only just been isolated (Sina et al., 1995). Since monoclonal antibodies directed against this protein block the invasion of hepatocytes *in vitro* this might indicate that this protein is also involved in the invasion of host cells (Sina et al., 1995).

Recognition of erythrocytes by merozoites of Plasmodium.

Over 40 years ago, it was suggested that the invasion of erythrocytes by merozoites is receptor-ligand mediated (McGhee, 1953). Although the process of merozoite attachment and invasion has been described in detail at the light microscope level (Dvorak et al., 1975) and by ultrastructural studies with the electron microscope (reviewed by Bannister and Dluzewski, 1990; Aikawa et al., 1978; Perkins, 1989), the molecules involved and their roles are still poorly understood. One of the main candidate ligands on the merozoite surface is the MSP-1 protein. This antigen is processed and the resulting polypeptides are found uniformly distributed across the surface of the merozoite (Holder et al., 1987, McBride and Heidrich, 1987). Perkins and Rocco (1988) have demonstrated that this protein binds to erythrocytes and they speculated that the polypeptide binds to sialic acid on the erythrocyte surface. When the structure of MSP-1 was investigated, two epidermal growth factor (EGF) modules were found in the 19 kDa C-terminal region of the antigen (Blackman et al., 1991a). The 19 kDa polypeptide is linked to a GPI-anchor and remains on the surface of the merozoite during the invasion process (Blackman et al., 1990a; Blackman et al., 1991b). Subsequently it has been shown that this cleavage is intrinsic to successful erythrocyte invasion and that this process is facilitated by a serine protease (Blackman et al., 1993). The proteolytic cleavage process

of MSP-1 by a serine protease has also been confirmed *in vivo* (Odea et al., 1995). Further evidence that this C-terminal region is involved in the invasion of erythrocytes was obtained by the use of antibodies in invasion blocking studies. A monoclonal antibody which binds to the first EGF module was shown to block merozoite binding to erythrocytes (Chappel & Holder, 1993: Su et al., 1993). Other monoclonal antibodies which react with the 19 kDa C-terminal fragment also block invasion (Blackman et al., 1990a; Cooper et al., 1992). Therefore, this C-terminal polypeptide seems important in the invasion process and the role of the EGF modules in this process needs to be evaluated further. Binding studies have revealed that the MSP-1 protein also binds to sugar residues found on the surface of erythrocytes (Qazi et al., 1994).

It seems likely that the MSP-1 protein is one of the first receptor-ligand interactions between the merozoite and the erythrocyte. This interaction might induce the next step in the cascade, i.e. the release of EBA-175 from the micronemes. EBA-175 was shown to belong to a family of microneme proteins (Camus and Hadley, 1985; Haynes et al., 1988; Wertheimer and Barnwell, 1989; Adams et al., 1992; Peterson et al., 1995) which can bind to the sialic acid moiety on glycophorin A (Orlandi et al, 1990; Klotz et al 1992; Orlandi et al., 1992; Sim et al., 1994). Sequence analysis of the microneme proteins of the above gene family reveals that these proteins are related and exhibit dimorphism (Klotz et al., 1992; Prickett et al., 1994). Similar to MSP-1, EBA-175 is proteolyticly cleaved during the invasion process into a 65-kDa fragment whose binding to erythrocytes is sialic acid independent (Kain et al., 1993). It is predicted that this cleavage process is involved in the formation of the moving junction between the merozoite and the erythrocyte membrane and thereby permitting membrane fusion with the host cell (Kain et al., 1993). The region involved in the interaction with the host cell membrane has been mapped to a 42 amino acid peptide in the C-terminus of EBA-175 (Sim et al., 1994a; Sim et al., 1994b). It has been proposed that EBA-175 is the most important ligand for binding of merozoites to glycophorin A on erythrocytes (Sim, 1995). However, other evidence exists that *P. falciparum* can utilize other receptors, such as glycophorin B (Dolan et al., 1994). Therefore it seems likely that parasite has multiple genes with different erythrocyte-receptor specificities which can be switched on and off (Miller et al., 1994).

Another merozoite protein has been isolated and is called MCP-1 (merozoite cap protein-1). This protein follows the distribution of the moving junction during the invasion of erythrocytes. Sequence conservation between this protein and bacterial and eukaryotic proteins has indicated that it has oxidereductase activity as well as another domain which may mediate interactions of MCP-1 with the cytoskeleton in erythrocytes (Hudson-Taylor, et al., 1995). During this and later stages of the invasion process, other proteins will be released from the merozoite such as further erythrocyte membrane binding proteins (Elmoudni et al., 1993), serine-proteases and phospholipases (Braunbreton et al., 1994; Florent et al., 1994), protein kinases (Ward et al., 1994) and various rhoptry proteins (Grellier et al., 1994: Samyellowe et al., 1995; Ndengele et al., 1995).

A different receptor on the erythrocyte has been implicated in merozoite invasion of *P. knowlesi* and *P. vivax*. This is the Duffy blood group antigen (Miller et al., 1975; Miller et al.,1979; Barnwell et al., 1989). One line of evidence for this was provided by monoclonal antibodies raised against the Duffy blood group antigen, as they block merozoite invasion (Barnwell et al., 1989). When the invasion of *P. falciparum* merozoites was investigated it was shown that there was no correlation of the expression of the Duffy blood group antigen and susceptibility of target erythrocytes (Holder, 1994). It has been shown that the Duffy blood group antigen consists of the IL8 receptor and MGSA (melanoma growth stimulatory activity) receptor (Horuk et al., 1993). Mutations in the MGSA receptors resulted in the loss of infectivity of the mutated cell lines (Hesselgesser et al., 1995). The family of Duffy blood group antigen binding proteins was isolated from *P. knowlesi* and *P. vivax*, was shown to contain homology to EBA-175 and their binding domain was mapped to a cystein-rich domain (Chitinis and Miller, 1994). This binding domain is hypervariable but this variability is limited (Tsuboi et al., 1994).

Although different *Plasmodium* species might utilise different ligand-receptor interactions for both the initial and subsequent invasion steps the general principal is the same and the parasite proteins involved seem to be related. Therefore, for invasion to be successful a cascade of parasite and host cell protein-protein

interactions has to take place and as a result the invading parasite manages to hide inside the host cell where it is protected from most of the hosts immune responses.

1.7.2. Host cell attachment and invasion by *Toxoplasma* tachyzoites.

Toxoplasma tachyzoites exhibit a very broad host cell range. They can invade virtually all nucleated cell types from warm-blooded vertebrate hosts (Werk, 1985). The invasion event is not due to phagocytosis as previously thought but is an active process. During this process no host cell membrane ruffling, actin microfilament reorganisation or tyrosin phosphorylation of host proteins takes place (Morisaki et al., 1995). These effects have, however, been seen during phagocytosis (Morisaki et al., 1995). The notion that trachyzoites can invade host cells by phagocytosis was based on the observation that phagocytic cells take up *Toxoplasma* parasites and that the parasite can survive in these cells. This however, is due to some parasites escaping from the phagosome by a process analogous to invasion (Morisaki et al., 1995). The invasion process of tachyzoites is very similar to the invasion process of *Plasmodium* because a cascade of interactions between the parasite and the host cell has to take place. Initially the tachyzoite loosely attaches to the host cell and rearranges itself to bring the apical pole in close apposition to the membrane (Wong and Remington, 1993; Werk, 1989). These processes involve receptor-ligand interactions and are followed by the protrusion of the conoid. Then a moving junction is formed between the parasite and the host cell membrane which is accompanied by microneme exocytosis. Rhoptry exocytosis is the next step and when the parasite is engulfed completely the prasitophorous vacuole is formed and transformed into an environment suitable for the parasite with the help from the dense granules (Joiner and Dubremetz, 1993; Dubremetz and Schwartzman, 1993). The complete process is extremely fast lasting between 25 and 40 seconds (Morisaki et al., 1995).

The attachment process of the tachyzoite to the host cell and the initiation of the invasion process involves an array of receptor-ligand interactions which have partly been identified. One of the these

interactions is between parasite laminin and the laminin receptor on host cells (Furtado et al., 1992) and it was also shown that the parasite laminin interacts with the β1 integrin receptor and colagen type IV (Joiner et al., 1990). Tachyzoite binding to host cells can be blocked by the addition of monoclonal antibodies against laminin and the α6 chain of the integrin family (Furtado et al., 1992). Two further parasite proteins have been identified which interact with the host cell membrane. These are parasite lectin like molecules (Robert et al., 1991) and major parasite protein SAG-1. SAG-1 was shown to bind glycosylated host cell receptors (Mineo et al., 1993; Grimwood& Smith, 1992) and therefore is important for parasite attachment. The importance of SAG-1 was supported by data which shows that antibodies against SAG-1 but not SAG-2 block tachyzoite invasion. (Kasper &Mineo 1994).

The first set of proteins which originate from organelles and are of importance during the invasion process are the microneme proteins. Three microneme proteins have been identified in *Toxoplasma* (MCI1: 60 kDa, MCI2: 120 kDa, MCI3: 90 kDa) but their role during the invasion process still needs to be identified (Achbarou et al., 1991)

During the next stage the content of the rhoptries is released. The main components of these are lipids (Foussard et al., 1991) but at least two rhoptry proteins have been cloned and studied. The first is ROP-1 which has one acidic and one basic domain that may confer a large spectrum of binding capabilities (Schwartzman 1986; Ossorio et al., 1992). It has been suggested that this protein may facilitate parasite motility during invasion (Kasper and Mineo, 1994). The second rhoptry protein is ROP-2 which is inserted into the parasitophorous vacuole membrane and is thought to allow interactions of the parasite with the cytoplasm of the host cell (Beckers et al., 1994). It has been shown that when the parasitophorous vacuole membrane is formed it contains either channels or transporters which regulate the trafficking between the parasite vacuole and the cytoplasm (Schwab et al., 1994: Demelo et al., 1992).

Finally there is another group of proteins which play an important role during the later stages of the invasion process. These are the proteins released from the dense granules and they are thought to be

responsible for the remodelling of the parasitophorous vacuole. Five of these proteins have been cloned and sequenced but the function of all five needs to be further studied. GRA-1 is a 24 kDa soluble calcium binding protein (Cesbron-Delauw et al., 1989; Sibley et al., 1995), while the function of the second protein GRA-2 is not yet known. However, it is soluble and has a molecular mass of 28.5 kDa (Prince et al., 1989; Mercier et al., 1993). The third, GRA-3, is a soluble 30 kDa protein which is thought to interact with the parasitophorous vacuole membrane (Bermudes et al., 1994; Ossorio et al., 1994). Both the function of GRA-4 (40 kDa) and GRA-5 (21 kDa) is not known but both proteins contain a transmembrane domain which might allow interactions of the parasite with the host cell (Mevelec et al., 1992; Lecordier et al., 1993).

1.7.3. The process of host cell recognition and invasion by *Theileria* sporozoites.

Host cell recognition and invasion by *Theileria* sporozoites is only poorly understood. It became possible to study this process after the infection and transformation of host cells by *Theileria parva* sporozoites was achieved *in vitro* (Brown et al., 1971) and the first insight into the invasion process was based on observations by electron microscopy (Fawcett et al., 1982). It was shown that the interaction of the sporozoite with the membrane occurs in two steps (Webster et al., 1985) as illustrated in Figure 5. First, the sporozoite loosely binds in any orientation to the surface of the host cell. This interaction is either terminated and the sporozoite is released or the parasite becomes more closely attached, leading to invasion (Webster et al., 1985). The membrane of the sporozoite comes into very close apposition with the membrane of the host cell. 'Zippering up' of the membranes then occurs and the sporozoite sinks into the host cell until the rim of the invagination pocket closes and fuses over the parasite (Fawcett et al., 1982). This process of internalisation is very fast, taking less than three minutes (Shaw et al., 1991). The invasion process is ligand-receptor mediated and it was thought that the sporozoite enters by passive endocytosis (Fawcett et al., 1984). It has been demonstrated that a surface antigen, p67 of *T. parva*, which is thought to be a candidate for host cell recognition and invasion is shed from the sporozoite during the

invasion process (Dobbelaere et al., 1985). During the entry process of the host cell by the sporozoite the micronemes and rhoptries discharge their contents and the enveloping host cell membrane is lysed (Fawcett et al., 1984 and Shaw et al., 1991). Thus the parasite comes to lie free in the cytoplasm.

The initial findings for *T. parva* attachment and invasion were largely confirmed for *T. annulata* (Jura et al., 1983). However conflicting results were obtained for the two species regarding the energy requirements of the sporozoite during the invasion process. In contrast to Fawcett's conclusion with *T. parva*, the process of invasion by *T. annulata* sporozoites is an active process. It is initiated by the sporozoite and involves either proteins or glycoproteins as receptors on the host cell (Jura, 1984). Although it is theoretically possible that the sporozoites of *T. annulata* and *T. parva* use very different methods for host cell invasion, this is thought to be very unlikely. The issue was resolved much later when the entry of *T. parva* sporozoites into bovine lymphocytes was re-examined. It was shown that whilst the initial interaction with the host cell is a chance event independent of temperature (Shaw et al., 1991), all subsequent stages of invasion are temperature dependent, requiring the participation of live, intact sporozoites and host cells. The process involves some rearrangement of the host cell surface membrane and cytoskeleton (Shaw et al., 1991). As yet none of the receptors involved in the invasion process of *T. annulata* are known, but in *T. parva* there is evidence that the MHC class I molecules is essential in the process of sporozoite entry into bovine leucocytes (Shaw et al., 1995).

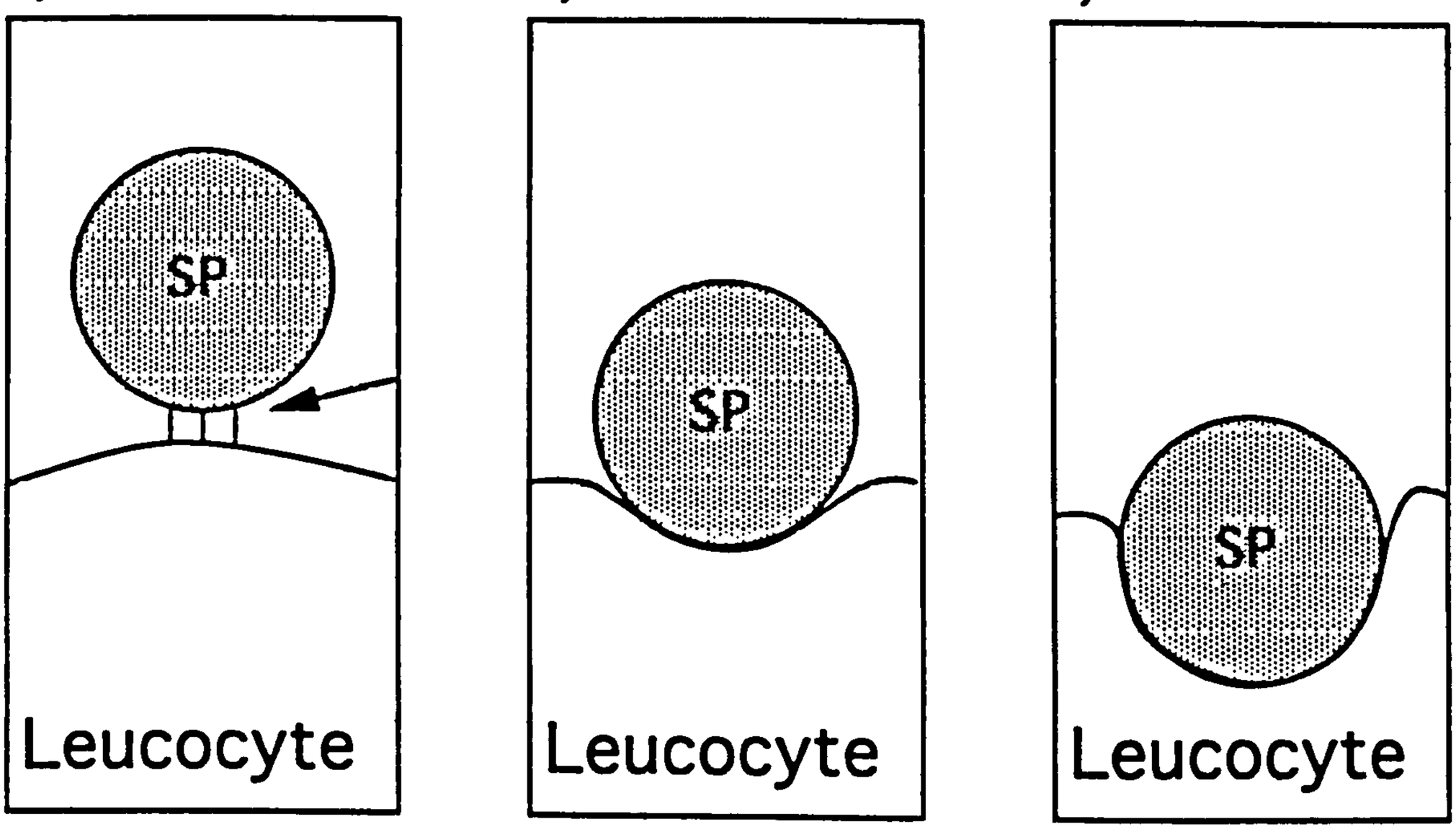


Figure 5: Diagrammatical representation of the invasion of a leucocyte by a *Theileria* sporozoite. Panel a) shows the initial loose interaction of the sporozoite surface coat with the surface of leucocytes (arrow). Panels b) and c) show the entry of the sporozoite into the leucocyte: the parasite and lymphocyte membrane are in close apposition; as the entry progresses, the sporozoite plasmalemma is 'zippering up' with the membrane of the lymphocyte.

1.7.4. Identification of target cells for invasion by *Theileria* sporozoites.

The method of *in vitro* infection and transformation of host cells by *Theileria* sporozoites (Brown et al., 1971) has also allowed us to study and characterise the target host cells for sporozoite invasion and transformation. The first indication that *T. parva* sporozoites infect a specific population of leucocytes was obtained when Stagg et al. (1981) found that the sporozoites only bind to and invade 25% of purified leucocytes. Later it was shown that T cells are the likely target for *T. parva* sporozoites since the transformed cell lines express T cell markers as defined by specific monoclonal antibodies (Pinder et al., 1981). Similar results to Stagg et al. (1981) were also observed for *T. annulata* (Jura et al., 1983). Thus *T. annulata* sporozoites commonly bind to only 10% - 20% (but occasionally up to 40%) of purified leucocytes (Jura et al., 1983). There were indications that there might be different target populations within leucocytes, since sporozoites bind to some cells only at one pole, but evenly across the surface of other cells (Jura et al., 1983). It was also shown that *Theileria* sporozoites are species-specific in the host cell invasion and transformation process. *T. parva* infects and transforms only peripheral blood leucocytes of cattle (ie *Bos taurus* and *Bos indicus*) and buffalo (*Syncerus caffer*) (Stagg et al., 1983) while *T. annulata* infects cells from a wider host range. *T. annulata* infects peripheral blood leucocytes from the Bovidae family ie. cattle (*Bos taurus*) and buffalo (both *Syncerus caffer* and *Bubalus bubalis*), goats (*Capra hircus*) and sheep (*Ovis aries*) (Steuber et al., 1986). However, *T. annulata* does not transform peripheral blood leucocytes from buffalo *in vitro* although these cells can be infected by *T. annulata* sporozoites (Steuber et al., 1986 and Chaudhri and Subramanian 1992).

The identification of cells infected by *Theileria* sporozoites has been made more difficult since it has become apparent that when the parasite transforms a leucocyte it changes the phenotype of the cell (Stagg et al., 1981). The changes include an increase in the amount of cytoplasm and alterations in the surface of the parasitised cell. The composition of the surface molecules changes so that for example, a T cell infected with *T. parva*, which did not express any MHC class II molecule prior to infection, expresses this marker once the cell is

transformed (Morrison et al., 1986 and Lalor et al., 1986). Therefore, to identify target cells for *Theileria* invasion, one cannot rely on the phenotype of the transformed cells as this might give misleading results. The target cells for *T. parva* invasion were investigated by attempts to infect subpopulations of sorted peripheral blood mononuclear cells *in vitro*. The identified target cells are T cells, B cells and Null cells while neither monocytes nor neutrophils could be infected (Baldwin et al., 1988). Other targets for *T. parva* sporozoites include eosinophils and fibroblasts (Morrison et al., 1986). It was also observed that, generally, all transformed B cells lose the expression of their surface immunoglobulin. Further, the transformed cells started to express T cell markers such as CD2 and CD8 (if they did not express them already (Baldwin et al., 1988)). The up-regulation of CD8 expression was also seen after sorted $CD4^+$ $CD8^-$ cells were infected *in vitro* (Emery et al., 1988). Conrad et al. (1989) observed that genotypically distinct *T. parva* parasites induce different levels of expression of CD6, CD8 and Null cell markers after the transformation of a cloned host cell line. It has also been observed that cells after infection by *T. annulata* have an altered membrane composition. This includes increased MHC class II expression (Glass and Spooner, 1990) and expression of parasite specific antigens (Shiels et al., 1986 and Shiels et al., 1989).

When the target cells for *T. annulata* sporozoites were compared to those of *T. parva* it became apparent that both parasites infect and transform different cell types as shown in Table 5 (Spooner et al., 1988, Glass et al., 1989 and Spooner et al., 1989). Thus *T. annulata* sporozoites preferentially infect MHC class II positive cells, whilst T cells were only infected at a very low frequency. In complete contrast, *T. parva* preferentially infects T cells and does not infect monocytes at all, although the latter are the main target for *T. annulata*. Furthermore, B cells are infected much more efficiently by *T. annulata* than *T. parva* (Glass et al., 1989 and Spooner et al., 1989). In summary the main targets for *T. annulata* infection are MHC class II positive cells, macrophages, monocytes and B cells. It is questionable whether T cells are a target for the *in vivo* infection by *T. annulata* sporozoites, since they can be infected only at a very low frequency *in vitro*, and if infected they rapidly lose their T cell markers such as CD4, CD8 (Innes et al., 1989, Spooner 1988). However, some evidence obtained from *in vivo* derived *T.*

annulata infected cell lines indicates that the sporozoites might infect a wider range of cells *in vivo*, than *in vitro*, since one such cell line expressed CD2, CD3 and the γ/δ T cell receptor indicating that this cell line is of T cell origin. But it did not express the other T cell markers CD4 or CD8 (Howard et al., 1993).

	T. annulata	*T. parva.*
MHC class II^+ cells	+++	+
MHC class II^- cells	-	+++
Monocytes	+++	---
B cells	++	+
T cells	+/-	+++
Macrophages	+	+/-
Fibroblasts	+	+

Table 5: Comparison of target cells for sporozoite invasion by *T. annulata* and *T. parva.*

1.8. Objectives of my work.

The aim of this thesis is to further the understanding of the functional and practical importance of the sporozoite antigens SPAG-1 and SPAG-2. Unfortunately not the whole SPAG-2 gene is available as discussed in section 1.4.2. and the part of the gene which had been cloned was made available only during the latter stages of my D.Phil. Therefore, most of my work concentrates on SPAG-1.

When I started my D.Phil, some preliminary evidence was shown for the existence of polymorphism in the SPAG-1 gene (Dr. R. Hall, personal communication). Therefore one of my objectives was to first confirm that SPAG-1 shows antigenic polymorphism and then to establish the degree of polymorphism. A full length cDNA sequence of SPAG-1 was available and a genomic SPAG-1 clone also existed, but only a very limited amount of sequence data was available for this clone. My primary objective was to sequence the genomic clone of SPAG-1 and to establish the level of polymorphism between the cDNA and genomic copy of SPAG-1. It was predicted that this sequence comparison might reveal important information for future vaccine development.

The next objective was to establish the degree of polymorphism of SPAG-1 at the DNA and protein levels between and within isolates. A SPAG-1 polymorphism was detected previously by RFLP analysis of piroplasm DNA digested with *Eco* RI and probed with the SR1 region of SPAG-1 (Williamson, 1988; see section 1.5.2.). To characterise this polymorphism a PCR based cloning and sequencing approach was chosen. The template DNA investigated was isolated from cloned macroschizont-infected cell lines which originated from two independent parasite isolates: one from Hisar and the other from Ankara.

SPAG-1 is a stage specifically regulated gene which is only expressed during sporozoite development (Williamson et al., 1989). A genomic SPAG-1 clone was available which contained the 5' untranslated region of approximately 1 kb. Therefore, one of my objectives was to map the mRNA initiation site in an attempt to locate the promoter region involved in the stage specific regulation of SPAG-1 transcription.

Further, attempts were made to isolate DNA binding proteins which are involved in this regulation process, as this might reveal novel transcription factors in *Theileria* as well as novel DNA binding sites.

Hall and co-workers (1992) observed three VGVAPG hexa-peptides within the predicted SPAG-1 amino acid sequence. This hexa-peptide is also found in bovine elastin and there it was identified to be the region to which the elastin receptor binds (Mecham et al., 1989). This observation and the fact that 1A7, a monoclonal antibody against SPAG-1, blocks invasion of sporozoites into host cells, led to the theory that SPAG-1 might be a ligand on the sporozoite (Hall et al., 1992). It was further proposed that SPAG-1 binds to the elastin receptor on the host cell surface during the recognition event of host cell invasion (Hall et al., 1992). Therefore, my aims were to show that recombinant SPAG-1 and subsequently SPAG-2 are ligands which are involved in host cell invasion. The approach chosen was to use recombinant parasite proteins in binding assays in an attempt to show that these two antigens bind specifically to putative host cells. Another of my objectives was to investigate the role of the elastin receptor for SPAG-1 binding and its involvement in host cell invasion.

Chapter 2

Materials and Methods

2.1. Materials.

2.1.1. Buffers.

The abbreviations given for the buffers listed in this section are used throughout the thesis. Unless otherwise stated, all buffers were made up in ultrapure water prepared by reverse osmosis and deionisation using the Purity-Labwater water purifier.

0.5 M EDTA pH 8.0

Disodium ethelenediaminetetra-acetate.$2H_20$ was made up to 0.5 M and the pH was adjusted to 8.0 with NaOH pellets.

10 x PBS (phosphate buffered saline) was as follows:

1.5 M NaCl

160 mM $Na_2HPO_4.2H_2O$

40 mM $NaH_2PO_4.H_2O$

The pH was adjusted to 7.3 with HCl.

1x TE comprises:

10 mM Tris base

1 mM EDTA, diluted from the 0.5 M EDTA pH 8.0 stock, pH was adjusted to 7.5 with HCl.

1.5 M Tris-HCl, pH 6.8

1.5 M Tris base was made up in H_20 and the pH was adjusted to 6.8 with HCl.

1 M Tris-HCl, pH 7.5

1 M Tris base was made up in H_20 and the pH was adjusted to 7.5 with HCl.

1 M Tris-HCl, pH8.0

1 M Tris base was made up in H_20 and the pH was adjusted to 8.0 with HCl.

1.5 M Tris-HCl, pH 8.8

1.5 M Tris base was made up in H_20 and the pH was adjusted to 8.8 with HCl.

Competent cell buffer I

30 mM KCOOH

50 mM $MnCl_2$

100 mM KCl

10 mM $CaCl_2.2H_2O$

15 % (v/v) glycerol

Competent cell buffer II

10 mM Na-MOPS, pH 7.0

75 mM $CaCl_2.6H_2O$

10 mM KCl

15 % (v/v) glycerol

2.1.2. Bacterial culture media.

All bacterial culture media were made up using deionised water and were sterilised by autoclaving for 20 minutes at 15 lb/square inch, 120°C. The following media were used:

Luria-Bertani (LB) medium

10 g NaCl

5 g Bacto yeast extracts

10 g Bacto-tryptone

per litre

When growing *E.coli* for the preparation of recombinant proteins, $MgSO_4$ and glucose was added to 10 mM and 0.2% respectively before use.

Ampicillin stock

100 mg ml^{-1} ampicillin (sodium salt) was made up in H_20, filter sterilized and stored at -20°C. Amplicillin was added to culture media at a final concentration of 100 µg ml^{-1} when *E. coli* harbouring recombinant plasmids carrying amplicillin resistance genes were grown.

LB agar

15 g Difco Agar was added per litre of LB medium prior to autoclaving.

TYM medium

2 % Bacto tryptone

0.5 % Yeast extracts

0.1 M NaCl

10 mM $MgSO_4.7H_20$

X-gal stock

20 mg ml^{-1} 5-bromo-4-chloro-3-indolyl-β-D-galactoside was made up in dimethylformamide and it was stored at -20°C wrapped in aluminium foil.

1 M IPTG stock

238 mg ml^{-1} isopropylthio-β-D-galactoside was made up in H_20, filter sterilized and stored at -20°C.

2.1.3. Solutions for extracting Plasmid DNA.

Solution I

50 mM glucose

25 mM Tris-HCl pH 8.0, diluted from the 1 M stock

10 mM EDTA, diluted from the 1M stock

sterilized by autoclaving.

Solution II

0.2 M NaOH

1 % SDS

filter sterilized.

Solution III

5 M potassium acetate

pH adjusted to 4.5 with CH_3COOH and sterilized by autoclaving.

Chloroform/isoamyl alcohol (IAA)

1 part IAA was equilibrated with 24 parts chloroform.

Phenol

Distilled aqua phenol (Applegen) was equilibrated with TE buffer and stored a 4°C in a dark bottle.

Phenol/chloroform

50 % equilibrated phenol and 50 % chloroform/IAA were mixed and stored at 4°C.

2.1.4. Solutions for restriction digests.

React buffer I

500 mM Tris-HCl, pH 8.0

100 mM $MgCl_2$

React buffer II

500 mM Tris-HCl, pH 8.0

100 mM $MgCl_2$

500 mM NaCl

React buffer III

500 mM Tris-HCl, pH 8.0

100 mM $MgCl_2$

1 M NaCl

React buffer IV

200 mM Tris-HCl, pH 7.4

50 mM $MgCl_2$

500 mM KCl

2.1.5. Solutions for agarose gel electrophoresis.

Ethidium bromide stock solution

This was made up in H_2O at 10 mg ml^{-1} and stored protected from light.

Low and high melting point electrophoresis agarose were purchased from BRL.

Molecular weight markers

1 kb ladder (BRL), fragment sizes from 75 bp-12.2 kb

These were made up according to manufacturer's instructions.

10 x TBE

0.9 M Tris base

0.9 M Boric acid

20 mM EDTA (diluted from the 0.5 M EDTA pH 8.0 stock solution)

This solution was made up in H_20.

2.1.6. Solutions for preparing GST fusion proteins.

MTPBS

150 mM NaCl

16 mM Na_2HPO_4

4 mM NaH_2PO_4

Sterilized by autoclaving, and stored at 4°C.

MTPBS, 1 % Triton

5 ml Triton X-100 were added to 500 ml MTPBS.

MTPBS,NaCl

3.15 M NaCl

16 mM Na_2HPO_4

4 mM NaH_2PO_4

Sterilized by autoclaving, and stored at 4°C.

1 M 1,10-Phenanthroline

1 M 1,10-phenanthroline monohydrate was made up in ethanol and stored at -20°C.

100 mM PMSF

100 mM PMSF was dissolved in DMSO and stored at 4°C.

DNase I stock

10 mg ml^{-1} DNase I was made up in H_20, filter sterilized and stored at - 20°C.

Elution buffer

50 mM Tris-HCl pH 8.0, diluted from 1 M stock

0.15 % reduced-glutathione

Made up immediately before use.

Reaction buffer for Factor Xa

20 mM Tris-HCl pH 8.0, diluted from the 1 M stock

100 mM NaCl

2 mM $CaCl_2$

Made up immediately before use.

2.1.7. Solutions for SDS polyacrylamide gel electrophoresis.

30 % Acrylamide

30 % (w/v) acrylamide was made up in H_20, deionized, filtered and stored in a dark bottle at 4°C.

1 % Bisacrylamide

1 % (w/v) bisacrylamide was made up in H_20, deionized, filtered and stored in a dark bottle at 4°C.

25 % APS

25 % (w/v) ammonium persulfate was made in H_20, and stored at -20°C.

Coomassie blue R250 stain

0.5 mg ml^{-1} Coomassie blue R250 powder was dissolved in 30 % methanol, 10 % acetic acid in H_20.

3 x Sample buffer

6 % SDS

3 % glycerol

0.05 % bromophenol blue

1.5 % 2-mercaptoethanol

187 mM Tris-HCl, pH 6.8, diluted from the 1.5 M stock solution This was made in H_20 and stored at RT.

5 x Gel running buffer

7.7 M glycine

1 M Tris base

2 % SDS

40 mM EDTA

This was made up in H_20.

Destaining solution

30 % methanol

10 % acetic acid

This was made in H_20.

Molecular weight markers

BDH high molecular weight markers were used, They have a molecular weight of 200 kDa, 116 kDa, 97 kDa, 77 kDa, 55kDa and 43 kDa. The low molecular weight markers were obtained from Sigma and they have a molecular weight of 66 kDa, 45 kDa, 36 kDa, 29 kDa, 24 kDa, 20 kDa and 14 kDa.

2.1.8. Solutions for Western blotting.

Transfer buffer

192 mM glycine

25 mM Tris base

20 % methanol

Made up immediately before use in H_20.

10 x TS

20 mM Tris-base

150 mM NaCl

10 x TS/Triton

0.5 % Triton X-100 in 10 x TS.

Blocking buffer

5 % non-fat skimmed milk powder in 1 x TS.

NBT

5 % nitro blue tetrazolium in 70 % dimethylformamide.

BCIP

5 % bromochloroindolyl in 100 % dimethylformamide.

A-P buffer

100 mM NaCl

5 mM $MgCl_2$

100 mM Tris base

pH was adjusted to 9.5 with HCl, autoclaved and stored at 4°C.

2.1.9. Solutions for biotinylation and flow cytometry.

10 % NMS

10 % normal mouse serum was made up in 1 x Dulbecco's PBS (Gibco).

Biotinylation buffer

3 mg ml^{-1} protein, which will be biotinylated

1 mg ml^{-1} NHS Biotin

10 % DMSO

0.1 M sodium hydrogen carbonate

This was made up in H_20.

FITC

The stock bought from Vector Laboratories was used at a 1/100 dilution in 1 x Dulbecco's PBS (Gibco).

Streptavidin/Phycoerythrin

10 µg ml^{-1} R-Phycoerythrin Streptavidin (Vector Laboratories) was made up in 1 x Dulbecco's PBS (Gibco).

Protein dilutions

x µg Protein

0.5% BSA

These were made in 1 x Dulbecco's PBS (Gibco) and stored at -20°C.

2.1.10. Solutions for protein iodination and binding assays.

TBS

20 mM Tris base

500 mM NaCl

This was made up in H_20 and the pH was adjusted to 7.5 with HCl.

Iodogen/chloroform

1 mg ml^{-1} iodogen (Pierce) was dissolved in chloroform.

Binding buffer

500 µg ml^{-1} Bovine serum albumin (BSA)

2 mM $MgCl_2$

2 mM $CaCl_2$

5 mM glucose

Made up in TBS, immediately before use.

2.1.13. Solutions for Southern blotting.

Depurination solution

0.25 M HCl.

Denaturing solution

1.5 M NaCl

0.5 M NaOH

Neutralizing solution

1 M Tris base

1.5 M NaCl

pH adjusted to 8.0 with HCl.

Hybridisation buffer

7 % SDS in 125 mM phosphate bufffer, pH 7.5

Wash buffer

0.2 % SSC

0.1 % SDS

2.1.14. Solutions for RNA extractions and primer extensions.

H_20, DEPC treated

0.1 % DEPC (v/v) in H_2O

This was left overnight and then autoclaved to remove remaining DEPC.

4 M Guanidium isothiocyanate

4 M guanidium isothiocyanate was made up in H_2O and stored at -20°C.

0.5 M EDTA, DEPC treated

The EDTA was made as described in 2.1.1. but before autoclaving it was treated with DEPC.

4 M NaCOOH, pH 6.0, DEPC treated

4 M NaCOOH was made up in H_20, the pH was adjusted to 6.0 with NaOH pellets, and then DEPC treated and autoclaved.

3 M NaCOOH pH 6.2, DEPC treated

3 M NaCOOH was made up in H_20, the pH was adjusted to 6.2 with NaOH pellets, then was DEPC treated and autoclaved.

10 x One Phor ALL buffer (Pharmacia)

100 mM Tris-acetate pH 7.5

100 mM Magnesium acetate

500 mM Potassium acetate

3 mM ADP

3 mM ADP was made up in DEPC treated H_20.

0.5 M PIPES, DEPC treated

0.5 M PIPES in H_2O, DEPC treated and autoclaved.

3 M NaCl, DEPC treated

3 M NaCl in H_2O, DEPC treated and autoclaved.

Hybridisation buffer

40 mM PIPES, pH 6.4, diluted from 0.5 M DEPC treated stock

1 mM EDTA pH 8.0, diluted from 0.5 M DEPC treated stock

0.4 M NaCl, diluted from the 3 M DEPC treated stock

Made up in formamide, giving a final concentration of 79 % formamide.

0.1 M DTT

0.1 M DTT was made in DEPC treated H_2O and stored at -20°C

5 x Super Script buffer (Gibco BRL)

250 mM Tris-HCl, pH 8.3

375 mM KCl

15 mM MgCl2

10 mM dNTP stock

2.5 mM dATP

2.5 mM dCTP

2.5 mM dGTP

2.5 mM dTTP

in DEPC treated H_2O, stored at -20°C.

TE, DEPC treated

This was prepared as described in 2.1.1 but before autoclaving it was DEPC treated.

RNase A (DNase free)

20 mg ml^{-1} stocks were made in H_2O, boiled for 15 minutes to destroy DNases and stored at -20°C.

1 x Loading dye (Pharmacia)

0.3 % Bromophenol blue

0.3 % Xylene cyanol FF

10 mM EDTA, pH 7.5

97.5 % deionized formamide

2.1.15. Solutions for working with bacteriophage λ.

SM phage dilution buffer

0.1 M NaCl

8 mM $MgSO_4$

50 mM Tris-HCl, pH 7.5, diluted from the 1 M stock

0.001 % gelatin

Sterilized by autoclaving.

TNE-50

10 mM Tris-HCl (pH 7.5)

50 mM NaCl

1 mM EDTA

1 mM DTT

Made up immediately just before use.

SW-block

2.5 % (w/v) dried milk powder

25 mM Hepes (pH 8.0)

1 mM DTT

10 % (v/v) glycerol

50 mM NaCl

1 mM EDTA

Made up immediately just before use.

Top agarose

0.7 % agarose in LB medium, and was sterilized by autoclaving.

2.2. Methods.

2.2.1. Tissue culture.

Culture of macroschizont infected cell lines from stocks cryopreserved in liquid nitrogen.

Vials containing cryopreserved macroschizont cell lines were thawed at 37 °C. The contents were added to 10 ml prewarmed culture medium (see section 2.1.11.) and the cells were pelleted at RT, 1100x g for 5 minutes. The cells were then resuspended in 5 ml culture medium, checked under a bifocal invertal optical microscope and incubated in a 50 cm^3 tissue culture flask in an incubator at 37 °C in 5 % CO_2. The next day another 5 ml of culture medium were added. The method for culturing macroschizonts was adapted from Brown et al. (1987). Generally the cultures were passaged three times a week by diluting the cultures down to 2 x 10^5 cells ml^{-1} with TBL medium.

Cloning of macroschizont-infected cell lines by limiting dilution.

The method for cloning macroschizont infected cell lines was adapted from Wathanga (1984) and Williamson (1988). Cells from an exponentially growing culture were diluted to an estimated 10 cells in 10 ml conditioned culture medium (see section 2.1.11.). The 10 ml of medium containing the cells was mixed and 100 µl aliquots were plated out in 96 well tissue culture plates. The plates were checked to see whether any of the wells contained more than one cell per well, and any that did were discarded. The cells were allowed to grow up and then the process was repeated. Finally, the resulting cloned macroschizont cultures were expanded up to 10 ml cultures and treated as described above.

Cell lines.

The cloned cell lines used in this thesis and their origins are listed in Table 6. All cloned cell lines with the exception of TaHBL3b were kindly given to me by Duncan Brown of the CTVM in Edinburgh. Other cell lines used were BL3 as described by Theilen et al. (1968), TaHBL3 (Baylis et al., 1992), TaH BL20 and TaH 46 (Shiels et al., 1986).

Table 6 Origins of cloned macroschizont cell lines

Cloned macroschizont cell line	Origin of cell line
TaA 46A	Wathanga (1984)
TaA 139D4	Wilkie (pers. comm.) / Williamson (1988)
TaA 139D7	Wilkie (pers. comm.)
TaA 139E3	Wilkie (pers. comm.)
TaA 139E5	Wilkie (pers. comm.) / Williamson (1988)
TaA 46.2	Wathanga (1984)
TaH 46.2	Wathanga (1984)
TaH 46.3	Wathanga (1984)
TaH 46.4	Wathanga (1984)
TaA 139D6	Wilkie (pers. comm.)
TaA 46.3	Wathanga (1984)
TaHBL3b	Katzer

2.2.2. Preparation of DNA.

Plasmid mini preparation.

This method was taken from Sambrook et al. (1989). 5ml bacterial overnight cultures were spun down for 10 minutes at 1000 g. The resulting cell pellet was resuspended in 100 µl of cold solution I (see section 2.1.3). 200 µl of freshly made solution II (see section 2.1.3) was added and the solution was mixed, before 150 µl of solution III (see section 2.1.3) was added. This was mixed again and spun down (10,000x g

for 10 minutes). The DNA was extracted from the supernatant by phenol and phenol/chloroform extraction and was precipitated with ethanol. The DNA pellet was resuspended in TE.

Wizard minipreps.

The Wizard DNA Miniprep kit was purchased from Promega and the DNA extraction was conducted according to the manufacturer's instructions. For this method a 1.5 ml bacterial overnight culture was used to extract DNA, The cells were lysed using the buffers provided and the DNA was bound to a DNA binding resin, provided by the kit. The resin was purified on Wizard Miniprep columns and the DNA was eluted with TE which had been heated to 55°C.

Preparation of single stranded M13.

An overnight *E. coli* culture was diluted (1/100) with fresh 2 x TY medium (1.5 ml - 5 ml total volume). This culture was inoculated with a single plaque of bacteriophage M13 using a sterile wooden cocktail stick. This culture was grown for 5 hours at 37 °C. After the incubation 1.5 ml of the culture were spun down in an Eppendorf centrifuge (10,000x g for 10 minutes) and the supernatant was recentrifuged in order to pellet all bacterial cells. 200 µl of 20 % PEG/2.5 M NaCl was added to the supernatant, mixed and incubated for 15 minutes at RT to precipitate the M13 phage. The phage were spun down by centrifugation. Single stranded DNA was isolated by phenol/chloroform extraction, ethanol precipitated and resuspended in 20 µl TE.

Genomic DNA preparations.

For DNA preparations, macroschizont cultures were grown up in 100 ml of TBL medium (see 2.1.11.) to a cell density of 2.0 x 10^6 cells per ml (2 x 10^8 cells in total). The cells were spun down for 10 minutes at 500x g, washed twice in PBS and once in 1 x SSC. The cells were spun down again and resuspended in 4 ml 1 x SSC, 4 ml TNE and 1 ml 10 % sarkosyl. The suspension was mixed gently, Proteinase K was added to a final concentration of 100 ng/µl and was incubated for 2 hours at 55°C. Following this incubation, the lysate was extracted once with phenol,

once with phenol/chloroform and once with chloroform, in order to remove any protein and other contaminants. The aqueous layer was dialysed against several changes of TE buffer at 4°C. Finally, the DNA concentration was estimated by measuring the optical density (OD) at 260 nm.

Piroplasm DNA.

T. annulata piroplasm DNA extracted from cultures which originated from field isolates from Ankara, Turkey (Schein et al., 1975) and Hisar, Morocco (Gill et al., 1976) was kindly given to me by Dr. R. Hall.

2.2.3. Electrophoresis and Southern blotting.

Restriction digests.

All restriction digests were carried out using the manufacturer's restriction enzymes (BRL), the appropriate buffer and recommended temperature. Plasmid DNA was digested for an average of two hours while genomic DNA was left to digest overnight.

Agarose gel electrophoresis.

DNA digested by restriction endonucleases was analysed by agarose gel electrophoresis as described by Sambrook et al. (1989) in TBE buffer (see section 2.1.5). Plasmid DNA was electrophoresed for a period of 1-4 hours, while cut genomic DNA was run out at a low voltage over night. Generally 1 µg plasmid DNA, 8-10 µg macroschizont cell line DNA or 2 µg piroplasm DNA was loaded per track. Molecular weight markers, described in section 2.1.5, were also loaded to allow the estimation of size of the DNA fragments. The DNA samples were diluted in type III gel loading buffer (Sambrook et al., 1989). Ethidium bromide at 1 µg ml^{-1}, was included in the gel so that the DNA could be visualised on a UV transiluminator.

Recovery of DNA fragments from low melting point agarose gels.

DNA fragments were run out on agarose gels made with low melting point agarose. The required fragment was cut out of the gel and the agarose was melted with 3 volumes of H_20 at 65 ^oC. This was then phenol/chloroform extracted until all the agarose had been removed. Finally the DNA fragment was precipitated with ethanol.

DNA recovery from agarose gels via Geneclean.

The Geneclean II kit, purchased from BIO 101, was also used, following the manufacturer's instructions, to purify DNA fragments from agarose gels. In this method the band was exised from an agarose gel, 3 volumes of NaI stock solution was added and the agarose was dissolved at 50^oC for up to 1 hour. Glassmilk was then added to the suspension and incubated, with occasional mixing, at RT for up to another hour allowing the DNA to bind to the glassmilk. The glassmilk/DNA complex was washed three times with New wash and finally the DNA was eluted from the glassmilk using TE (see section 2.1.1), warmed to 50^oC.

Random prime-labelling.

DNA fragments for making labelled probes were cut out of low melting point agarose gels, melted in 3 volumes of H_20 and made single stranded by boiling for 5 minutes. Approximately 25 ng of DNA were labelled with 50 μCi $\alpha^{32}P$ dCTP using the Random Priming DNA Labelling Kit (Boehringer-Mannheim Pharmaceuticals) according to the manufacturer's instructions. After labelling, the reaction mix was run through a spin column, as described by Sambrook et al. (1989), to separate the labelled probe from unincorporated radioactive nucleotides. Before use, the probe was boiled at 95^oC for 5 minutes to make the DNA single stranded.

Southern blotting.

This method was adapted from the standard Southern blotting method (Sambrook et al., 1989 and Southern, 1975). After electrophoresis the gel was incubated for 15 minutes in the depurination solution (see section 2.1.13.), followed by 20 minutes in denaturing solution (see section 2.1.13.) and finally 30 minutes in neutralizing solution (see section 2.1.13.). After a brief wash in H_2O the DNA was transferred from the gel onto a nylon membrane (Hybond N, Amersham) overnight as described by Southern (1975). The membrane was washed in 2 x SSC and then the DNA was fixed to the membrane by UV cross-linking for 7 minutes.

Once the DNA was fixed to the membrane, the membrane was prehybridised in hybridisation buffer (see section 2.1.13.) for 1 hour at 65°C. The probe prepared as discussed previously, was boiled and then added to the hybridisation buffer. This was followed by overnight incubation at 65°C with constant agitation. The membrane was then washed once with wash buffer (see section 2.1.13.) at RT, followed by 2-3 further washes at 65°C. The excess fluid was removed and the membrane was exposed to Kodak X-OMAT S film for several days in a film cassette containing an intensifying screen at -70°C. Finally the film was developed in a Kodak X-OMAT Developer.

2.2.4. Molecular cloning techniques.

Competent cells.

20 ml of TYM broth were inoculated with the *E. coli* strain tg1 recO and the bacteria were grown overnight. The culture was then diluted with 100 ml of warm TYM broth and the cells were grown until an OD of 0.5-0.9 at 600 nm. The culture was diluted again with 500 ml of warm TYM and was grown until an OD of 0.6 at 600 nm. The culture was then rapidly chilled and the bacteria were harvested by centrifugation. The cells were resuspended in 100 ml of cold competent cell buffer I and were left for 5 minutes. The cells were collected again by centrifugation and

resuspended in 20 ml of cold competent cell buffer II. Aliquots of 0.5 ml were pipetted into pre-chilled microfuge tubes and frozen in liquid nitrogen. These aliquots were stored for up to 2 month at -70°C.

Oligonucleotide primers.

Table 7 shows all oligonucleotide primers used both for PCR cloning and generation of PCR fragments as well as oligonucleotide primers used for sequencing.

Name of primer	Sequence of primer 5' to 3'	Origin of Sequence	Position on Sequence
420	ctgccaattcttccggtttg	SPAG-1, cDNA	1919-1900
646	ggagtagacttggcctagg	SPAG-1, cDNA	344-326
647	cttgctgaggatcctcctcc	SPAG-1, cDNA	1158-1139
710	ccaaggaagccatggtacttc	SPAG-1, cDNA	1450-1470
711	gaagaagttttgaaagttttgg	SPAG-1, cDNA	2667-2688
815	ggtctacagggaccaggagg	SPAG-1, cDNA	784-804
932	gtggactatgctgaatatgatt	SPAG-1, cDNA	2722-2700
FK1	gtataatattcatcgttgagtgttgg	SPAG-1, cDNA	-13 to +13
pGEX For	gcatggcctttgcaggg	pGEX Vector	855-871
pGEX Rev	gctgcatgtgtcagaggttttcaccg	pGEX Vector	1014-988
Sp6	gatttaggtgacactatag	pGEM-T	143-126
T7	taatacgactcactataggg	pGEM-T	2987-3
Y2	gtgtaccaggaggaaaggc	SPAG-1, cDNA	945-963
Y3	gcggacaagatgcctgcggg	SPAG-1, cDNA	53-83
Y11	ccagatacaaaaatgaccgg	SPAG-1, cDNA	31-33
Y50	aatattatctcaaaacgagtgtg	SPAG-1, gDNA	-371 to -347
Y51	cagctgtataagaccgtgtttagtc	SPAG-1, gDNA	-241 to -263
Y82	ggatcctcgagagatagcgccaag	SPAG-2, cDNA	4-27
Y83	gaattctactatccgaagttccctg	SPAG-2, cDNA	367-353
Y84	gaattccaaattggctccccttgttg	SPAG-2, cDNA	667-651

Table 7: The oligonucleotide primers used for sequencing and cloning. The name of the primers, their sequence, the origin of their sequence and the position they map to are listed.

Transformation.

In general 1/3 of a ligation mix was used to transform 200 µl of competent *E. coli* strain tg1 recO. The ligation mix and freshly thawed cells were mixed and incubated on ice for 30 minutes. Then 400 µl of LB was added to the suspension, mixed and the cells heat shocked for 2 minutes at 42°C. After heat-shocking, the cells were incubated for 1 hour at 37 °C and aliquots of 40 µl, 75 µl, 150 µl and 300 µl were plated onto LB agar plates containing ampicillin.

DNA amplification by polymerase chain reaction (PCR).

The primers used for PCR are listed in the above table. In all cases the PCR reaction mix was made up to 50 µl containing 50 µmoles of each primer, 100 ng genomic DNA, 10 µM dNTP, 3 units of Taq XL and 1 x Taq Buffer (Northumbria Biologicals Ltd). The reaction mix was incubated at 96°C for 5 minutes and this was followed by 30 cycles of 1 minute at 94°C, 1 min at 60°C and 2 minutes at 70°C after which the reaction was left at 70°C for 10 minutes. The PCR products were separated by agarose gel electrophoresis, from which fragments were purified using the Geneclean method (see above for method).

PCR cloning into pGEM-T.

The PCR products were electrophoresed on agarose gels, the required DNA fragments were cut out, purified via Geneclean and cloned into pGEM-T (Promega) following the manufacturer's instructions.

PCR cloning into pGEX-2T.

PCR fragments were first cloned into pGEM-T (as above) and were sequenced to check for PCR errors. Correct fragments were cut out of pGEM-T using convenient *Bam* HI and *Eco* RI restriction sites. The DNA fragment was purified from an agarose gel using Geneclean (as above). The purified fragment was ligated overnight into a *Bam* HI / *Eco* RI digested pGEX-2T vector with a 3:1 insert to vector ratio. 1/3 of the ligation mix was used to transform *E. coli*.

2.2.5. DNA sequencing.

T7 DNA polymerase sequencing of double-stranded and single-stranded DNA.

Sequencing was carried out using the dideoxy chain termination method of Sanger and co-workers (1977). A T7 DNA sequencing kit (Pharmacia) was used for all sequencing reactions. Double stranded DNA was prepared using the Wizard Miniprep method (see section 2.2.2.), denatured with NaOH and ethanol precipitated. The primer was annealed to the denatured DNA at 37°C. Single stranded DNA was prepared from M13 as discussed in section 2.2.2. and the primer was annealed to the DNA at 65°C. The remaining sequencing reactions were conducted according to the manufacturer's instructions using ^{35}S dATP. 2.5 µl of the reaction was loaded onto 6 % acrylamide, 0.5 x TBE non-gradient gels (Sambrook et al., 1989).

The gels were fixed in 20 % ethanol, 20 % acetic acid for 30 minutes before being dried down onto 3MM paper at 80 °C for 2 hours. Finally the gels were exposed to Kodak X-OMAT S film in a film cassette at RT for 24 hours.

2.2.6. Preparation of recombinant proteins.

Expression of GST-fusion proteins.

100 ml overnight cultures of *E. coli* tg1recO containing pGEX plasmids were grown up in LB medium (see section 2.1.2.) and diluted by adding 1 litre of prewarmed LB. The cells were allowed to grow for 2 hours before protein expression was induced by adding IPTG to a final concentration of 1µM. The culture was then grown for another 2 hours at 37°C. The cells were then collected by centrifugation, resuspended in MTPBS/Triton (see section 2.1.6.), containing 10 mM 1,10-phenanthroline, 10 mM PMSF and 10 µg DNase I. The cells were lysed by sonication and insoluble proteins and cell debris were removed by

centrifugation. The supernatant was passed through a glutathione 4B Sephadex column (Pharmacia) in order to affinity purify the fusion protein. The column was washed several times with MTPBS/Triton and with MTPBS. Finally the bound fusion protein was eluted with the elution buffer (see section 2.1.6.). The eluted protein was collected in 1 ml fractions and subsequently visualised on SDS-PAGE (using a method adapted from Laemmli (1970)). The fractions containing protein were pooled and the protein concentration was estimated via the Bradford assay.

Bradford assay.

The protein concentration was estimated via the Bradford assay (Bradford, 1976), by mixing diluted protein, as well as known bovine serum albumin (BSA) standards, with Bradford reagent. These solutions were kept at RT for 15 minutes and then their OD was measured at 595 nm. The readings of the protein standards were converted into a standard curve which was used to estimate the protein concentration of the sample in question.

Thrombin cleavage.

After the protein concentration was estimated, the fusion protein was cleaved using thrombin to separate the parasite protein from its GST fusion partner. The protein was dialysed against 1 x PBS (see 2.1.1.) containing 2.5 mM $CaCl_2$. 2 µg thrombin (Sigma) was added for every mg of fusion protein (Knight, 1993). The reaction mix was incubated for 2 hours at RT. The cleaved protein was then purified by running the reaction mix through a glutathione 4B sepharose column (Pharmacia) to bind the GST and the run-through was collected. The run-through was poured over the same regenerated column twice more, to make sure that all the uncleaved protein, as well as the GST, was removed. The cleaved recombinant protein was then requantified and dialysed against the required buffer for biotinylation or iodination as necessary.

Factor Xa cleavage

For Factor Xa cleavage the GST fusion protein was dialysed against the appropriate reaction buffer (see section 2.1.6.). 5-10 mg of fusion protein were incubated with 50 µg activated Factor X for up to 12 hours at RT. The cleaved protein was purified using the same method as described in Thrombin cleavage (see above).

2.2.7. Protein analysis.

SDS-polyacrylamide gel electrophorsis.

The method for SDS-PAGE was adapted from Laemmli (1970). The mini Protean II gel apparatus from BioRad was used. A resolving gel was poured first consisting generally of 10 % acrylamide and 0.13 % bisacrylamide in a 3.75 M Tris-HCl pH 8.8 buffer and polymerized with 0.05 % TEMED and 0.05 % APS. The stacking gel consisted of 5 % acrylamide, 0.135 % bisacrylamide, 125 mM Tris-HCl pH 6.8, polymerised with 0.05 %APS and 0.1 % TEMED. Protein samples were diluted in Loading dye (see 2.1.7.) and before application to the gel, and were heated for 5 minutes at 95°C. Generally, the gels were run at 200V for a period of about 45 minutes. The gels could then either be stained in Coomassie blue stain and destained in destaining solution (see 2.1.7.), or were used for Western blotting.

Western blotting.

The method for Western blotting was adapted from Hunt and Hall (1993) and Towbin et al. (1979). Protein bands (0.5-1 µg) from SDS-PAGE gels were transferred onto nitrocellulose membrane (Sartorius) using a BioRad Protean II Western blotting apparatus. The SDS-PAGE gel was placed on the nitrocellulose membrane and placed in the Western blotting tank with transfer buffer (see section 2.1.8.). The protein was usually transfered for 75 minutes at 150 mA. The membrane was blocked in blocking buffer (see section 2.1.8.) for 30 minutes and was then incubated for 1 hour in primary antibody diluted in blocking buffer. The

antibody dilution chosen was according to Knight (1983). After this incubation the membrane was washed once in 1 x TS/Triton and three times in 1 x TS. The membrane was then incubated in the secondary antibody diluted in blocking buffer for a further hour. The membrane was then washed as before and bands reacting with the antibodies were detected using in A-P buffer with 0.06 % NBT and 0.03 % BCIP (see section 2.1.8.).

2.2.8. Flow cytometry.

Preparation of PBM cells.

Bovine peripheral blood mononuclear cells (PBM) were prepared from fresh heparinised blood from calves. 10 ml of blood was mixed with an equal volume of Dulbecco's PBS and placed on a 10 ml histopaque 1086 (Sigma) bed. This was centrifuged at 1500x g for 30 minutes at RT to separate the PBM cells. After the spin, the PBM were pipetted off the histopaque/PBS interphase. They were then washed three times in Dulbecco's PBS and counted using a haemocytometer.

Single colour Flow cytometry.

PBM cells were prepared as above and 2 x 10^5 cells were used per reaction. The cells were plated out in 96 well microtitre plates and were pelleted by centrifugation for 2 minutes. The cells were then resuspended in a volume of 25 µl containing either a diluted biotinylated antibody or a biotinylated protein. They were incubated at 4°C for 1 hour, washed three times in Dulbecco's PBS, and incubated in 25 µl of 1/100 diluted phycoerythrin streptavidin for 15 minutes. The cells were then washed twice in Dulbecco's PBS and analysed using a Becton and Dickinson FACScan flow cytometer.

Two colour Flow cytometry.

The method for 2-colour flow cytometry was adapted from Sopp et al. (1991). PBM cells were prepared as above and 2 x 10^5 cells were used per reaction. The cells were plated out in 96 well microtitre plates and were pelleted by centrifugation for 2 minutes. The cells were resuspended and incubated in 25µl of diluted monoclonal antibodies for 30 minutes at RT. The cells were washed 3 times in Dulbecco's PBS and incubated for 30 minutes in 1/100 diluted FITC-conjugated secondary antibody. After this incubation the cells were washed again in Dulbecco's PBS and incubated in 1 % normal mouse serum for 15 minutes. This was followed by another wash and the cells were then incubated in either biotinylated proteins or biotinylated antibodies (30 minutes at RT). The cells were washed three times again and incubated in 1/100 diluted phycoerythrin streptavidin for 15 minutes. Then they were washed once more and analysed using a Becton and Dickinson FACScan flow cytometer.

2.2.9. Iodination and ^{125}I labelled protein binding assay.

Iodination of proteins.

This method was adapted from Sambrook et al. (1989) and Wrenn et al. (1988). 20 µl of 1 mg/ml iodogen (Pierce) in chloroform was pipetted into an eppendorf microcentrifuge tube which had been washed previously in ethanol. The iodogen was dried in an air stream. In a well ventilated fume hood, 40 µl of protein at a concentration of 1 mg/ml was added to this iodogen covered tube, together with 10 µl 0.3 M phosphate buffer, pH 6.8, and 500 µCi of Na-I^{125}. This was incubated for 10 minutes at RT. Then 750 µl of TBS was added to the reaction and the mixture was loaded on an equilibrated PD-10 G-25 (Sephadex) column (Pharmacia). The eppendorf tube was rinsed once more with 750 µl TBS which was also added to the column. The column was allowed to run dry before 5 ml of TBS was added to it. Fractions of 1 ml were collected and checked for radiation using a Geiger counter. The fractions containing most

radiation were pooled and their protein content as well as the success of the iodination reaction were tested by running a fraction of the sample onto a SDS PAGE gel.

Binding assay.

BL3 cells were grown in 100 ml cultures which were kept growing exponentially. These cells were harvested by centrifugation and were washed three times in Dulbecco's PBS. The cell number was estimated by counting a proportion on a haemocytometer. On average 4×10^5 cells were used per reaction. These were incubated with a dilution series of iodinated protein. Each reaction with a given protein concentration was performed in triplicate and another identical triplicate was set up with a fifty-fold excess of unlabelled protein as a competitor. The cells were incubated for 1 hour at RT. The cells were then washed three times with TBS and tranferred to a fresh eppendorf tube. The cells were then spun down and a count of emitting gamma radiation was taken using a γ radiation counter.

2.2.10. Sporozoite RNA extraction.

Extraction of RNA from infected-tick salivary glands.

The salivary glands of *T. annulata* infected ticks, which had been fed for two days on the ears of laboratory rabbits, were dissected by Lesley Bell-Sakyi. These salivary glands were homogenised in a sterile homogeniser containing 4 ml of 4 M Guanidium isothiocynate (see section 2.1.14.) and ethanol precipitated. The RNA, DNA and protein was pelleted by centrifugation, and the pellet was dissolved in 5 ml of 50 mM EDTA (DEPC treated). This was followed by a phenol/chloroform extraction including 2 back extractions with 2.5 ml 50 mM EDTA (DEPC treated). 30 ml of 4 M NaCOOH pH 6.0 was added and the solution was incubated overnight at 4^oC, to precipitate the RNA. The RNA was pelleted by centrifugation and the very small pellet was dried and resuspended in 100 µl H_2O (DEPC treated). The RNA was then quantified as described below.

RNA Quantification.

The RNA was diluted 1/100 in DEPC-treated H_2O and OD readings were taken against a blank containing DEPC-treated H_2O at 260 and 280 nm. The RNA concentration was calculated knowing that 1 OD unit at 260 nm equals 40 µg ml^{-1} RNA and the purity of the RNA sample was determined by the 260/280 nm ratio.

2.2.11. S1 intron mapping.

The *Bam* HI / *Acc* I fragment of the gH3.4 allele (position 1201-2179 on the genomic SPAG-1 sequence) was cloned into M13 mp18 and M13 mp19. The two resulting vectors were grown up (as described in section 2.2.2.) with the addition of 120 µl inorganic ^{32}phosphate (10mCi/ml orthophosphoric acid). The resulting radioactively labelled single-stranded M13 DNA was extracted as described in section 2.2.2. 50 µg of RNA, extracted from infected-tick salivary glands (see section 2.2.10.) was added to 2 µl of the single-stranded M13 DNA. The RNA and the DNA were denatured in formamide and resuspended in hybridisation buffer and annealed at 50 °C for 14 hours. The RNA-DNA hybrid was ethanol precipitated and resuspended in S1-buffer. 600 units of S1 nuclease were added to the suspension and incubated for 45 minutes at 37 °C to digest any single stranded DNA or RNA. After this incubation the suspension was phenol/chloroform extracted and the resulting RNA-DNA hybrid was ethanol precipitated and resuspended in loading dye. The product was denatured by heating to 95°C for 5 minutes and run out on a 6 % sequencing gel and visualised by autoradiography.

2.2.12. Primer extension.

Endlabelling of primer.

This method has been adapted from Sambrook et al. (1989). 0.5 µg of double stranded oligonucleotide were labelled in 1 x One-Phor-All Buffer in the presence of 10 units of T4 polynucleotide kinase, 300 µM ADP and 50 µCi γ-^{32}P ATP, in a final volume of 10 µl. This reaction mix

was incubated for 1 hour at 37°C after which the unincorporated radioactive nucleotides were separated from the labelled primer using sephadex spin columns (Sambrook et al., 1989).

Primer extension.

This method was also adapted from Sambrook et al. (1989). 30 µg of infected tick salivary gland RNA and 1% of the endlabelled oligonucleotide were precipitated together in the presence of ethanol and sodium acetate. The pelleted RNA and primer were dried briefly and resuspended in hybridisation buffer (see secetion 2.1.14.). The hybridisation buffer was heated to 85°C for 10 minutes and the primer was annealed to the RNA overnight at 37°C. The annealed primer and RNA were precipitated in ethanol, and then the primer was extended using reverse transcriptase in the appropriate buffer in the presence of RNA Guard, 1 mM dNTP and 10 mM DTT. The remaining RNA was digested with RNase I which was DNase free. The solution was then phenol/ chloroform extracted and the DNA was precipitated in ethanol. The DNA was collected by centrifugation, air dried and resuspended in 10 µl of 1 x loading dye. Half of the extended primer was run out on a 6 % sequencing gel.

2.2.12. λgt11 library screening.

Plating out bacteriophage λ.

Bacterial strain Y1090R$^-$ was grown up overnight in LB medium containing 0.2 % maltose, 10 mM $MgSO_4$ and 50 µg ml^{-1} ampicillin. The bacteria were harvested by centrifugation and resuspended in 1/4 of the volume of the overnight culture of 10 mM $MgSO_4$. The bacteria could then be stored at 4°C for up to a week. LB plates containing ampicillin (50µg ml^{-1}) were dried and prewarmed to 37°C. Top agarose was melted and cooled to 42°C. Plating bacteria and phage were mixed and incubated for 20 minutes at 37°C. Then the phage and bacteria were mixed with 3 ml of top agar and poured onto the LB plates. The plates were incubated overnight at 42°C.

Preparation of filters for library screening.

Plating bacteria were prepared as above. 21 cm x 21 cm sterile plates containing 250 ml LB agar containing ampicillin were used for library screening. It was aimed to plate out 5 x 10^4 phage per plate with 1.5 ml of plating cells. These were incubated for 20 minutes at 37°C and were plated out with 30 ml of top agarose. The plates were left at RT for 5 minutes to allow the top agar to set and were then incubated for 3 to 4 hours at 42°C. In the meantime, nitrocellulose membrane (Sartorius), measuring 20 cm x 20 cm, was soaked in 10 mM IPTG for 30 minutes and then dried. Once the phage plaques became visible on the LB plate, the membrane was placed on top of the top agarose. The phage were then left to grow up overnight. The following morning the filter was removed from the plate and was ready to be screened.

Filter screening for DNA binding proteins.

The method for screening the filters was adapted from Latchman (1993) and Singh et al. (1990). The filter was blocked in 500 ml of SW-block for 1 hour. Then the filter was incubated with 100 ml of TNE-50 containing an endlabelled double stranded DNA probe and non-specific competitor DNA. The filter was incubated at RT for 1 hour and was then washed in 3 to 4 washes of TNE-50, after which the filter was exposed to X-Ray film.

Chapter 3

Sequence comparison of SPAG-1 alleles: practical and functional importance.

3.1 Introduction.

It has been proposed that SPAG-1 might function in evasion of the host's immune system and as a ligand for host-cell recognition (Hall et al., 1992). Due to selection pressure, sequences specifying structures involved in invasion of host cells by parasites will tend to be conserved. Structures mediating immune evasion could also be conserved in some situations e.g. where they are involved in molecular mimicry, or at the other extreme selection may produce extreme antigenic polymorphism. When such polymorphism occurs in immunologically relevant antigens, it poses a serious problem for the development of vaccines based on these molecules. This situation is widespread amongst infectious agents (Mendis et al., 1991). With the above considerations in mind I have obtained and compared the sequences of SPAG-1 alleles and identified polymorphic and conserved regions. The implications of the results of this analysis for the function of SPAG-1 and for vaccine development are discussed in this chapter.

3.1.1. The sequence analysis of SPAG-1 with respect to functional importance.

The regions of SPAG-1 which may be involved in the invasion process of host cells, and those that are required for maintaining the conformation, and processing of the protein would be expected to be conserved among *T. annulata* strains. Other regions which are involved in the expression of the molecule on the surface of the sporozoite and its membrane anchor regions are expected to be conserved between species in related proteins, such as p67 of *T. parva* (Nene et al., 1992). Therefore I decided to study SPAG-1 sequences from different isolates in order to identify conserved regions of this molecule. The sequences of these

regions, if homologous to previously described motifs, might help to elucidate their function and that of the SPAG-1 protein. These results might provide further support for the theory that SPAG-1 is involved in the process of host cell recognition and invasion. Nevertheless these results will indicate a) whether more than one SPAG-1 allele exists, b) the extent of the SPAG-1 polymorphism and c) which parts of the protein are most conserved.

3.1.2. Sequence analysis of SPAG-1 with respect to vaccine development.

The identification of variable and constant parts of SPAG-1 will be of importance for the future development of sub-unit vaccines against *T. annulata* because SPAG-1 is a candidate for inclusion in a sub-unit vaccine against tropical theileriosis. Determining the variablility of particular sequences will permit the inclusion of the optimal recombinant SPAG-1 components, so that a future vaccine could consist of constant, cross-protective regions of the protein and possibly a cocktail of all identified polymorphic regions. Although a single variant of one of the polymorphic regions will not induce cross-protective immune responses, it might nonetheless contain essential B cell and/or T cell epitopes for the induction of strain specific immune responses.

3.2 Results

3.2.1. Allelic RFLP analysis demonstrates that SPAG-1 is a single copy gene.

Previous data demonstrated that the SR1 fragment hybridised to 3 *Eco*RI restriction fragments (3.4, 4.8 and 6.0 kb) in piroplasm DNA (Williamson et al., 1989). These data are confirmed and extended in the Southern blots shown in Figure 6. It can be seen that the piroplasm DNA of Ankara origin displays three bands, one at 3.4-kb, one at 4.8-kb and one at 6.0-kb (Figure 6, lanes 1 and 18). The DNA of Hisar origin shows

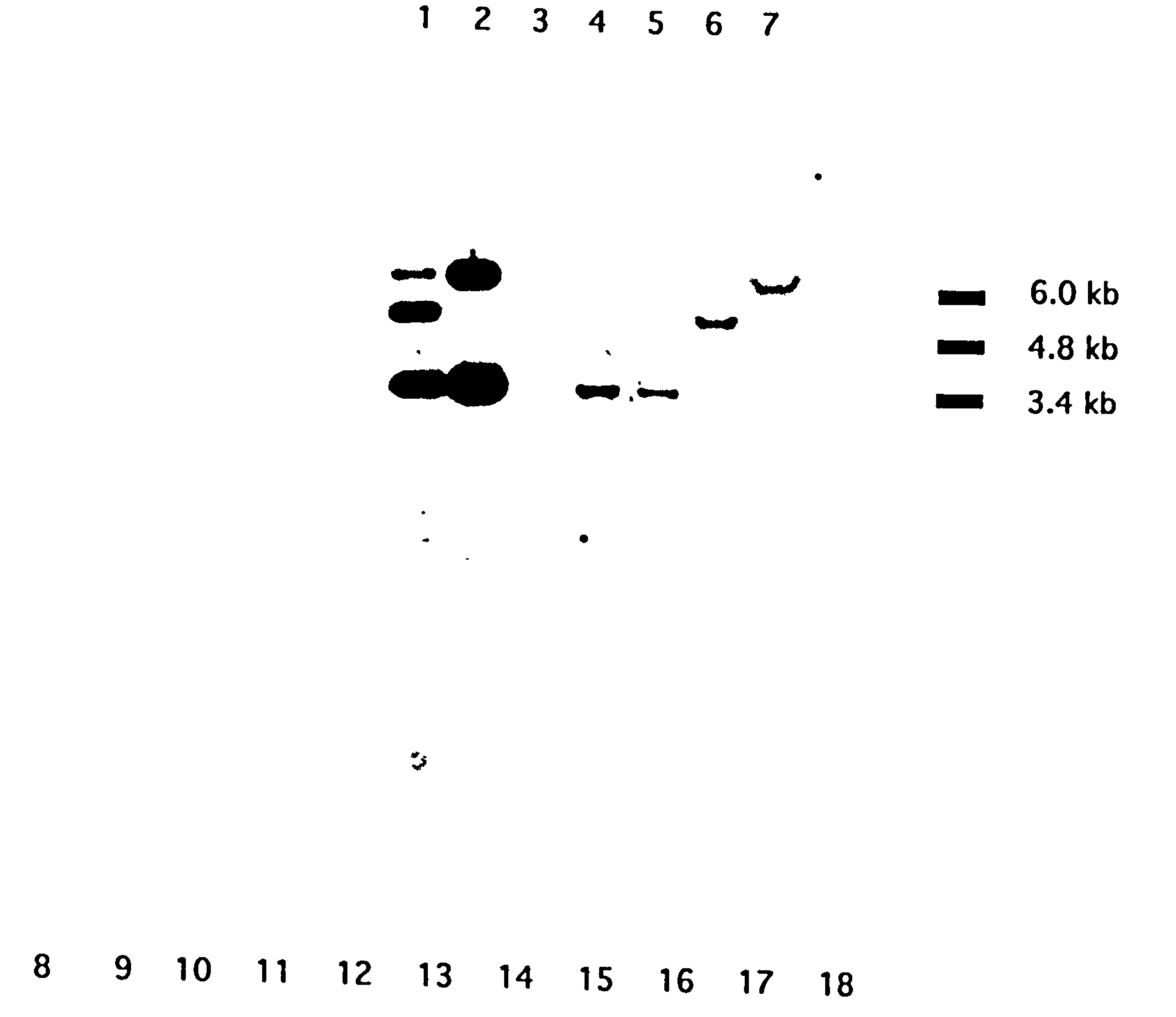


Figure 6: Southern blot analysis of the SPAG-1 associated RFLPs. Genomic DNA was digested with *Eco* RI and the blot was probed with the SPAG-1 insert derived from λgt11-SR1 (Williamson et al., 1989). Lane 3 contains 15 µg DNA from uninfected BL3 cells, lanes 1 and 18 contain 2 µg and 20 µg piroplasm DNA extracted from Ta Ankara, respectively and lanes 2 and 17 contain an equal amount of Ta Hisar stocks piroplasm DNA. Lanes 4-16 contain 15 µg of DNA extracted from the following cell lines: TaA 139D4 (lane 4), TaH 46.2 (lane 5), TaA 139D6 (lane 6), TaHBL3b (lane 7), TaA 46A (lane 8), TaA 139D7 (lane 9), TaA 139E3 (lane 10), TaA 46.2 (lane 11),TaA 139E5 (lane 12), TaA 46.3 (lane 13), TaHBL3 (lane 14), TaHBL20 (lane 15) and TaH46 (lane 16). The bars mark the segregating RFLPs and give their sizes in kb.

only 2 bands, one at 3.4-kb and one at 6.0-kb (Figure 6, lanes 2 and 17) whereas Williamson et al. (1989) found that the 4.8 kb band was weakly visible in the Hisar piroplasm DNA. The presence of the 4.8-kb fragment appears to fluctuate during passage of the Hisar stock. This observation alone suggests that the multiple bands result from mixtures of distinct parasite stocks.

To ascertain whether SPAG-1 is encoded by a single copy gene, an RFLP analysis was performed on DNA extracted from cloned macroschizont-infected cell lines (Table 8). In preliminary experiments I found that none of the available 11 clones contained the 6.0 kb *Eco*R I RFLP. I therefore isolated a twelfth clone called TaHBL3b from the TaHBL3 line (Table 8, see Materials and Methods for details). The extracted DNA was digested with *Eco* RI and subjected to Southern blotting with the SR1 probe. The result of this analysis is shown in Figure 6. Uninfected BL3, a bovine negative control (lane 3), shows, as expected, no band which indicates that the SR1 probe does not hybridise to bovine DNA. Lanes 4 to 13 contain the DNA of different cloned macroschizont-infected cell lines and in each track only a single band is visible. Each band co-migrates with one of the RFLPs found in the piroplasm DNA and each of the RFLP types is represented i.e. 3.4 kb in tracks 4, 5, 8-10, 12, 13; 4.8 kb in tracks 6 and 11; and 6.0 kb in track 7. Thus these RFLPs segregate clonally and this provides strong evidence that the SPAG-1 gene exists as a single copy. Each of the three restriction fragments behaves as alternative forms of a gene at a single locus and therefore I will henceforth adopt the term allele. The different alleles are named according to their isolate name and the restriction fragment size. Thus the alleles from the Hisar and Ankara isolates are prefixed by an H and an A respectively e.g. H3.4 is the Hisar allele marked by the 3.4 kb *Eco*RI fragment. A summary of the 12 cloned lines, their RFLP type and their allele designation is provided in Table 8.

Table 8 **Cloned macroschizont infected cell lines, SPAG-1 alleles and *Eco* RI-SR1 fragments they contain.**

Cloned macroschizont cell line	Size of *Eco* RI-SR1 fragments	Allele
TaA 46A	3.4 kb	A 3.4
TaA 139D4	3.4 kb	A 3.4
TaA 139D7	3.4 kb	A 3.4
TaA 139E3	3.4 kb	A 3.4
TaA 139E5	3.4 kb	A 3.4
TaA 46.2	3.4 kb	A 3.4
TaH 46.2	3.4 kb	H 3.4
TaH 46.3	3.4 kb	H 3.4
TaH 46.4	3.4 kb	H 3.4
TaA 139D6	4.8 kb	A 4.8
TaA 46.3	4.8 kb	A 4.8
TaHBL3b	6.0 kb	H 6.0

TaA and TaH denote *T.annulata* Ankara and Hisar, respectively.

3.2.2. Sequence analysis of a genomic copy (allele gH3.4) of SPAG-1.

Since the above data demonstrate that there is polymorphism at the DNA level associated with the SPAG-1 locus, I decided to establish to what extent this was reflected in sequence variation of the structural gene. Therefore two pUC18 plasmids, containing a partially sequenced SPAG-1 genomic clone, were kindly given to me by Dr. R. Hall. The inserts of these plasmids originated from a phage clone which was isolated from a λEMBL3 library containing genomic Ta Hisar piroplasm DNA (Katzer et al., 1994). The phage clone contains a 13 kb insert which hybridises to the SR1 probe (Williamson et al., 1989). One plasmid (pSPAG3.4) contains a 3.4-kb *Eco*RI fragment, which hybridises to the SR1 probe on Southern blotting. The other plasmid (pSPAG1.4) contains a 1.4-kb *Eco*RI fragment, which is located adjacent to the 3.4-kb fragment in the original λEMBL3 clone and codes for the 5' region of the SPAG-1 gene (Figure 7).

A panel of 182 m13mp18 clones containing random fragments (average size 300 bp) generated by sonicating pSPAG3.4 was also made available by Dr. R. Hall. The sequencing of these clones was at a very preliminary stage of analysis when I received them. The limited

information that was available indicated that this genomic SPAG-1 clone is not identical to the published SPAG-1 cDNA sequence (Hall et al., 1992). For ease of description these different copies will be referred to hereafter as gH3.4 (genomic Hisar 3.4 kb *Eco*RI RFLP) and cH (cDNA from Hisar). Compilation of my data from the 3.4 kb genomic fragment demonstrated that the first 297 bases of the coding sequence were lacking as deduced by comparison with the cH sequence. I subsequently obtained this information from a 600 bp *Hind III-Eco RI* fragment of pSPAG1.4 (see Figure 7). This was sub-cloned into m13mp18/19 and sequenced. The genomic SPAG-1 sequence data to this point was assembled using the GelAssemble programme of the UWGCG computer package (Devereux, 1989). A diagram of the alignment of 23 of the sequences (thin arrows) obtained from the m13 clones is shown in Figure 8. The thin arrows depicted in this figure represent only those sequences which provide all the essential information that could be derived from the analysis at this stage. For the sake of clarity the other 162 m13 clones which duplicated and consolidated this sequence information are not included in this diagram. Figure 8a shows the sequence information obtained in the 5' to 3' orientation and Figure 8b shows the sequence information obtained in the 3' to 5' orientation. In three locations sequences were only obtained in one direction and these are marked as shaded boxes (Figure 8). These sequence gaps were filled by double stranded DNA sequencing using specially designed oligonucleotides (Y2, Y3 and Y4,). The sequence obtained with these primers is shown in Figure 8 by the thicker arrows. The final DNA sequence and its predicted amino acid sequence is shown in Figure 9.

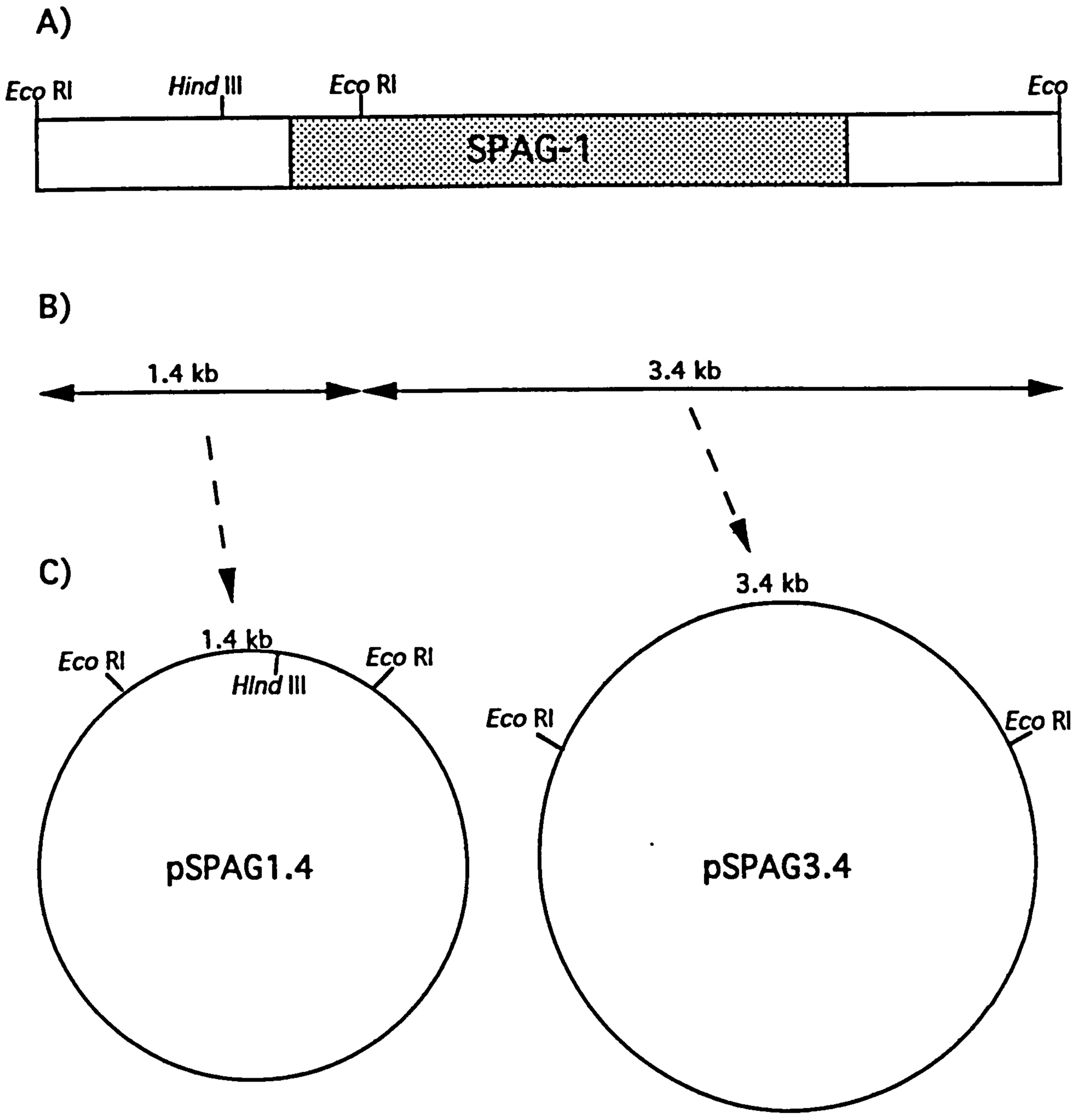


Figure 7: A diagrammatic representation of the sub-cloning gH3.4 into plasmid pSPAG1.4 and pSPAG3.4. A) shows a representation of the genomic SPAG-1 fragment from the original EMBL3 clone; B) shows how the SPAG-1 gene was subcloned in two fragments, one was 1.4 kb and the other 3.4 kb in length; C) depicts the final pUC plasmids, pSPAG1.4 and pSPAG3.4, containing the genomic SPAG-1 allele gH3.4.

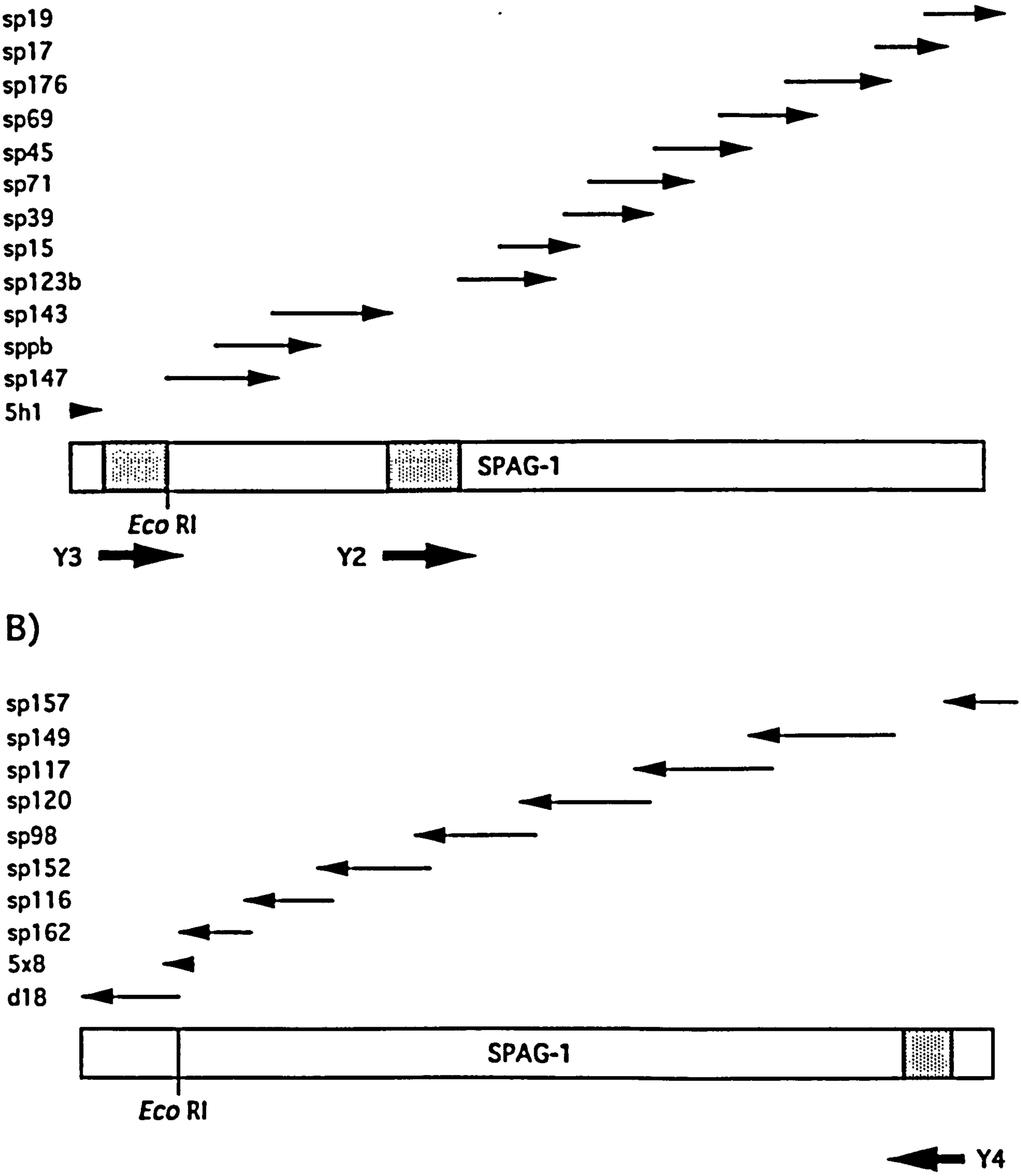


Figure 8: Diagrammatic representation of the SPAG-1 gH3.4 sequence obtained for each DNA orientation. The large bar with SPAG-1 written in it represents the SPAG-1 gene. The internal *Eco* R1 site is marked. The thin arrows represent the sequence information obtained for the m13mp18 clones named at the left hand site. The boxed in areas represent region for which sequence information was obtained using one of the two DNA strands. The darker arrows represent the sequence obtained by double stranded DNA sequencing with SPAG-1 specific oligonucleotides. The name of the oligonucleotides used is written next to the beginning of the sequence. A) The sequence information for the 5' to 3' DNA orientation. B) The sequence information for the 3' to 5' DNA orientation.

```
   1  CTTATGCTTAGGAGTAACATTGAAATTTAAATTTCATTTTCCAAAACTCAACGATGAACATTTTACACTTTCTGTTGACCATTCCGGTCA   90
                                                                M  N  I  L  H  F  L  L  T  I  P  V  I
  91  TTTTTGTATCTGGAGCGGACAAGATGCCTGCGGGAGAAAGTTCTAGAACCTCTAAACCCAGTCCCCTAGTAACCCTAGAATCAGCGGTAA   180
       F  V  S  G  A  D  K  M  P  A  G  E  S  S  R  T  S  K  P  S  P  L  V  T  L  E  S  A  V  T
 181  CACAACCTTCAAAAGACCCATTCAAGACAATTAGTGCCTTGTCAAAAGCAACAAAAGTATGGAAGTCAGCGGTATCAGTATCAGGTGACT   270
       Q  P  S  K  D  P  F  K  T  I  S  A  L  S  K  A  T  K  V  W  K  S  A  V  S  V  S  G  D  S
 271  CTAAGACTGTTCCTACTCCAGTTTCGGAACCAATTATTACTCGATCTTTTCAAGAACCAGTATCTCAAGAACTTGAATTCCAATCAGATA   360
       K  T  V  P  T  P  V  S  E  P  I  I  T  R  S  F  Q  E  P  V  S  Q  E  L  E  F  Q  S  D  T
 361  CTGAAATTAATGAGTCAGGATCCGGTTCAGATGAGGATGACGATGACGATGAGGAGGAGGAAGAAGAAGACGATGGATCCGGCTCATCTA   450
       E  I  N  E  S  G  S  G  S  D  E  D  D  D  D  D  E  E  E  E  E  E  D  D  G  S  G  S  S  K
 451  AAGGTGCAAAAGGAAGCCCAAAAGCTCAGGCTGCAGTATCTTCAAGCAGTACATCCACAGCAAGTCCAACATCTCCAACTACAACATCAT   540
       G  A  K  G  S  P  K  A  Q  A  A  V  S  S  S  S  T  S  T  A  S  P  T  S  P  T  T  T  S  S
 541  CACAACCTGGCTTGGGATCAAGTGGTTCACACGGTCAACAAGTTCCAGGTGTAGGTGTTCCAGGTGTAGGAGTTCCAGGTGTAGGAGTTC   630
       Q  P  G  L  G  S  S  G  S  H  G  Q  Q  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P
 631  CAGGTGTAGGAGTTCCAGGTGTAGGAGTTCCAGGTGTAGGAGTTCCAGGTGTAGGAGTTCCCGGTGTAGGTGTTCCCGGTGTAGGTGTTC   720
       G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P
 721  CAGGGGTAGGCGTTCCAGGGGTAGGCGTACCAGGGGTTGGAGCTGTACCTGGAGTGGGGGGGCTTGGAGATGATAGTAGTGCATTGCCTG   810
       G  V  G  V  P  G  V  G  V  P  G  V  G  A  V  P  G  V  G  G  L  G  D  D  S  S  A  L  P  G
 811  GAAGTGGTGGTCTTGGAGCAGGAGCAAAGGCTGGGAAAGGTCCAGGATCTGGTTTACAGGGACCAGGAGGTGTTGGACCAGGAGTACCTG   900
       S  G  G  L  G  A  G  A  K  A  G  K  G  P  G  S  G  L  Q  G  P  G  G  V  G  P  G  V  P  G
 901  GTGTAGGTGATGCAGCTTCTTCTTCTTTACCAGGAAAACCTCCAGGTGTAGGAGTTCCTGGAGCAGTAGAACCTGGATTACCAGGAGCAG   990
       V  G  D  A  A  S  S  S  L  P  G  K  P  P  G  V  G  V  P  G  A  V  E  P  G  L  P  G  A  A
 991  CAGGTGTACCAGGAGGAAAGGCAGGAAAATCGAATAAATCTTCCGACCCTGAATTAGATTTTGGAGATGAATCTGATGGGTCAGGATCCG  1080
       G  V  P  G  G  K  A  G  K  S  N  K  S  S  D  P  E  L  D  F  G  D  E  S  D  G  S  G  S  G
1081  GTTCAGCAGGGGACGACGATGACAATGACGATGAAGAGGAAGAAGACGATAAATCTACCTCATCTAAAGGAGCAGGAGCAAAGGCTGGGA  1170
       S  A  G  D  D  D  D  N  D  D  E  E  E  E  D  D  K  S  T  S  S  K  G  A  G  A  K  A  G  K
1171  AAGGTCAAGGATCTGTATCACCAGGAGGAGGATCCTCAGCAAGTCAAACATCTTCAACTACAACATCATCACAATCTGGCTTGGCACCAA  1260
       G  Q  G  S  V  S  P  G  G  G  S  S  A  S  Q  T  S  S  T  T  T  S  S  Q  S  G  L  A  P  S
1261  GTGGTTCTCACCCTCAACAAGTTCCTCAACAAGATCCAGCGCTTAGTCAACCTAGTGGAGGAGGTGTGCCCGGAGTTGGAGTTCCCGGAG  1350
       G  S  H  P  Q  Q  V  P  Q  Q  D  P  A  L  S  Q  P  S  G  G  G  V  P  G  V  G  V  P  G  V
1351  TTGGAGTTCCTGGAGTTGGAGTTCCTGGAGTTGGAGTTCCTGGAGTTGGAGTACCCGGTGTTGGGGGTGCAACAACTTCATCATCATCAA  1440
       G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  V  P  G  V  G  G  A  T  T  S  S  S  S  T
1441  CAACTTCAACTACTACTACTACTACAACATCATCTTCACCTGGAAAACCTTCAAACCAAGGAAGCCATGGTACTTCTCCAAGAAAGCTAG  1530
       T  S  T  T  T  T  T  T  T  S  S  S  P  G  K  P  S  N  Q  G  S  H  G  T  S  P  R  K  L  V
1531  TAACCAGACAAACTGACTCAATATCAGGACCCATACCATCACCAGGAGATCCAAGAGCAATTACTGGACAAATGGGTTTGTTATATTCAA  1620
       T  R  Q  T  D  S  I  S  G  P  I  P  S  P  G  D  P  R  A  I  T  G  Q  M  G  L  L  Y  S  S
1621  GTAAATCTTTTGTAGGTGAAGGAGAAAGGTTTGCTGCACAGTTCTTGGGAGATTTTAAACCAAAACCAAGGAGATATGAAGGACAAGAAA  1710
       K  S  F  V  G  E  G  E  R  F  A  A  Q  F  L  G  D  F  K  P  K  P  R  R  Y  E  G  Q  E  T
1711  CAGATGCAGTAAAACTAAAACAATTCATTTTTGAAGAGGTCAAATCGCTGGTGCAAACGTTAATAAACCTTAAATTAGCAATTGCAAACG  1800
       D  A  V  K  L  K  Q  F  I  F  E  E  V  K  S  L  V  Q  T  L  I  N  L  K  L  A  I  A  N  D
1801  ACTTTGTTGAAATCAGTGAAAAGTTGAAAAAGAAAAATCAAAATTACGTACCGAAATTAAAGTTGTTAAAAGGAGAACAATTTGACACCA  1890
       F  V  E  I  S  E  K  L  K  K  K  N  Q  N  Y  V  P  K  L  K  L  L  K  G  E  Q  F  D  T  K
1891  AACAGAAGGTAGCCAACGTGCTAAAGGGGTTCAATTCTCTGTACTTCGTATTTTTTATGAACCTTAACCTAGCGAAAGAAGTAAACAAAC  1980
       Q  K  V  A  N  V  L  K  G  F  N  S  L  Y  F  V  F  F  M  N  L  N  L  A  K  E  V  N  K  P
1981  CGGAAGAATTGGCAGAATTTCTTTGGAAACTAAATACAATCCCAGATAAAGTAGGAAGAGAATTTGAGTTAGCAATAGAAAAAACTAAAG  2070
       E  E  L  A  E  F  L  W  K  L  N  T  I  P  D  K  V  G  R  E  F  E  L  A  I  E  K  T  K  G
2071  GTTCAGAGAAAAAGAAGGAATTAGAAGAAGCATTTAATTCAATAGGGTTAGGTTTCAAAATAGCACAGTACGCAACAAATGACATCCTCT  2160
       S  E  K  K  K  E  L  E  E  A  F  N  S  I  G  L  G  F  K  I  A  Q  Y  A  T  N  D  I  L  S
2161  CAAGTATAACAAATTCAGTCTACTCCCTGATAAAACTAAAGAATTTTGGAGATGATTTTATTACCGAAGTAAGAAAGTCGCTGCAAATGG  2250
       S  I  T  N  S  V  Y  S  L  I  K  L  K  N  F  G  D  D  F  I  T  E  V  R  K  S  L  Q  M  V
2251  TTCCACACCAAAAGAACCTAAACGGATCAGCGTTTATAGTCAAAATCTCAGAAATAATCAACAAAAAAGGAACAGAAGATGAGGATCAAA  2340
       P  H  Q  K  N  L  N  G  S  A  F  I  V  K  I  S  E  I  I  N  K  K  G  T  E  D  E  D  Q  T
2341  CATCAGGAAGTGGGTCAAAAGGGACAGAAGGAGTATCACTAAGGGGGCAAGATTTGACAGAAGAAGAAGTTTTGAAAGTTTTGGATGAAC  2430
       S  G  S  G  S  K  G  T  E  G  V  S  L  R  G  Q  D  L  T  E  E  E  V  L  K  V  L  D  E  L
2431  TAGTGAAGGATGTAAGCGAAGAACAGGTTGGAATAGGAGATTTAAGTGACCCGAATAGCAGAACACCAAATGCAAAACCAGCCGAACTTG  2520
       V  K  D  V  S  E  E  Q  V  G  I  G  D  L  S  D  P  N  S  R  T  P  N  A  K  P  A  E  L  G
2521  GACCTTCACTAGTGATACAAAATGTACCATCAGACCCCTCAAAAGTGACACCAACACAGCCTTCAAATTTGCCACAAGTACCAACAACAG  2610
       P  S  L  V  I  Q  N  V  P  S  D  P  S  K  V  T  P  T  Q  P  S  N  L  P  Q  V  P  T  T  G
2611  GGCCGGGGAACGGGACGGATGGAACAACAACAGGACCAGGTGGAAACGGGGAAGGAGGCAAAGATTTGAAGGAAGGAGAAAAGAAAGAAG  2700
       P  G  N  G  T  D  G  T  T  T  G  P  G  G  N  G  E  G  G  K  D  L  K  E  G  E  K  K  E  G
2701  GATTATTTCAAAAGATCAAAAACAAACTCTTGGGCTCAGGATTCGAAGTCACAAGTATTATGATACCAATGACAACAATCATATTCAGTA  2790
       L  F  Q  K  I  K  N  K  L  L  G  S  G  F  E  V  T  S  I  M  I  P  M  T  T  I  I  F  S  I
2791  TAGTCCACTAAAACTAAAAACACAACTAACCACACTAATTTATAATATACACAATAAAT  2849
       V  H  *
```

Figure 9: DNA sequence of the gH3.4 allele of SPAG-1. The predicted amino acid sequence is shown below.

3.2.3. The comparison of the cDNA (cH) and genomic (gH3.4) sequence of SPAG-1.

The amino acid sequences predicted from the gH3.4 and cH copies of the SPAG-1 gene are compared in Figure 10a. These sequences are clearly different due to multiple point mutations, deletions and/or insertions. Therefore it is apparent, unless one invokes unusual post-transcriptional processing mechanisms, that the cH sequence is not the product of the RNA transcript of the gH3.4 gene. There is evidence, presented in chapter 4, consistent with one of the gaps (marked as a series of asterisks in Figure 10a) being an intron. The extent of the polymorphism is shown diagrammatically in Figure 10b. Overall, the degree of identity is 92% with the most conserved regions lying in the C terminal half (97% identity) and the N terminal quarter (92% identity). The second quarter of the molecule is the most variable region and can be divided into three sub-regions with values of 81%, 60% and 86% identity.

A number of features highlighted in the original cH sequence (Hall et al., 1992) are conserved. Thus the two blocks of striking PGVGV pentapeptide elastin repeats (Raju and Anwar, 1987) are present; as are the glutamate/aspartate (D/E) and the threonine/serine rich (T/S) motifs. Interestingly the VGVAPG hexapeptide, which is identical to the elastin receptor ligand (Mecham et al., 1989), found three times in the cH sequence, is absent from the gH3.4 sequence (see Discussion and Chapter 5). The putative N terminal signal peptide (amino acids 1-18) and C terminal membrane anchor are well conserved, howerver.

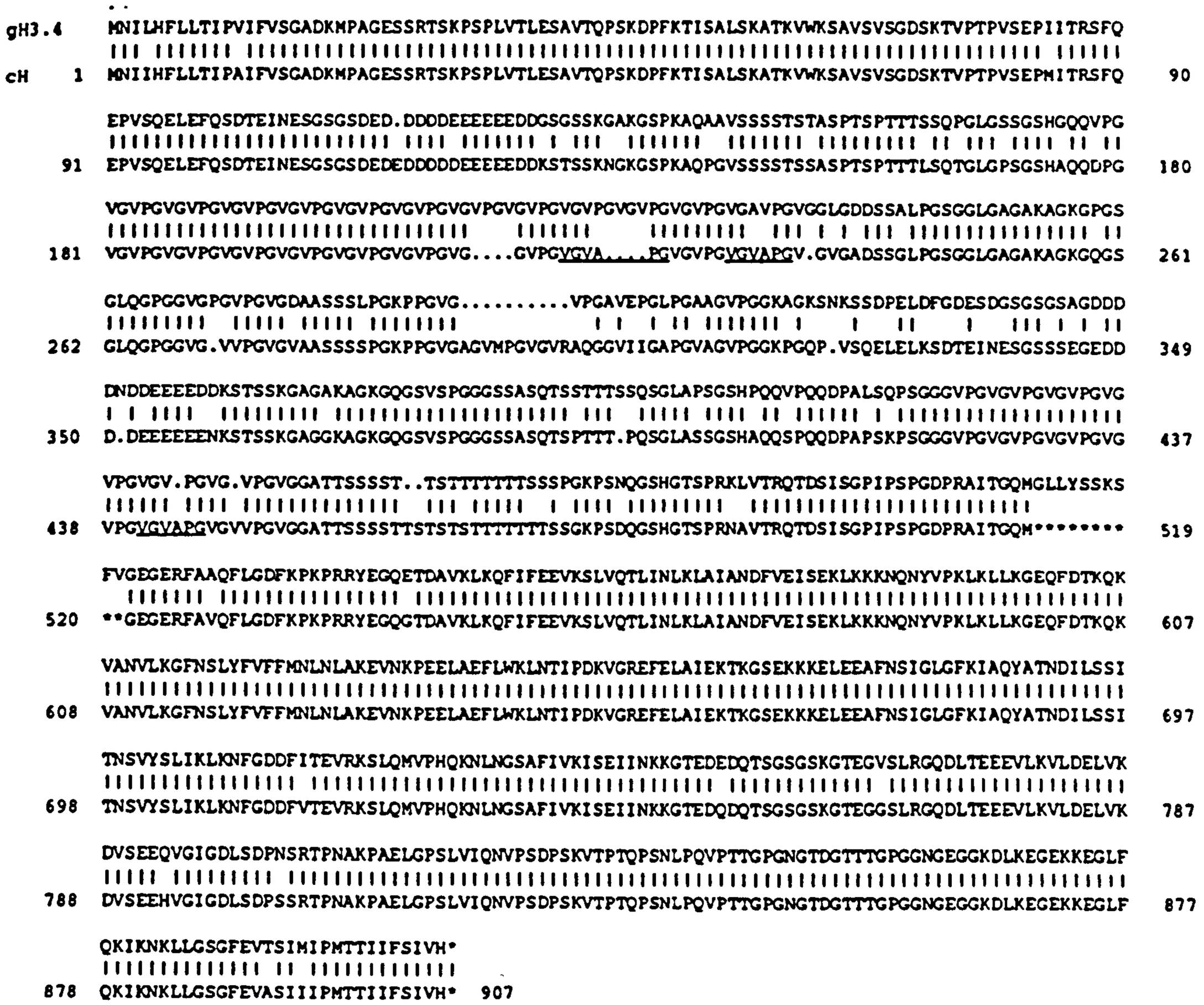
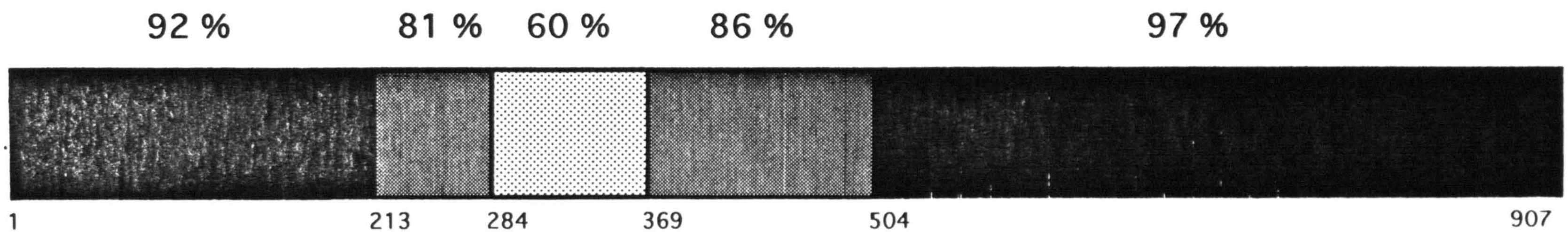


Figure 10: Comparison of SPAG-1 protein sequences, derived from the DNA sequences of alleles gH3.4 and cH. a) The polypeptide sequences are compared using the GAP programme on the UWGCG analysis software package (Devereux, 1989). The VGVAPG motifs in cH are underlined and the region spanned by the putative intron (see chapter 4) is shown with asterisks. **b)** Schematic representation of the comparison shown in Figure 10a. The figures above the blocks represent the percentage identity over each segment calculated by scoring a gap as one change. The numbers below designate the amino acid residues at the boundaries of the sequence blocks based on the cH molecule shown in Figure 10a.



OVERALL IDENTITY 92 %

3.2.4. Further analysis of variable and constant regions of SPAG-1 alleles.

The above sequence analysis of the cH and the gH3.4 alleles identified a highly polymorphic region towards the second quarter of the SPAG-1 molecule (Figure 10b). The C-terminus containing the 1A7 epitope (Williamson et al., 1989; Boulter et al., 1994) is highly conserved between these two alleles (see Figure 10b). Therefore a PCR based strategy was adopted to study the extent of sequence variation and conservation across these regions. The template DNA used for these PCR reactions was extracted from the macroschizont-infected cell line clones containing the identified segregating *Eco* RI RFLPs (see Table 8).

3.2.4.1. Sequence analysis of the most polymorphic region of SPAG-1.

A DNA fragment of about 1860 bp (positions 53-1919 according to the cH sequence, Hall et al 1992), containing the polymorphic region, was amplified by PCR using primers 420 and Y3 (Materials & Methods and Figure 11) from the DNA of all 12 available cell line clones (see Table 8). Figure 12 shows a representative gel of four of these PCR products. The relevant bands were purified by Geneclean and cloned into pGEM-T. DNA minipreps were prepared using the Wizard reagents (Promega) and clones containing inserts were identified by agarose gel electrophoresis. A representative gel containing the DNA of 11 such pGEM-T clones and a negative control is shown in Figure 13. The series of clones generated in this way are called "series 1" clones, as depicted in Figure 11. Using these clones as templates, I sequenced the most polymorphic region from bases 850 to 1107 (corresponding to amino acids 284 to 369 of SPAG-1). The primers used for sequencing are 815, Y2 and 647 as shown in Figure 11. Since Taq polymerase is error-prone, I sequenced three independent pGEM-T clones from each PCR reaction.

The sequence comparison of all "series 1" clones, based upon 12 cloned macroschizont cell lines, revealed 4 DNA sequences which are highly polymorphic. These four different sequences, as well as the cH sequence and the gH3.4 sequence are shown in Figure 14a and their corresponding translated amino acid sequences are shown in Figure 14b.

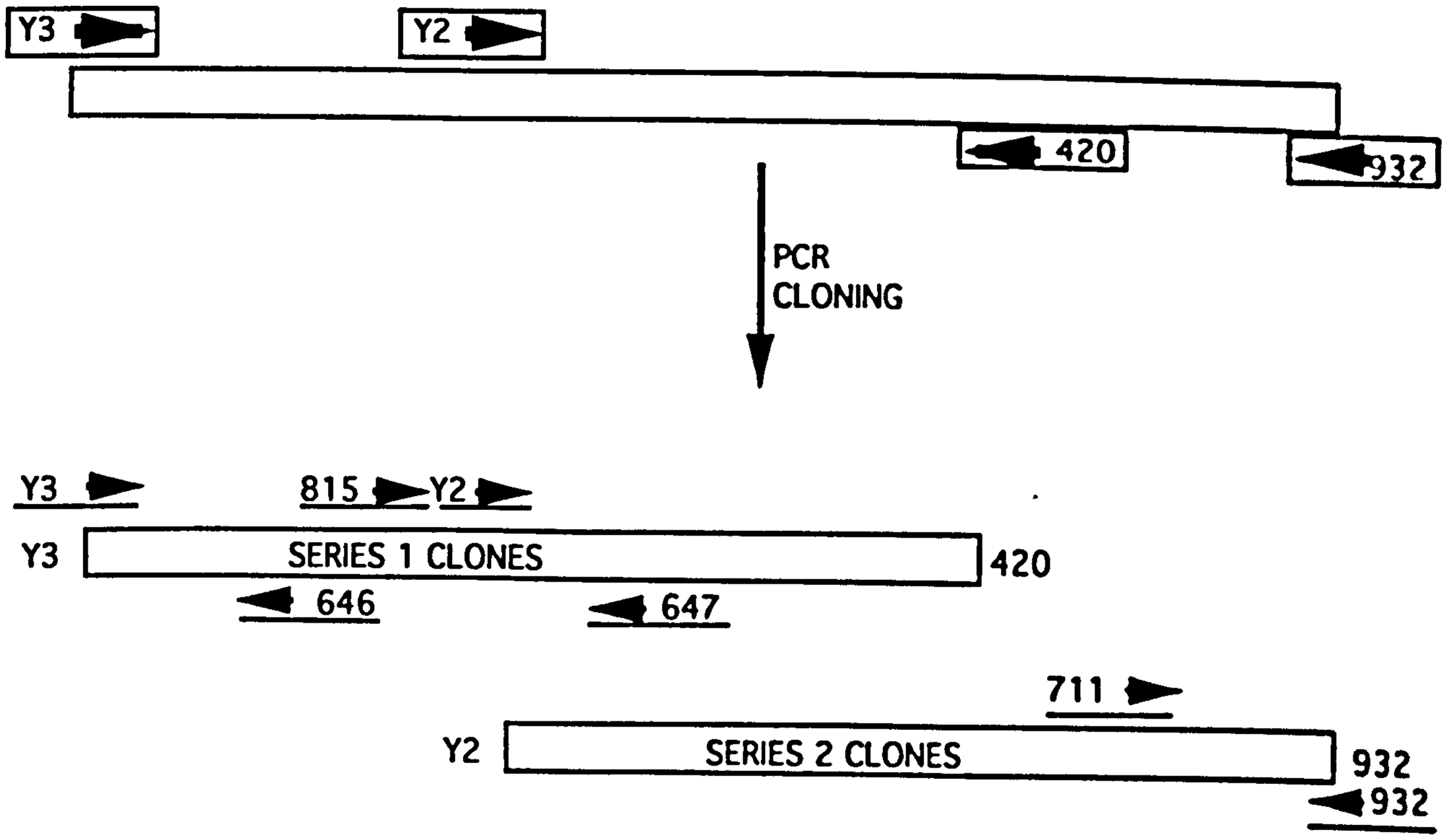


Figure 11: Schematic representation of the PCR based cloning and sequencing strategy. The location of the primers used to generate the series 1 and series 2 fragments is shown by the boxed arrows, whilst those used for sequencing are shown by the smaller, underlined arrows. For precise details for each primer see Table 6 in the Materials and Methods section 2.2.4.

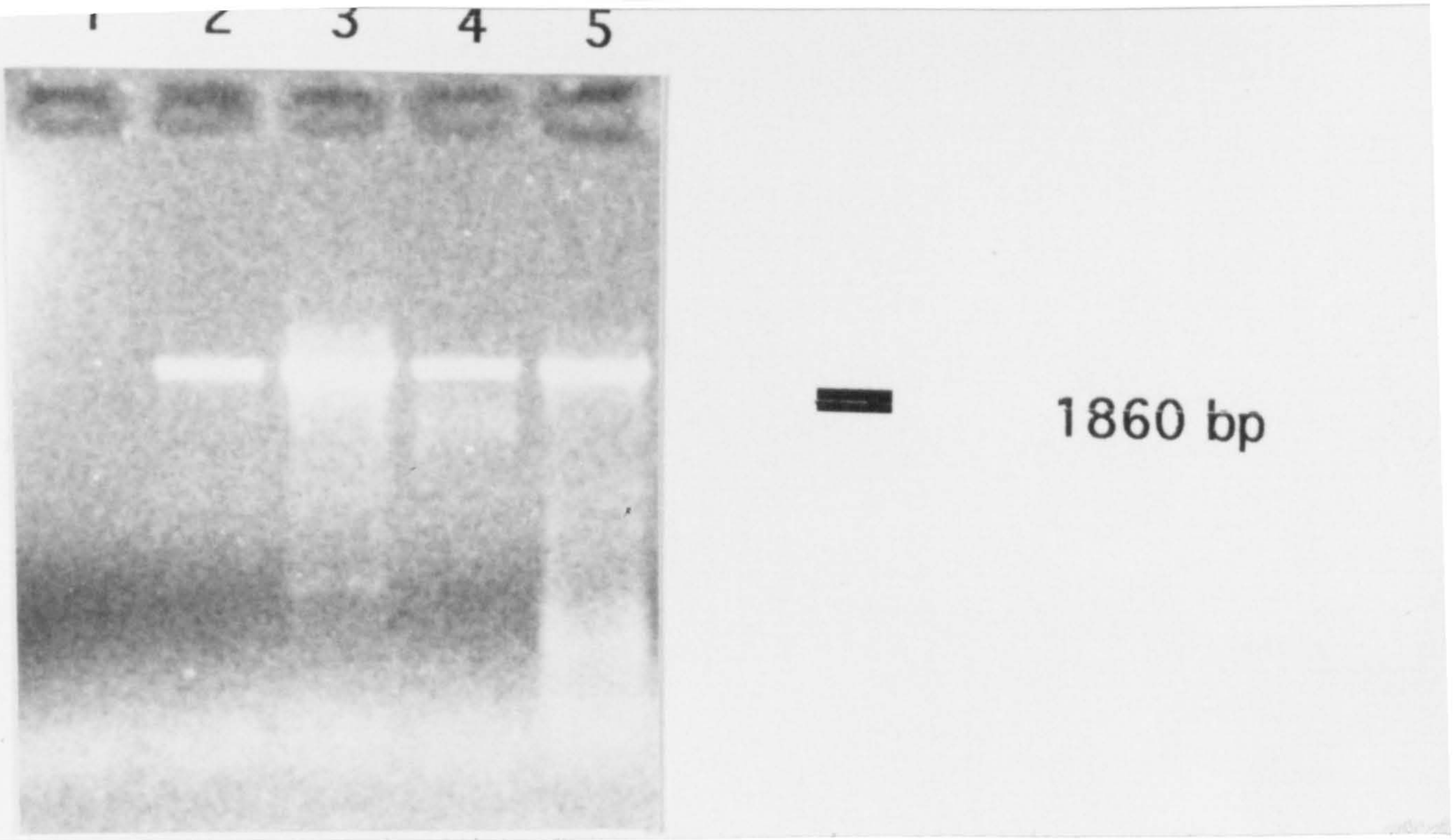


Figure 12: Agarose gel with PCR products, stained with ethidium bromide. The PCR products, shown in lanes 1-5, were obtained using the primers 420 and Y3 and the following cloned macroschizont-infected cell line DNA as template: no DNA (lane 1),TaA 46A (lane 2), TaA 139D7 (lane 3), TaA 139E3 (lane 4) and TaA 139E5 (lane 5).

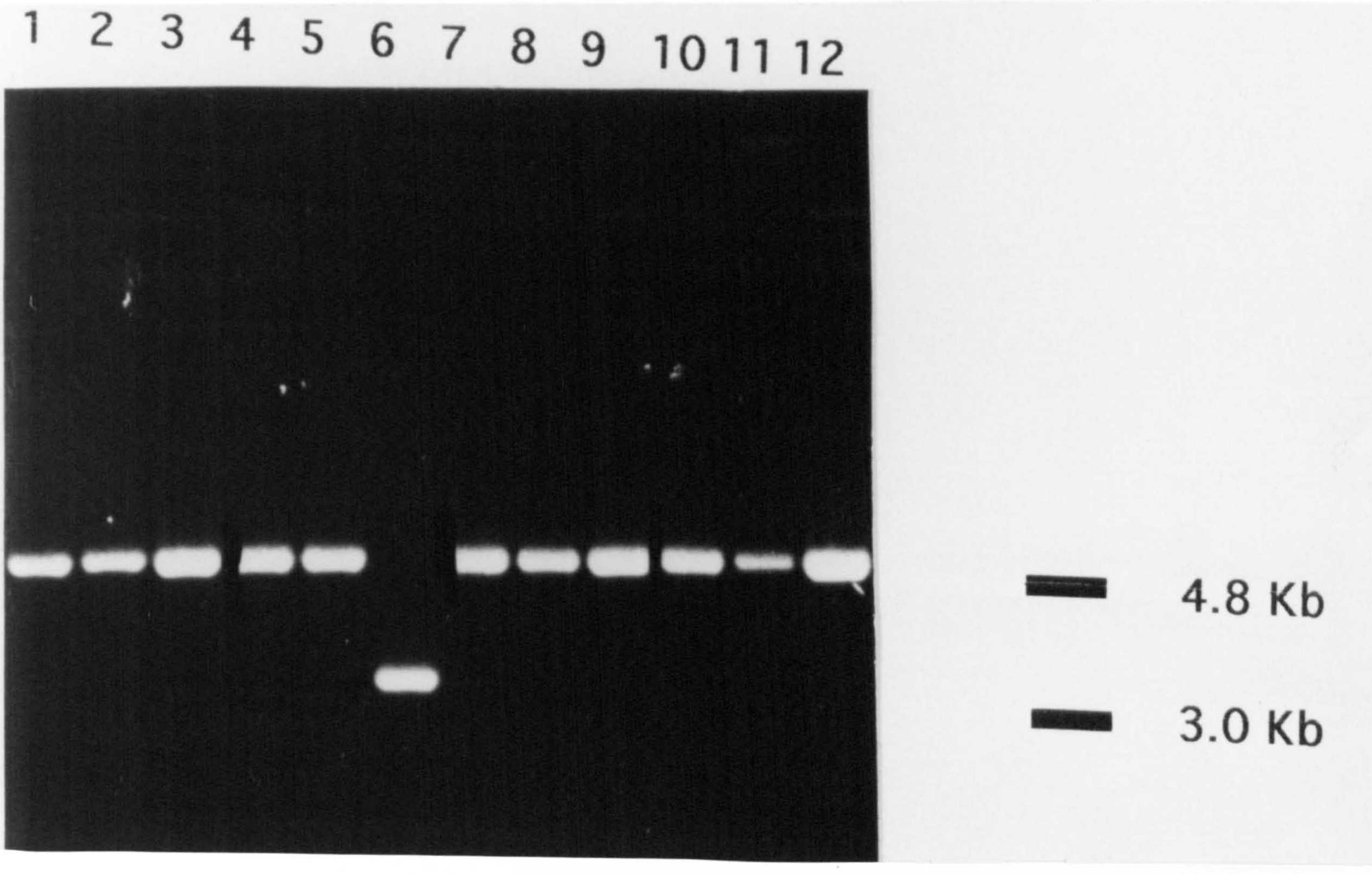


Figure 13: Agarose gel with DNA from pGEM-T clones, stained with ethidium bromide. DNA of three putative pGEM-T clones was extracted, using the Wizard miniprep method, for the PCR products of the following parasite clones : TaA 46A (lane 1-3), TaA 139D7 (lane 4-6), TaA 139E3 (lane 7-9) and TaA 139E5 (lane 10-12). The pGEM-T clone in lane 6 has no insert. The bars mark the size of the bands in kb.

a)

```
cH          CCAGGAAAACCTCCAGGAGTAGGAGCAGGAGTTATGCCTGGAGTTGGTGTACGAGCA
A 3.4       ------------------------------------------------------------
A 4.8       ----------A-------------------------------------------------
gH 3.4      ----------------------.............................-TT
H 3.4       ----------------------.............................-TT
H 6.0       ----------A---T-C-----------------------------------A----

cH          CAAGGAGGAGTAATAATTGGTGCGCCAGGAGTAGCAGGTGTG.........CCA...
A 3.4       ------------------------------------------.........---...
A 4.8       ---------------------------------C---------A.............---...
gH 3.4      -CT----C----GA-CC---ATTA-------C---------A.........---...
H 3.4       -CT----C----GA-CC---ATTA-------C---------A.............---...
H6.0        ---------C----------------------C----A-A...TCTGGATTA---AGG

cH          ..............GGAGGAAAGCCAGGACAACCA...GTATCTCAAGAACTTGAA
A 3.4       ..............--------------------..----------------
A 4.8       ..............---------G-----A--T-GAATAA----TCC--C-C----
gH 3.4      ..............---------G-----A--T-GAATAA----TCC--C-C----
H 3.4       ..............---------G-----A--T-GAATAA----TCC--C-C----
H6.0        AGGAGCAGGTGTACC--GA-----G-----A--T--AAAAA----TT---C-C----

cH          CTGAAATCAGACACTGAAATTAATGAGTCAGGTTCCAGTTCAGAAGGGGAAGACGAT
A 3.4       ------------------------------------------------------------
A 4.8       T-A...-T--GAGA----TC-G----C-----A----------G------C----T-
gH 3.4      T-AG-T-TT-GAGA----TC-G---G------A---G------C------C------
H 3.4       T-AG-T-TT-GAGA----TC-G---G------A---G------C------C------
H6.0        TCAG-T-TT-GAGA----TC-G----C-----A----------G------C------

cH          GAC...GATGAAGAAGAGGAAGAAGAAAATAAATCTACCTCATCTAAAGGAGCAGGAGGA
A 3.4       ---...------------------------------------------------------
A 4.8       ---AAT--C--T--------------CG--------------------------------
gH 3.4      ---AAT--C--T--------------CG--------------------------------
H 3.4       ---AAT--C--T--------------CG--------------------------------
H6.0        ---...-----------A--C-------------------------------------
```

Figure 14: Comparison of the SPAG-1 alleles over the most polymorphic region (bases 850-1107, amino acids 284-369). PCR derived sequences are shown for each allele from series 1. Templates were derived from the same macroschizont clones as used in Figure 6. In addition the same region from alleles cH and gH3.4 is presented for further comparison. The data are given in (a) as the DNA sequence and in (b) as the derived peptide sequence. All alignments are made to the cH sequence and - represents identity, . represents a gap and any substitutions are shown by the relevant letter.

b)

```
cH          PGKPPGVGAGVMPGVGVRAQGGVIIGAPGVAGV...P......GGKPG
A 3.4       ------------------------------...-......-----
A 4.8       ---H--------------------------A---...-......---A-
gH 3.4      --------.........VP-A-EP-L--A---...-......---A-
H 3.4       --------.........VP-A-EP-L--A---...-......---A-
H6.0        ---H-A-----------QAQ--L------A-R.SGL-RRSRCT-R-A-


cH          QP.VSQELELKSDTEINESGSSSEGEDDD.DEEEEEEENKSTSSKGAGG
A 3.4       --.-------------------------.------------------
A 4.8       KSNK-SDP--.LGD-SDD-----G-D-V-N-D----DD----------
gH 3.4      KSNK-SDP--DFGD-SDG---G-A-D---N-D----DD---------A
H 3.4       KSNK-SDP--DFGD-SDG---G-A-D---N-D----DD---------A
H6.0        KSKK-LDP-SDFGD-SDD-----G-D---.----D------------
```

Six cloned macroschizont cell lines of Ankara origin with the 3.4-kb *Eco* RI fragments all possessed the same sequence which therefore defines the A3.4 allele (Figure 14a and Table 8). Interestingly and unexpectedly this A3.4 allele matches the cH sequence totally over this stretch (see Discussion). The 3 cell line clones of Hisar origin with the 3.4-kb *Eco* RI fragment were identical to each other but distinct from A3.4 and thus define the allele H3.4 (Figure 14a and Table 8). The sequence of H3.4 is identical as expected to gH3.4, Two cloned macroschizont-infected cell lines of Ankara origin with the 4.8-kb *Eco* RI fragment are identical to each other and different to any other sequence and thus define the A4.8 allele (Figure 14a and Table 8). The sequence derived from the only cloned macroschizont-infected cell line of Hisar origin with an 6.0-kb *Eco* RI RFLP is named H6.0 and is also shown in Figure 14a.

3.2.4.2. Sequence analysis of the C-terminal constant region of SPAG-1.

The C-terminal constant region of SPAG-1 was amplified by PCR using DNA obtained from the cloned cell lines, TaA 139D4, TaH 46.2, TaA 139D6 and TaHBL3b. These cell lines were selected as they represent the four different sequence variants identified in section 3.2.4.1. The PCR primers used were Y2 and 932 (Table 7) and the fragments of about 1780 bp generated were cloned into pGEM-T giving rise to the 'series 2' clones (see Figure 11). Three independent pGEM-T clones for each of the cell line clones were sequenced using the primers 711 and 932 (Table 7 and Figure 11).

The region of the "series 2" clones which was sequenced corresponds to most of the original SR1 region. The observed DNA sequences and the predicted amino acid sequences are shown in Figures 15a and 15b respectively. The comparison of the "series 2" clones revealed that this region is highly conserved with very few base pair changes which resulted in even fewer amino acid changes. Again four distinct sequences can be seen, where cH is identical to A 3.4, gH3.4 is identical to H 3.4, and A 4.8 and H 6.0 are different to any other sequence. Thus the analyses of the polymorphic and SR1 regions are in agreement in that they indicate the existence of at least 4 SPAG-1 alleles.

a)

```
cH          ATAGGAGATTTAAGTGACCCAAGTAGCAGAACACCAAATGCAAAACCAGCCGAACTT
A 3.4       ------------------------------------------------------------
A 4.8       ------------------------------------------------------------
gH 3.4      --------------------G-A---------------------------------------
H 3.4       --------------------G-A---------------------------------------
H6.0        ------------------------------------------------------------
p67         C-C-----C---------------G-A---T--T-...C-A--G--A-C--TC---C
```

```
cH          GGACCTTCACTAGTGATACAAAATGTACCGTCAGACCCCTCAAAAGTGACACCAACA
A 3.4       ------------------------------------------------------------
A 4.8       ----------------------------------T-------------------------
gH 3.4      ------------------------------A-----------------------------
H 3.4       ------------------------------A-----------------------------
H6.0        ----------------------------------T-------------------------
p67         ----------T--A---ACTG---G--AAG---GA---A---T---AT-T------
```

```
cH          CAGCCTTCAAATTTGCCACAAGTACCAACAACAGGGCCGGGGAACGGGACGGATGGA
A 3.4       ------------------------------------------------------------
A 4.8       ------------------------------------------------------------
gH 3.4      ------------------------------------------------------------
H 3.4       ------------------------------------------------------------
H6.0        ------------------------------T----A-----------------------
p67         GG---CA---TAGCAG-TGG--G-GA-CA-C--CCTTC......A-CTC-TA-----
```

```
cH          ACAACAACAGGACCAGGTGGAAACGGGGAAGGAGGCAAAGATTTGAAGGAAGGAGAA
A 3.4       ------------------------------------------------------------
A 4.8       ------------------------------------------------------------
gH 3.4      ------------------------------------------------------------
H 3.4       ------------------------------------------------------------
H6.0        ------------------------------------------------------------
p67         --CG----G--G--..............--C---A-C-C-ACCTG---A------G
```

```
cH          AAGAAAGAAGGATTATTTCAAAAGATCAAAAACAAACTCTTGGGCTCAGGATTCGAAGTC
A 3.4       ------------------------------------------------------------
A 4.8       ------------------------------------------------------------
gH 3.4      ------------------------------------------------------------
H 3.4       ------------------------------------------------------------
H6.0        ------------------------------------------------------------A
p67         --------------GA-A--G---C----G--A------C----G--T------------
```

Figure 15: Comparison of the SPAG-1 alleles over the most constant region (bases 2419-2706, amino acids 806-902). PCR derived sequences are shown for each allele from series 2. Templates were derived from the same macroschizont clones as used in Figure 6. In addition the same region from alleles cH and gH3.4 and the p67 gene of *T. parva* is presented for further comparison. The data are given in (a) as the DNA sequence and in (b) as the derived peptide sequence. The 16 amino acids containing the 1A7 epitope (Boulter et al., 1994) are underlined. All alignments are made to the cH sequence and - represents identity, . represents a gap and any substitutions are shown by the relevant letter.

b)

```
cH          IGDLSDPSSRTPNAKPAELGPSLVIONVPSDPSKVTPTQPSNLPQVPT
A 3.4       ---------------------------------------------------
A 4.8       ----------------------------------S----------------
gH 3.4      -------N---------------------------------------------
H 3.4       -------N---------------------------------------------
H6.0        ----------------------------------S----------------
p67         L-------G-SS.ERQPS-------TDGQAG-TI-S--G-TIAAGGEQ


cH          TGPGNGTDGTTTGPGGNGEGGKDLKEGEKKEGLFQKIKNKLLGSGFEV
A 3.4       ---------------------------------------------------
A 4.8       ---------------------------------------------------
gH 3.4      ---------------------------------------------------
H 3.4       ---------------------------------------------------
H6.0        S---------------------------------------------------
p67         PPSAPN..--A---A-TQPEG.....--------I--L-K---------
```

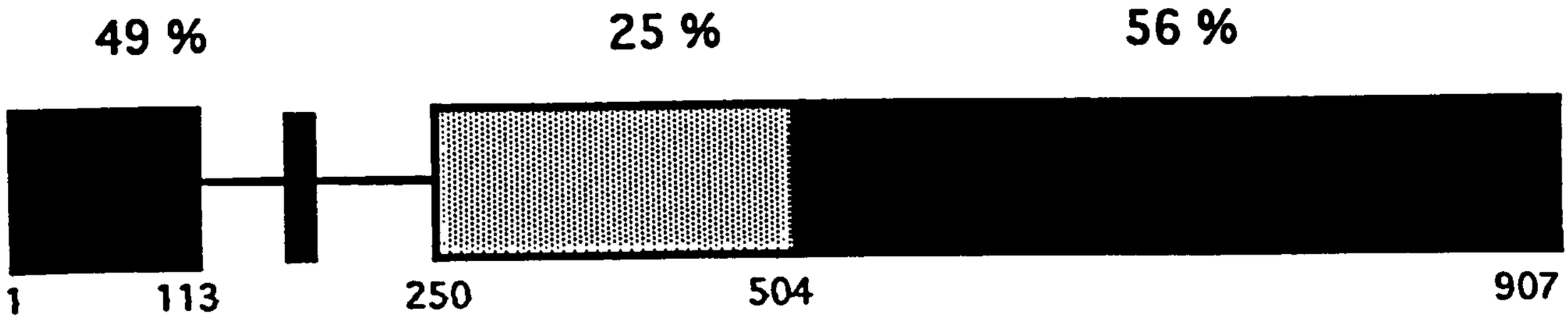
3.2.5. Comparison of SPAG-1 with the p67 antigen of *Theileria parva*: implications for the structure of the 1A7 epitope.

The sequence comparison of SPAG-1 and p67 (Nene et al., 1992) reveals that overall they are 47 % identical at the amino acid level (Figure16). Interestingly, the least similar region between these two antigens corresponds to the most polymorphic region within the SPAG-1 alleles. The most conserved parts are located at the N terminus (49% identity) and the C terminus (56% identity) again correlating well with the interallelic SPAG-1 comparison. The elastin repeat motif, PGVGV, only occurs once in p67 whereas as previously noted it is extensively repeated in SPAG-1.

The SR1 homologous regions of SPAG-1 and p67 are compared in Figure 15. The 16 amino acid sequence containing the neutralising 1A7 epitope (Boulter et al., 1994) is shown underlined. All alleles of SPAG-1 share this sequence but p67 has only 8 identical amino acids of which seven consecutive. This and the further observation that 1A7 and p67 weakly cross react (Knight, 1993) indicate that the 1A7 epitope is at least partially located in these shared residues.

3.2.6. The *Eco*RI RFLP pattern is based on the loss of an internal *Eco*RI site.

It was shown that both the sequences of cH and gH3.4 contain the internal *Eco*RI site but both sequences were matched to the 3.4 kb *Eco*R I RFLP band, as cH is identical to A 3.4 and gH3.4 is identical to H 3.4. Both A 3.4 and H 3.4 have an *Eco*RI RFLP band of 3.4 kb on a Southern blot probed with SR1. No information was available about the internal *Eco*R I site of A 4.8 and H 6.0 and whether the *Eco*RI RFLP is located in the coding region of the SPAG-1 gene. Therefore the "series 1" clones for the cell line clones TaA 139D4, TaH 46.2, TaA 139D6 and TaHBL3b were sequenced using the primers Y3 and 646. The resulting sequence information is shown in Figure 17a and the predicted amino acid sequence is shown in Figure 17b. The *Eco*RI site is underlined in the cH sequence. It can be seen that cH, gH3.4, A 3.4 and H 3.4 have an intact *Eco*RI site while no *Eco*RI site is present at this site for A 4.8 and H 6.0.



OVERALL IDENTITY 47%

Figure 16: Schematic representation of the comparison of the cH allele of SPAG-1 and p67 of *T. parva*. The figures above the blocks represent the % identity over each segment calculated by scoring a gap as one change. The numbers below designate the amino acid residues at the boundaries of the sequence blocks based on the cH molecule shown in Figure 10.

Therefore it can be concluded that the *Eco*RI RFLP pattern of 4.8-kb, seen in A 4.8, is due to a lack of the internal *Eco*RI site and that H 6.0 lacks this as well as a further *Eco* RI site outside the coding region of SPAG-1. The predicted amino acid sequence for this region shown in Figure 17b clearly shows that the change of the *Eco*RI site at DNA level also affects the predicted amino acid sequence. Therefore, the *Eco*RI RFLP is in this case a marker for polymorphism of the SPAG-1 gene at both the DNA and protein level.

	277	312	93	104
cH	GTATCTCAAGAACTTGAATTCCAATCAGATACTGAA		VSQELEFQSDTE	
A 3.4	----------	----------	----------	
A 4.8	----------C-GA---T---CC-----		------LKLCP-	
gH3.4	----------	----------	----------	
H 3.4	----------	----------	----------	
H 6.0	--------T--TGC-----..--T--GAGA---T		--HDA-L.LGDD	

Figure 17a **Figure 17b**

Figure 17. Comparison of the SPAG-1 alleles across the internal *Eco* RI site. (a) Sequence comparison at DNA level (bases 277-312); the *Eco* RI site is underlined. (b) Sequence comparison at amino acid level (amino acids 93-104); the amino acids encoded by the bases involved in the *Eco* RI site are underlined.

3.3 Discussion.

3.3.1. SPAG-1 is a single copy gene.

It has previously been shown that the piroplasm stocks of *T. annulata* contain a mixture of parasite strains (Wilkie et al., 1986). I have been able to confirm this by RFLP analysis. Further I provided evidence that the RFLPs behave as alternative forms of the same gene at a single locus and thus they behave as alleles. My data provides strong evidence in favour of there being one SPAG-1 gene per haploid genome.

3.3.2. Implications of the cDNA and genomic SPAG-1 sequence comparison.

I completed the sequencing of a genomic SPAG-1 clone and compared this to the cH allele (see Figure 10a). This comparison clearly indicates that these two sequences represent two different SPAG-1 alleles. Further, the comparison of their predicted amino acid sequence shows that the two protein sequences are 92% identical (see Figure 10b). More interestingly, the C-terminal half and the N-terminal segment of the molecule are most conserved (97% identity and 92% identity respectively). The polymorphism is greatest in the central region of the molecule (only 60% identity). The variation is due to multiple gaps/insertions and amino acid substitutions. The presence of conserved and polymorphic parts are important for the elucidation of the molecules function as well as for sub-unit vaccine development. These implications are discussed below.

Both alleles contain the PGVGV repeats which are homologous to the repeats found in bovine elastin (Raju and Anwar, 1987). Interestingly, unlike the cDNA allele cH, the gH3.4 allele product does not contain any VGVAPG motifs which considerably weakens the hypothesis that this represents a ligand for host cell recognition (Hall et al., 1992). The lack of the VGVAPG motifs in one allele and a lack of correlation between cells expressing the elastin receptor and their susceptibility to infection by sporozoites (Campbell et al., 1994) indicate that the elastin receptor is an unlikely candidate receptor for host cell

recognition. This suggests that the main function of the VGVAPG and the PGVGV motifs is most likely attributed to mimicry of the host and thereby the process of immune evasion (Hall et al., 1992: Hall, 1994).

3.3.3. Implications of the sequence comparison of the four SPAG-1 alleles.

A number of conclusions can be drawn from the sequence analysis across the most polymorphic region of SPAG-1 from 12 macroschizont-infected cloned cell lines. First of all it is clear that at least 4 distinct SPAG-1 alleles exist. These are A3.4 (this is identical to cH), H3.4 (which is identical to gH3.4), A4.8 and H6.0. The second observation is that the 3.4-kb RFLP is composed of two distinct alleles, A3.4 and H3.4. Unfortunately no cloned macroschizont cell lines were available of Hisar origin with an 4.8-kb *Eco*RI restriction fragment or of Ankara origin with an 6.0-kb *Eco*RI restriction fragment. These would indicate the degree of polymorphim between the two isolates. Interestingly, the sequence of the Ta Hisar cH allele is identical to the Ta Ankara allele A3.4 over all regions investigated. Since the cDNA is of *T. annulata* Hisar origin (Hall et al., 1992) this observation implies that the Hisar stock also contains the A3.4 allele but tests to confirm this are still required.

3.3.4. Implications for vaccine development.

The finding that the C terminal and N terminal regions of SPAG-1 are the most conserved regions is important information from the point of view of sub-unit vaccine design. This is reinforced by the recent epitope mapping studies which have located neutralising determinants recognised by bovine sera between residues 784 and 892 of the molecule (Boulter et al., 1994). Therefore, it can be predicted that these regions of the protein are likely, if included in a sub-unit vaccine, to induce cross protective immune responses. It is also of considerable interest that the most conserved region between *T. parva* p67 and SPAG-1 is in the C-terminal half of the molecule with 56% identity over amino acids 504-907 (Nene et al., 1992) as depicted in Figure 16. In addition the sequence

comparison of the four SPAG-1 alleles and p67 showed that a continuous stretch of 7 out of 16 residues containing the 1A7 epitope is conserved between SPAG-1 and p67. Since the monoclonal antibody 1A7 reacts weakly with recombinant p67 (Knight, 1993) it can be concluded that the epitope is at least partially located in these shared residues. This indicates that the 1A7 epitope is a cross reacting neutralising epitope for all isolated SPAG-1 alleles. This raises the idea of a common vaccine to *T. annulata* and *T. parva* designed on homologous regions in the C-terminus.

In the future these results might assume more importance once more information about immunodominant structures on SPAG-1 is accrued. When T cell epitopes are located it will be possible to assign some form of rating to them, according to their general usefulness depending upon whether they locate to polymorphic or conserved segments. Further studies investigating the optimal sub-unit vaccine delivery system must be conducted to form a basis for the rational design of future vaccine constructs.

3.3.6. Implications for the functional importance of SPAG-1.

The comparison of the the cH and gH3.4 allele indicated that the region towards the middle of the SPAG-1 molecule is highly polymorphic with only 60% identity at the amino acid level. While both the C terminus and the N terminus are highly conserved, 97% and 92% identity respectively between the two alleles. This might indicate that these regions are conserved for functional purposes. Since the 1A7 epitope is located in the C terminus of SPAG-1 and 1A7 blocks invasion of sporozoites into host cells *in vitro* as well as the fact that the C terminus is highly conserved might indicate that this part of SPAG-1 is involved either in the recognition event or in the subsequent invasion events into host cells. The predicted C-terminal membrane anchor (Hall et al., 1992) is highly conserved with only two conservative changes. The N terminus might be conserved for other purposes such as to facilitate the processing and folding of the protein and the predicted leader peptide (Hall et al., 1992) encoded by the first 18 amino acids is also found in

gH3.4 with only two conservative changes. Although the middle part of the protein is highly polymorphic, it might still have a very important function, but this does not involve sequence conservation. The middle part of SPAG-1 is most likely to be involved in the immune evasion process. This might also explain the finding of the PGVGV repeats which are conserved between gH3.4 and cH, as these mimic the host's own elastin molecule which has 11 such repeats. It was shown by Boulter (personal communication) that cattle infected with *T. annulata* or vaccinated with SPAG-1 do not develop antibodies against the PGVGV repeat structure. It can therefore be postulated that the elastin-like repeats make the molecule less immunogenic.

Chapter 4

Regulation of the expression of SPAG-1.

4.1. Introduction.

The expression of the SPAG-1 gene is stage-specifically regulated. mRNA transcription is initiated when an infected tick starts feeding on its bovine host. No SPAG-1 mRNA is detectable in any of the later life cycle stages of the parasite (Williamson et al., 1989). Since a clone was available containing the 5' DNA sequence of SPAG-1 it was thought that this gene might make a good candidate to study transcriptional regulation during the sporozoite stage in *Theileria*.

4.2. Results.

4.2.1. Sequencing of the 5' untranslated region of SPAG-1.

A pUC 18 clone containing a 1437 bp insert, containing the first 297 bp of the SPAG-1 gH3.4 allele and 1140 bp of the 5' untranslated region of SPAG-1 was given to me by Dr. R. Hall. The origin of this clone and the sequencing strategy of the SPAG-1 component is described in chapter 3. In order to investigate how SPAG-1 is regulated I decided to sequence the 5' region of SPAG-1 to identify sequences that might be involved in the stage-specific gene regulation of this gene. The insert of the pUC 18 clone was cut out of the vector and cut with *Hind* III which revealed two fragments. The first of these is 601 bp long and contains the first 297 bp of the SPAG-1 gH3.4 allele, and the other fragment of 836 bp in length as depicted in Figure 7 in chapter 3. Both these fragments were cloned into M13 mp18 and M13 mp19 for single stranded DNA sequence

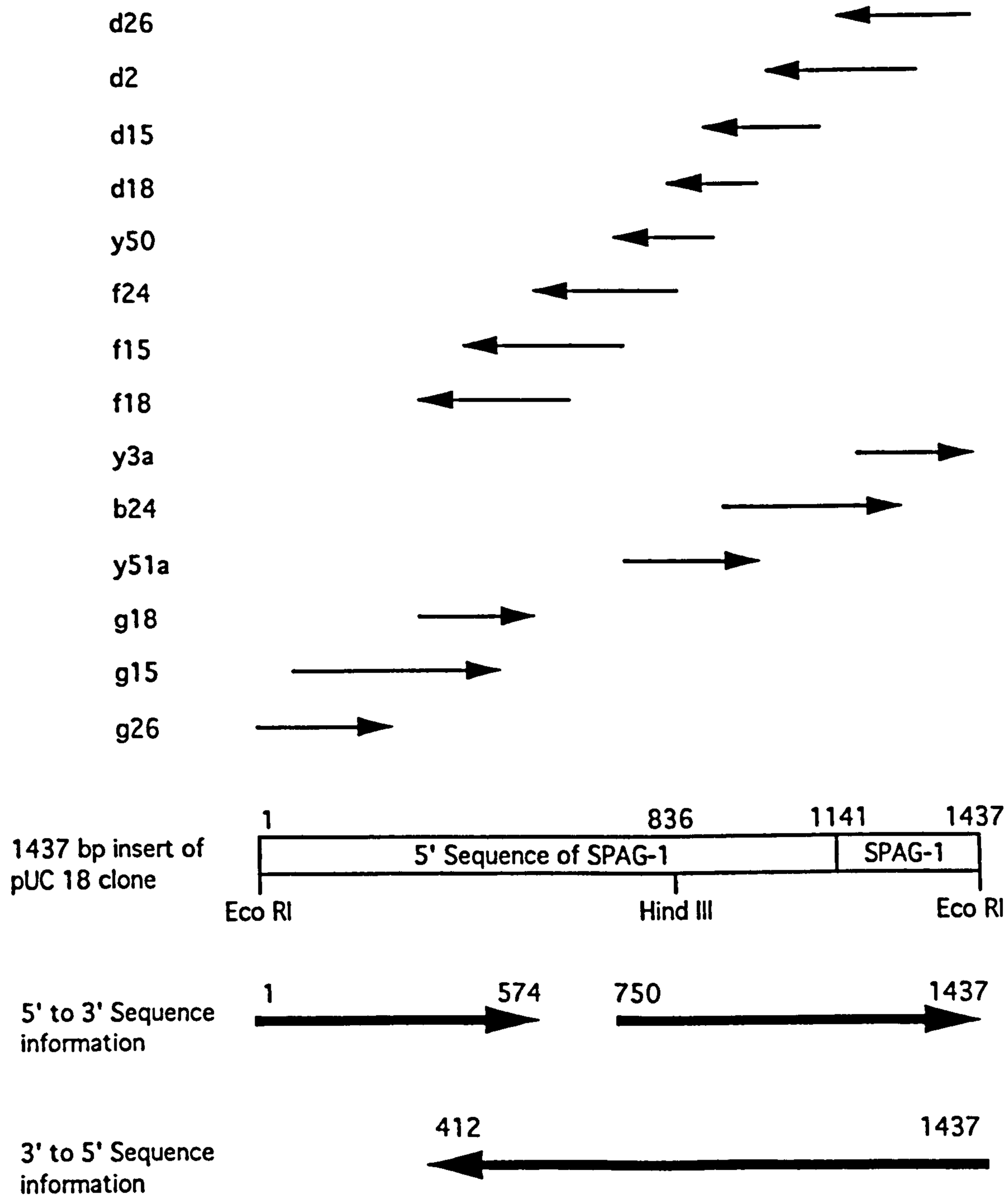


Figure 18: Diagrammatic representation of the sequencing information for the 5' untranslated region of SPAG-1. The thin arrows represent the length and position of individual sequencing reactions and their names are on the left hand side of the figure. The block in the middle of the figure represents the insert of a pUC 18 clone containing the first 1141 bp of the 5' untranslated region of SPAG-1 and the first 297 bp of the SPAG-1 gene. The large arrows at the bottom of the figure represent the sequence information obtained for each of the two DNA orientations.

analysis. The process of the sequence analysis is summarised in Figure 18. The first part of the figure shows the result of individual sequence reactions of the M13 clones and also the resulting three sequences using specific primers (Y3, Y50 and Y51) and double-stranded template DNA. The sequence was aligned using the UWGCG (Devereux, 1989) package on the SERC facility at the Daresbury laboratory. The summary of sequence information obtained for both DNA orientations is shown at the bottom of this figure. In the 5' to 3' DNA orientation there is a sequencing gap of 175 bases towards the middle of the 5' untranslated region. In the 3' to 5' DNA orientation there is a sequencing gap of 411 bases which are the most 5' bases of the pUC 18 clone. The resulting DNA sequence is shown in Figure 19 and the ATG start codon as well as the *Hind* III site are highlighted.

4.2.2. Sequence comparison of p67 and gH3.4.

The 5' sequence of the gH3.4 allele of SPAG-1 was compared to the 5' sequence of p67, the *T. parva* homologue of SPAG-1, and this comparison is depicted in Figure 20. This comparison revealed that the 5' untranslated region of both genes is 75.6% identical. It is noteworthy that the level of identity in the 5' untranslated region between p67 and SPAG-1 is higher than in the coding region, since the level of identity at DNA level between the two genes is 52%. Since these two regions are so highly conserved it might be an indication that these regions have regulatory functions. Both genes are transcribed in a stage-specific manner, both during sporozoite development, and therefore it can be assumed that the DNA sequences for regulation of their expression are conserved. However, since the whole of the upstream region is very conserved it is not possible to pinpoint particular regions which might be involved in gene regulation. It has been proposed, since the genome of *Theileria* is very A/T rich, that regulatory elements might contain a higher percentage of G's and C's. One of the larger stretches which is conserved between p67 and SPAG-1 also contains a higher than average percentage of G's and C's. It is located 490 to 535 bp upstream from the the ATG start codon. This region contains two interesting features 1) a sequence, GCTAGC, which is repeated twice and 2) GAGCTC. Both sequences are palindromic. Since both these sequences are G/C rich and

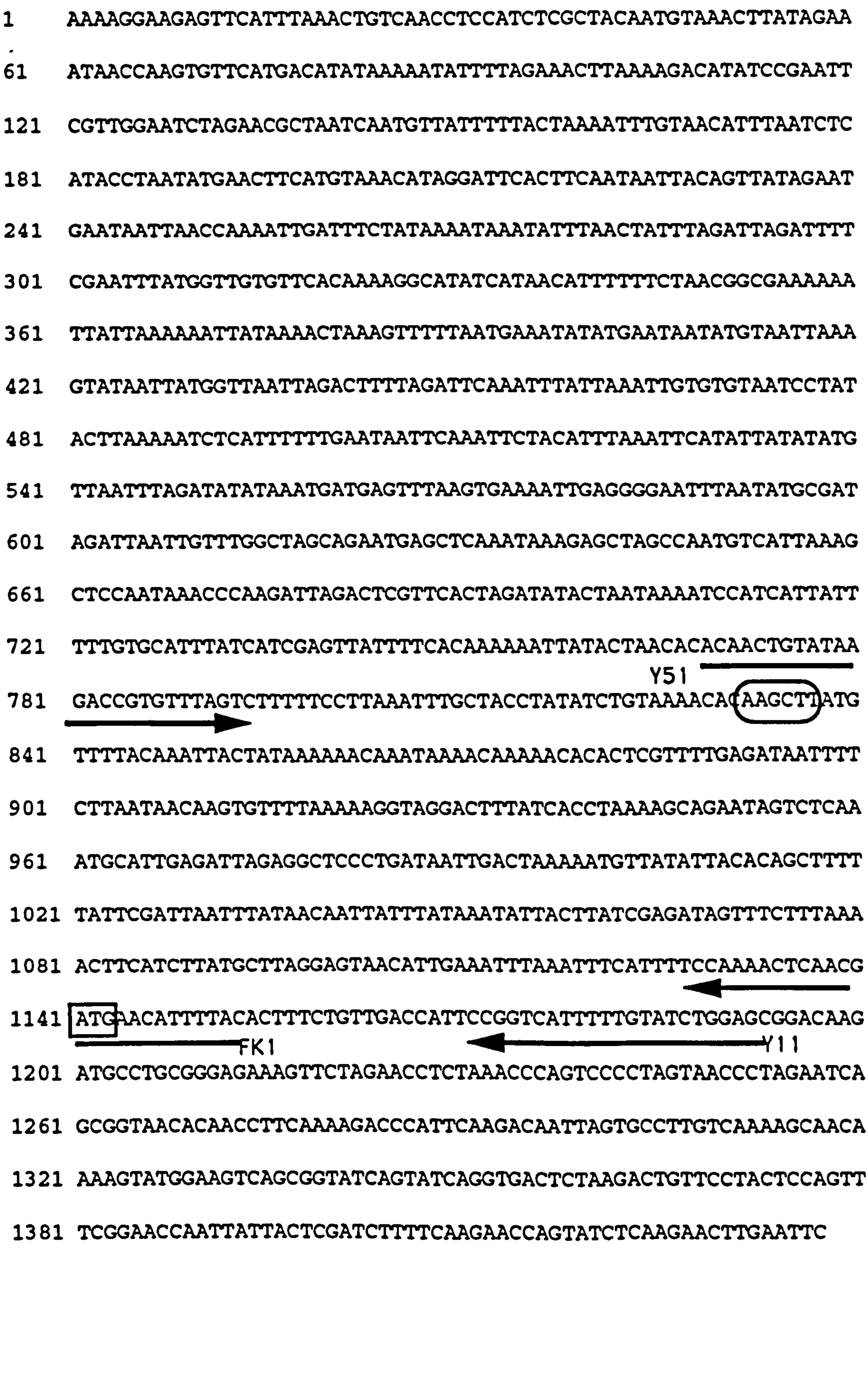


Figure 19: The sequence of 1141 bp of the 5' untranslated region of SPAG-1 and the first 297 bp of the SPAG-1 gene. The ATG start codon is boxed and the *Hind* III restriction site for sub-cloning the two fragments for sequencing is circled. The arrows represent primers which have been used for primer extension analysis and sequencing. The arrow heads are at the 3' end of the primers and the name of each primer is written at the 5' end of the arrow.

Figure 20: Sequence comparison of the 5' untranslated sequence of SPAG-1 of *T. annulata* and p67 of *T. parva*. gH3.4 represents the SPAG-1 sequence, where g stands for genomic DNA, H indicates that this DNA is of Hisar origin and the 3.4 shows that this allele has a 3.4 kb *Eco* RI restriction fragment associated with it. The p67 marks the DNA sequence of the *T. parva* sequence. The figures at the end of each row represents the sequence position with reference to the ATG start codon. A line indicates that the bases compared are identical and a space indicates that they are different and a dot represents a sequence gap. There are two GCTAGC sequences which are boxed and one GAGCTC sequence which is circled. these sequences are palindromic and conserved between SPAG-1 and p67 and their significance is discussed in the text.

```
gH3.4 AAAAGGAAGAGTTCATTTAAACTGTCAACCTCCATCTCGCTACAATGTAAACTTATAGAAATAACCAAGTGTTCA -1053
      ||| |||||||||||||| ||||| || |||||||| | |||||||||||| ||||||   ||       | |
 p67  AAAGGGAAGAGTTCATTTGAACTGCCAGCCTCCATCGCAATACAATGTAAACTCATAGAATCTACTCTAAAATGA -1071

gH3.4 TGACATATAAAAATATTTTAGAAACTTAAAAGACATATCCGAATTCGTTGGAATCTAGAACGCTAATCAATGTTA -978
        | || |||| ||||||||||||||||||||||||||| || ||||||||| | |||||||||| | ||  |
 p67  CCAAATCTAAACATATTTTAGAAACTTAAAAGACATATTCGTCTTCGTTGGAGTTTAGAACGCTAAATATTGAAA -996

gH3.4 TTTTTACTAAAATTTGTAACATTTAA.........TCTCATACCTAATATGAACTTCATGTAAACATAGGATTCA -912
      |||  ||| | |||||||||| |||         | |||||||||| ||| |||||||| |  |    ||||
 p67  TTTACACTCACATTTGTAACATATAAATCACATAATTTCATACCTAAAATGTGCTTCATGTTATTACTCAATTCT -921

gH3.4 CTTCAATAATTACAGTTATAGAATGAATAATTAACCAAAATTGATTTCTATAAAATAAATATTT.AACTATTTAG -838
        | ||||||||| ||||| ||| ||| |||||  |||||| | |||||| | |||||||||| || ||||| |
 p67  AATTCATAATTACATTTATATAATTAATTATTAATTAAAATTCAATTCTAT.ATATAAATATTTTAATTATTTGG -847

gH3.4 ATTAGATTTTC.GAATTTATGGTTGTGTTCACAAAAGGCATATCATAACATTTTTTC.TAACGGCGAAAAAATTA -765
        | || ||| ||||| ||    || || ||||  |   |  || | | |||||||  | |  | | ||||||  |
 p67  GCCATATATTCCGAATT.ATAAACGTATTAACAATGGATGTGCCACAGCGTTTTTTCATGAATGAGTAAAAATAA -773

gH3.4 TTAAAAAATTATAAAACTAAAGTTTTTAATGAAA..TATATGAATAATATGTAATTAAAGTATAATTATGGTTAA -692
      |||||||  | |||||  |||||||||||| |||  |||||||||| |||  || |||| ||||| | |||| ||
 p67  TTAAAAACGTGTAAAATCAAAGTTTTTAATAAAAATTATATGAATATTATACAAATAAACTATAAATTTGGTAAA -698

gH3.4 TTAGACTTTTA.GATTCAAATTTATTAAATTGTGTGTAATCCTATACTTAAAAATCTCA.TTTTTTGAATAATTC -619
      |||  ||| |||  ||||  |||||||| || ||||||| | |||  |  ||||||  | | | ||||||  || | |||
 p67  TTAAACTATTATGATTACAATTTATTCAAGTGTGTGTTAACCTCTGGTTAAAATTTTTAATTTTTTTTAAAATTTC -623

gH3.4 AAATTCTACATTTAAATTCATATTATATATGTT.AATTTAGATATATAAATGATGAGTTTAAGTGAAAATTGAGG -545
        |  ||||  |  |||  || |||| | |   | | ||  |  ||  | ||||  |    ||||||||||||| |||
 p67  AG.TTCTTTAACTAATTTTATATAAAAATAGATCAAAATCTAT.GAAAAATCAGATTTTTAAGTGAAAATT.AGG -551

gH3.4 GGAATTTAA.TATGCGATAGATTAATTGTTTGGCTAGCAGAATGAGCTCAAATAAAGAGCTAGCCAATGTCATTA -471
      ||||  ||  |   |     ||  |||||||||||||||||||  |||||||  ||  |||  ||||||  ||||  || |
 p67  GGAAATTGAGCAAATCGTAAATTAATTGTTTGGCTAGCAGAATGAGCTCCAAAAAACGGCTAGCTAATGCTATGA -476

gH3.4 AAGCTCCAATAAACCCAAGATTAGACTCGTTCACTAGATATACTAATAAAATCCATCATTATT.TTTGTGCATTT -397
      |||   |||||| |||||| |||||  ||  |||| |||| |||||||||   ||   | ||   ||  ||||||||||
 p67  AAGGGGCAATAACCCCAAGGTTAGAGTCATTCAGTAGAAATACTAATACGATTAAGCAAATTTATTTGTGCATTT -401

gH3.4 ATCATCGAGTTATTTTCACAAAAAATTATACTAACACACAACTGTATAAGACCGTGTTT.AGTC.TTTTTCCTTA -324
      |   ||  |  |  |||||  |  |||  ||||  |  ||||  |  |  |  |  ||||||  ||  ||  ||||  ||||||  |||
 p67  ACTGTCAAATCATTTTGATAAATAATTGTTTTAACCTATAGCCCTGTAAGACTGTATTCAAGTCATTTTTCGTTA -326

gH3.4 AATTTGCTACCTATATCTGTAAAACACAAGCTTATGTTTTACAAATTACTATAAAAAACAAATAAAACAAAAACA -249
      ||||||  ||  |  |  |  |||||||||  ||  ||  ||||||||||  |||||  ||||  |||||||  ||||  |||
 p67  AATTTGTCACATTTTTATGTAAAACAGAAACTCATGTTTTACACATTACCGTAAATAACAAATTAAACCAAAGTG -251

gH3.4 CACTCGTTTTGAGATAATTTTCTTAATAACAAGTGTTTTAAAAAGGTAGGACTTTATCACCTAAAAGCAGAATAG -189
      ||||||||||||||||||||||||  ||||||||||  |||||||||||||||  ||  |  |||  |  ||||  |
 p67  TACTCGTTTTGAGATAATTTTCTTATTAACAAGTGTTTCAAAAAGGTAGGACTTAATTAGTTAAGACCAGATTTA -191

gH3.4 TCTCAAATGCATTGAGATTAGAGGCTCCCTGATAATTGACTAAA......AATGTTATATTACACAGCTTTTTAT -120
      ||||||||  |  ||||||||   |  ||   |||  ||  |  ||  |      |||||||   |||  |  |||||||
 p67  TCTCAAATTCTCTGAGATTAAGCGGTCTTGGATTATAAAGTAGAATCACTAATGTTA...TACGAA..TTTTTAT -121

gH3.4 TCGATTAA.TTTATAACAATTATTTATAAATATTACTTATCGAGATAGTTTCTTTAAAACTTCATCTTATGCTTA -47
      ||  ||||  ||||||  ||||||||||||||||  ||  ||||  |  ||  |||||||||  |  |||||||  |||||||||
 p67  TCACTTAATTTTATATAAATTATTTATAAATATCACCTATCAAACTAATTTCTTTACACCTTCATCCTATGCTTA -47

gH3.4 GGAGTAACATTGAAATTTAAATTTCATTTTCCAAAACTCAACGATG
      ||  ||||||||||||||  |||||||  |||||  |  |||  |||||||
 p67  GGTGTAACATTGAAATTTGAATTTCACTTTCCGACACTGAACGATG
```

are conserved between the two species they are candidate sequences which might be involved in the transcriptional regulation of both SPAG-1 and p67.

4.2.3. Sequence comparison of the 5' untranslated region of 4 alleles of SPAG-1.

In chapter 3, four cloned macroschizont infected cell lines were identified which contain 4 different SPAG-1 alleles. DNA from these cell line clones was used to PCR amplify a 374 bp fragment which maps to the position of -350 to +24 of the gH3.4 sequence. The primers used for this analysis are Y11 and Y51 and are shown in Figure 19. The PCR products were cloned into pGEM-T and sequenced. The sequence comparison over the first 350 bp upstream of the ATG start codon are shown in Figure 21. The 5' region of H3.4 is identical to that of gH3.4 which is in accordance with the results shown in chapter 3, but of interest is that the upstream sequence of alleles A4.8 and H6.0 is also identical. Only the upstream region of A3.4 is different from that of the other alleles by 8 bases. This sequence comparison indicates that the upstream region of the 4 alleles is highly conserved, probably for gene regulatory purposes. Since this region is conserved at such a high level no DNA sequence motifs could be identified which might be involved in the transcriptional regulation of SPAG-1.

4.2.4. Mapping the beginning of the SPAG-1 mRNA.

To gain more information about the stage-specific regulation of SPAG-1, the beginning of the mRNA transcript of SPAG-1 was mapped using the primer extension method. For this method two primers were designed which are located around the ATG start codon. The first primer, FK1, maps to position bp 13 to -13 and the second primer Y11 maps to bp 46 to 24, as shown in Figure 19. These primers were end-labelled with ^{32}P, annealed to mRNA extracted from infected tick salivary glands and were reverse transcribed. The products were separated by acrylamide gel electrophoresis and visualised by autoradiography. The result is shown in Figure 22. The samples were run adjacent to ^{32}P labelled 100 bp

```
gH3.4    -350   AGTCTTTTTCCTTAAATTTGCTACCTATATCTGTAAAACACAAGCTTATGTTTTACAAAT
A3.4            ------------------------------------------------------------
A4.8            ------------------------------------------------------------
H3.4            ------------------------------------------------------------
H6.0            ------------------------------------------------------------

gH3.4    -291   TACTATAAAAAACAAATAAAACAAAAACACACTCGTTTTGAGATAATTTTCTTAATAACA
A3.4            ------------------------------------------------------------
A4.8            ------------------------------------------------------------
H3.4            ------------------------------------------------------------
H6.0            ------------------------------------------------------------

gH3.4    -231   AGTGTTTTAAAAAGGTAGGACTTTATCACCTAAAAGCAGAATAGTCTCAAATGCATTGAG
A3.4            ----------------------------------------------------C-------
A4.8            ------------------------------------------------------------
H3.4            ------------------------------------------------------------
H6.0            ------------------------------------------------------------

gH3.4    -171   ATTAGAGGCTCCCTGATAATTGACTAAAAATGTTATATTACACAGCTTTTTATTCGATTA
A3.4            -C----------------------------------------G-------------------
A4.8            ------------------------------------------------------------
H3.4            ------------------------------------------------------------
H6.0            ------------------------------------------------------------

gH3.4    -117   ATTTATAACAATTATTTATAAATATTACTTATCGAGATAGTTTCTTTAAAACTTCATCTT
A3.4            ----------------------A----------T------------------------C-
A4.8            ------------------------------------------------------------
H3.4            ------------------------------------------------------------
H6.0            ------------------------------------------------------------

gH3.4     -58   ATGCTTAGGAGTAACATTGAAATTTAAATTTCATTTTCCAAAACTCAACGATG
A3.4            --------------------G----------------------C-----------
A4.8            ------------------------------------------------------
H3.4            ------------------------------------------------------
H6.0            ------------------------------------------------------
```

Figure 21: Sequence comparison of the first 350 bp of the 5' untranslated sequence of SPAG-1 from 4 different alleles. gH3.4 indicates the DNA sequence from Figure 19. A3.4, A4.8, H3.4 and H6.0 represent the 5' DNA sequence obtained from four cloned macroschizont cell lines, representing 4 different SPAG-1 alleles. The number represents the DNA position with reference to the ATG start codon. A line indicates that the base at a given position is identical to the gH3.4 sequence and if a base is different the appropriate base is given. The ATG start codon is underlined.

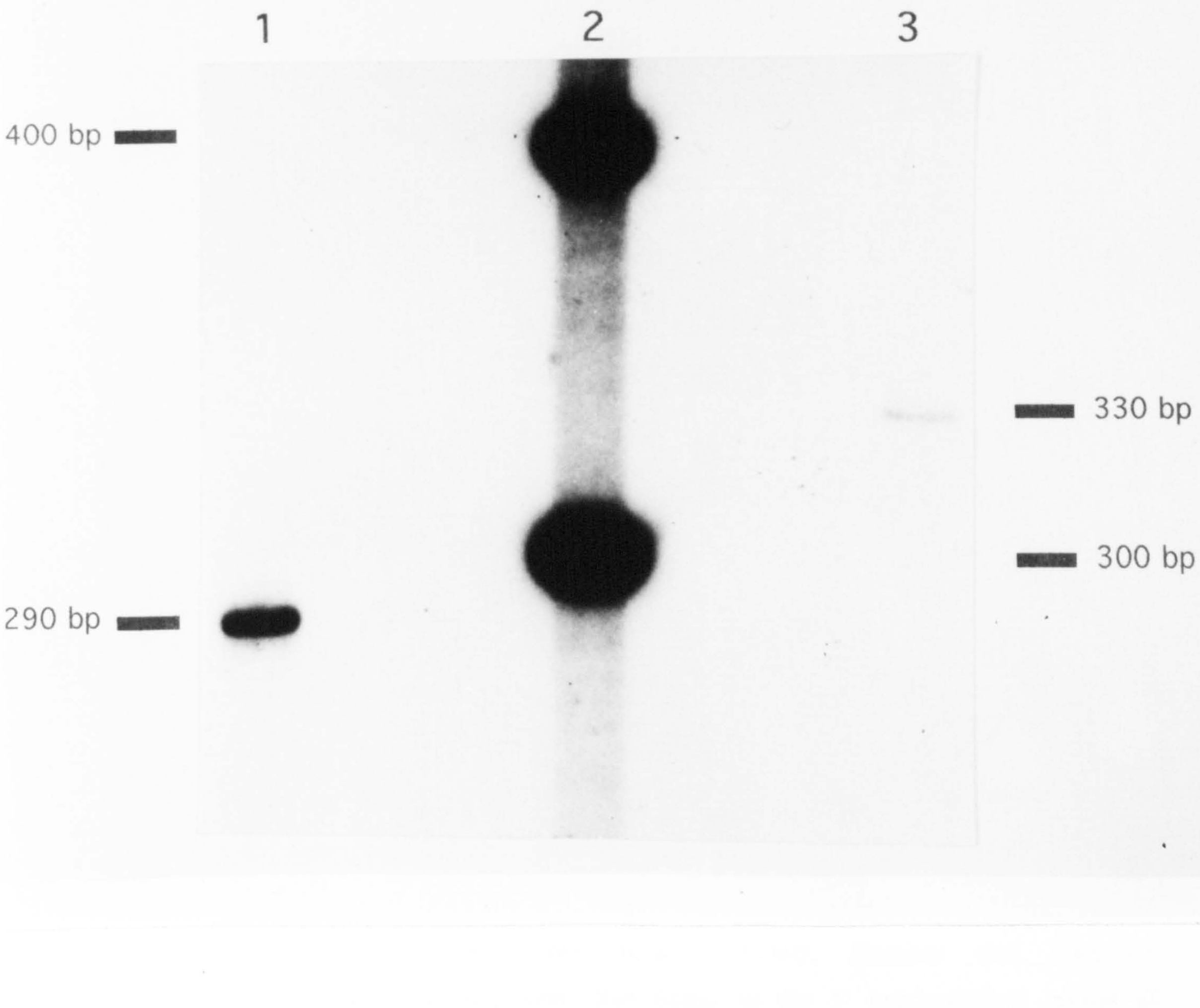


Figure 22: Autoradiograph of the primer extension analysis. Lane 1 shows one band representing the primer extension product using primer FK1. Lane 2 contains molecular weight markers and lane 3 is the primer extension product using primer Y11. The bars next to lanes indicate the size of the observed bands in bp.

markers, to allow an estimation of size of the resulting bands. For each of the reverse transcription reactions there is only one band visible. The size of the band for the reaction containing the primer FK1 is about 290 bp and the band for the reaction containing the primer Y11 has a size of about 330 bp. Taking into account the starting position of each primer, the beginning of the mRNA for SPAG-1 was mapped to a C at 278 bp 5' of the ATG start codon. The transcription initiation site is highlighted by an arrow in Figure 23.

4.2.5. Putative promoter binding sites in the 5' untranslated region of SPAG-1.

Once the beginning of the mRNA for SPAG-1 was mapped it became of interest to look for putative DNA binding sites of proteins which might be involved in the regulation of SPAG-1 expression. The 5' untranslated region of SPAG-1 was searched for an array of DNA sequences which have been shown to be DNA binding sites of proteins and which are involved in the regulation of transcription in other organisms. A list of these sequences and their DNA binding proteins with relevant references is shown in Table 3. Three of these sequences were found in the 5' untranslated region of SPAG-1. A total of 8 TATA boxes were found in the region investigated. The presence of these is not very surprising due to the A/T richness of the *Theileria* genome, but of interest is the finding of the sequence, CAACTG. This sequence in chicken has been shown to be the site to which the Myb oncogene product can bind (Biedenkapp et al., 1988) and it has been shown that the Myb protein is a transcriptional activator which functions by binding to this sequence (Weston and Bishop, 1989). Another two identical sequences, ATGAATAA, were also found in the 5' untranscribed region of SPAG-1. ATGAATAA has been shown to be the site to which Pit-1 binds (Ingraham et al., 1988). The Pit-1 sequence is found in two genes which are expressed specifically in the anterior pituitary gland and it has been shown that the Pit-1 protein binds to this sequence (Ingraham et al., 1988). Subsequently, transformation experiments showed that the Pit-1 sequence is an effective promoter sequence in pituitary cells (Esholtz et al., 1990). The 5' sequence of p67 was searched for the existence of these

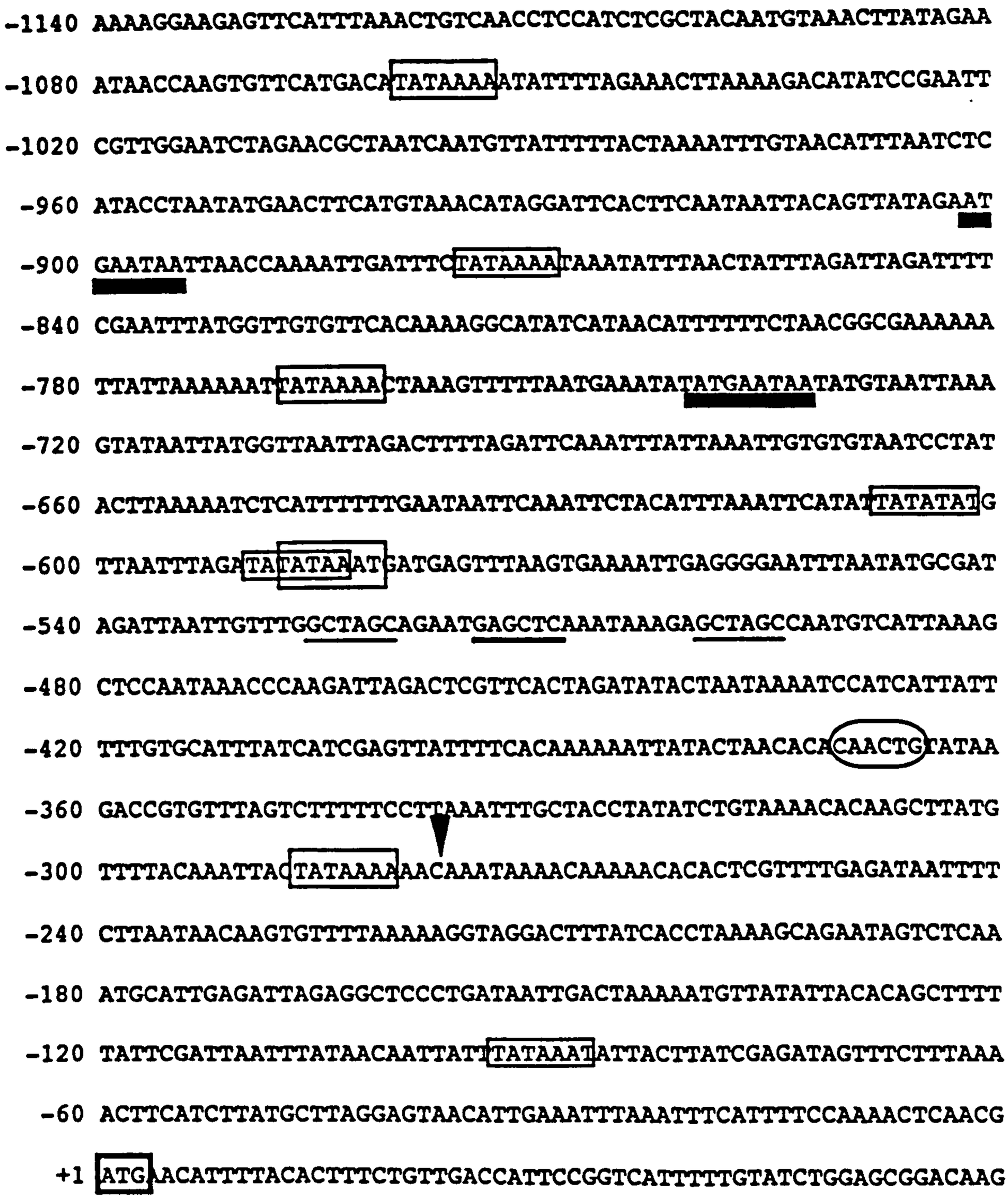


Figure 23: Sequence motifs in the 5' untranslated region of SPAG-1. The 5' untranslated region of SPAG-1 gH3.4 is shown. The numbers represent the sequence position with reference to the ATG start codon. The ATG start codon is in a thick lined box. The transcription initiation site is marked by an arrow head. 8 sequences which fulfil the criteria of a TATA box are boxed with thin lines. The sequence which is in accordance to the consensus sequence of the Myb binding site is circled and two sequences identical to the consensus sequence for a Pit-1 binding site are underlined with a thick line. The two sequences which are underlined with a thin line and the sequence which is underlined with a medium thickness line were identified in a sequence comparison of 5' region of p67 and SPAG-1, their significance is discussed in the text.

sequence motifs. None of the Pit-1 sequences were intact nor was the Myb binding site and only one of the eight TATA boxes was conserved between p67 and SPAG-1. This TATA box lies between the transcription initiation site and the ATG start codon, therefore it is unlikely to be involved in the regulation of SPAG-1 expression.

Attempts to isolate DNA binding proteins that bind to the first 371 bases of the 5' untranslated region of SPAG-1 by screening a λgt11 expression library containing *T. annulata* genomic DNA failed. Therefore no novel or known DNA binding proteins were identified which bind to the 5' untranslated region of SPAG-1 and which could be involved in the regulation of SPAG-1 transcription. The regulatory mechanism of SPAG-1 and p67 transcription therefore remains unknown.

4.2.6. Evidence for an intron in SPAG-1.

When the SPAG-1 sequence, gH3.4, which is derived from a genomic *T. annulata* Hisar DNA stock, was compared to the published cDNA sequence, cH, of SPAG-1 in the previous chapter, a putative intron was recognized and marked as such in Figure 7. The putative intron is based on the observation of a sequence gap of 30 bp in the cH sequence in the comparison between the two sequences. The presence of these 30 bp in the gH3.4 could of course be explained by an insertion in the gH3.4 sequence or a deletion in the cH sequence. As the length of this sequence difference is a multiple of 3 base pairs, it does not interrupt the open reading frame, even if it is an intron. If this sequence difference is due to an intron, it is therefore a cryptic intron. It was decided to investigate further whether this sequence difference is due to a cryptic intron.

Four different SPAG-1 alleles were identified in chapter 3 and one of these, allele A3.4, was shown to be identical to the published cDNA sequence, cH, over the two regions investigated. The pGEM-T clone, a series 1 clone from chapter 3, containing genomic DNA for the A3.4 allele was sequenced across the region containing the putative intron

using site specific primers, 420 and 710. The sequence of the putative intron from the A3.4 allele was aligned to the cH sequence and is shown in Figure 24. The series 1 pGEM-T clones for the other 3 alleles were also sequenced across the putative intron and their sequences are also shown in Figure 24. It can clearly be seen that all sequences, based on genomic DNA, contain the sequence of the putative intron and that cH does not contain this sequence. A3.4 is identical to cH in all sequence comparisons made so far apart from the presence of the putative cryptic intron. It therefore provides strong supporting evidence that the 30 bp comprises a cryptic intron. Interestingly, this intron is found in all 4 alleles and they are identical. This sequence conservation indicates that most bases in this short cryptic intron are probably needed for the excision of the intron during mRNA processing.

```
cH      1595   GGACAAATGG|.............................|GTGAAGGAGA   1614
gH3.4   1596   GGACAAATGG|gtttgttatattcaagtaaatcttttgtag|GTGAAGGAGA   1645
A3.4           GGACAAATGG|gtttgttatattcaagtaaatcttttgtag|GTGAAGGAGA
A3.8           GGACAAATGG|gtttgttatattcaagtaaatcttttgtag|GTGAAGGAGA
H3.4           GGACAAATGG|gtttgttatattcaagtaaatcttttgtag|GTGAAGGAGA
H6.0           GGACAAATGG|gtttgttatattcaagtaaatcttttgtag|GTGAAGGAGA
```

Figure 24: The SPAG-1 intron. A sequence comparison of the cDNA sequence cH of SPAG-1 and the corresponding sequence from 4 SPAG-1 alleles (A3.4, A4.8, H3.4 and H6.0) and the genomic SPAG-1 sequence gH3.4. The dots represent sequence gaps in the cDNA sequence and the sequence in lower case letters represents the sequence of the intron. The line indicates the start and the end of the intron. The numbers represent the position of the beginning and the end of the sequence shown in the cDNA and the genomic sequence respectively.

4.2.7. Confirmation of intron by S1 mapping.

To provide more supporting evidence that the 30 bp are a cryptic intron, the method of S1 mapping was chosen. A *Bam*HI-*Acc*I fragment of the pUC 18 plasmid, described in chapter 3, containing most of the genomic copy of the gH3.4 SPAG-1 gene was cloned into M13 mp18 and M13 mp19. Both these clones were grown in the presence of organic ^{32}P in an otherwise normal overnight culture. Subsequently, the single stranded DNA was extracted for both M13 clones. The M13 inserts map to the position 1201 to 2179 on the gH3.4 sequence and the putative intron maps to the position 1606-1635. The single stranded DNA was then annealed to sporozoite mRNA extracted from infected tick salivary glands. The RNA-DNA hybrid was treated with the S1 nuclease which digests single stranded DNA and single stranded RNA but not RNA-DNA hybrids. The products were denatured and separated on a 6% polyacrylamide sequencing gel next to a labelled marker track and the results were visualised by autoradiography. A picture of the autoradiograph is shown in Figure 25.

In Figure 25 it can clearly be seen that there are 2 bands in the track containing the S1 treated RNA-M13 mp18 hybrids while no bands were observed in the track containing the negative control, S1 treated RNA-M13 mp19 hybrids. If no intron was present in the DNA fragment one would expect only a single band but two bands indicate the presence of one intron. If one subtracts the total size of the two bands from the size of the initial DNA fragment one gets an estimated intron size of about 30 bp which coincides with the predicted intron size from the sequence comparison described above. Furthermore, by matching the sizes of the bands with the starting and finishing position of the initial DNA fragment one can predict the position of the intron to be in either of two sites. From this the intron is either at position 1606-1635 or at 1744-1773. The first of these is in accordance with the predicted intron found by sequence comparison of the cDNA and genomic DNA sequence of SPAG-1 as well as the predicted position from the DNA sequences of the macroschizont clones.

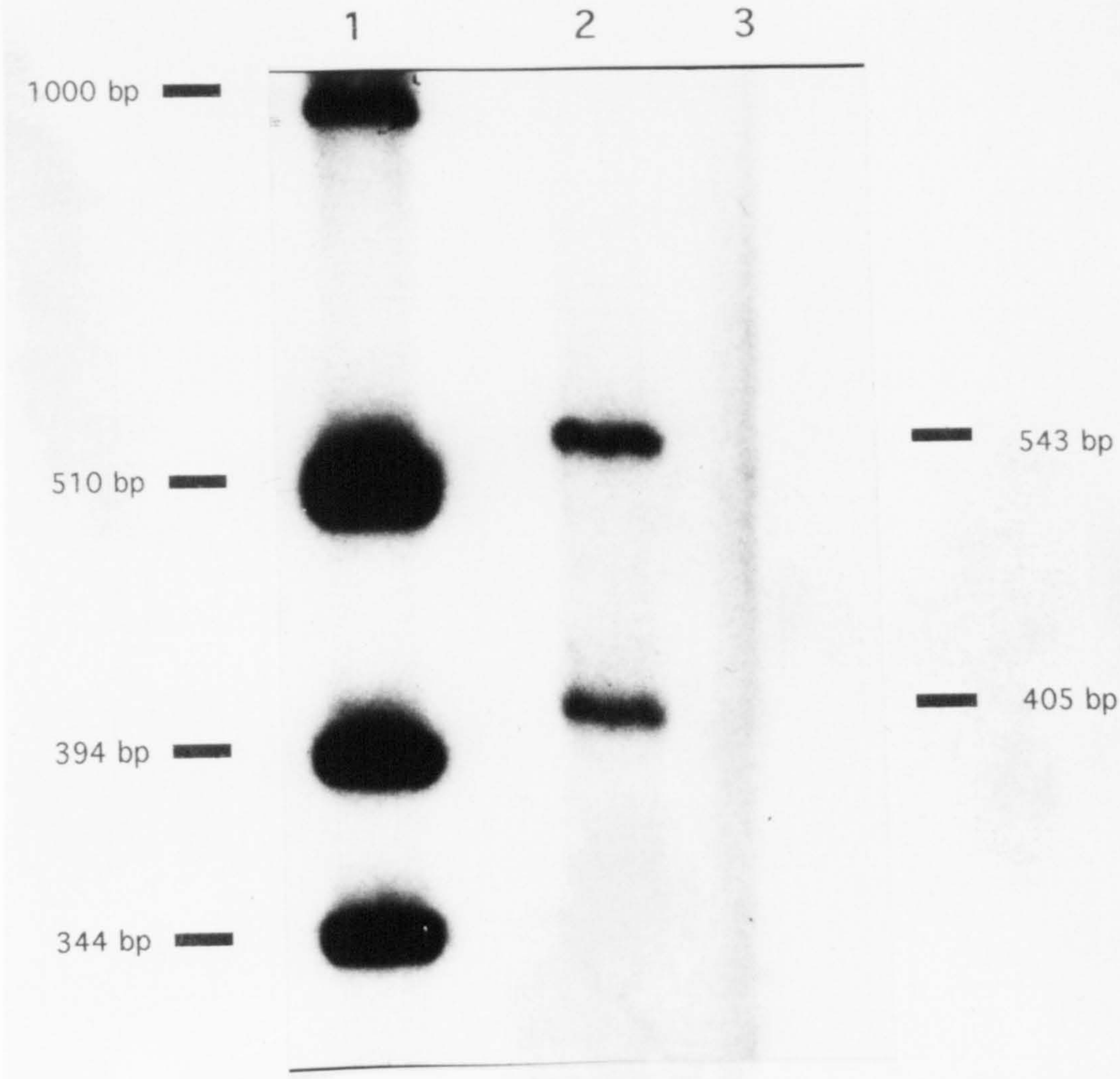


Figure 25: Autoradiograph of the S1 nuclease intron mapping experiment. Lane 1 contains molecular weight markers, lane 2 contains the product of S1 nuclease treated DNA-RNA hybrid and lane 3 contains the negative control. The bars next to lane 1 and 3 represent the molecular weight of the observed bands in bp.

4.2.8. Comparison of the SPAG-1 intron to other *Theileria* introns.

In total only three introns have been identified in *Theileria*. The reason for this is that only a few genes have been cloned and fully sequenced and in most cases the sequence information is based only on a cDNA sequence, and introns would not therefore have been detected. Two further genes have been analysed for introns, the hsp 70.1 gene (Mason et al., 1989) and the cysteine protease gene of *T. annulata* (Baylis et al., 1992) and it has been shown that neither contain introns. The sequence of the SPAG-1 intron has been aligned to the intron sequences of the two other identified *Theileria* introns, from the p67 gene (Nene et al., 1992) and the cysteine protease intron of *T. parva* (Nene et al., 1990). The sequence comparison can be seen in Figure 26. All three introns are short, varying from 28 to 32 bp. The comparison of these three introns allow the prediction of a short consensus sequence for the recognition and excision of introns in *Theileria*. The predicted consensus sequence is **G|gttngtt ... ttntag|G**. To verify this prediction more introns need to be identified in *Theileria* genes and these also need to be aligned to the existing sequences.

```
T.a.    SPAG-1                AAATGG|gtttgttatattcaagta.aat.cttttgtag|GTGAAG
T.p.    p67                   --TCA-|-------.-t--.g----.t--.ggg--t---|-CA-TA
T.p.    cysteine protease     ---A--|---a----c-ca--ct--at--ta----a---|----T-

Consensus                     G|gttngtt                              ttntag|G
```

Figure 26: Sequence comparison of introns found in *Theileria*. The sequence in capital letters represents the coding sequence flanking the intron, the bar marks the start and end of the intron and the sequence of the intron is written in lower case letters. The . represents a gap, the - represents identical bases, *T.a.* and *T.p.* represent *T. annulata* and *T. parva* respectively.

4.3. Discussion

4.3.1. Stage-specific regulation of SPAG-1.

It has been shown that SPAG-1 is transcriptionally regulated and is only expressed during sporozoite development when an infected tick is feeding on its host (Williamson et al., 1989). The SPAG-1 gene provides an opportunity for investigation of stage-specific gene regulation in the sporozoite stage of *Theileria*, since this gene is under tight stage-specific transcriptional control and no factors involved in regulation are known. Studying the 5' untranslated region of SPAG-1 was thought to be likely to reveal the identity of one or more of these factors.

The sequence comparison between the first 350 bases of the 5' untranslated region of the four SPAG-1 alleles did not reveal any motifs which might be involved in the transcriptional regulation of SPAG-1. The region analysed contains only about 60 bases upstream of the transcription initiation site since this has subsequently been mapped to position -278.

Both SPAG-1 and p67 are under transcriptional control and are expressed during the same life cycle stage, therefore it can be hypothesised that homologous transcription factors are involved in the regulation of both. If homologous transcription factors are involved, then the DNA sequences they bind to must also be conserved in the 5' untranslated region of both genes. The comparison of the 5' sequences of the two genes led to the identification of two G/C rich, palindromic sequences. The sequences are GCTAGC and GAGCAC. Unfortunately these sequences have not been identified previously to be sites which are involved in protein-DNA interactions. Therefore one can only speculate that these sequences might be involved in the stage specific regulation of SPAG-1 and p67. Further experiments are needed to investigate the role of these sequence motifs in the regulation of both genes.

It was thought to be likely that by identifying the site for transcriptional initiation in SPAG-1 one might gain more information about the process of stage specific transcriptional regulation during the sporozoite stage. The site of transcriptional initiation was mapped to 277

bases 5' of the ATG start codon and is in accordance with the transcription initiation site consensus sequence Py Py C A Py Py Py Py Py (Corden et al., 1980). The immediate 5' region to the initiation site does not bear any homology to other known promoter binding sites. A search for sequence motifs, which in other organisms have been shown to be sites recognized by transcription factors resulted in the identification of only one conserved TATA box between SPAG-1 and p67. As this TATA box is located between the transcription initiation site and the ATG start codon, it seems unlikely that even this motif is involved in the regulatory process.

Other TATA boxes are however present in the p67 5' untranslated region which are absent in SPAG-1. Since the genome of *Theileria* is A/T rich it is not surprising that such sequences would occur frequently and because of this it is very unlikely that any of these are involved in stage specific gene regulation. Similarly there are 7 TATA boxes, two Pit-1 and one Myb binding sites in SPAG-1 which are not conserved in p67. It therefore seems likely that none of these sites are of importance for the regulation of SPAG-1 unless p67 and SPAG-1 employ different transcription factors for the regulation of the two genes. This seems unlikely since these genes are closely related. This leaves only the two sequence motifs GCTAGC and GAGCTC as putative regulatory sequences.

In *Plasmodium*, a sequence comparison of the 5' region of the CS gene and that of the GBP130 gene revealed that these sequences contain a region which is related to the SV40 enhancer region (Lanzer et al., 1990; Lanzer et al., 1992). When the 5' region of the SPAG-1 gene was searched for this sequence no region with significant homology could be found.

It can be concluded that it is most likely that one or more novel transcription factors are involved in the regulation of SPAG-1 transcription. To identify which parts of the upstream region of the SPAG-1 gene are of importance in its regulation, one would ideally perform transfection studies with a range of deletions of the 5' untranslated region of SPAG-1 linked to a reporter gene. Unfortunately no transfection system is yet available for *Theileria*. Another possible way to identify binding regions is to use nuclear extracts from

sporozoites and conduct band shift assays to identify specific sequences to which these nuclear extracts bind. The limitation of these experiments is the supply of sporozoite material. To determine whether the two motifs, GCTAGC and GAGCTC, identified by sequence comparison of the 5' region of p67 and SPAG-1, are transcription factor binding sites, these could be used to screen an expression library to isolate sequence specific DNA binding proteins. The disadvantage of this approach is that if no proteins are isolated, no conclusion can be drawn from the experiments since the transcription factor may need processing, folding or polyadenylation to make it functional. Even if a protein which can bind to a certain motif has been isolated it does not provide evidence that it is actually a transcription factor. Ultimately, one has to rely on a transfection system to provide this evidence. Since stable transformation has been achieved in *Plasmodium* (Wu et al., 1995), it might not be much longer until stable transformation of *Theileria* will be achieved.

4.3.2. The SPAG-1 intron.

In this thesis I provide three independent pieces of evidence for the existence of a single, 30bp long, cryptic intron in the SPAG-1 sequence. The intron was mapped to the position 1616 to 1635 on the genomic DNA sequence gH3.4. The first indication for the existence of the intron is based on a sequence comparison of the published cDNA sequence of SPAG-1 to the gH3.4 sequence of SPAG-1. One of these experiments is the S1 nuclease mapping. The result of this indicates that the intron is not just only present in the gH3.4 allele but also in the H6.0 allele and if there is an H4.8 allele then also in this allele. The reason for this is based on the observation that there was no full-length band visible in Figure 25, therefore one can conclude that all SPAG-1 RNA species from this mixed parasite stock contain an intron at the predicted position.

Analysis of the flanking regions of the only three introns so far identified in *Theileria* resulted in a consensus sequence which is in accordance of that for other eukaryotes but is still more elaborate than

that of Perlman et al. (1984). The consensus sequence is in accordance with the GT-AG rule for splice sites found in introns isolated in *Plasmodium* (Ravetch et al., 1984; Wesseling et al., 1989; Adams et al., 1992). Therefore it can be concluded that it is unlikely that novel factors are involved in intron splicing in the apicomplexan parasites *Theileria* and *Plasmodium*.

Chapter 5

Binding studies with recombinant sporozoite antigens.

5.1 Introduction.

In order to develop vaccines or treatments for tropical theileriosis which are sporozoite based, it is of great importance to understand what interactions occur between the sporozoite and host cells and how the parasite invades a host cell. Thus by studying the function of sporozoite antigens one might provide some information which could be used in the prevention of host cell invasion.

5.1.1. Are SPAG-1 and SPAG-2 ligands for host cell recognition/invasion?

The first indication that SPAG-1 and SPAG-2 might play a role in the process of host cell recognition and invasion is based on the observation that sporozoite entry into host cells can be blocked *in vitro* if the parasite is incubated with either of the monoclonal antibodies 1A7 or 4B11 prior to invasion (Williamson, 1988). There are two possible explanations as to how the monoclonal antibodies could block the invasion of the sporozoite into the host cell: 1) the effect is purely steric in the sense that the sporozoite surface is covered by so many antibodies that it prevents its membrane binding closely to the host cell surface and thus inhibits the invasion process or 2) the antibody binds directly to the molecule involved in the binding process, either directly at the binding site or very close to it and thus blocks invasion. The observation that other monoclonal antibodies reacting with the middle region of SPAG-1 do not block the invasion process (Williamson, 1988) seems to indicate that SPAG-1 is directly involved in either the recognition or the invasion process. Bovine serum raised against the C-terminus of SPAG-1,

but which does not react with the 1A7 epitope, also inhibits sporozoite invasion of host cells (Boulter et al., 1994). This observation indicates that, if SPAG-1 is involved in the process of host cell recognition or invasion, its C terminus is important. A similar experimental approach, using monoclonal antibodies to block invasion, has highlighted the importance of the MHC class I molecule on the host cell surface in the invasion of *T. parva* sporozoites (Shaw et al., 1991).

Further support for the theory that SPAG-1 is a ligand for host cell recognition was based on the fact that there are three VGVAPG peptides in the predicted amino acid sequence of the SPAG-1 cH allele (Hall et al., 1992 and Hall, 1994), at least before the results of chapter 3 were obtained. The VGVAPG peptide is also found in bovine elastin (Raju and Anwar, 1987), in which it has been identified as the ligand which binds specifically to the elastin receptor (Blood et al., 1988; Mecham et al., 1989; Robert et al., 1989). This therefore led to the speculation that the sporozoite might recognise a host cell via the VGVAPG peptide in the SPAG-1 molecule (Hall et al., 1992). Interestingly, the elastin receptor has been found on macrophages (Huard et al., 1986) and monocytes (Senior et al., 1984; Blood et al., 1988; Mecham et al., 1989) which are the main target cells for *T. annulata* sporozoites (Glass et al., 1989 and Spooner et al., 1989), and the receptor is also found on fibroblasts (Hinek et al., 1988 and Wrenn et al., 1988) which can also be infected by *T. annulata* (Brown and Gray, 1981 and Morrison et al., 1986). However the suggested role of the VGVAPG motif was largely invalidated once the gH3.4 sequence was obtained and found not to contain it (Katzer et al., 1994; chapter 3).

SPAG-1 is stage-specifically regulated (Williamson et al., 1989). Its mRNA is only transcribed during the early stages of sporozoite development and transcription stops once the sporozoite is mature. In the mature sporozoite the processed SPAG-1 is found on the surface (Williamson et al., 1989 and Knight, 1993). Unfortunately nothing is known about the fate of SPAG-1 either during the entry process or once the entry process has been completed. The *T. parva* homologue, p67, has been shown to be shed from the sporozoite surface during the invasion process (Dobbelaere et al., 1985). But one can speculate that a gene which is under such tight regulation and is expressed on the surface of

the sporozoite might function in host cell recognition or invasion. Little is known about SPAG-2; as stated above, it has only been partially sequenced and one can only speculate about its function. But since the monoclonal antibody 4B11, which reacts with SPAG-2, blocks the invasion of sporozoites into host cells, this antigen might also be involved in the process of host cell recognition and/or invasion.

To investigate whether either SPAG-1 or SPAG-2 are involved in the recognition of host cells one has to ascertain if either molecule binds specifically to cells which can be invaded by *T. annulata* sporozoites but not to those which are not a target for invasion. In order to study whether SPAG-1 or SPAG-2 are involved in host cell recognition or invasion, recombinant proteins have been expressed and employed in binding studies.

5.2 Results

5.2.1. Cloning and sequence data for SPAG-1 and the SPAG-2 constructs.

To study the involvement of the SPAG molecules in the process of host cell recognition and invasion, recombinant protein was expressed in bacteria using the pGEX expression system (Smith and Johnson, 1988). The advantage of this system is that large quantities of protein can be expressed and easily purified. A further advantage for these particular studies was that the two genes coding for SPAG-1 and SPAG-2 were already cloned into this vector system for a different purpose.

The pGEX-2.7 plasmid clone was kindly given to me by Dr. Hall. This plasmid contains a 2667 bp insert derived from the SPAG-1 gene. pGEX-2.7 was cloned by introducing an artificial *Eco* RV site, via site directed mutagenesis, into the cDNA sequence of the SPAG-1 gene. The gene was then cut with *Eco* RV and *Eco* RI and cloned into a *Sma* I - *Eco* RI cut pGEX-3X vector (Boulter et al., 1994). The resulting gene product

lacks the first 19 amino acids of SPAG-1 and the twentieth has been mutated from lysine to isoleucine but is otherwise the complete protein. The first 18 amino acids which are not encoded by the pGEX-2.7 construct are predicted to be a signal peptide (Hall et al., 1992). The junctions of this construct have been sequenced in order to confirm the validity of the construct (Boulter et al., 1994).

Four plasmids containing parts of the SPAG-2 gene were given to me by Dr. Knight. These plasmid constructs are presented diagrammatically in Figure 27. The first plasmid contains the KP8 insert, which is the original SPAG-2 fragment (Knight et al., 1993), cloned into pGEX-1λT. KP8 is not the full length gene for SPAG-2 since it consists only of 980 bp while the whole gene product is predicted to be 150 kDa. The second plasmid contains the *Hinc* II fragment of SPAG-2, called KP27 and is cloned into Bluescript SK$^+$. This fragment maps to the middle of KP8 and is 515 bp long but due to the cloning procedure has lost its flanking restriction sites. The third plasmid contains the *Eco* RI - *Xba* I fragment (B14) of KP8 and is cloned into pGEM 5Zf$^+$. This fragment consists of the first 900 bp of KP8. Finally the last plasmid given to me contains the last 260 bp of KP8 (A9), and was made from a deletion of KP8, and is cloned into pGEM 5Zf$^+$.

At the time when KP8 was given to me, it had not been fully sequenced and only the first 252 bp were available (Knight, 1993). As a priority I thus decided to sequence completely KP8 using double stranded DNA from the four plasmids as templates. The sequence information obtained from these plasmids is shown diagrammatically in Figure 28. Apart from 95 bases within the *Hinc* II fragment the KP8 5' to 3' DNA strand was sequenced fully. Four hundred and seventy-eight bases were obtained from the 3' to 5' DNA direction. These overlapped with the sequence from the 5' to 3' direction. The KP8 sequence I obtained was compared to the KP8 sequence obtained subsequently on both strands by Dr. Knight at the University of Glasgow. The finalised DNA sequence based upon both sets of data and its predicted amino acid sequence is shown in Figure 29. At the same time as the sequence information of KP8 was obtained the junctions of the KP8 insert in pGEX-1λT were sequenced. The sequence information of the junctions of the KP8 insert

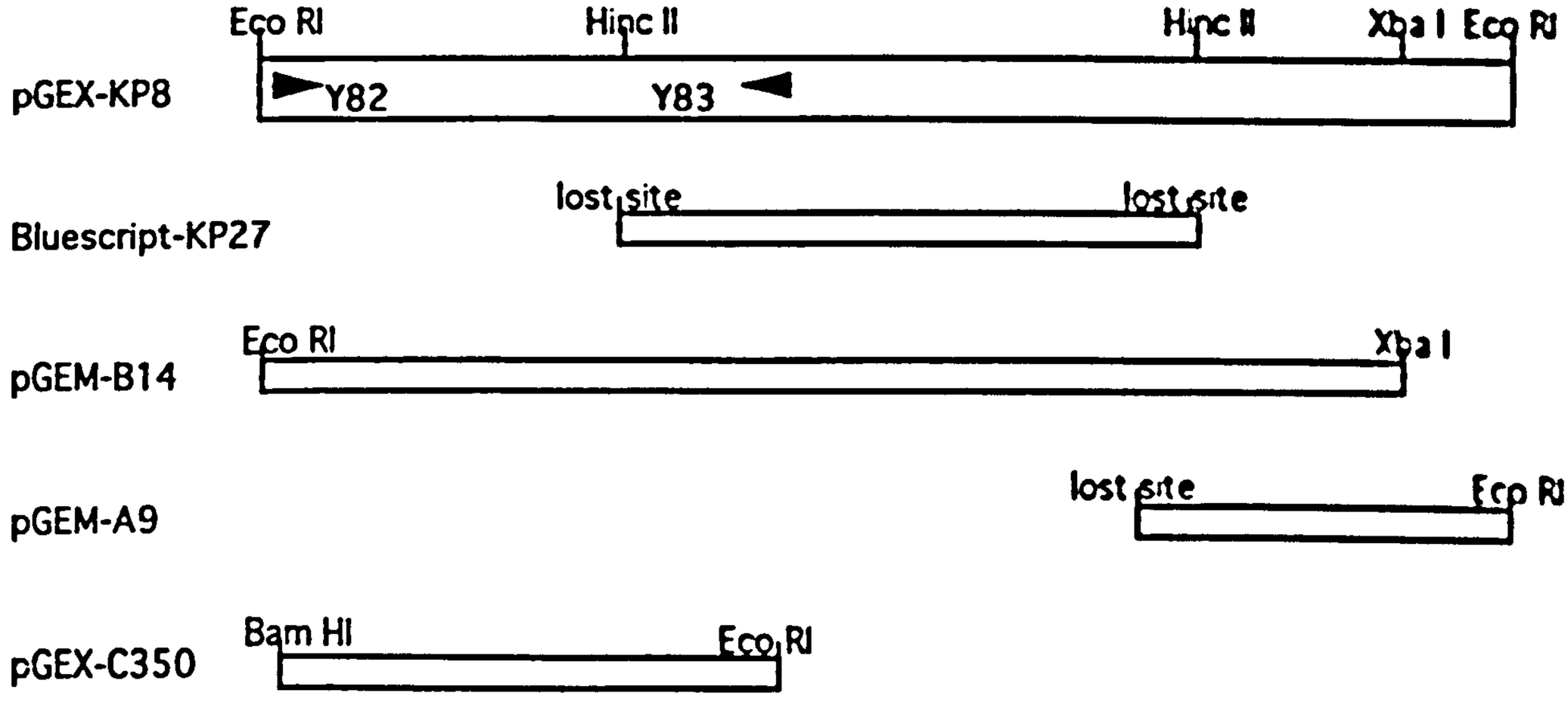


Figure 27: Diagrammatic representation of the SPAG-2 constructs. The name and vector for each of the constructs is shown at the left. KP-8 in pGEX is the largest isolated fragment of the SPAG-2 gene. Some of its restriction sites are marked. The other SPAG-2 constructs are shown according to their position, and their flanking restriction sites are shown. 'Lost site' indicates that these restriction sites were lost during the cloning event. The arrows, marked either Y82 or Y83, indicate the location of the primers used to PCR amplify the DNA fragment which was used to clone pGEX-C350.

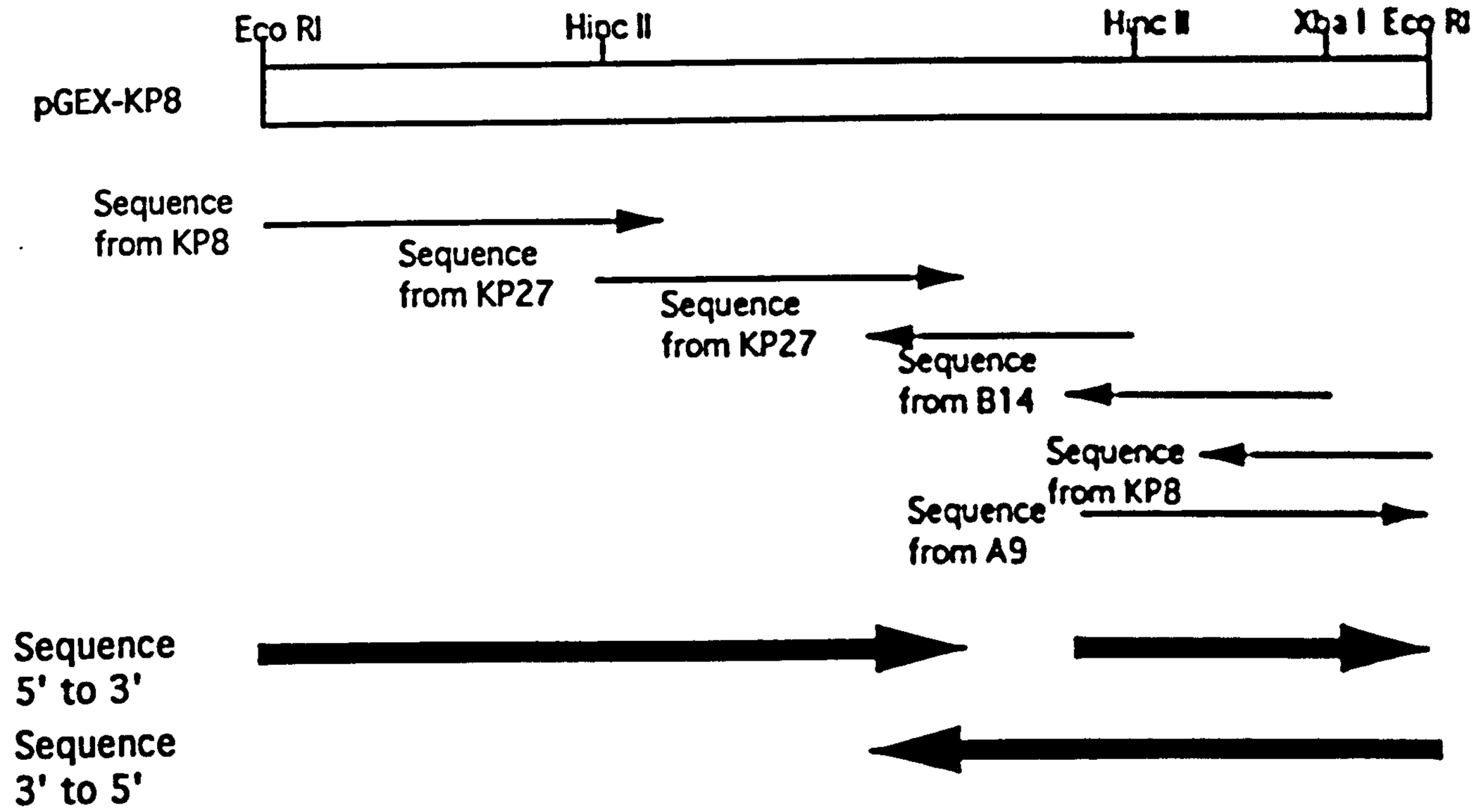


Figure 28: Diagrammatic representation of the SPAG-2 sequence data. KP-8 represents the largest isolated fragment of SPAG-2. The thin arrows beneath indicate the length and orientation of the sequence information obtained from the construct mentioned to its left. The dark arrows summarise the total sequence information obtained for each orientation.

```
   1  GAATTCGGCTCGAGAGATAGCGCCAAGAAAACAAGTGACCATGAGACAAAAAGAAAGTAAA
      E  F  G  S  R  D  S  A  K  K  T  S  D  M  E  T  K  E  S  E

  61  GACCATAGAGAAAGAGAGTACAAATACAATAACAAAGATGATAATTCCAAAGATTACGAA
      D  H  R  E  R  E  Y  K  Y  M  M  K  D  D  M  S  K  D  Y  E

 121  TGCATCGACTCTGAAGCAATCAAGGCAGTAGTGGAAAAGGCAGTTATAGAAGCATTTGAC
      C  I  D  S  E  A  I  K  A  V  V  E  K  A  V  I  E  A  F  D

 181  AAGTGCCTGTCAGAAAAAATTAAGGGTGAAGAAACATGTCTAACAGACTCAGTAAACCGA
      K  C  L  S  E  K  I  K  G  E  E  T  C  L  T  D  S  V  N  R

 241  GAGTACACTTTTGATAAAAATGACACAACACTAAGCTATCAGGAGCAAAGTCAACTTTAT
      E  Y  T  F  D  K  N  D  T  T  L  S  Y  Q  E  Q  S  Q  L  Y

 301  TCCCGTGATTATCTCCAAAGACTCACAAATGACCAAAATAACCAGGGAACTTCGGATATG
      S  R  D  Y  L  Q  R  L  T  N  D  Q  M  M  Q  G  T  S  D  M

 361  AACCGATATGGAAAAAATGATTTTAAAGAAAATAATGAAAATGATACCCATGGTCTAAAA
      N  R  Y  G  K  N  D  F  K  E  N  N  E  M  D  T  H  G  L  K

 421  CCATTTGAAAAAGATTTCTCAAACGTGGAGAATCGAGAATATTCTAAAGATAAAACAGTT
      P  F  E  K  D  F  S  N  V  E  N  R  E  Y  S  K  D  K  T  V

 481  GAAAGCAATAACTTTGAGTTTAAAGGGAAGATAATTTCAAATTCAACAGCATACGCAAAC
      E  S  N  N  F  E  F  K  G  K  I  I  S  N  S  T  A  Y  A  N

 541  ACACAAGAAAACCCATATGTAACTGATCATATGAGCCAGAATGTGTATGAGGAGATTTCA
      T  Q  E  N  P  Y  V  T  D  H  M  S  Q  N  V  Y  E  E  I  S

 601  TATGAGAATAAAACTGATGAATTTGATAGTGAATCGTTTAATTTTCAATTCAACAAGGGG
      Y  E  N  K  T  D  E  F  D  S  E  S  F  N  F  Q  F  N  K  G

 661  AGCCAATTTGAATACAATTTTGATGGAGTTGATTTGTTTTCAAATAGAGACCAAACAGAT
      S  Q  F  E  Y  N  F  D  G  V  D  L  F  S  N  R  D  Q  T  D

 721  AGTTTATTCTACCCAGATGAAGAACAATTCAATAGTTTGAATAATAATGTAGCTGTAAGA
      S  L  F  Y  P  D  E  E  Q  F  N  S  L  N  N  N  V  A  V  R

 781  GATGACTTCGCCCGTGCGAGAAATGTATGATAAGGAACAACAAGATAACTTCTCGGAACAG
      D  D  F  A  V  G  E  M  Y  D  K  E  Q  Q  D  N  F  S  E  Q

 841  TTGACCAACCAAGATGATAAATCGGAACTCTTATGCTCACAAACGTTTGAAGGAAATGAT
      L  T  N  Q  D  D  K  S  E  L  L  C  S  Q  T  F  E  G  M  D

 901  CTAGATGATAAAAATTTTGATCAAACATACACACTTGGAAACTGTTTCGCCAATTTCACT
      L  D  D  K  N  F  D  Q  T  Y  T  L  G  N  C  F  A  N  F  T

 961  CGAGCCGCTCGAGCCGAATTC  981
      R  A  A  R  A  E  F
```

x

Figure 29: DNA sequence of SPAG-2 (KP8). The predicted amino acid sequence is shown below. The sequence shown is based on the comparison of sequence information obtained by Dr. Knight and F Katzer.

in the pGEX vector confirmed that the predicted open reading frame is intact. Furthermore, by sequencing pGEX-KP8 using double stranded DNA the validity of the construct was confirmed.

Finally, another SPAG-2 construct was created for expression in the pGEX system using a PCR approach. This construct is called pGEX-C350 and was created to determine which part of SPAG-2 harbours the ligand responsible for binding to peripheral blood mononuclear cells. The locations of the primers used to PCR amplify bases 10 to 361 of the KP8 sequence are shown in Figure 27 and the sequence of the primers is shown in Table 7 in the Material and Methods section. Bases 10 to 361 of KP8 were amplified by PCR, cloned into pGEM-T, checked by sequencing, cut out of pGEM-T via *Bam* HI and *Eco* RI digestion (these restriction sites were built into the primers) and cloned into pGEX-2T. The junctions of the resulting expression plasmid pGEX-C350 were sequenced in order to confirm that the plasmid contains the correct insert and that the predicted open reading frame is intact.

5.2.2. Expression of SPAG-1 and the SPAG-2 constructs.

The three pGEX expression plasmids pGEX-2.7, pGEX-KP8 and pGEX-C350 were expressed in *Escherichia coli*, producing three GST fusion proteins, GST-SPAG-1, GST-KP8 and GST-C350 respectively. The fusion proteins were purified on glutathione 4B sepharose columns and were analysed by electrophoresis on SDS polyacrylamide gels. Gels showing the fusion proteins are shown in Figure 30. The GST-SPAG-1 fusion protein has an apparent molecular mass of 145 kDa although the expected size is only 112.8 kDa (Boulter et al., 1994). The GST-SPAG-2 protein expressed by pGEX-KP8 has a molecular mass of 58 kDa and the product of pGEX-C350 consisting of GST and the first 116 amino acids of the known SPAG-2, has a molecular mass of 41 kDa. It can be seen that all three constructs produce one major protein product, as visualised by SDS PAGE, at a size slightly larger than expected and they also contain smaller protein bands which may be degradation products. In all lanes there is another protein band visible at about 70 kDa which I believe is due to bacterial contaminantion.

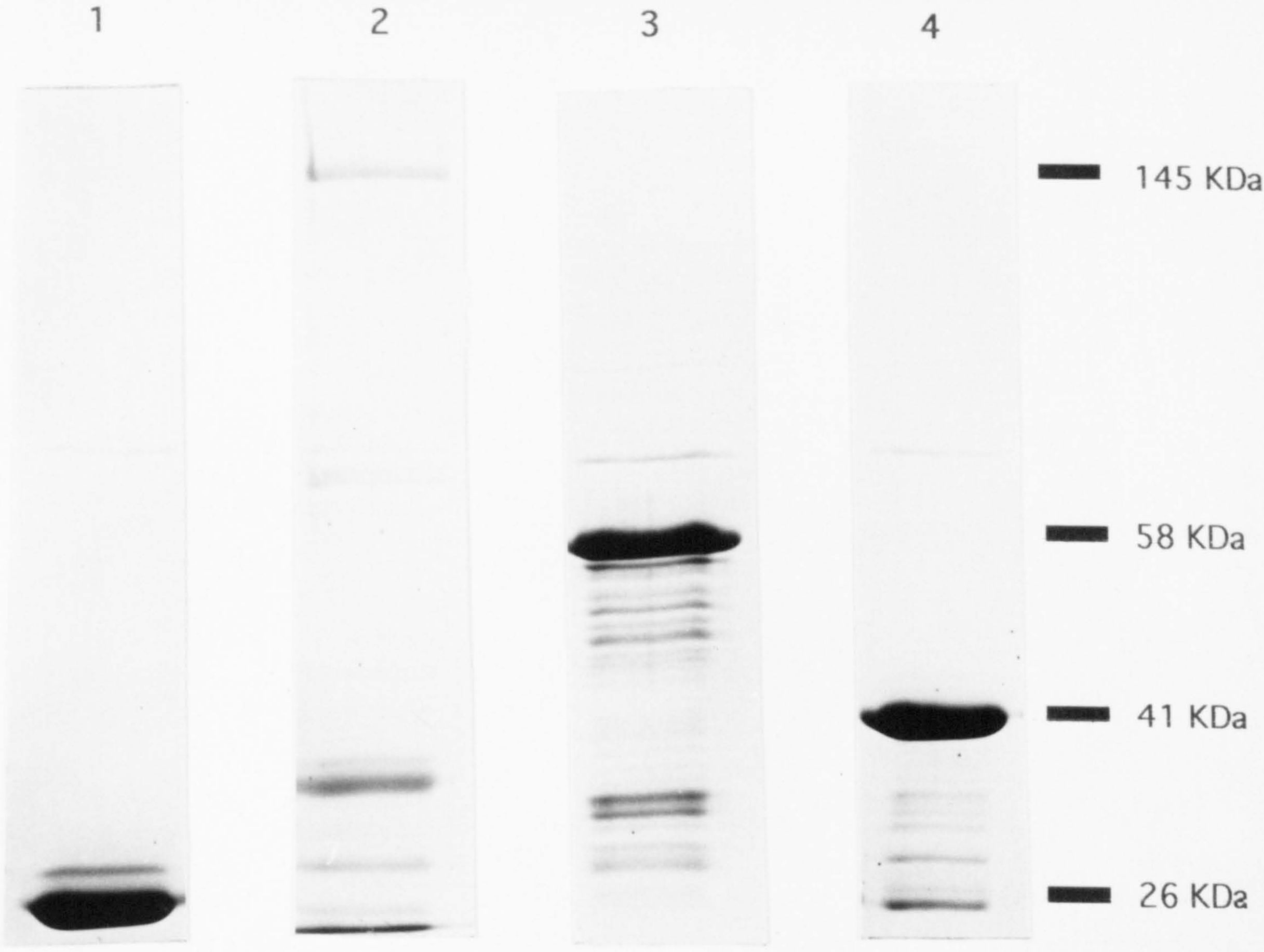


Figure 30: Coomassie stained SDS-PAGE gel of the GST-SPAG proteins. SPAG-1 and the two SPAG-2 fragments were expressed using the pGEX expression system and the purified proteins are shown. GST in lane 1, GST-SPAG-1 in lane 2, GST-KP8 in lane 3 and the GST-C350 in lane 4. The bars mark the size of the fusion proteins in kDa.

The validity of the protein construct GST-SPAG-1 was confirmed by Western blotting. A nitrocellulose membrane to which the GST-SPAG-1 fusion protein was transferred from SDS PAGE gels was probed with the monoclonal antibodies 1A7 (Williamson, 1988), BA4 (Wrenn et al., 1986) and as a negative control with 1E11 (anti *Theileria annulata* macroschizont antibody (Shiels et al., 1986)). 1A7 reacts with SPAG-1 and therefore is a positive control. BA4 is an anti-elastin antibody which has been shown to react with the peptide VGVAPG; as this peptide is found three times in the predicted SPAG-1 amino acid sequence, BA4 is another positive control for the GST-SPAG-1 fusion protein. The Western blots of GST-SPAG-1 probed with 1A7, BA4 and 1E11 (negative control) are shown in Figure 31. 1A7 and BA4 both react with GST-SPAG-1 as predicted, and 1E11 does not react with GST-SPAG-1; it can therefore be concluded the pGEX-2.7 vector expresses the correct protein. These results are in accordance with those of Boulter et al. (1994). Of note is that 1A7 reacts with some of the smaller SPAG-1 products. This is surprising, since it would expected that if these products are due to degradation this would occur from the C terminus as the GST moiety must be present because of the purification procedure. One possible explanation is that SPAG-1 oligomerizes through its C terminal region and that N terminally deleted GST fusion proteins co-purify via association with intact molecules (Boulter et al., 1994). Another possible explanation for some of the observed fragments, although extremely unlikely is that SPAG-1 is internally processed, so that N terminal and central regions of this molecule are deleted, leaving the GST moiety and the C terminal region of SPAG-1.

The validity of the pGEX-C350 construct was confirmed by sequencing the whole clone and its junctions. The sequence of the 5' junction of the pGEX-C350 clone is shown in Figure 32a and matches that obtained by Knight (1993). The expressed fusion protein has a molecular weight of 41 kDa as would be expected from the DNA sequence. The sequence identity of pGEX-KP8 was also confirmed by sequencing the 5' junction, as shown in Figure 32b. This sequence matches that of Knight (1993 and personal communication) and it has been shown that the pGEX-KP8 fusion protein product reacts with 4B11 (Knight, 1993). Therefore I believe that both pGEX-KP8 and pGEX-C350 express the correct fusion proteins as predicted.

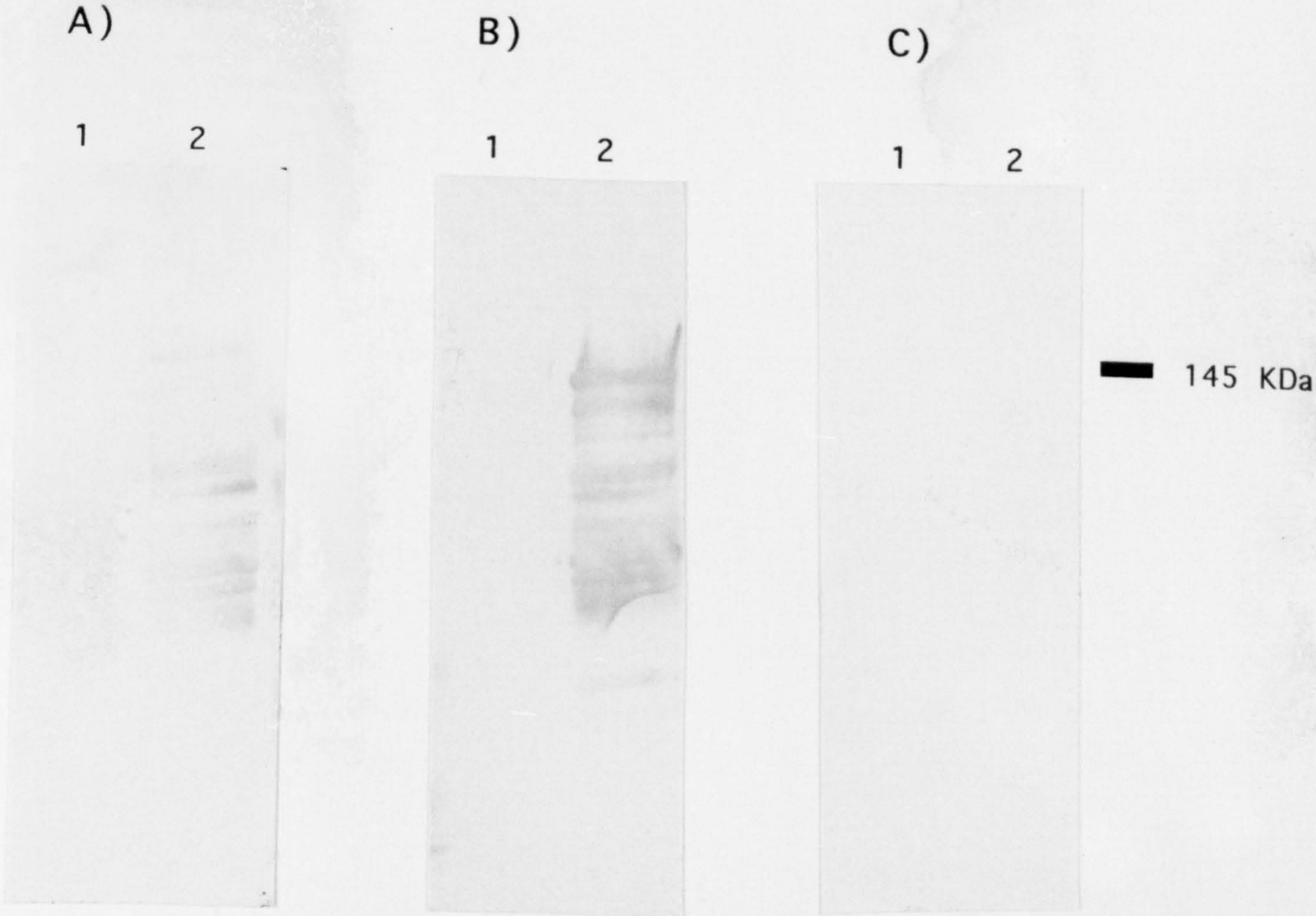


Figure 31: Western blot of GST-SPAG-1. GST (lane 1) and GST-SPAG-1 (lane 2) were run onto PAGE gels, transferred onto a nitrocellulose blots and probed with monoclonal antibodies 1A7 in (a), BA4 in (b) and with 1E11 as negative control in (c). The bar marks the size of the fusion protein in kDa.

a)

pGEX-C350 CTG GTT CCG CGT GGA TCC TCG AGA GAT AGC GCC AAC AAA

L V P R G S S R D S A K K

b)

pGEX-KP8 CTG GTT CCG CGT GGA TCC CCG GAA TTC GGC TCG AGA

L V P R G S P E F G S R

Figure 32: Sequence data of the junctions of the SPAG-2 clones in the pGEX vector system. The top sequences represent the DNA sequence of the vectors and inserts. The sequences below are the predicted amino acid sequences. The sequence of the vectors is underlined and the bold sequences shows the restriction sites used in cloning these fragments. (a) The 5' junction of the C350 insert in pGEX-2T, validating the expected open reading frame. The double underlined sequence represents the sequence of the primer Y82. (b) The 5' junction of the KP8 insert in pGEX-1λT, validating the expected open reading frame.

5.2.3. Cleavage of SPAG-1 and the SPAG-2 constructs.

The recombinant expressed fusion proteins were cleaved from their fusion partner, GST, for further studies. The protein samples needed to be as free from GST as possible, so that the results obtained in subsequent studies could be attributed fully to the parasite protein rather than leaving doubts about the involvement of GST.

The method for cleaving the three fusion proteins was adapted from Knight (1993) and is described in 2.2.6. To cleave GST-SPAG-1, activated factor Xa was used while thrombin was used to cleave the SPAG-2 fusion protein constructs. The method for separating the cleaved protein from GST and any remaining uncleaved protein was improved from that of Smith and Johnson (1988) so that the cleaved protein is run through a glutathione 4B sepharose column three times. The Figure 33 shows a coomassie stained SDS-PAGE gel, showing the GST-KP8 fusion protein in lane 2. In lane 3 is the cleaved protein with cleaved fusion

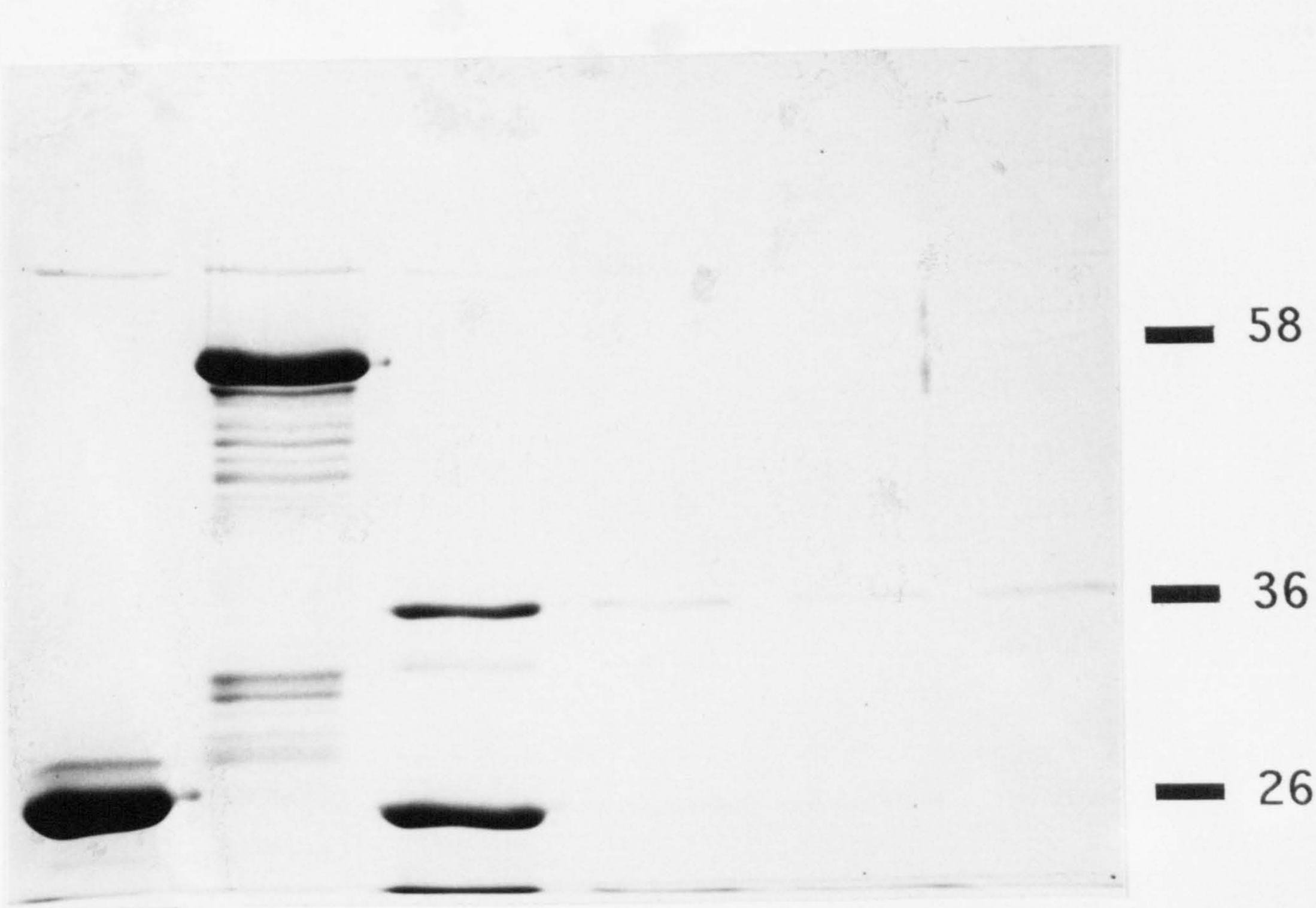


Figure 33: Coomassie stained SDS-PAGE gels of the cleaved SPAG-2 during different stages of purification. The GST control is shown in lane 1, the GST-SPAG-2 fusion protein (lane 2), the cleavage product (lane 3), cleaved SPAG-2 after the first purification step (lane 4), cleaved SPAG-2 after the second purification step (lane 5), cleaved SPAG-2 after the third purification step (lane 6).

partner, lane 4 shows the protein after the first purification step. lane 5 shows the protein after the second purification step and finally lane 6 shows the protein after the third purification step. From this it is clear that there is still some cleaved GST present after the first purification step and although none is visible after the second purification step a third purification step was used to be absolutely sure that there is no GST or uncleaved protein present in the final sample. All three protein constructs were purified with this adapted method after cleavage. The purified cleaved SPAG-1, KP8 and C350 were run out onto SDS PAGE gels next to their uncleaved partners to check the purity and the validity of the constructs. A SDS-PAGE gel containing the three constructs both cleaved and uncleaved is shown in Figure 34.

5.2.4. Do SPAG-1 and SPAG-2 bind to host cells?

To test whether SPAG-1 and SPAG-2 are involved in the process of host cell recognition two different types of binding studies were conducted to ascertain whether either molecule binds specifically to cell types which can be infected by *T. annulata* sporozoites. The first binding assay involved the binding of biotinylated protein to purified bovine peripheral blood mononuclear cells. The degree of binding was determined by flow cytometry.

First, the three purified and cleaved proteins were labelled with biotin as described in 2.2.7. Successful biotinylation was confirmed by Western blotting. To do this, the proteins were run onto SDS-PAGE gels, transferred to a nitrocellulose membrane and probed with Extravidine (Sigma), which binds to the biotin of the biotinylated protein. Extravidine is visualised using the alkaline phosphatase detection method as described in 2.2.7. A Western blot confirming the successful biotinylation of the three proteins can be seen in Figure 35. Once it had been shown that all three molecules were biotinylated, they could then be used in the binding studies. Peripheral blood mononuclear cells were purified from the blood of Friesian calves. Aliquots of 5×10^5 of these cells were incubated with a dilution series of either of the three molecules or biotinylated GST which was used as a negative control. After 30 minutes incubation and three washes, a streptavidin phycoerythrin

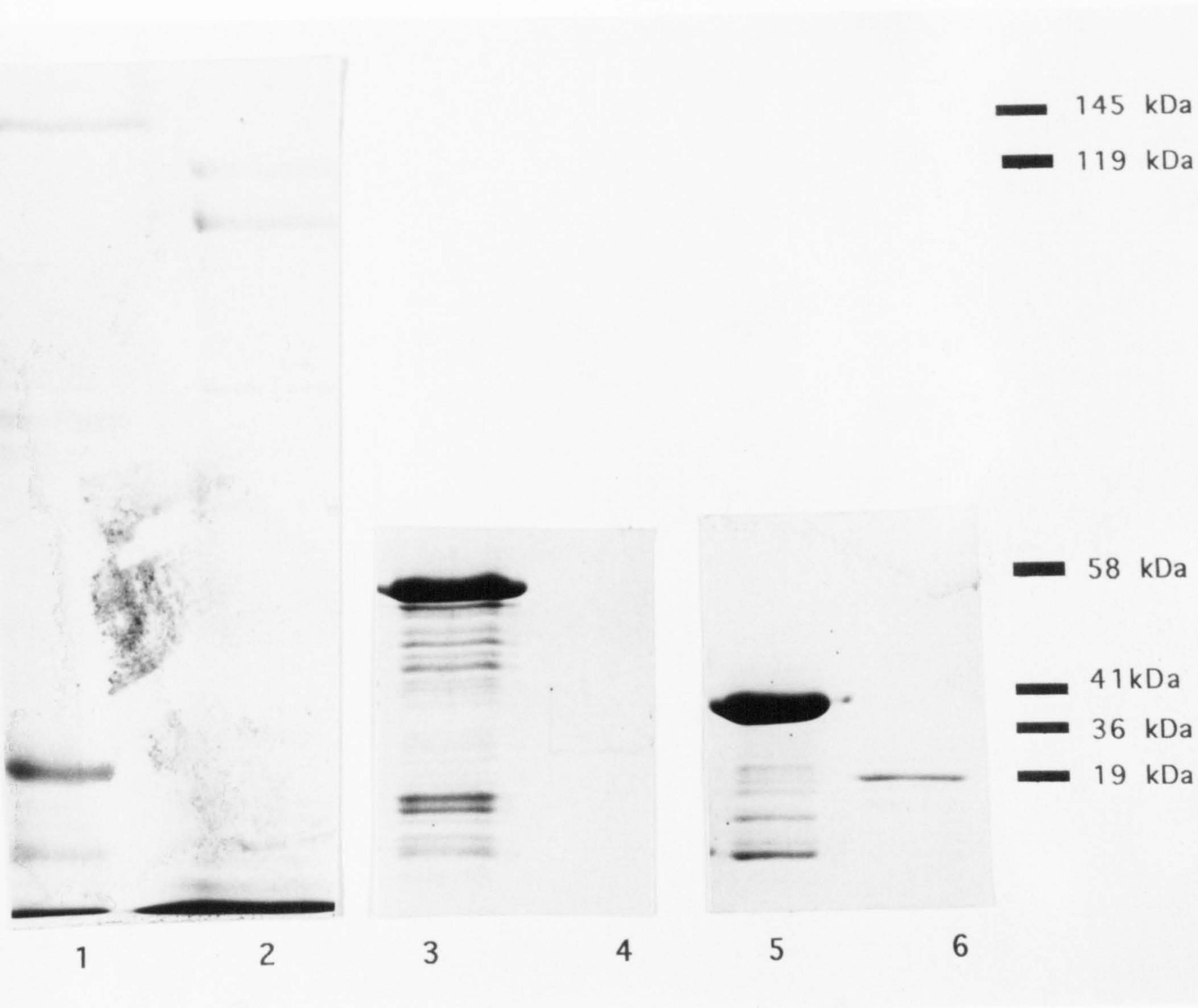


Figure 34: Coomassie stained SDS-PAGE gels of the cleaved SPAG proteins. GST-SPAG-1 is shown in lane 1 and the cleaved SPAG-1 in lane 2, GST-KP8 in lane 3 and the cleaved KP8 in lane 4 while GST-C350 is shown in lane 5 and the cleaved purified C350 is shown in lane 6. The bars mark the size of the proteins in kDa.

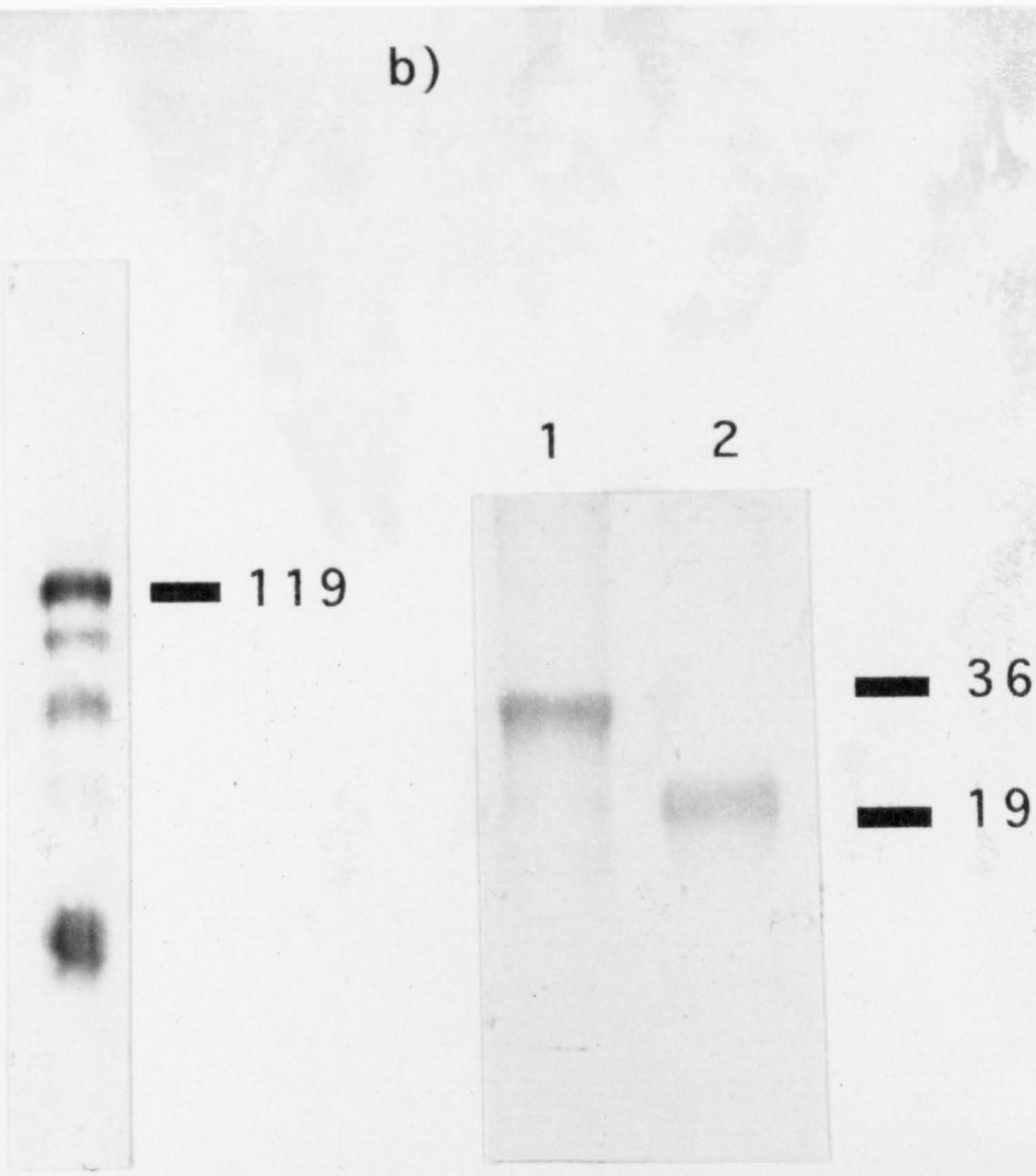


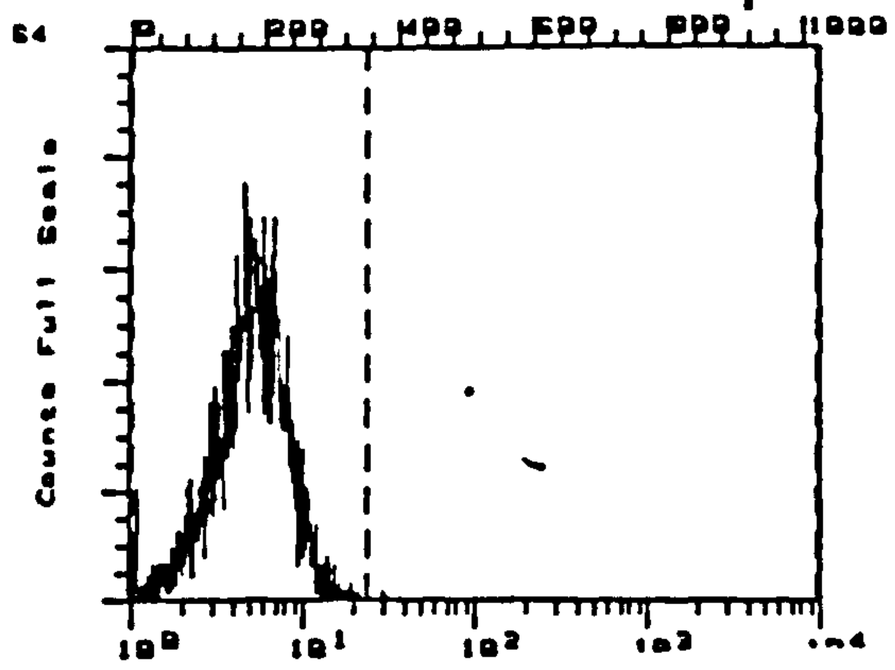
Figure 35: Western blot of biotinylated protein detected with Extravidine. The labelling of the protein with biotin was checked by western blotting using Extravidine as a detection agent. (a) Biotinylated cleaved SPAG-1 is shown and the size of the largest fragment is 119 kDa. (b) Biotinylated cleaved KP8 is shown in lane 1 and the biotinylated cleaved C350 is shown in lane 2. The bars mark the size of the proteins in kDa.

conjugate was bound to the biotinylated protein and was detected using a Facscan flow cytometer. The results for the binding of the SPAG-1 dilution series is shown in Figure 36a, that of the GST dilution series in Figure 36b, the KP8 dilution series in Figure 36c, and the dilution series of C350 is shown in Figure 36d. From these results it can be deduced that all three sporozoite molecules bind to a sub-population of the bovine PBM cells since for these three proteins there is a two-peak histogram present in which the peak nearest to the Y-axis represents the cell population which is negative for protein binding and the peak further away from the Y-axis represents the cells which are positive for protein binding. The binding of the GST dilution series is different in the sense that no clear two-peak histogram could be observed. Furthermore, there is almost no binding observed at a protein concentration of 250 ng and 125 ng, although a clear positive and negative cell population could still be seen for the three SPAG protein constructs at these concentrations. It can therefore be concluded that GST does not bind specifically to a sub-population of peripheral blood mononuclear cells. This experiment using the SPAG constructs indicates that all three molecules bind specifically to some of the PBM cells but not to all of them. The results of the dilution series for all three recombinant protein preparations indicated that a working concentration of 500 ng of protein results in a good separation of positive and negative cells. This working dilution was therefore adopted for further two-colour flow cytometry.

Figure 36: Single colour flow cytometry using a dilution series of biotinylated proteins and PBM cells. A method to establish the best protein concentration for two colour flow cytometry. (A) The dilution series of biotinylated cleaved SPAG-1 on PBM cells. Reagent control, cells incubated without protein or primary antibody (panel 1), positive control, anti MHC class II antibody (panel 2), 1.25 µg SPAG-1 (panel 3), 500 ng SPAG-1 (panel 4), 250 ng SPAG-1 (panel 5), 125 ng SPAG-1 (panel 6), 25 ng SPAG-1 (panel 7) and 12.5 ng SPAG-1 (panel 8). (B) The dilution series of biotinylated GST on PBM cells. Reagent control (panel 1), positive control, anti MHC class II antibody (panel 2), 1.25 µg GST (panel 3), 500 ng GST (panel 4), 250 ng GST (panel 5) and 125 ng GST (panel 6). (C) The dilution series for biotinylated cleaved KP8 on PBM cells. The reagent control (panel 1), positive control, anti MHC class II antibody (panel 2), 1.25 µg KP8 (panel 3), 500 ng KP8 (panel 4), 250 ng KP8 (panel 5), 125 ng KP8 (panel 6), 25 ng KP8 (panel 7) and 12.5 ng KP8 (panel 8). (D) The dilution series for biotinylated cleaved C350 on PBM cells. The reagent control (panel 1), positive control, anti MHC class II antibody (panel 2), 1.25 µg C350 (panel 3), 500 ng C350 (panel 4), 250 ng C350 (panel 5), 125 ng C350 (panel 6), 25 ng C350 (panel 7) and 12.5 ng C350 (panel 8).

Number of cells

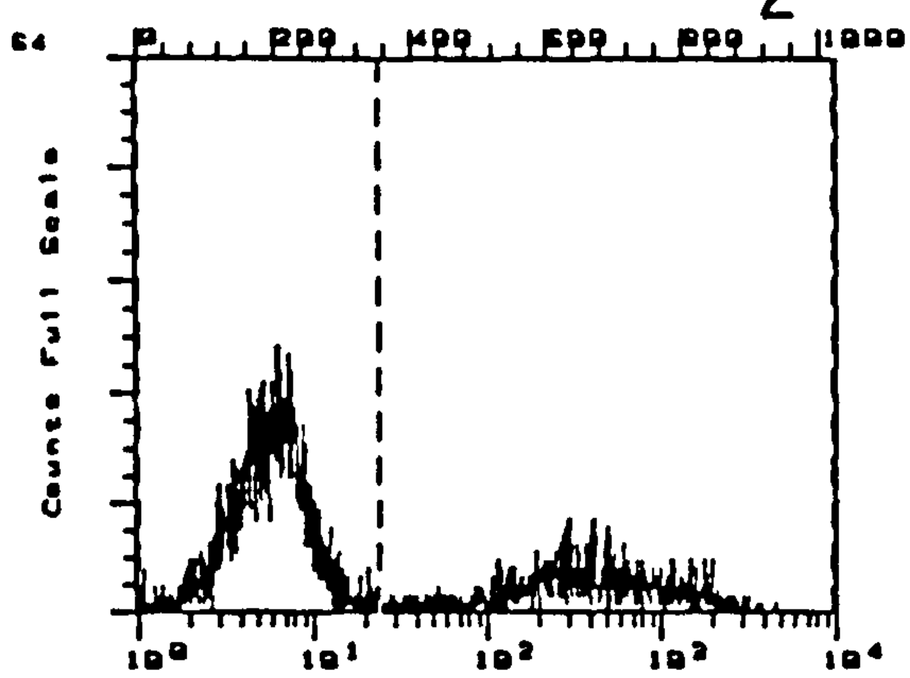
1



Reagent control (no protein)

Number of cells

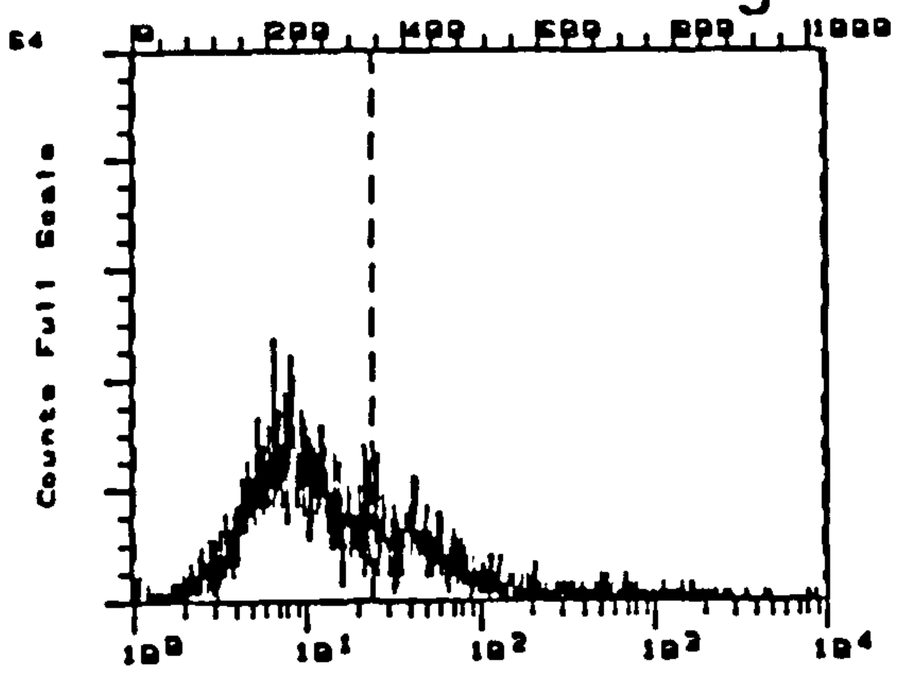
2



ILA 21

Number of cells

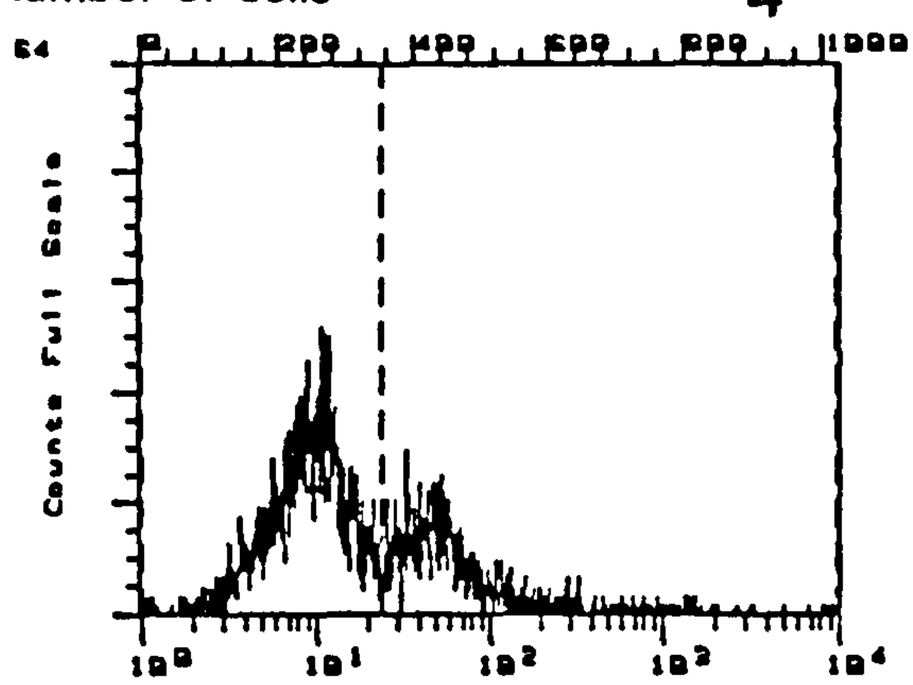
3



1.25 ug SPAG-1

Number of cells

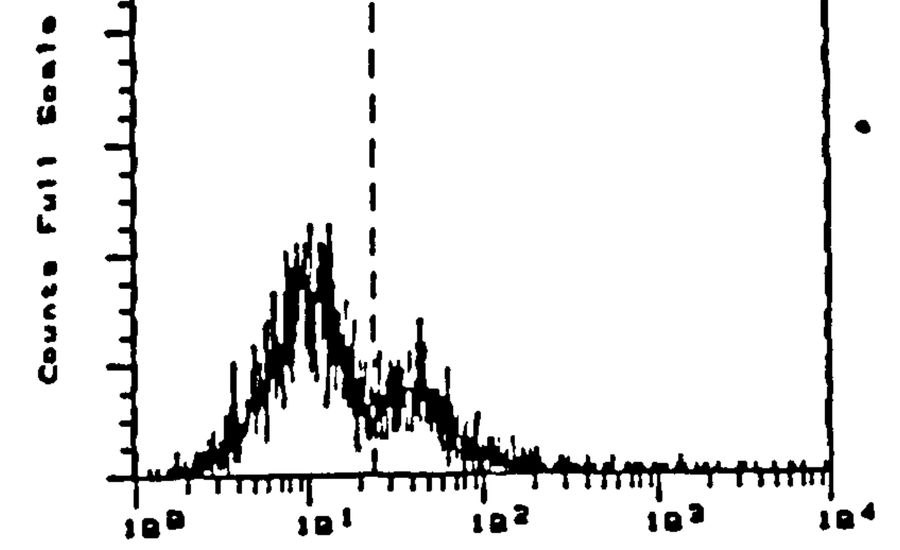
4



500 ng SPAG-1

Number of cells

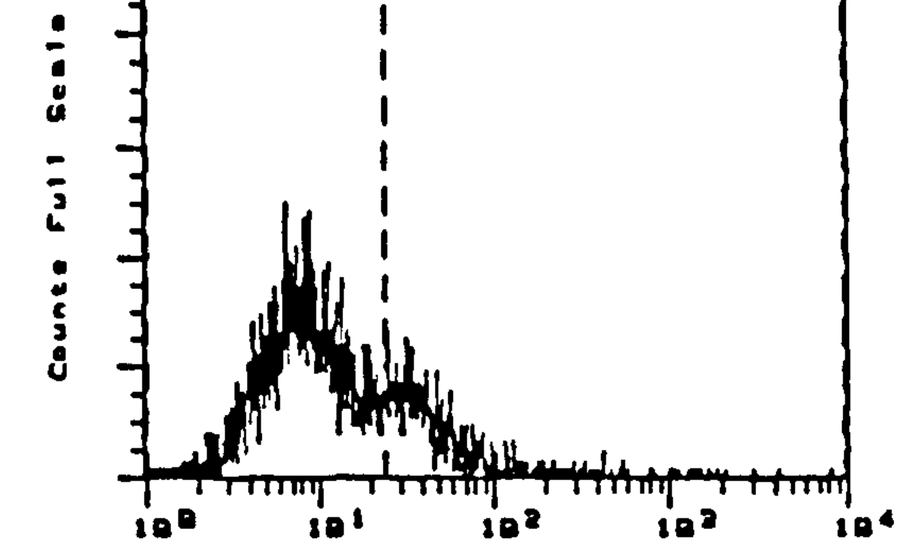
5



250 ng SPAG-1

Number of cells

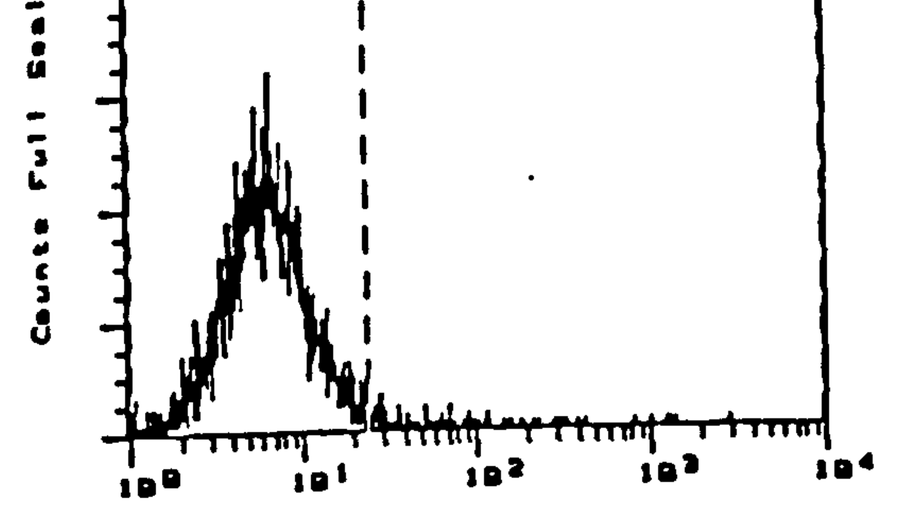
6



125 ng SPAG-1

Number of cells

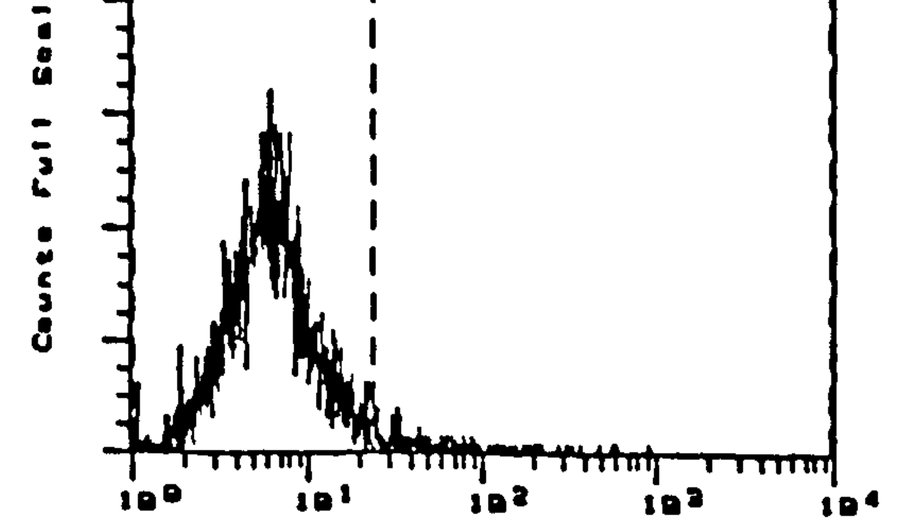
7



25 ng SPAG-1

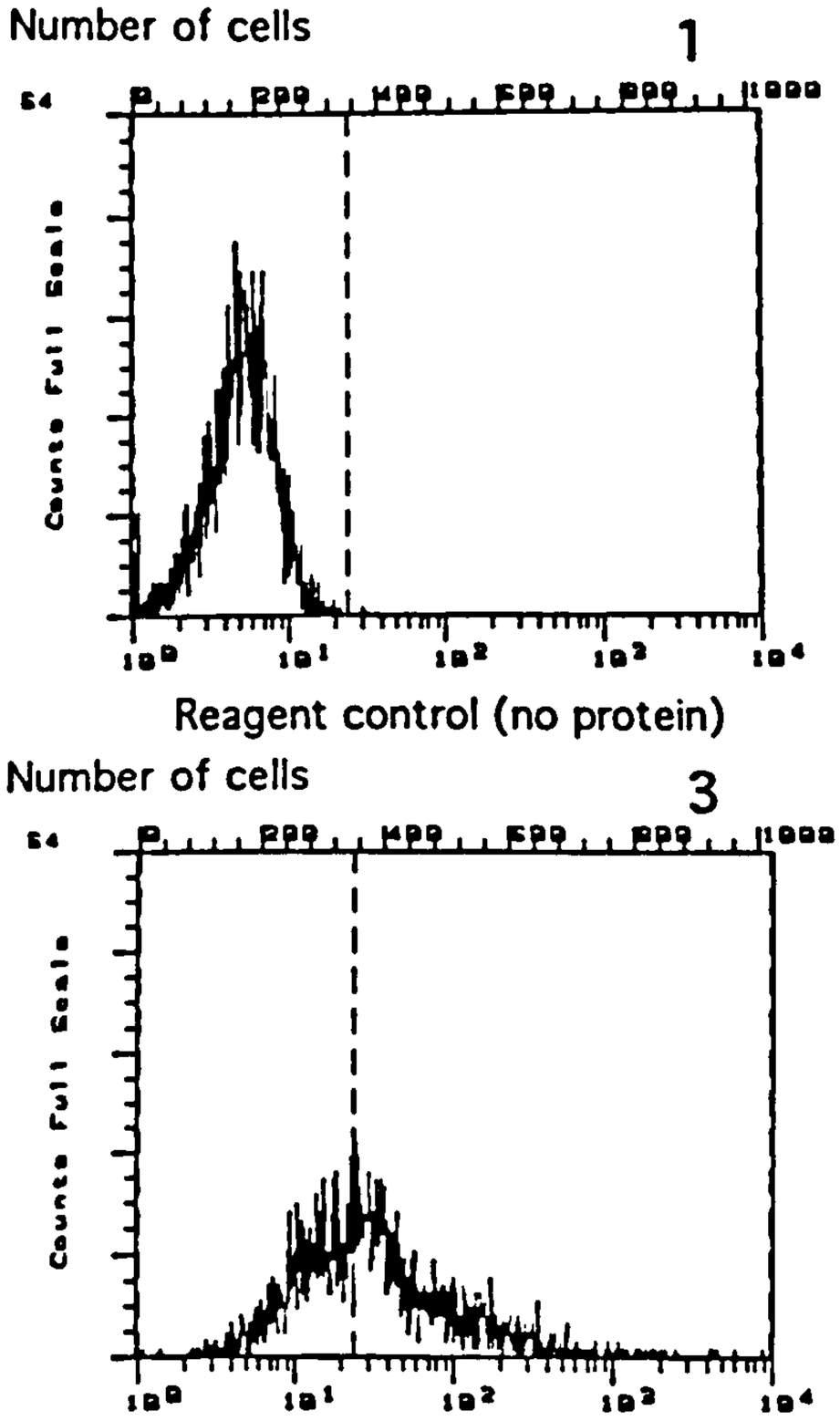
Number of cells

8

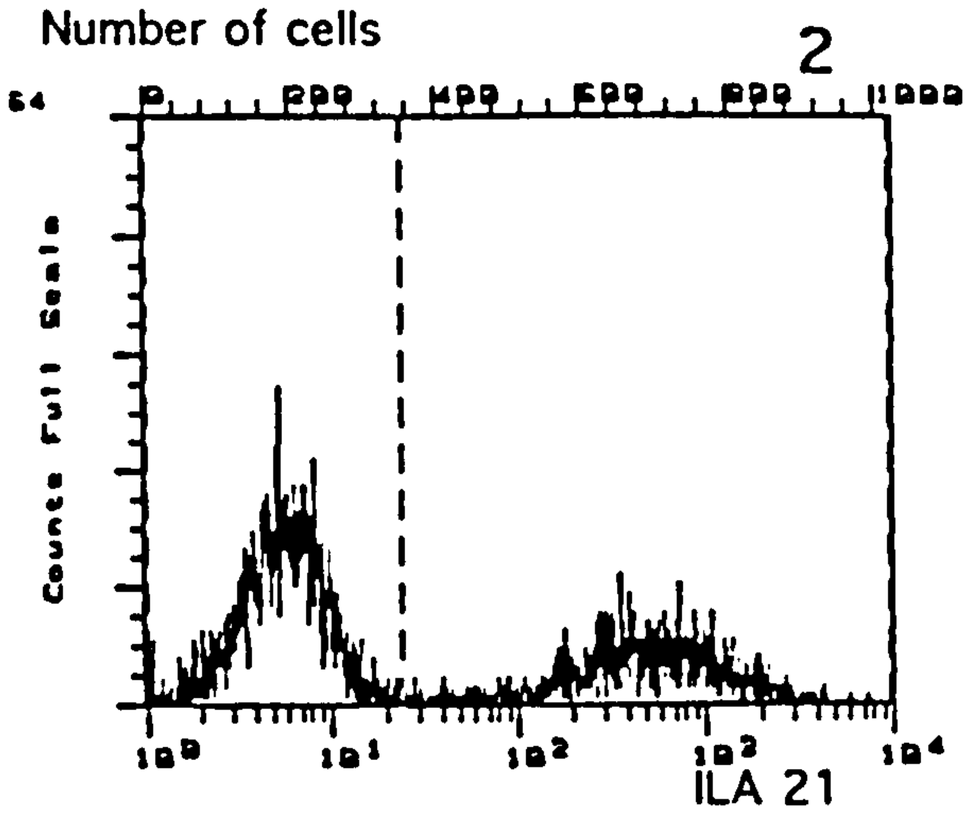


12.5 ng SPAG-1

B)

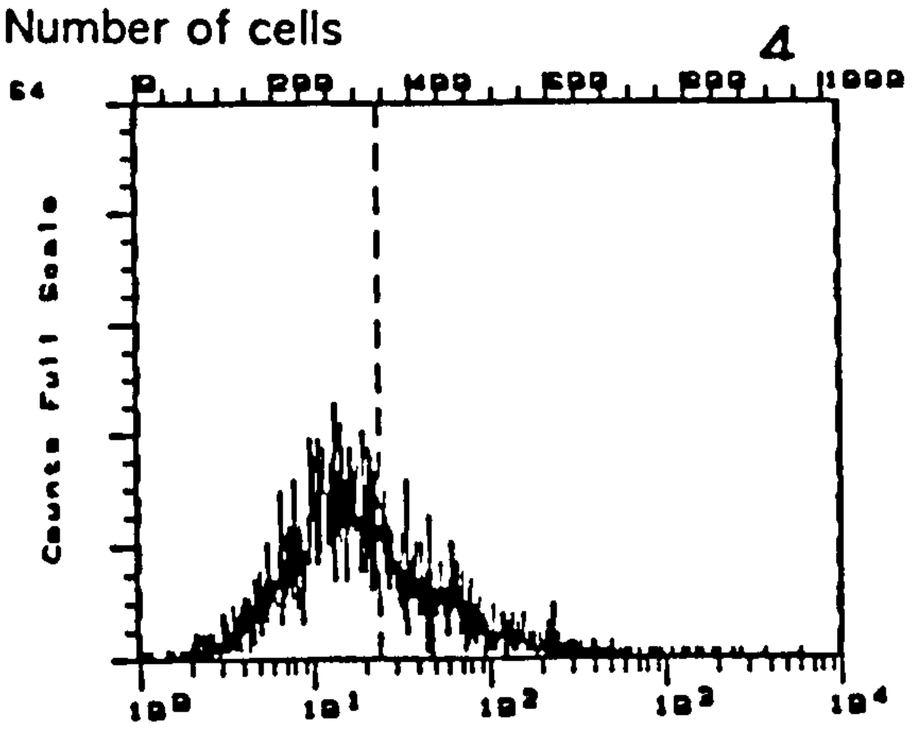


Reagent control (no protein)

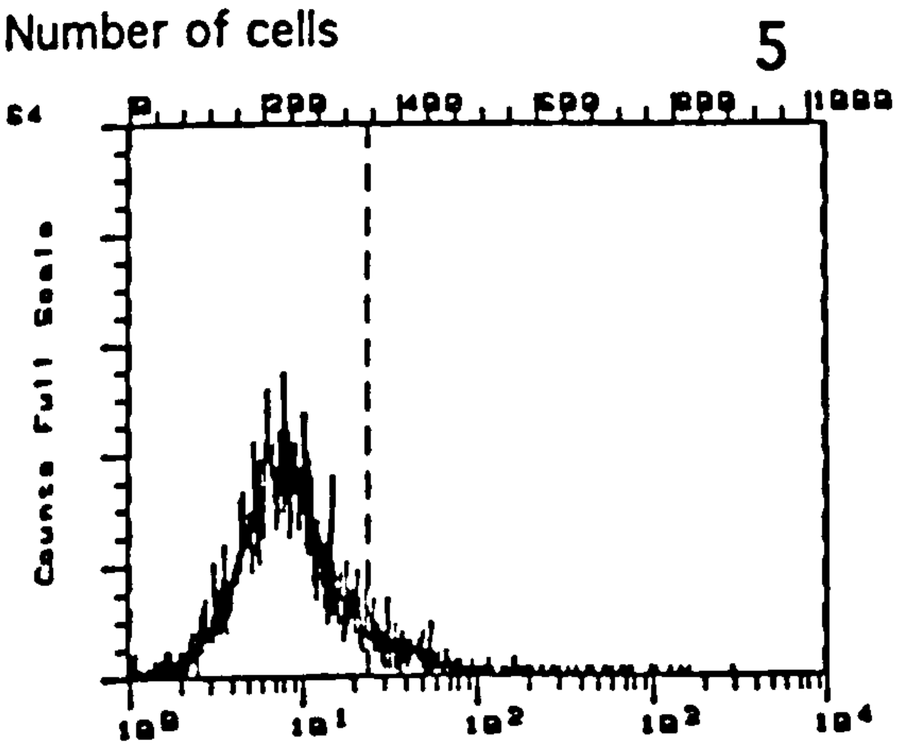


ILA 21

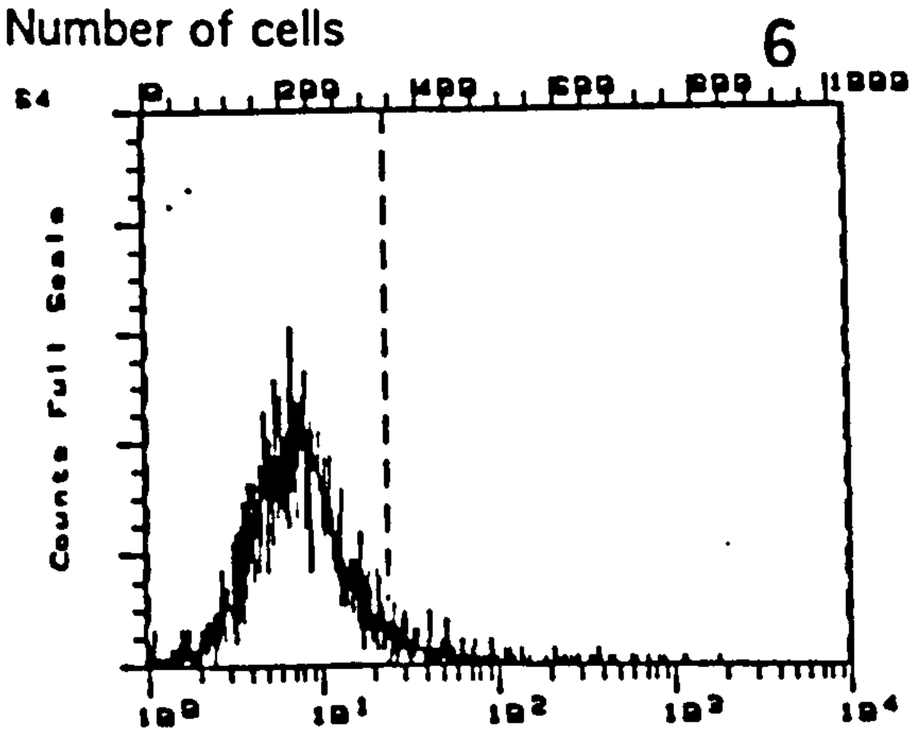
1.25 ug GST



500 ng GST



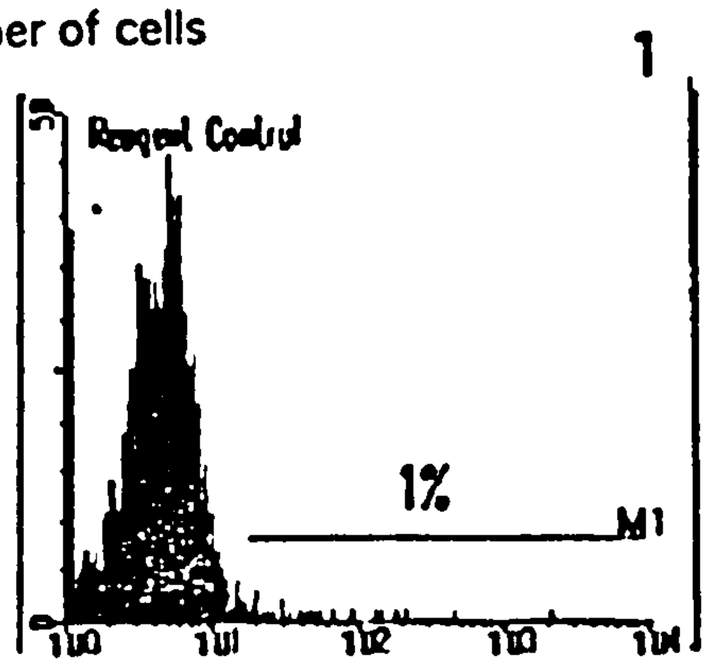
250 ng GST



125 ng GST

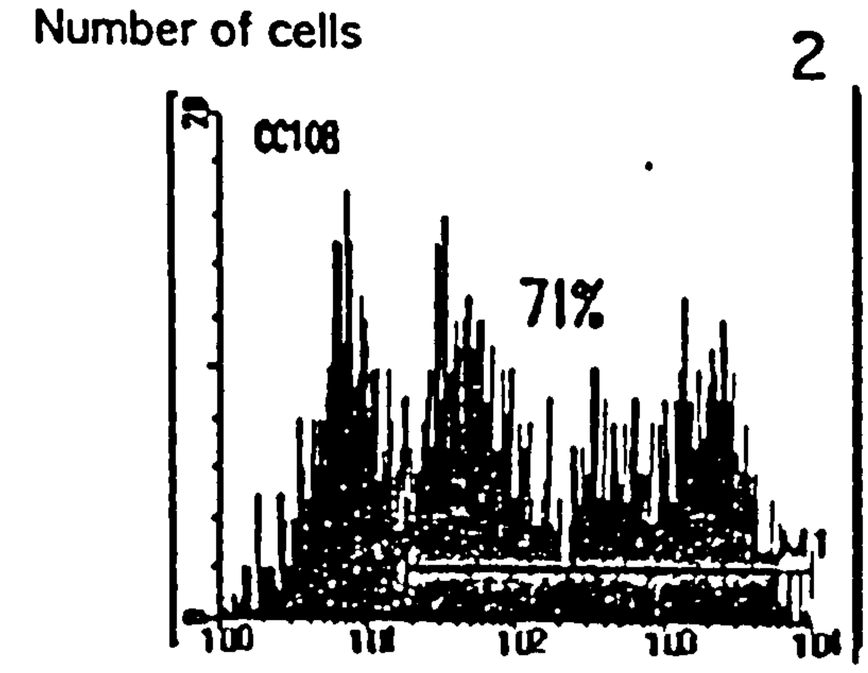
C)

Number of cells



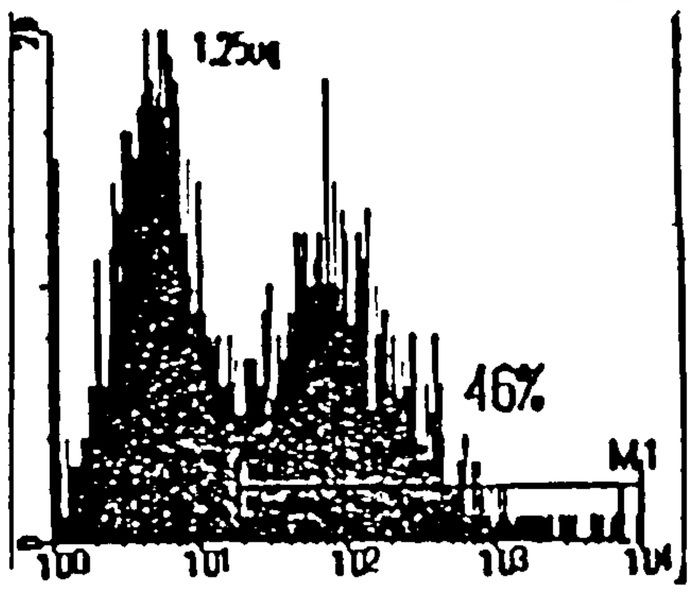
Reagent control (no protein)

Number of cells



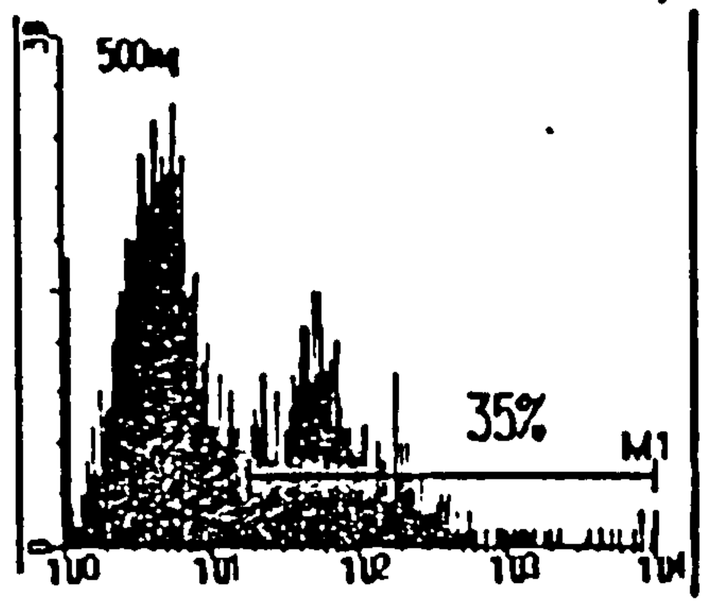
CC108

Number of cells



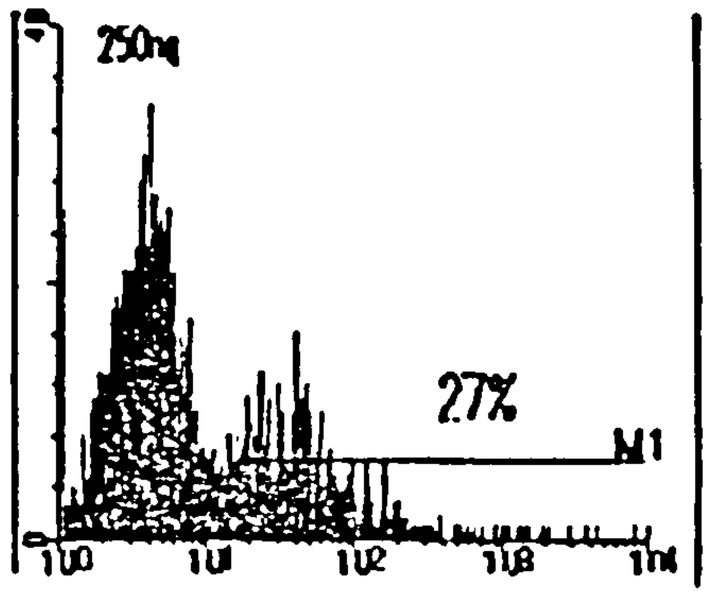
1.25 ug KP8

Number of cells



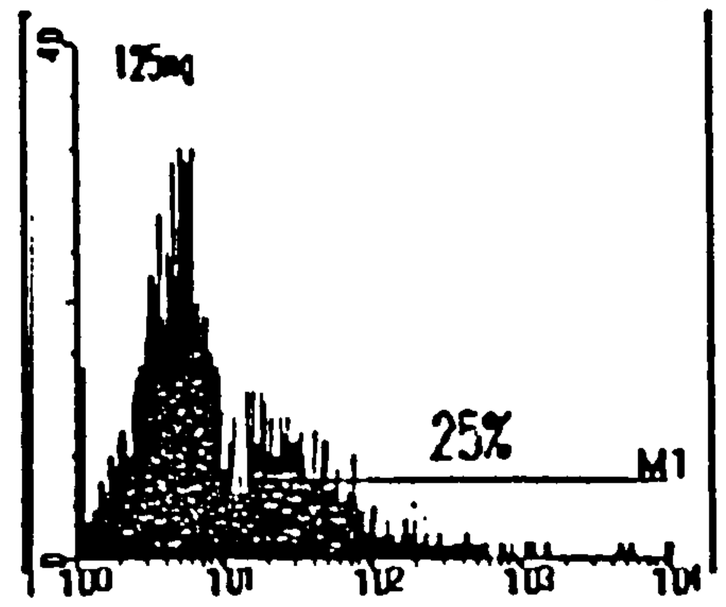
500 ng KP8

Number of cells



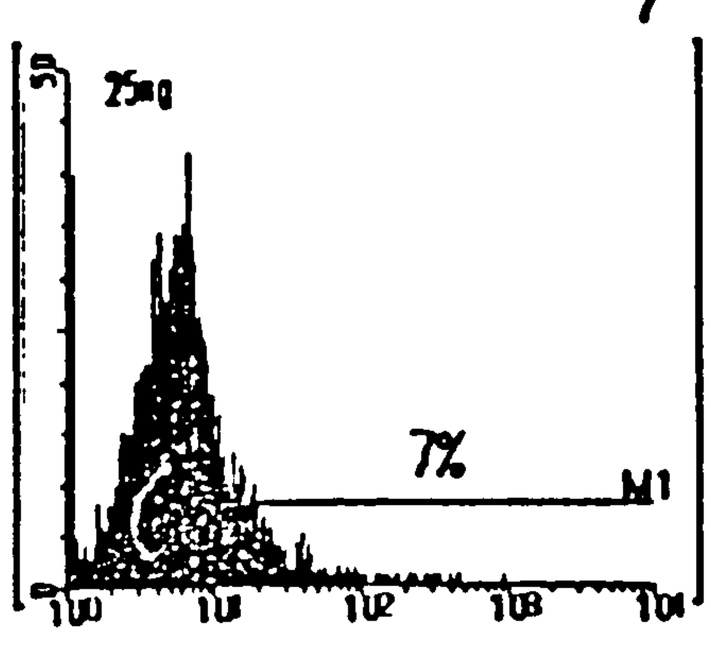
250 ng KP8

Number of cells



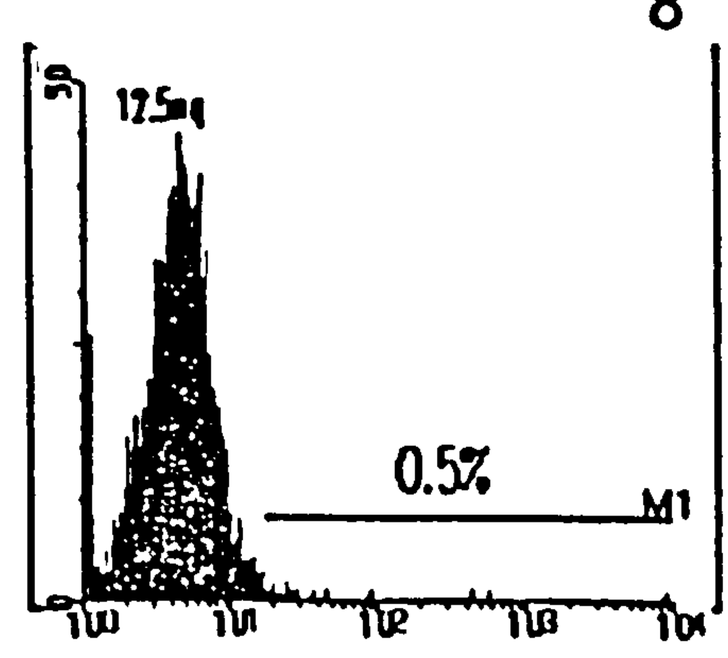
125 ng KP8

Number of cells



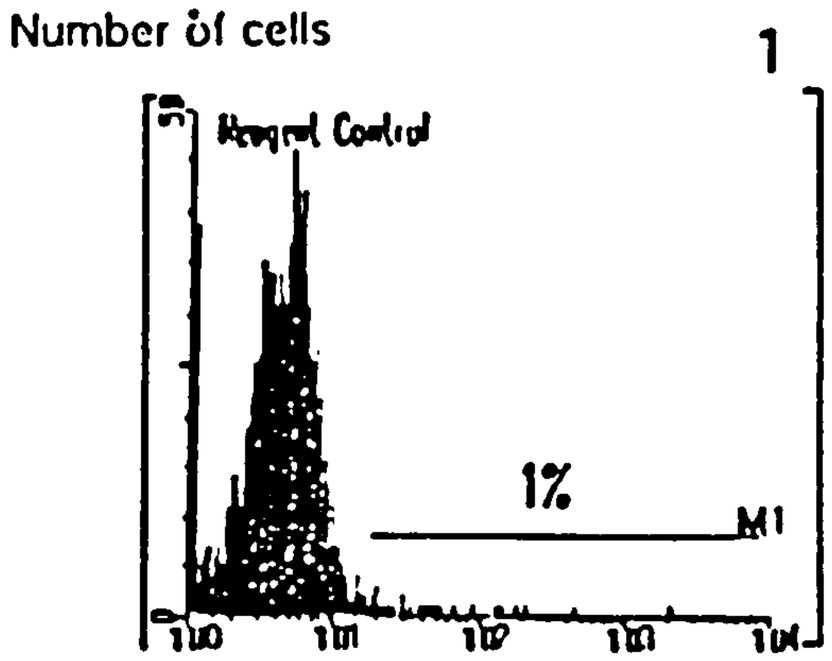
25 ng KP8

Number of cells

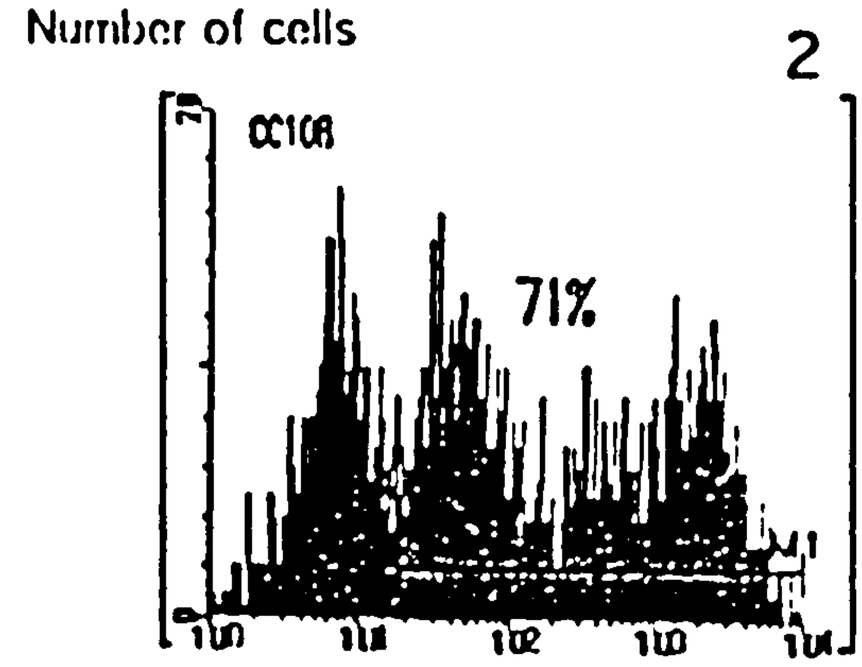


12.5 ng KP8

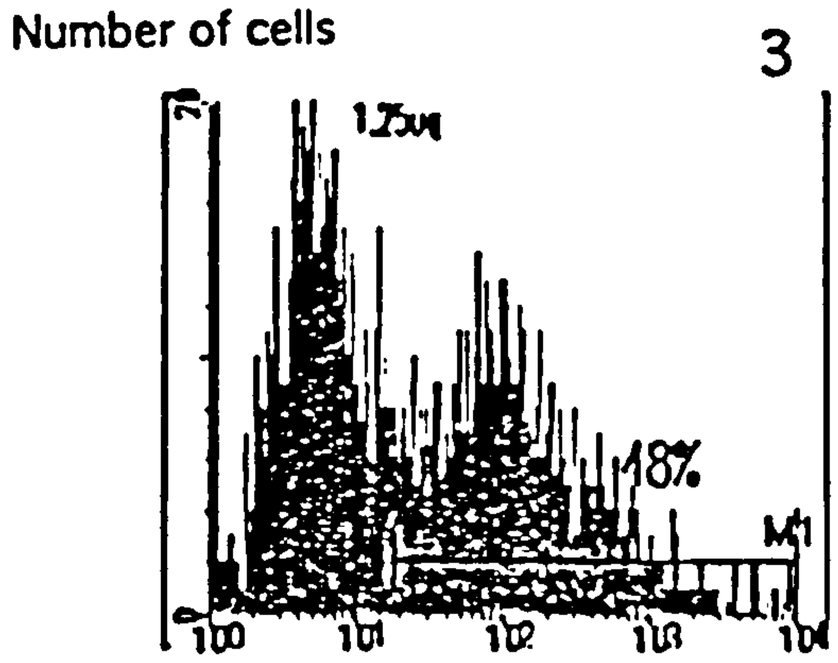
D)



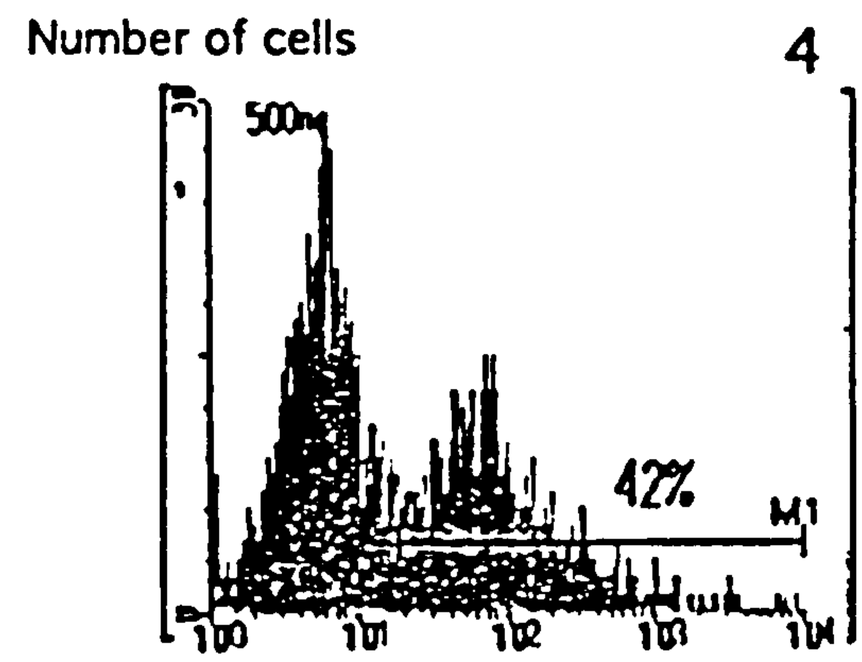
Reagent control (no protein)



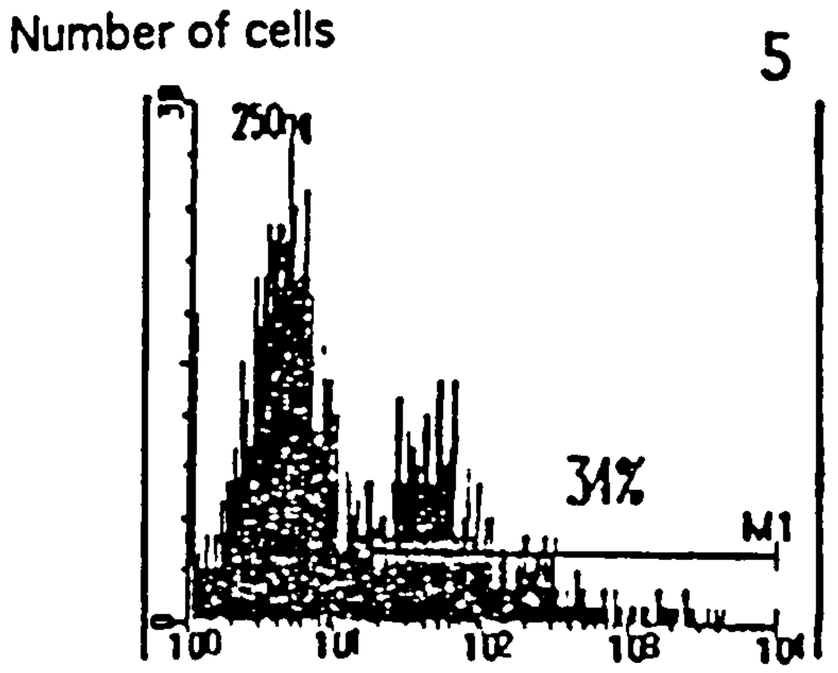
CC108



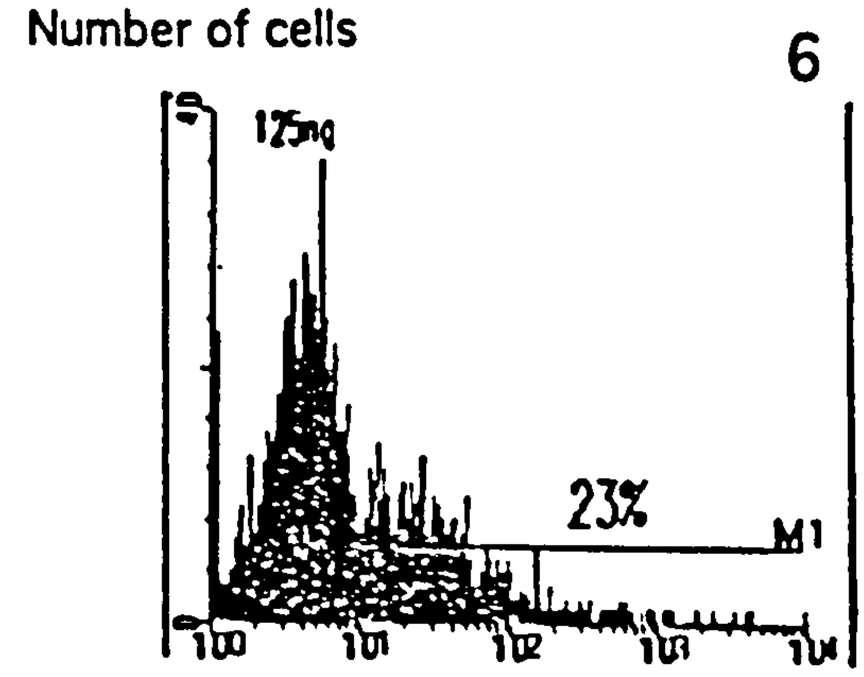
1.25 ug C350



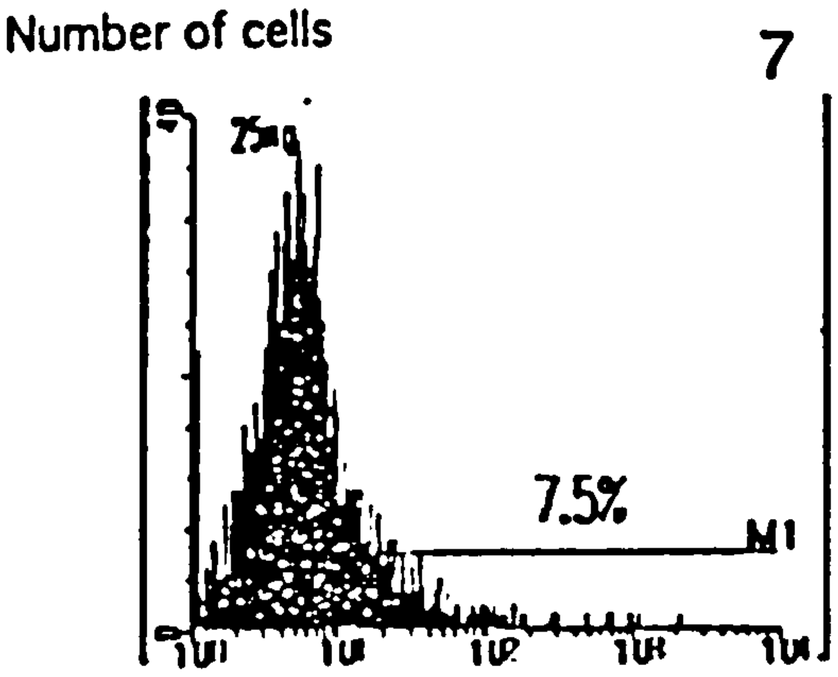
500 ng C350



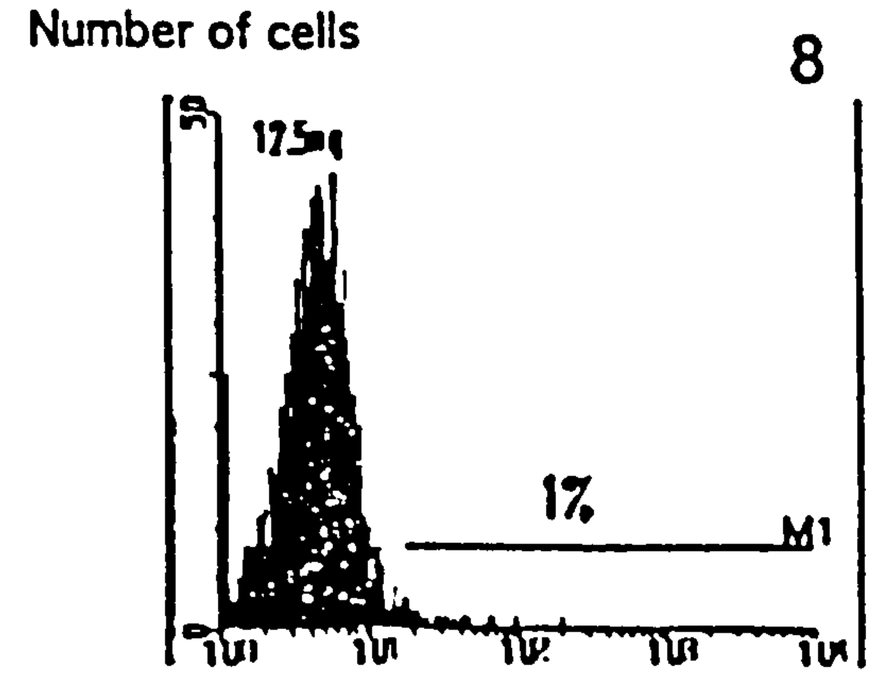
250 ng C350



125 ng C350



25 ng C350



12.5 ng C350

5.2.5. Which are the target cells for SPAG-1 and SPAG-2 binding?

The previous results do not indicate to which cell types the three sporozoite protein preparations bind. Two-colour flow cytometry has therefore been used to identify these sub-population of cells. Aliquots of bovine PBM cells were stained with a panel of monoclonal antibodies. These antibodies react with surface molecules of specific PBM subpopulations (such as B cells, T cells, and monocytes) as listed in Table 9. Each of the three biotinylated protein constructs was then incubated with these PBM cells. The binding of the monoclonal antibodies was visualised using the flow cytometer with green fluorescence while the protein binding was visualised by red fluorescence. The binding results for the three proteins is shown in Figures 37-39. Each of the squares represents one of the initial PBM aliquots. The X-axis reflects the binding of the monoclonal antibodies and the Y-axis represents the binding of the protein. The results for SPAG-1 binding are shown in Figure 37, the results for KP8 in Figure 38 and those for C350 are shown in Figure 39.

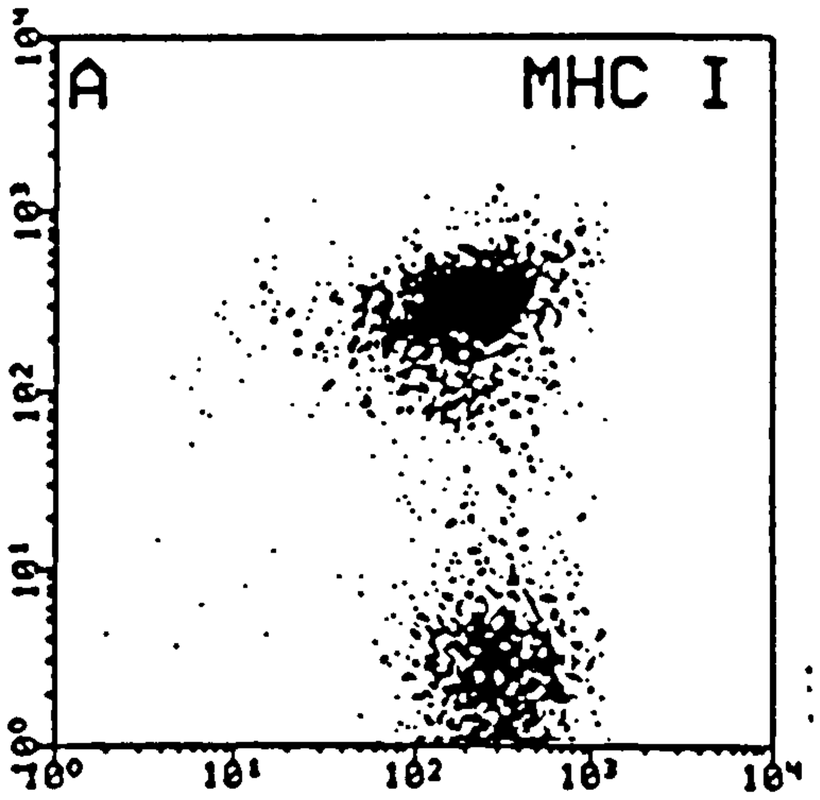
Two-colour flow cytometry using biotinylated cleaved SPAG-1 and a panel of 6 monoclonal antibodies, as described in Table 9, shows that SPAG-1 strongly binds to a subpopulation of 50% of peripheral blood mononuclear cells as shown in Figure 37A. In the next panels, Figure 37B-F, it is shown that SPAG-1 binds to 76% of CD2 positive T cells, 44% of γ/δ T cells, 19% of B cells, 27% of monocytes and 27% of MHC class II positive cells.

Name of antibody	Target molecule	Target cell type
CC15	WC1	γ/δ T cells
CC21	WC3	B cells
CC42	CD2	T cells
CC94	CD 11b	mainly Monocytes
ILA 19	MHC class I	all PBM cells
ILA 21	MHC class II	Monocytes and B cells

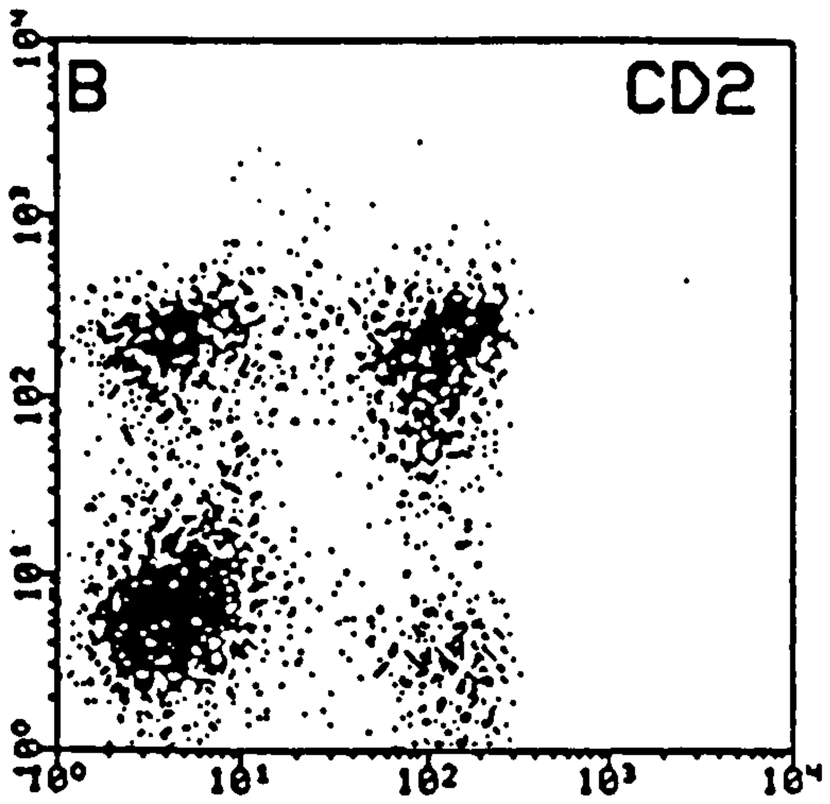
Table 9: Monoclonal antibodies used in the 2 colour flow cytometry., The table lists the monoclonal antibodies used, the surface molecules they react with and the cell types on which those molecules are found.

Figure 37: Two colour flow cytometry of PBM cells using biotinylated cleaved SPAG-1 and a panel of monoclonal antibodies. 500 ng SPAG-1 were used for each of the results shown in panel A-F and of a series of monoclonal antibodies. Panel A shows the result for ILA-19, anti MHC class I; CC 42, anti CD2 (T cells) in (B); CC15, anti WC1 (γ/δ T cells) in (C); CC21, anti WC3 (B cells) in (D); CC94, anti CD11b (mainly monocytes) in (E) and ILA-21 anti MHC class II in (F).

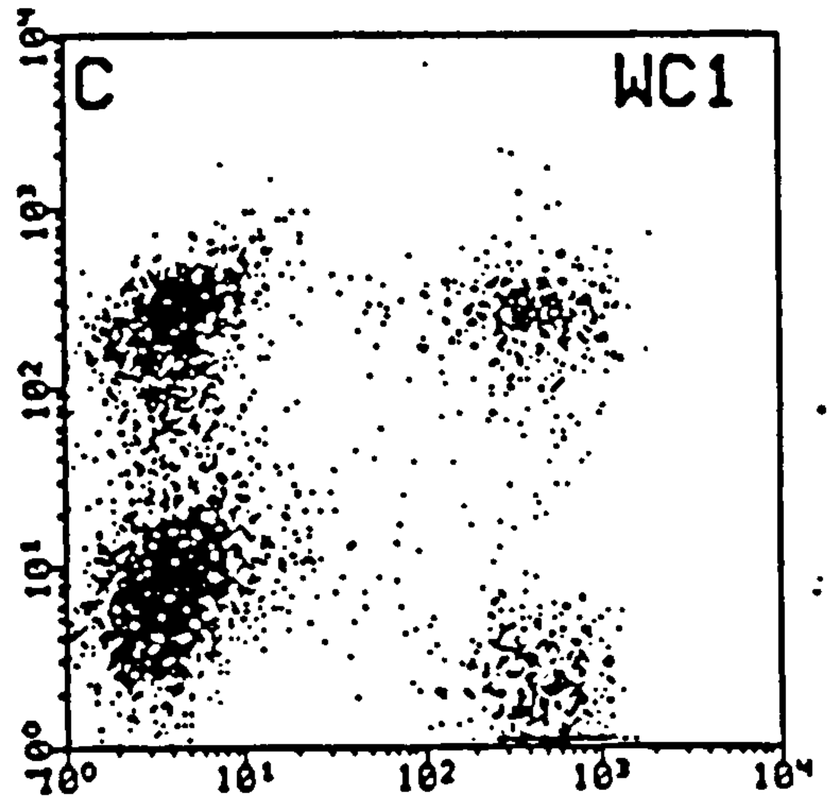
500ng SPAG-1



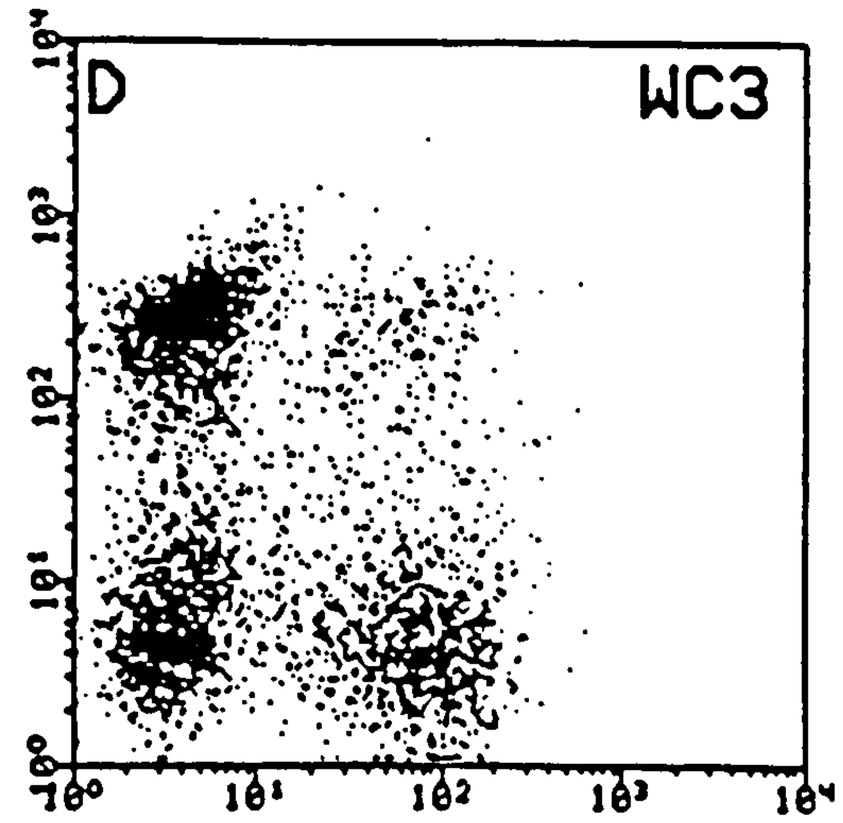
500ng SPAG-1



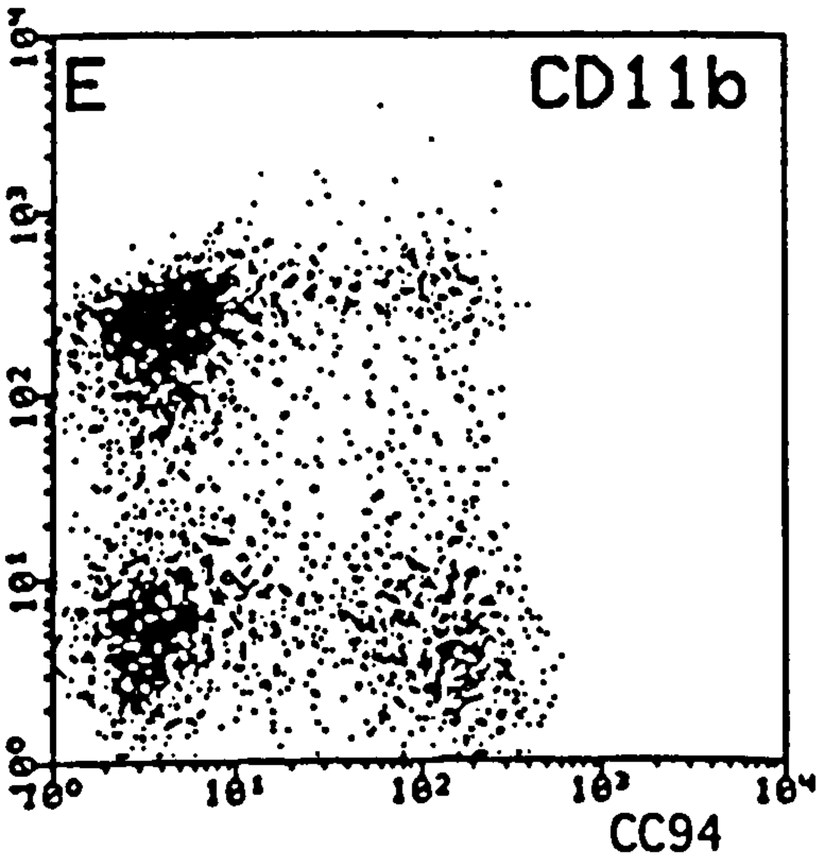
500ng SPAG-1



500ng SPAG-1



500ng SPAG-1



500ng SPAG-1

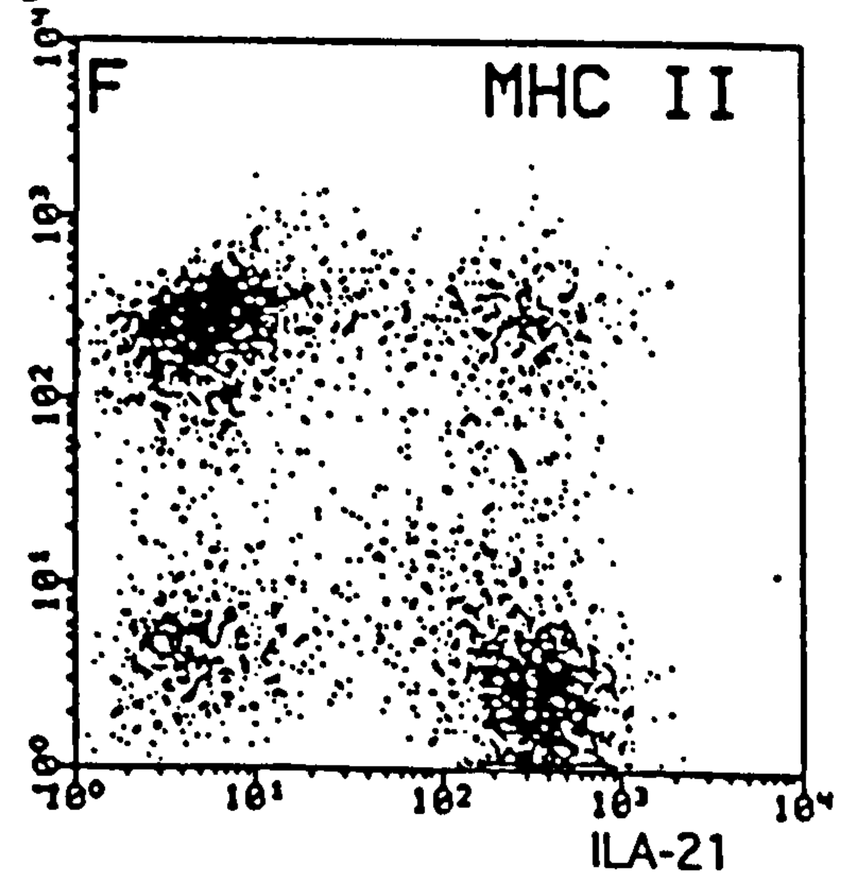
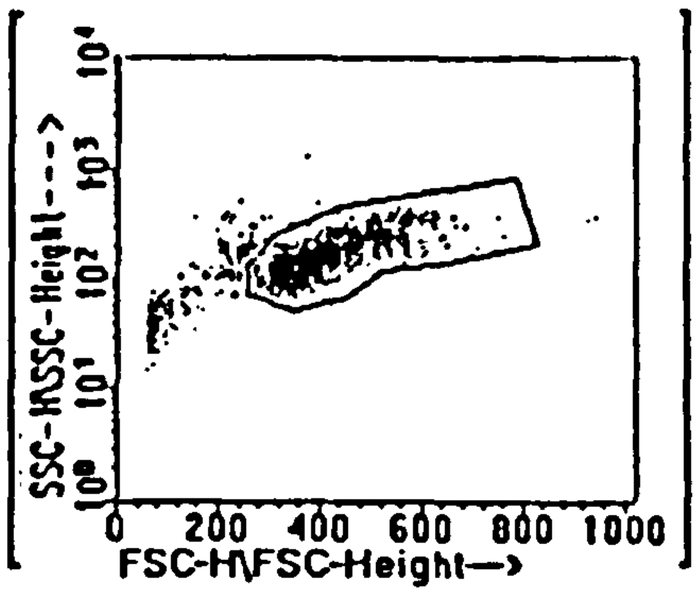
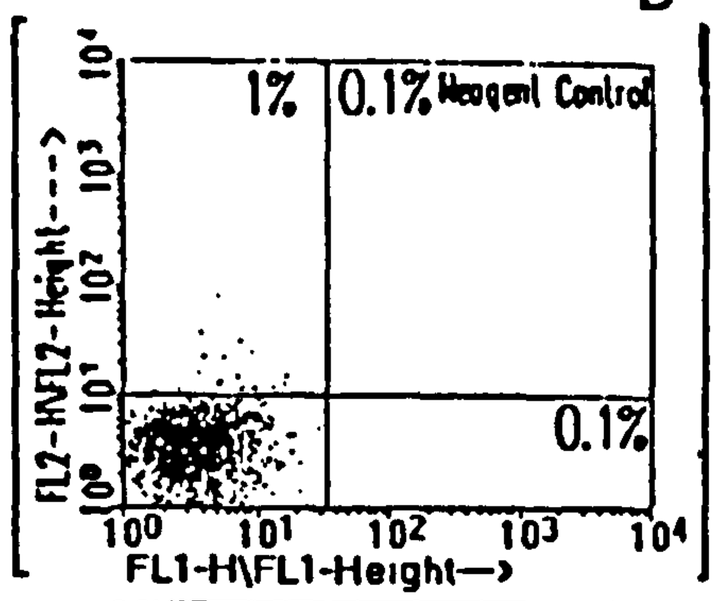


Figure 38: Two colour flow cytometry of PBM cells using biotinylated cleaved KP8 and a panel of monoclonal antibodies. 500 ng KP8 were used for each reaction of panel C-H. Panel A shows the cell population investigated. The reagent control is shown in (B). A combination of KP8 and one of a series of monoclonal antibodies were used for panels C-H. ILA-19 anti MHC class I in (C); CC42, anti CD2 (T cells) in (D); CC15, anti WC1 (γ/δ T cells) in (E); CC21, anti WC3 (B cells) in (F); CC94, anti CD11b (mainly monocytes) in (G) and ILA-21 anti MHC class II in (H).

A



B

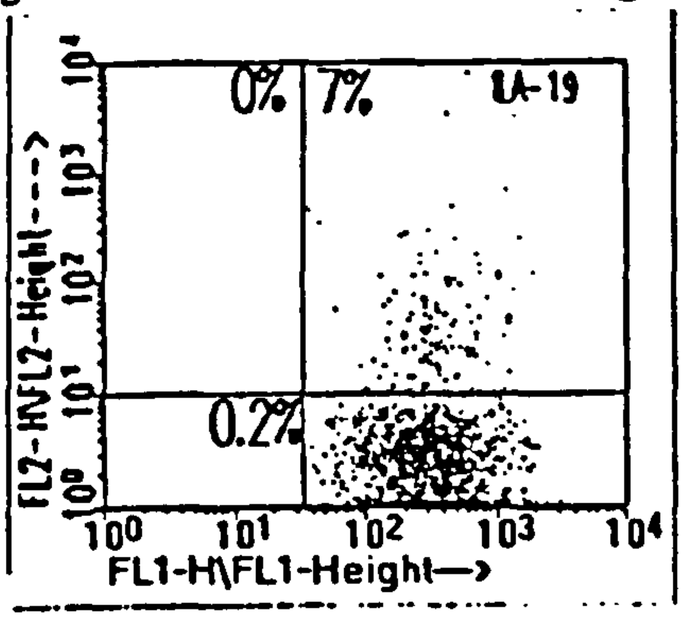


Cell population analysed

Reagent control (no protein, no antibody)

500 ng KP-8

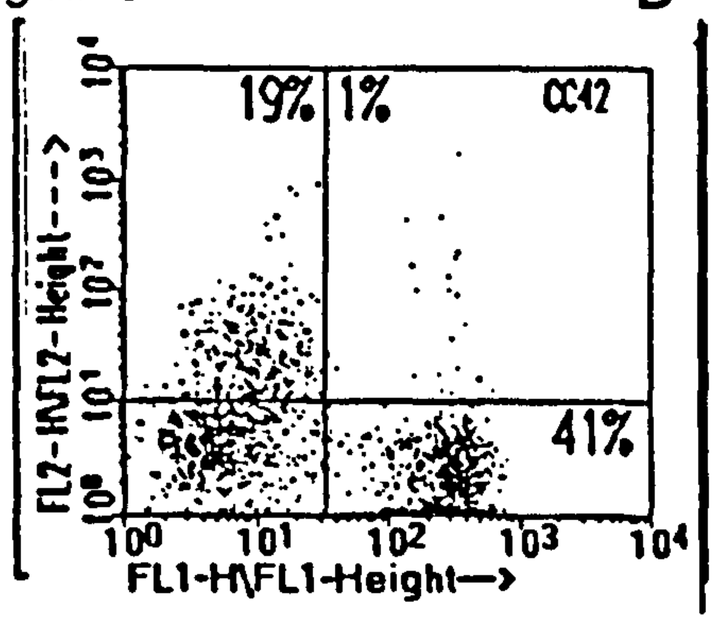
C



MHC I

500 ng KP-8

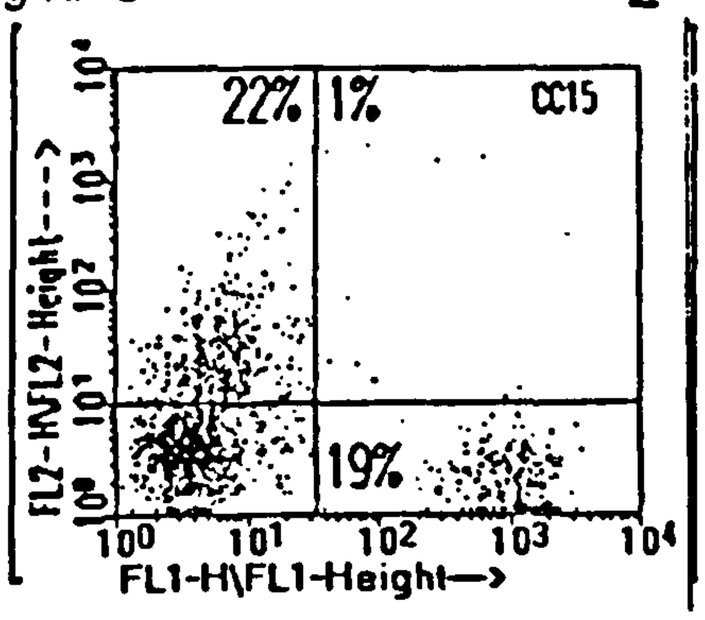
D



CD2

500 ng KP-8

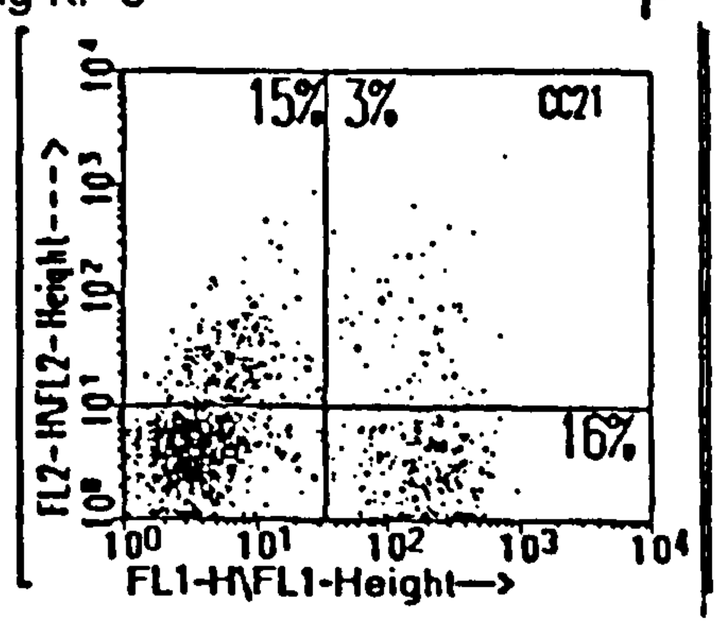
E



WC1

500 ng KP-8

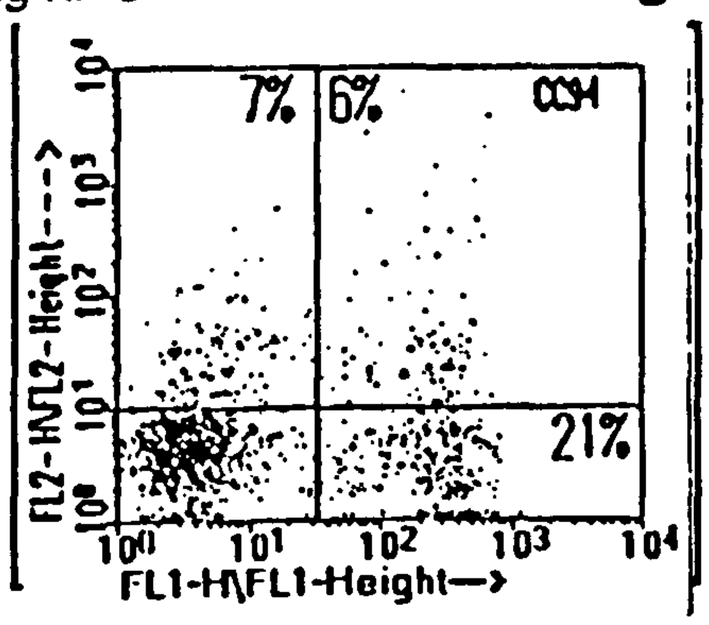
F



WC3

500 ng KP-8

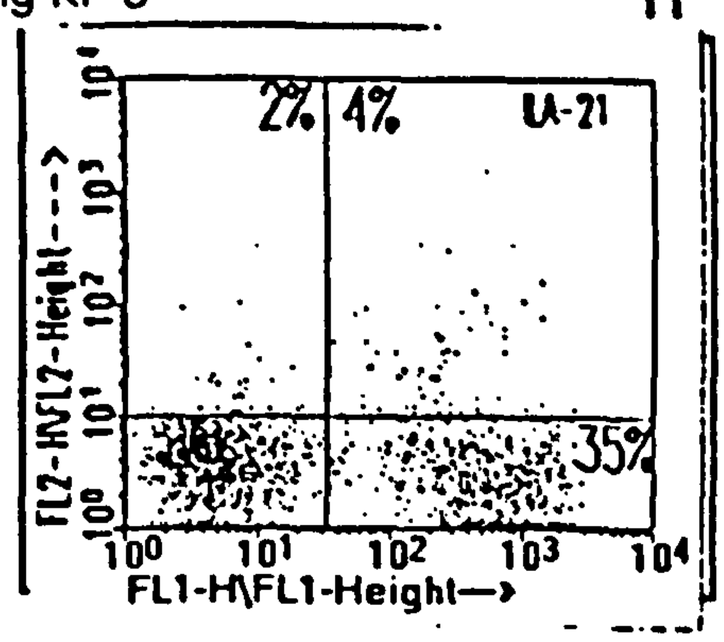
G



CD11b

500 ng KP-8

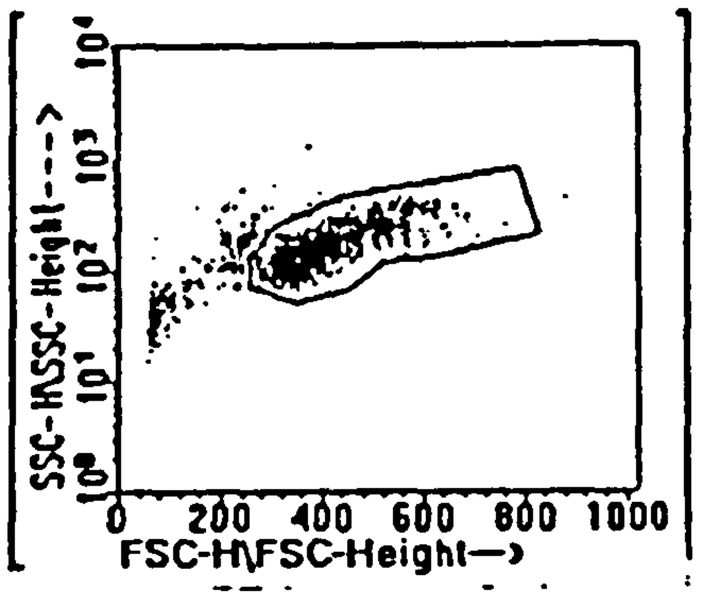
H



MHC II

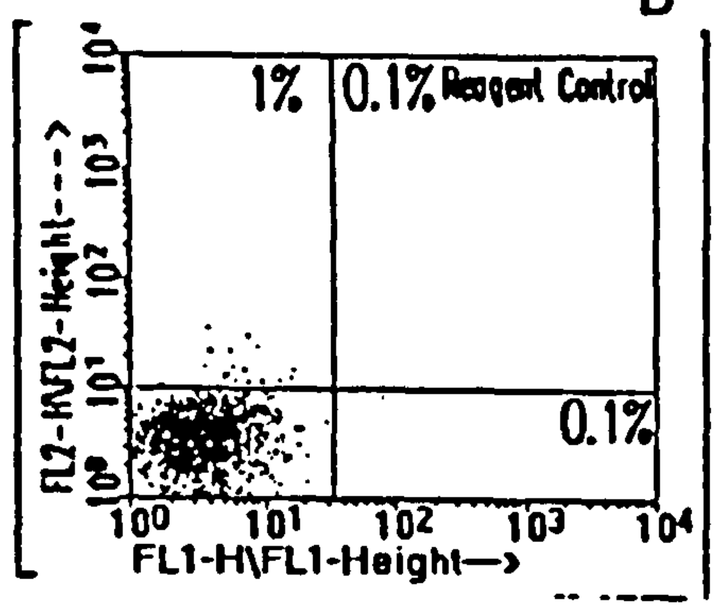
Figure 39: Two colour flow cytometry of PBM cells using the biotinylated cleaved 350 and a panel of monoclonal antibodies. 500 ng of C350 were used for each reaction of panel C-H. Panel A shows the cell population investigated. The reagent control is shown in (B). A combination of C350 and one of a series of monoclonal antibodies were used for panels C-H. ILA-19 anti MHC class I in (C) CC42, anti CD2 (T cells) in (D); CC15, anti WC1 (γ/δ T cells) in (E); CC21, anti WC3 (B cells) in (F); CC94, anti CD11b (mainly monocytes) in (G) and ILA-21 anti MHC class II in (H).

A



Cell population analysed

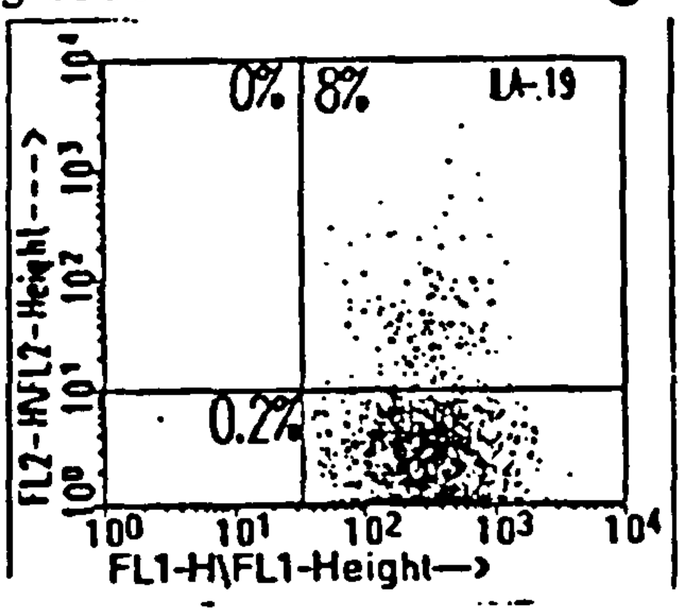
B



Reagent control (no protein, no antibody)

500 ng C350

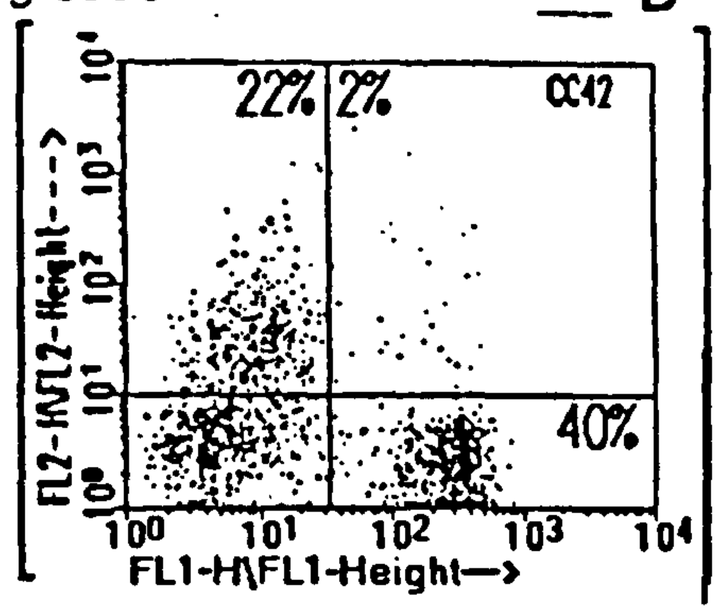
C



MHC I

500 ng C350

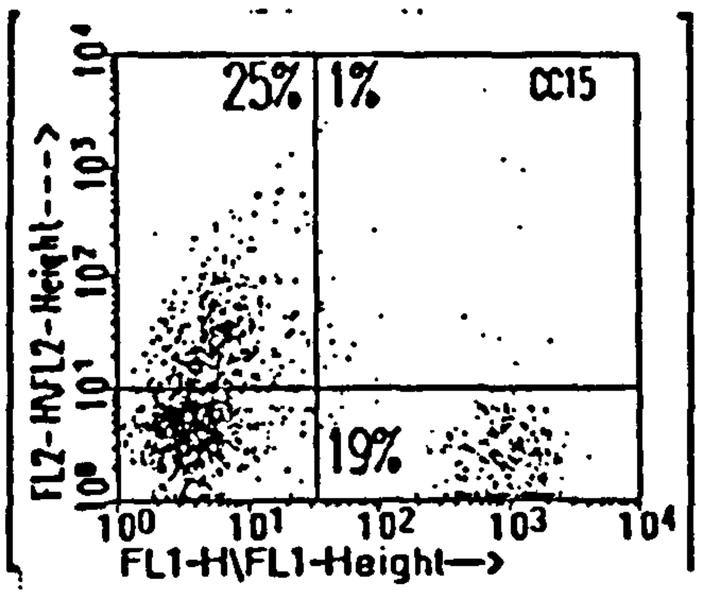
D



CD2

500 ng C350

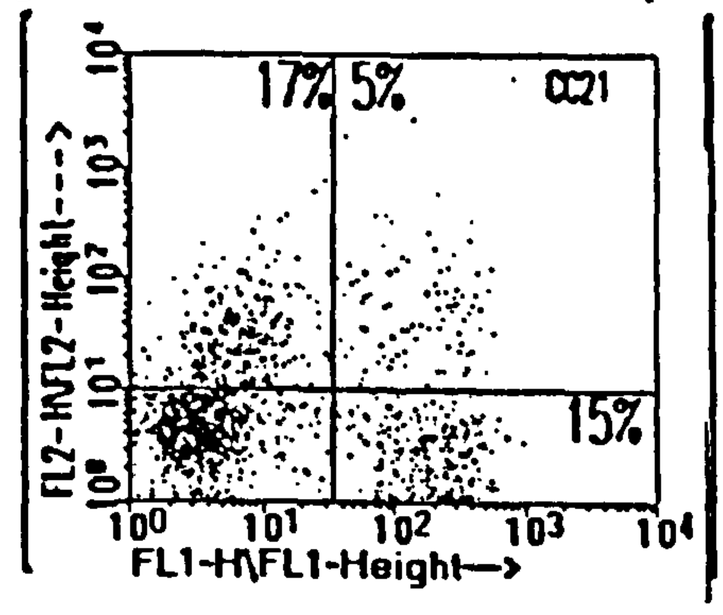
E



WC1

500 ng C350

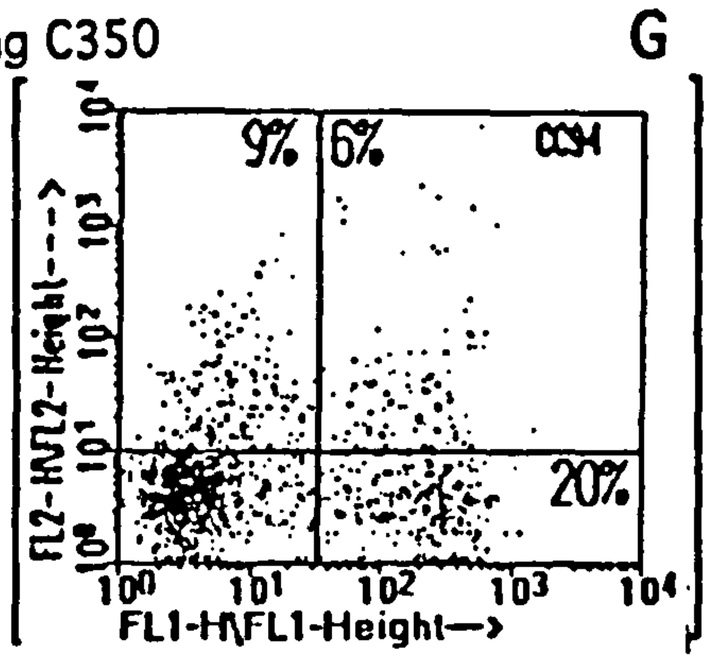
F



WC3

500 ng C350

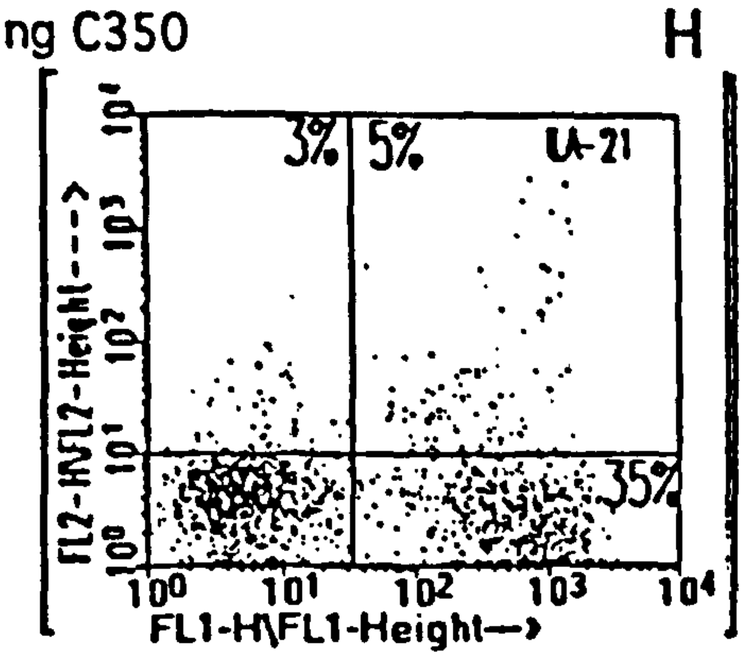
G



CD11b

500 ng C350

H



MHC II

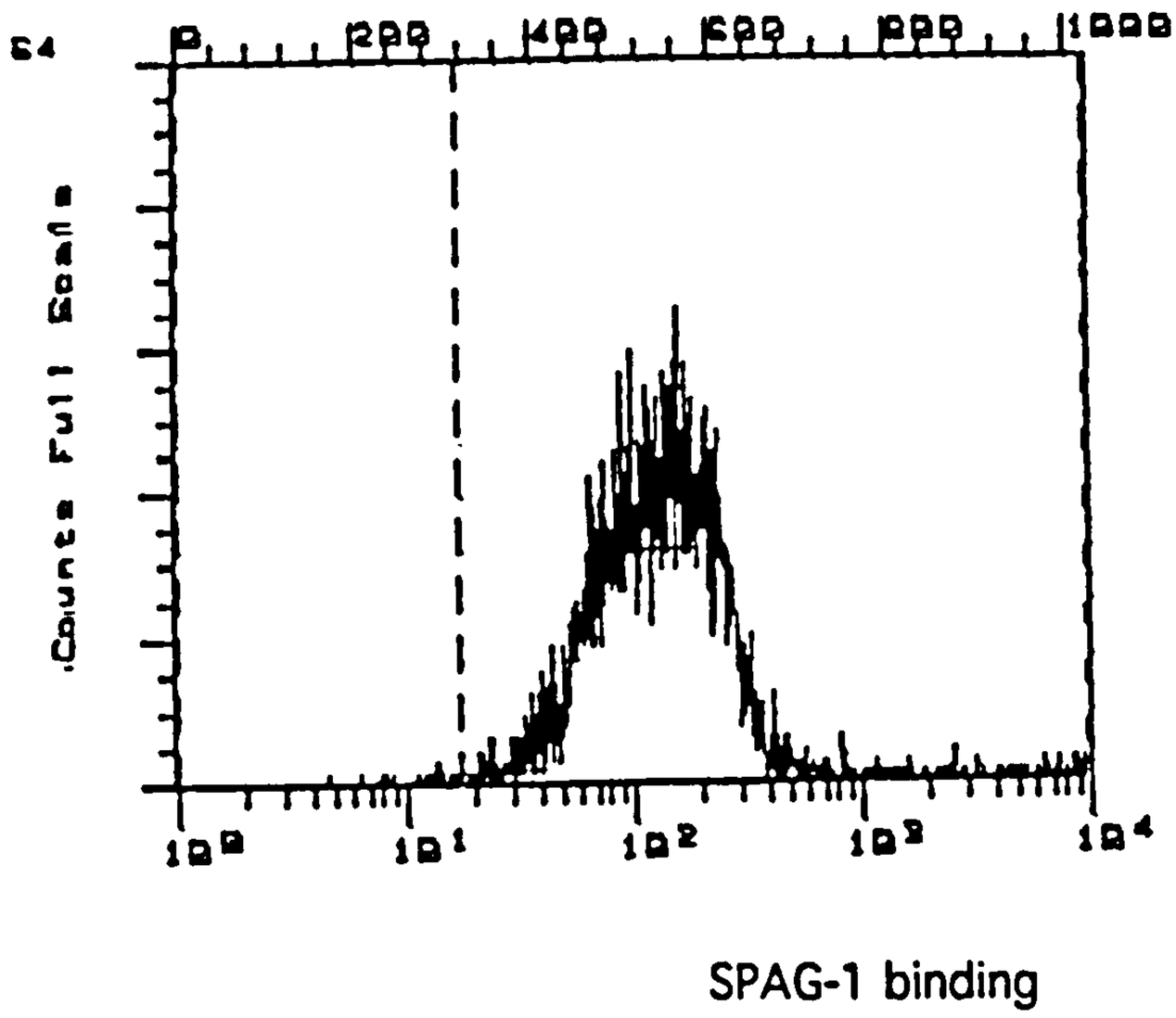
From Figure 38 and Figure 39 it becomes apparent that KP8 and C350 behave in an identical manner, and that the binding results of the SPAG-2 constructs are different to those of SPAG-1. It can be seen in Figure 38 that KP8 binds to approximately 2.4% of CD2 positive T cells, 5% of γ/δ T cells, 16% of B cells, 22% of monocytes, 10% of MHC class II positive cells and 7% of MHC class I positive cells. For C350 the figures are as follow: 5% of T cells, 5% of γ/δ T cells, 25% of B cells, 23% of monocytes, 12% of MHC class II positive cells and 8% of MHC class I positive cells. Interestingly the total number of cells to which both KP8 and C350 bind reduces dramatically if the cells were incubated with CC21, CC94, ILA-19 or with ILA-21 before the cells were incubated with the recombinant proteins. This might indicate that these cell populations contain the target cells for SPAG-2 binding but that competition with these monoclonal antibodies prevents the SPAG-2 molecule from binding.

In summary it can be concluded that both SPAG-2 constructs behave in a very similar manner to each other and that the main targets for SPAG-2 binding are B cells and monocytes, while only a very small number of T cells were bound by SPAG-2, which is of considerable interest since B cells and monocytes are the target for *T. annulata* sporozoite invasion. SPAG-1, on the other hand, binds a large proportion of T cells, which are very unlikely to be the target for sporozoite invasion. SPAG-1 also binds to B cells and monocytes.

5.2.6. SPAG-1 binds to BL3 cells, but SPAG-2 does not.

It would be of interest to find out whether either of the SPAG molecules bind specifically to a cell line which is known to be susceptible to invasion by *T. annulata* sporozoites, and BL3 is such a cell line. Flow cytometry has therefore been employed to test whether recombinant SPAG-1 and KP8 bind specifically to these cells. The concentration of the recombinant proteins was chosen to be the same as for the two-colour flow cytometry in the above experiment (500 ng). The result for SPAG-1 binding is shown in Figure 40. Here it can be seen clearly that, at a concentration of SPAG-1 which revealed specific binding to a subpopulation of PBM cells, almost all BL3 cells are positive

A)



B)

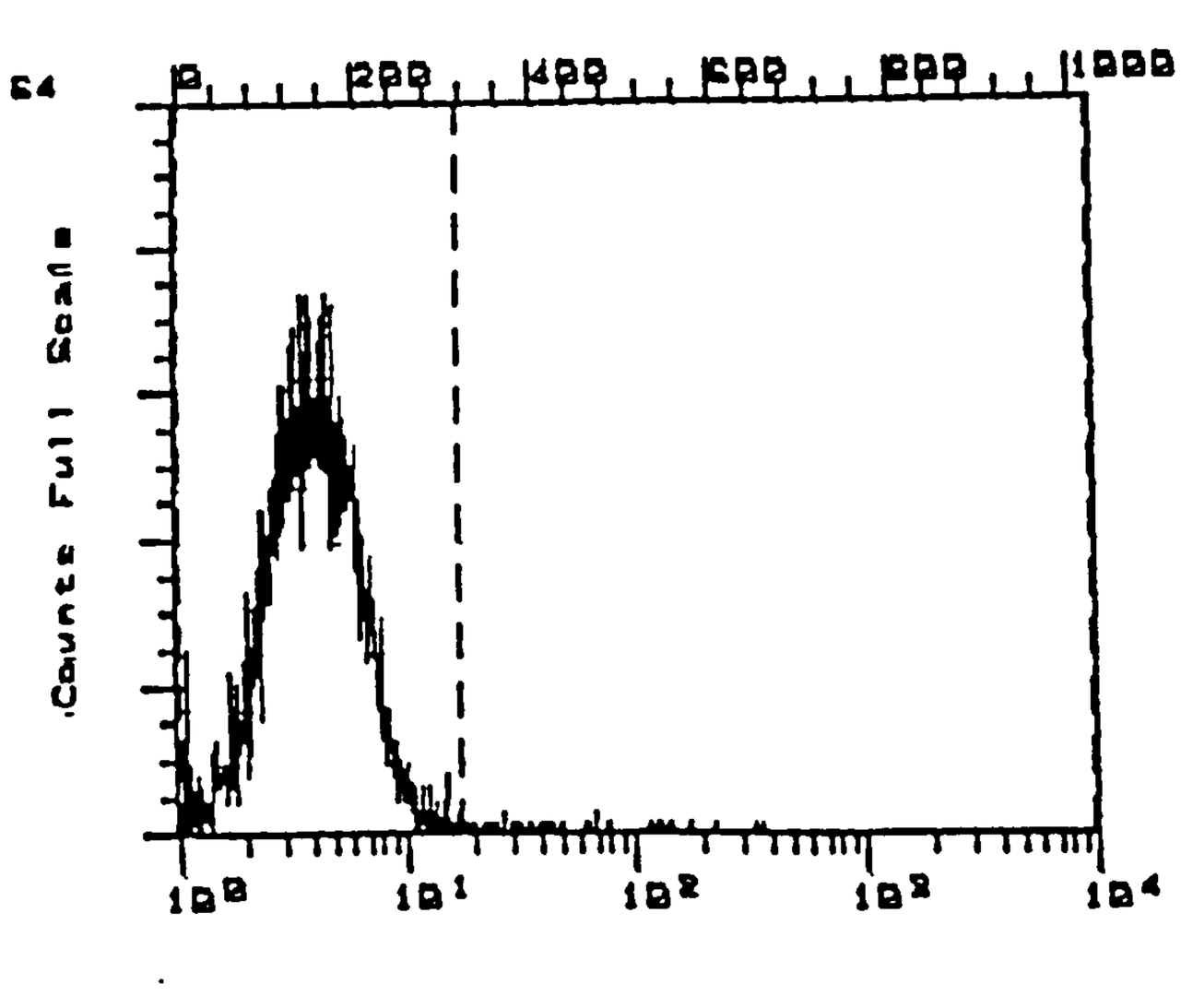


Figure 40: One colour flow cytometry of BL3 cells using biotinylated cleaved SPAG-1. The combination of 500 ng cleaved SPAG-1 and BL3 cells are shown in (a) and the reagent control is shown in (b).

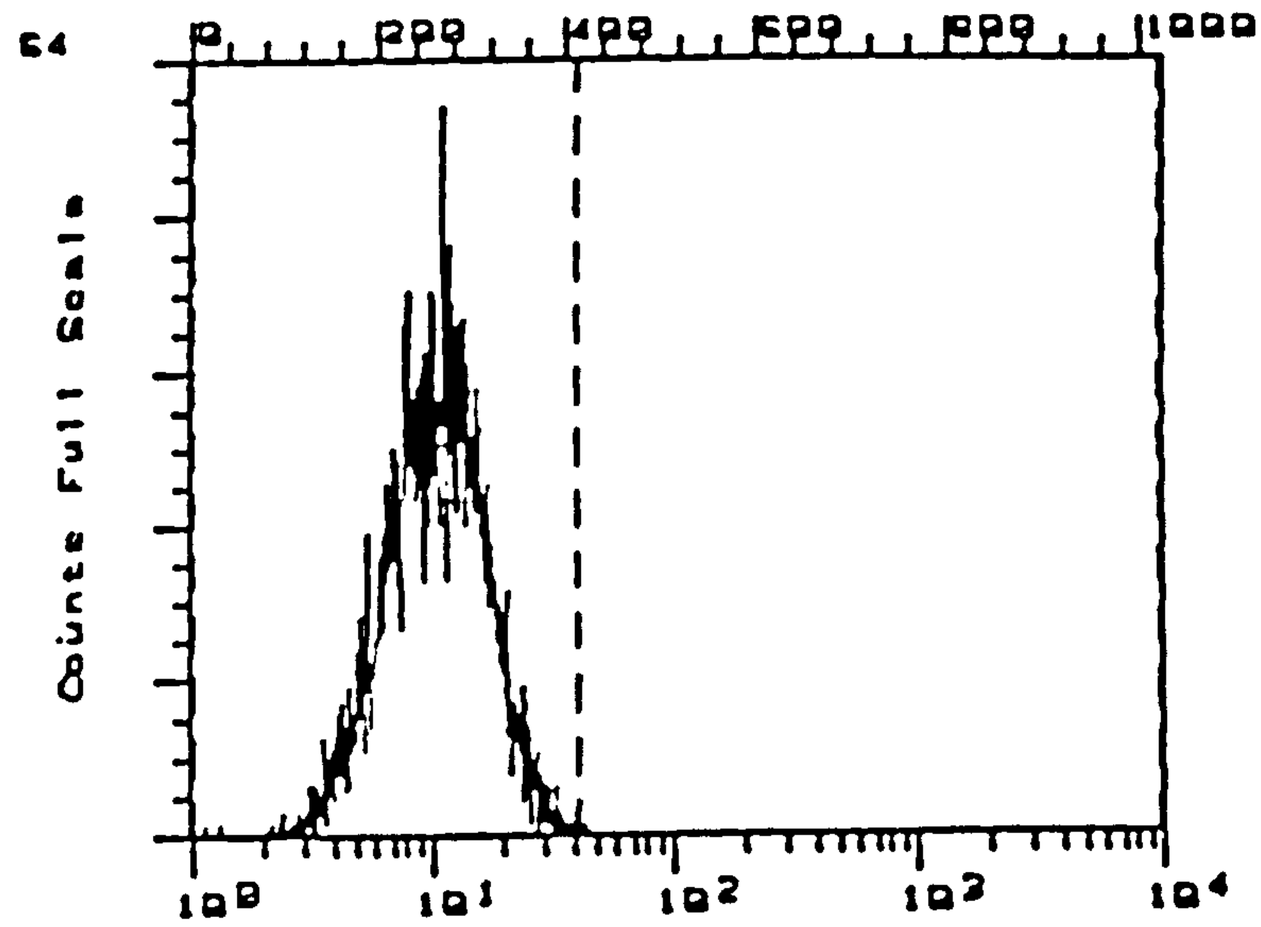
for SPAG-1 binding. This is clear evidence that recombinant SPAG-1 binds specifically to cells which can be invaded by *T. annulata* sporozoites. This observation indicates that the SPAG-1 molecule might be of importance in recognising BL3 cells as host cells. On the other hand KP8 did not bind to BL3 cells at this concentration. This indicates that SPAG-2 or at least the KP8 part of SPAG-2 has no importance in recognising BL3 cells as host cells.

5.2.7. BL3 cells do not express the elastin receptor.

It has been proposed that the elastin receptor might be one of the target receptors on the surface of host cells to which SPAG-1 binds. Through this binding the host cell would be specifically recognised and the invasion process of the *T. annulata* sporozoite be initiated. It is therefore of interest to find out whether the BL3 cell line, which can be invaded by *T. annulata* sporozoites, express the elastin receptor. If these cells expresses the elastin receptor it would support the theory that both SPAG-1 and the elastin receptor are involved in the recognition or invasion process of host cells. On the other hand, if the BL3 cells do not express the elastin receptor it will show that the elastin receptor is probably not of importance for the invasion of BL3 cells, and that SPAG-1 binds to these cells via a different receptor.

Single-colour flow cytometry has been employed to study whether BCZ, an anti-elastin receptor antibody (Mecham et al., 1989), binds to BL3 cells. The antibody concentration was chosen to be the same as for antibodies used in the two-colour flow cytometry and the negative control was TRT1, a mouse monoclonal antibody, which reacts with a surface molecule of a turkey virus (Howard, personal communication). The flow cytometry result is shown in Figure 41. It can clearly be seen that neither of the two monoclonal antibodies bind to BL3 cells. It can therefore be concluded that BL3 cells do not express the elastin receptor.

B)



A)

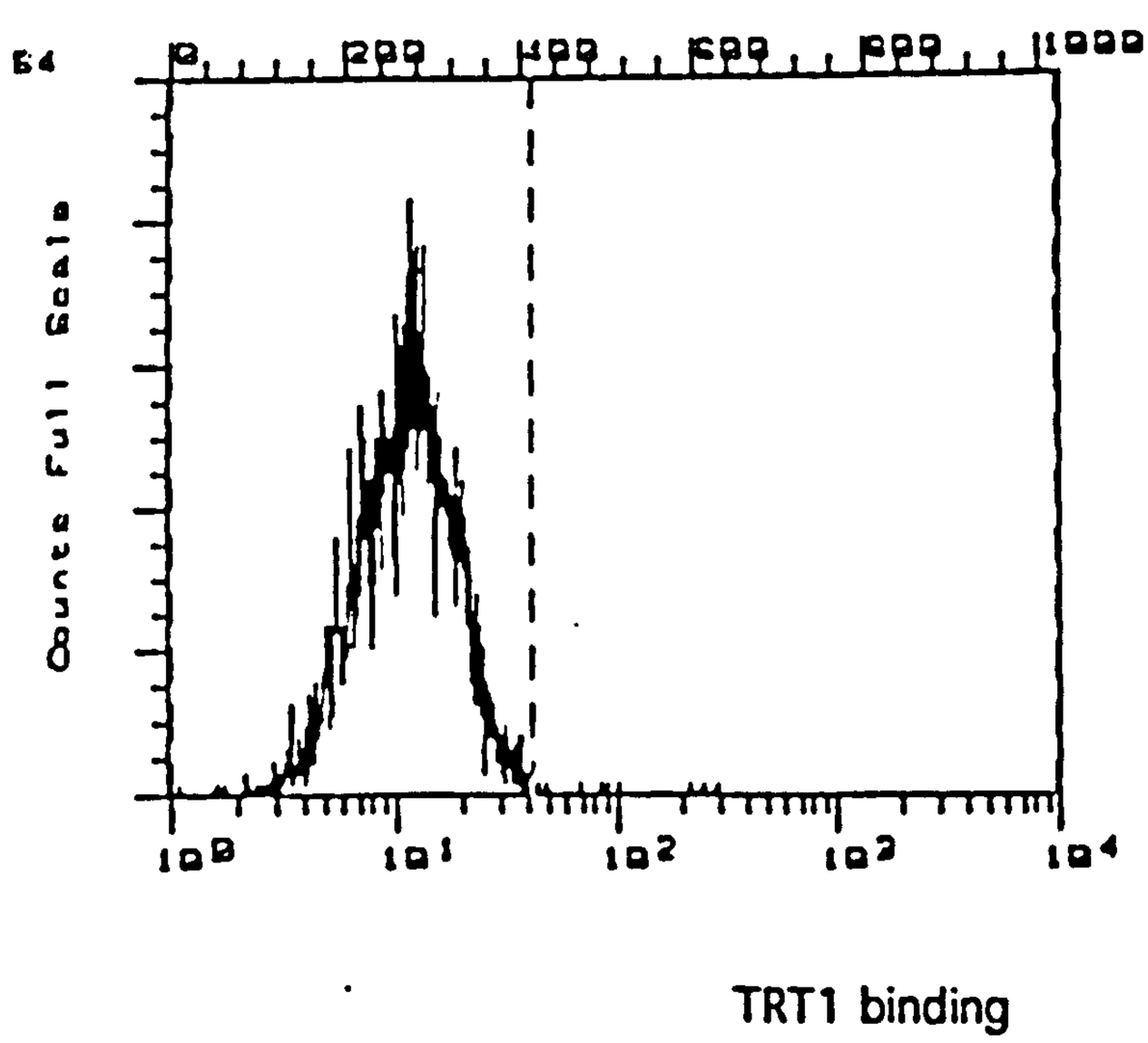


Figure 41: One colour flow cytometry of BL3 cells using the anti elastin receptor antibody BCZ. The combination BCZ and BL3 cells is shown in (a) and a negative control using the TRT1 monoclonal antibody is shown in (b).

5.2.8. How many SPAG-1 receptors are there on BL3 cells?

Once it had been established that SPAG-1 binds specifically to BL3 cells, it led to the question of how many receptors for SPAG-1 binding are expressed on the surface of each BL3 cell. In order to answer this question, a different binding assay was used. SPAG-1 and KP8 were labelled with ^{125}I. An autoradiograph of an SDS-PAGE gel containing both labelled SPAG-1 and KP8 is shown in Figure 42, confirming that the proteins were labelled successfully. The labelled protein was incubated with an aliquot of BL3 cells for 30 minutes at room temperature. A control reaction was performed with a 50-fold excess of unlabelled protein included in addition to the ^{125}I-labelled protein. Each sample point was done in triplicate in order to reduce the potential for error. After incubation the unbound protein was washed off and a count of the gamma radiation emitted from the cells was determined. The samples with the 50-fold excess of unlabelled protein give a value for non-specific binding and if this is subtracted from the total binding observed for each given protein concentration it gives the level of specific binding.

The results for SPAG-1 binding are shown in Figure 43. Figure 43a shows three plots, where the the top graph represents the total counts read for each protein concentration, as shown on the X-axis. The bottom graph reflects the non-specific binding and the solid line (centre) shows the specific binding which was calculated by subtracting the values used to plot the bottom graph from those of the top graph. In the graphs in Figure 43b the counts were converted into ng of SPAG-1 bound to the BL3 cells and again the dark graph in the middle reflects the specific binding of SPAG-1 to the BL3 cells. Finally the graph representing the specific binding in ng, from Figure 43b, was converted into molecules of SPAG-1 bound specifically to each BL3 cell in each reaction. From the observed plateau in the graph of Figure 43c, one can estimate that 70,000 to 75,000 SPAG-1 molecules bind to each BL3 cell. Assuming that only one SPAG-1 molecule binds to each receptor, there are 70,000 to 75,000 SPAG-1 receptors on each BL3 cell.

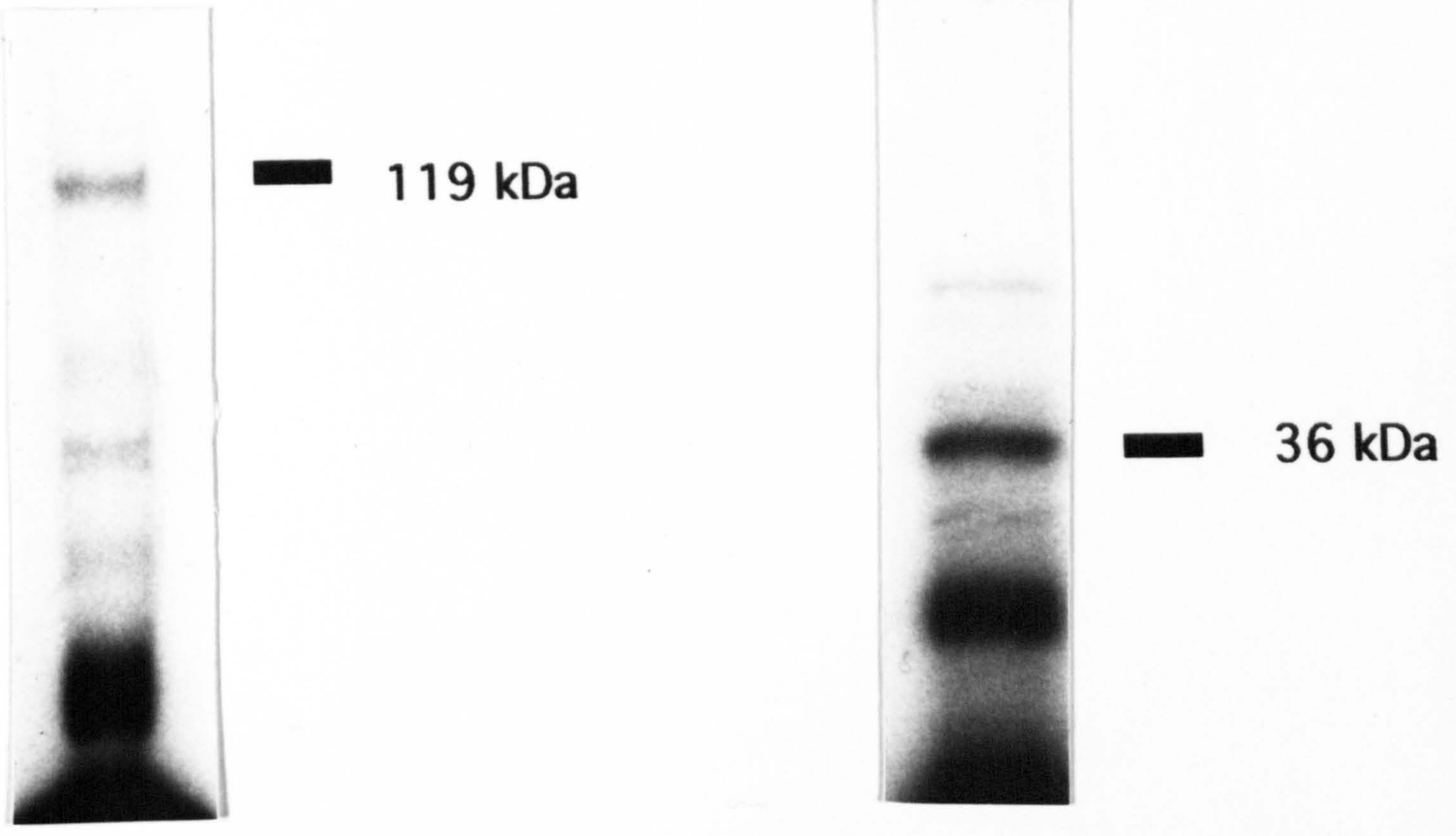


Figure 42: Autoradiograph of ^{125}I labelled SPAG-1 and KP8 on a SDS PAGE gel. ^{125}I labelled cleaved SPAG-1 is shown in (A), ^{125}I labelled cleaved KP8 is shown in (B) and the bars mark the size of the proteins in kDa.

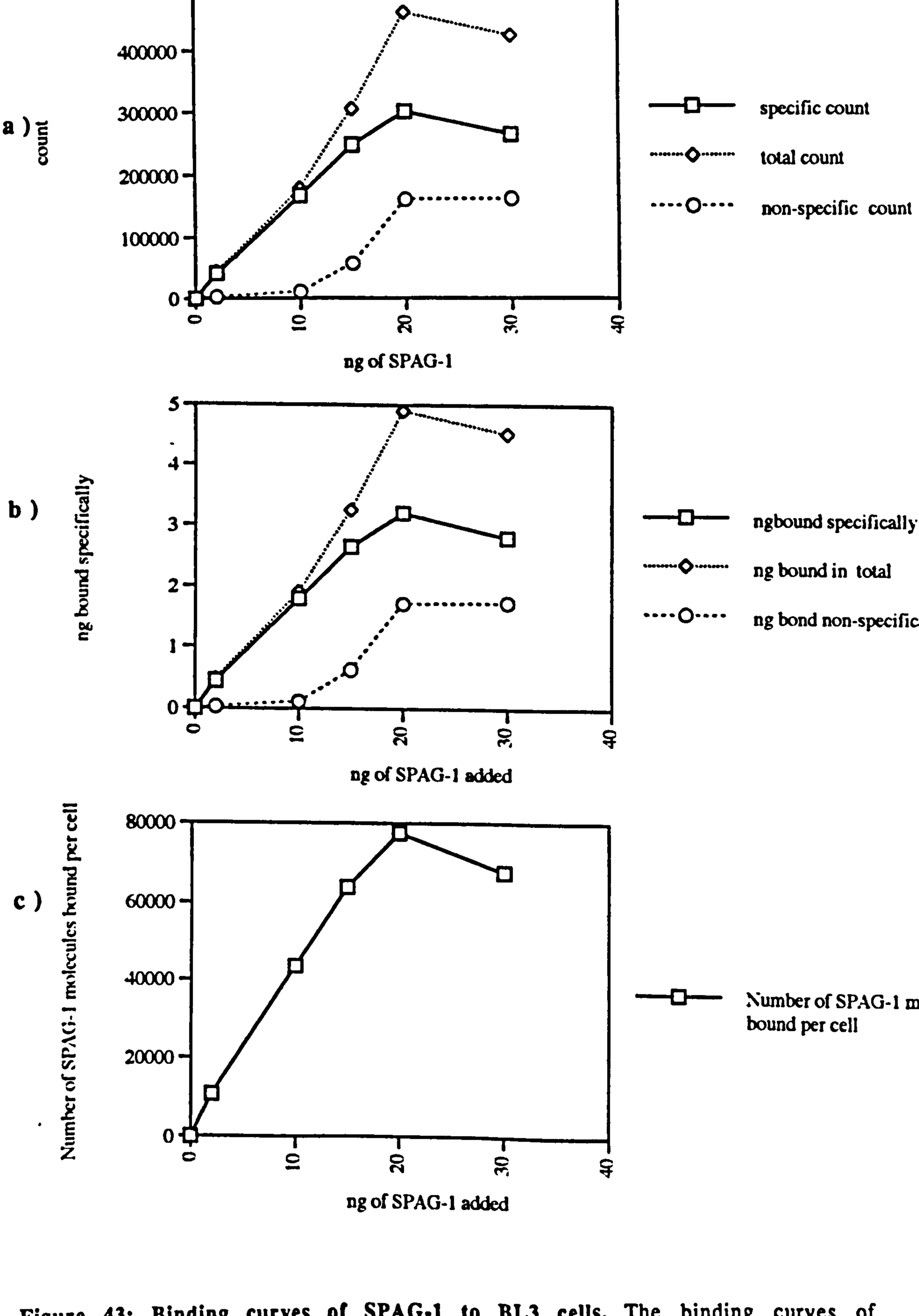


Figure 43: Binding curves of SPAG-1 to BL3 cells. The binding curves of SPAG-1 to BL3 cells with the observed counts are shown in (a). The X axis reflects the amount of labelled protein added and the Y-axis shows the observed counts of labelled protein bound to BL3 cells for each point. The top curve represents the total counts for each point, the bottom curve shows non-specific binding. The curve in the middle represents specific binding of SPAG-1 to BL3 cells and is the difference between total count and non-specific counts. (b) The same data as in (a) but the observed counts are converted into ng of protein bound to BL3 cells. (c) The same data as shown in (a and b) but the ng of protein bound were converted into SPAG-1 molecules bound specifically to each BL3 cell.

When the binding assay was repeated using ^{125}I labelled KP8, no specific binding by this molecule could be detected. The readings obtained were almost as low as the experimental background level and the same readings were obtained with a 50 fold excess of unlabelled KP8. These results are not shown and I conclude that ^{125}I labelled recombinant KP8 does not bind to BL3 cells. This result is in accordance with the binding data of biotinylated KP8 and BL3 cell binding using flow cytometry.

5.3. Discussion

5.3.1. Confirmation of the validity of the protein constructs.

Ideally experiments would have been conducted using live sporozoites of *T. annulata* to study the process of host cell recognition and invasion, but the laboratory in York has no license to keep the sporozoite stage and only a very limited supply of sporozoites is available at the CTVM in Edinburgh. Recombinant proteins were therefore the basis for my investigations of the role of SPAG-1 and SPAG-2 in the process of host cell recognition and invasion by sporozoites. Recombinant proteins, expressed in a prokaryotic expression system, present an obvious limitation since the proteins may not be folded correctly, nor are they processed or glycosylated. Therefore the molecules are only likely to be successful in demonstrating binding to host cells if linear regions are responsible for the binding process. The work presented here demonstrates specific binding of the recombinant proteins showing that linear binding regions may indeed be responsible for the ligand receptor interaction. This has previously been shown to be the case for the CSP-1 protein of *Plasmodium falciparum* in which the recombinant protein, and subsequently peptides derived from it, bind specifically to host cells (Cerami et al., 1992).

5.3.2. Are SPAG-1 and SPAG-2 involved in host cell recognition and invasion?

Binding assays using recombinant SPAG-1 revealed that SPAG-1 binds specifically to a large population of T cells, but more interestingly, to a sub-population of monocytes and B cells. Monocytes and B cells are the main target cells for the sporozoites of *T. annulata* (Spooner et al., 1989 and Glass et al., 1989). It has been argued that T cells are not invaded by *T. annulata* sporozoites, but it has been shown that they can be invaded at a low frequency *in vitro* (Spooner et al., 1989). Another observation was that T cells transformed by *T. annulata* sporozoites lose their T cell surface markers rapidly (Innes et al., 1989). It is therefore difficult to assess which cells are infected and transformed *in vivo* by *T. annulata* sporozoites. Some evidence has been found suggesting that T cells might be a target for *T. annulata* sporozoites *in vivo* since an *in vivo*-derived transformed cell line was obtained which expressed three different T cell markers (Howard et al., 1993). Although these results give no clear evidence that SPAG-1 binds to cells which can be invaded by *T. annulata* sporozoites, there is some correlation of SPAG-1 binding to definite target such as B cells and monocytes. However, SPAG-1 also binds to T cells and one can only speculate whether these are the target cells for the invasion of *T. annulata* sporozoites *in vivo*. The observation that SPAG-1 binds to some cells that are target for invasion and the fact that it is expressed only on the surface of the sporozoite stage, in addition to the blocking data, make SPAG-1 a serious candidate for involvement in the host cell invasion process.

Recombinant KP8 binds to a low proportion of cells, some monocytes and B cells, which have been shown to be the main target for *T. annulata* sporozoite invasion (Spooner et al., 1989 and Glass et al., 1989) and only to a small number of T cells. The role of T cells as putative targets is still questionable. This demonstration of selective binding to the main target cell types for *T. annulata* sporozoite invasion gives support to the theory that SPAG-2 is also involved in the process of host cell recognition and invasion. SPAG-2, however, binds to such low numbers of target cells that it is unlikely that this antigen is soley responsible for the recognition of target cells. Another very interesting finding is that C350 reacts in an almost identical manner to the

recombinant KP8. This indicates that the region which is responsible for SPAG-2 binding must be located within the first 116 amino acids of SPAG-2, as only this region of the protein is common to the two constructs.

The results of the binding assay using SPAG-1 and BL3 cells are of particular interest. These give direct evidence that SPAG-1 binds specifically to a cell line which can be invaded by *T. annulata* sporozoites. This further supports the theory that SPAG-1 is involved in the process of host cell recognition and invasion. The observation that KP8 does not bind to BL3 cells may indicate that the sporozoite might express several different ligands which are involved in the invasion process of different cells. This result does not rule out SPAG-2 as a ligand for host cell recognition and invasion, especially since the binding results obtained by KP8 on peripheral blood mononuclear cells indicate that SPAG-2 binding shows a closer correlation to target cells than that of SPAG-1. Taking into account the blocking data and binding data for both molecules it seems likely that neither of the SPAG molecules alone is sufficient for the sporozoite to identify a host cell. It seems far more likely that both molecules play a part in the recognition and invasion process, but that there are probably other molecules on the surface of the sporozoite which are also involved. Therefore I would propose that a combination of surface molecules of the sporozoite have to bind to a putative host cell in order to initiate the invasion process.

5.3.3. Is the elastin receptor involved in the process of host cell recognition?

The finding that SPAG-1 contains three VGVAPG motifs (Hall et al., 1992) gave rise to the theory that SPAG-1 might bind to the elastin receptor on host cells to initiate invasion by the parasite. Bovine elastin contains VGVAPG in its amino acid sequence and the elastin receptor binds to this peptide. My work provided no further evidence that the elastin receptor is involved in sporozoite invasion. In contrast, evidence was found to suggest that the elastin receptor is very unlikely to be involved in the invasion process. The first indication that the elastin receptor and the VGVAPG molecule of SPAG-1 are not involved in host

cell recognition is based on the finding that BL3 cells which can be infected by *T. annulata* sporozoites do not express the elastin receptor but recombinant SPAG-1 still binds to them. The binding mechanism must therefore involve a different receptor on the surface of the BL3 cell. This, of course, is not conclusive evidence that the elastin receptor is not involved; it might indicate that sporozoites utilise different receptors on the surface of different cell types for invasion.

Further evidence suggesting that the elastin receptor is not involved in host cell invasion is based on an observation that PBM cells sorted into a population which is negative for the expression of the elastin receptor could be infected by *T. annulata* sporozoites equally well as PBM cells which express the receptor (Campbell et al., 1994). More evidence that the elastin receptor is unimportant for the invasion of host cell is derived from infection studies in which the monoclonal antibodies BCZ (anti-elastin receptor (Mecham et al., 1989)) and BA4 (anti VGVAPG (Wrenn et al., 1986)), as well as an excess of elastin peptides, failed to block invasion of sporozoites into host cells *in vitro* (Brown, personal communication). Finally, in chapter 3, I have shown that one of the SPAG-1 alleles does not contain the elastin receptor ligand VGVAPG. I therefore conclude that the elastin receptor is not involved in SPAG-1 binding. It is far more likely that the purpose of the elastin repeats, including VGVAPG, found in SPAG-1 is centred around the evasion of the host immune response (Hall, 1994).

5.3.4. Future work and unanswered questions.

Although I have been able to provide evidence to support the theory that SPAG-1 and SPAG-2 are involved in the process of host cell recognition and invasion, there are still many unanswered questions which might be investigated in the future. These investigations could provide a better understanding of the involvement of these molecules during host cell recognition and invasion of bovine cells by *T. annulata* sporozoites. For example it has been shown that both 1A7 and 4B11 block invasion of sporozoites into host cells, but how exactly do they achieve this? Do they prevent the sporozoite from attaching to the host cell, or can the sporozoite still bind to host cells but the invasion process is

blocked at a later stage? Further questions which are of interest are; how and when are the proteins processed, how many of the processed SPAG-1 and SPAG-2 products are expressed on the surface of the sporozoite and what happens to SPAG-1 after the sporozoite has entered the host cell. To answer these questions a plentiful supply of live sporozoites is needed. The answers might reveal which parts of the SPAG molecules are involved and further elucidate the mechanism of invasion. Recombinant constructs of various parts of the SPAG-1 molecule might be employed in binding studies to map the region of the protein which is involved.

Another major area of research lies in the cloning and sequencing of the full length SPAG-2 gene, which I believe is currently being addressed by Dr. Knight. This might yield more information about the involvement of SPAG-2 in the invasion process through homology to other molecules. Other major areas to follow up are the identification of further molecules expressed on the sporozoite surface which might also be involved in the invasion process, and molecules which are expressed by the target cell and allow sporozoite invasion. Such sporozoite molecules might be included in future sub-unit vaccines which could be developed to protect against tropical theileriosis. Information gained about the ligand-receptor interactions which are involved in the invasion process might, in future, yield the basis for the development of preventative treatments which inhibit sporozoites from binding and invading host cells.

Chapter 6.

General Discussion.

The work described in this D.Phil thesis is aimed to further the understanding of the biology and practical applications of SPAG-1 in particular and, to a lesser extent, SPAG-2. The three main areas of study attempt to elucidate: a) information about SPAG-1 which is of importance for sub-unit vaccine development, b) the process of stage-specific regulation of the expression of the SPAG-1 gene and c) the involvement of the SPAG molecules during host cell invasion. The results obtained in these areas are discussed below.

6.1. SPAG-1 and sub-unit vaccine development.

It has been suggested that SPAG-1 is a candidate for the inclusion in a sub-unit vaccine (Hall et al., 1992). This prediction is based on two observations. The first is that the monoclonal antibody 1A7 blocks sporozoite invasion into host cells *in vitro* (Williamson et al., 1989). The 1A7 epitope has subsequently been mapped to the SR1 region of SPAG-1 (Boulter et al., 1994). Secondly, a trial using recombinant SR1 indicated that vaccination with SPAG-1 induces partial immunity in cattle *in vivo* (Boulter et al., 1995). Further support for the potential of SPAG-1 as a vaccine component is gained from its homology to p67, a sporozoite antigen of *T. parva*. I have shown that p67 and SPAG-1 are 47 % identical at the protein level and that they share parts of the 1A7 epitope. p67 has also shown promising results in vaccination trials providing partial and full protection from *T. parva* infection. The high homology between these two antigens may allow the development of a combined *T. annulata* and *T. parva* vaccine.

If SPAG-1 is to be included as a component in a sub-unit vaccine, then it is important to know as much about the protein as possible. It has been shown that the SPAG-1 gene is linked to an *Eco*RI RFLP (Williamson et al., 1989), but the extent of the SPAG-1 polymorphism was not known. My studies identified the existence of four polymorphic SPAG-1 alleles from two distinct geographical isolates. Analysis of the protein sequence revealed that the C- and N-termini are highly conserved (92 and 97 % identity respectively) and that the central region of the antigen is highly polymorphic. This finding suggests that SPAG-1 sub-unit vaccine components should encompass the conserved N- and C-terminal regions as well as a cocktail of sequences from the polymorphic region of the middle of SPAG-1. The inclusion of this SPAG-1 cocktail would increase the probability of inducing cross-protective responses in the immunised cattle. It might also be important to eliminate the elastin repeats from the vaccine components. This is because they might induce auto-immune responses of the host, although to date, no humoral auto-immune responses against these regions have been reported either in immunised or naturally infected cattle (Boulter, personal communication).

The immunogenicity of the SPAG-1 antigen still need further characterisation as only one protective B-cell epitope has been found in SPAG-1. To optimise the SPAG-1 component in a sub-unit vaccine it is necessary to identify more B-cell and T-cell epitopes. A recent vaccination trial using recombinant SPAG-1 has indicated that there is at least one T-cell suppressor epitope located in the SR1 region of SPAG-1 (Boulter et al., 1995). It is thus important to locate such suppressor epitopes in order to eliminate these from the sub-unit vaccine components. Furthermore, it is also of importance to identify protective T-cell epitopes from polymorphic SPAG-1 antigens in order to optimise the vaccine components and thus to maximise the immune responses. Another important area for further study is the optimisation of the vaccine delivery system. This includes research on antigen expression systems as well as adjuvants.

6.2. Stage-specific regulation of the SPAG-1 gene.

It has been shown that SPAG-1 is only found during the sporozoite stage of the *T. annulata* life cycle (Williamson, 1988). It was also shown that SPAG-1 expression is regulated at the transcriptional level (Williamson et al., 1989). This finding is based on Northern blot analysis of RNA extracted from infected tick salivary gland, macroschizonts and piroplasms. SPAG-1 specific mRNA was only observed during sporozoite development.

In order to investigate the mechanism of transcriptional regulation of the SPAG-1 gene, the 5' region of the gene was sequenced and subsequently the beginning of the mRNA was mapped. The analysis of the 5' region of the mRNA initiation site of SPAG-1 and the homologous region of p67 showed that no conserved sites for known transcription factors are present. But the analysis revealed two palindromic nucleic acid sequences, 6 bp in length, which were conserved. The function of these sequences is unknown but it can be speculated that they might be involved in the stage-specific transcription initiation of both SPAG-1 and p67. To prove this theory one would have to conduct band-shift assays. Attempts to isolate DNA binding proteins which bind to the 5' untranslated region of SPAG-1 failed. If the band-shift assays reveal the involvement of this region, attempts could be made to isolate transcription factors from sporozoite material using affinity columns.

The analysis of the mRNA, using S1 nuclease, confirmed the existence of a 30 bp cryptic intron in the genomic SPAG-1 gene sequence. The existence of this intron was predicted during the comparison of the two full length SPAG-1 gene sequences, cH and gH3.4. The S1 nuclease analysis confirmed the existence of the intron as well as its position. The intron also follows the GT-AG rule for intron splice sites found in eukaryotes (Perlman et al., 1984).

6.3. Functional importance of the SPAG molecules.

It has been suggested that SPAG-1 and SPAG-2 are ligands which are involved in host cell invasion (Hall et al., 1992; Knight, 1993). These suggestions are based on results from sporozoite inhibition assays. For SPAG-1, these assays indicate that the C-terminus is of importance during sporozoite invasion of host cells. Further data indicating the importance of the C-terminus are presented in chapter 3. Sequence analysis of four SPAG-1 genes has indicated that the C-terminus is highly conserved in the alleles investigated, while the middle part of the gene and its predicted amino acid sequence are highly polymorphic. This indicates that the C-terminus is probably conserved for functional purposes such as host cell invasion. The polymorphism of the middle region suggests that the antigen is under selection pressure to change. A possible explanation for the polymorphism is the evasion of the immune responses of the bovine host. It is probable that the PGVGV elastin-like repeats, which are found in the four alleles investigated, have the same function, i.e. immune evasion (Hall, 1994).

The sporozoite inhibition assays however do not indicate during which stage of the invasion process SPAG-1 is involved, and even less is known about the role of SPAG-2. The results of my binding studies using recombinant SPAG molecules have shown that both SPAG-1 and SPAG-2 bind specifically to a sub-population of bovine peripheral blood mononuclear cells. Neither, however, showed enough specificity to recognise only cells which have been identified as targets for *T. annulata* sporozoites. I would thus predict that both antigens are involved in the process of host cell invasion. Further, I suggest that both antigens, and possibly other sporozoite antigens which have not yet been isolated, act together in the recognition of a host cell and are subsequently involved in the invasion process. The process of host cell recognition and invasion needs to be further investigated in order to substantiate this prediction. For example, more data is needed about the fate of SPAG-1 during the invasion process, i.e. is it shed during invasion like p67 in T. *parva*, and which regions of SPAG-1 are involved in the invasion process?

Another very important area of research is the identification and isolation of receptors on the host cell to which the sporozoite ligands bind during the invasion process. The elastin receptor has been suggested to fulfil this role as SPAG-1 contains three VGVAPG hexapeptides (Hall et al., 1992). Data presented in this thesis, however, suggest that the elastin receptor is not of importance during sporozoite invasion. The evidence for this is two-fold. Firstly, the genomic SPAG-1 allele, gH3.4, does not contain any VGVAPG hexapeptides. Since SPAG-1 is a single copy gene, sporozoites expressing the gH3.4 SPAG-1 allele would not be able to invade host cells as they do not express the amino acid sequence recognised by the elastin receptor.

The second line of evidence is based on flow cytometry data using BL3 cells and the anti-elastin receptor antibody BCZ. BL3 cells are targets for sporozoite invasion *in vitro* (Baylis et al., 1992b), but my data shows that these cells do not express the elastin receptor. The two sets of data, therefore, suggest that the elastin receptor cannot be the sole receptor for sporozoite invasion, but does not prove that the receptor is not involved in the invasion process at all. However, there is evidence suggesting that the elastin receptor is not involved, provided independently by Campbell et al. (1994) and Wilkie (personal communication). Campbell and co-workers sorted bovine peripheral blood mononuclear cells into two populations; one that expresses the elastin receptor and one that does not. Both these populations could be infected by sporozoites equally well. The second set of experiments by Wilkie showed that sporozoite invasion of host cells could not be blocked by the addition of the anti-elastin receptor antibody, BCZ, nor with BA4, the anti-VGVAPG antibody (Wilkie, personal communication).

These four independent experiments show that the elastin receptor is not of importance in sporozoite invasion of host cells. This means, therefore, that no host cell receptor for the invasion of host cells by *T. annulata* sporozoites has been identified.

In contrast, the MHC class I molecule has been shown to be involved during the invasion process of *T. parva* (Shaw et al., 1991; Shaw et al., 1995). However, the MHC class I molecule is found on every blood

cell and therefore cannot be responsible for specificity of *T. parva* sporozoites for T cell invasion. As a result, it seems more likely that the MHC class I molecule is involved in the invasion process subsequent to host cell recognition.

References

Achbarou, A., Mercereau_puijalon, O., Autheman, J.M., Fortier, B., Camus, D. and Dubremetz, J.F. (1991) Characterization of microneme proteins of *Toxoplasma gondii*. Molecular and Biochemical Parasitology 47, 223-233.

Adams, J. H., Lee-Sim, B.K., Dolan, S.A., Fang, X., Kaslow, D.C. and Miller, L.H. (1992a) A family of erythrocyte binding proteins of malaria parasites. Proceedings of the National Academy of Sciences, U.S.A. 89, 7085-7089.

Adams, J. H., Fang, X., Kaslow, D.C. and Miller, L.H. (1992b) Identification of a cryptic intron in the *Plasmodium vivax* Duffy binding protein gene. Molecular and Biochemical Parasitology 56, 181-184.

Aikawa, M., Miller, L.H., Johnson, J. and Rabbege, J. (1978) Erythrocyte entry by malarial parasites. The Journal of Cell Biology 77, 72-82.

Aikawa, M., Yoshida, N., Nussenzweig, R.S. and Nussenzweig, V. (1981) The protective antigen of the rodent malaria sporozoites (*Plasmodium berghei*) is a differentiation antigen. Journal of Immunology 126, 2494-2495.

Allsopp, B. A., and Allsopp, M.T.E.P. (1988) *Theileria parva* : genomic DNA studies reveal intra-specific sequence diversity. Molecular and Biochemical Parasitology 28, 77-84.

Allsopp, B., Carrington, M., Baylis, H., Sohal, S., Dolan, T. and Iams, K. (1989) Improved characterisation of *Theileria parva* isolates using the polymerase chain reaction and oligonucleotide probes. Molecular and Biochemical Parasitology 35, 137-148.

Baldwin, C. L., Black, S.J., Brown, W.C., Conrad, P.A., Goddeeris, B.M., Kinuthia, S.W., Lalor, P.A., MacHugh, N.D., Morrison, W.I., Mozaria, S.P., Naessens, J. and Newson, J. (1988) Bovine T cells, B cells, and Null cells are transformed by the protozoan parasite *Theileria parva*. Infection and Immunity 56, 462-467.

Bannister, L. H. and Dluzewski, A.R. (1990) The ultrstructure of red cell invasion in malaria infections: a review. Blood Cell 16, 257-292.

Barnwell, J. W., Nichols, M.E. and Rubinstein, P. (1989) *In vitro* evaluation of the role of the Duffy blood group erythrocyte invasion by *Plasmodium vivax*. Journal of Experimental Medicine 169, 1795-1802.

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

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

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

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

Beckers, C.J.M., Dubremetz, J.-F., Mercereau-Puijalon, O. and Joiner, K.A. (1994) The *Toxoplasma gondii* rhoptry protein ROP2 is inserted into the parasitophorous vacuole membrane, surrounding the intracellular parasite, and is exposed to the host cell cytoplasm. Journal of Cell Biology 127, 947-961.

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

Bermudes, D., Dubremetz, J.F., Achbarou, A. and Joiner, K.A. (1994) Cloning of a cDNA encoding the dense granule protein GRA3 from *Toxoplasma gondii*. Molecular and Biochemical Parasitology 68, 247-257.

Biedenkapp, H., Borgmeyer, U., Sippel, A.E. and Klempnauer, K.-H. (1988) Viral myb encodes a sequence-specific DNA-binding activity. Nature 335, 835-837.

Bienz, M. and Pelham, H.R.B. (1986) Heat shock regulatory elements function as an inducible enhancer when linked to a heterologous promoter. Cell 45, 753-760.

Bishop, R. P., Sohanpal, B.K., Allsopp, B.A., Spooner, P.R., Dolan, T.T. and Morzaria, S.P. (1993) Detection of polymorphism among *Theileria parva* stocks using repetitive, telomeric and ribosomal DNA probes and anti-schizont monoclonal antibodies. Parasitoligy 107, 19-31.

Blackman, M. J., Heidrich, H.G., Donachie, S., McBride, J.S. and Holder, A.A. (1990) A single fragment of a malaria merozoite surface protein remains on the parasite during red cell invasion and is the target of invasion-inhibiting antibodies. Journal of Experimental Medicine 172, 379-382.

Blackman, M. J., Whittle, H. and Holder, A.A. (1991a) Processing of the *Plasmodium falciparum* major 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., Ling, I.T., Nicholls, S.C. and Holder, A.A. (1991b) Proteolytic processing of *Plasmodium falciparum* merozoite surface protein-1 produces a membrane-bound fragment containing two epidermal growth factor-like domains. Molecular and Biochemical Parasitology 49, 29-34.

Blackman, M.J., Chappel, J.A., Shai, S. and Holder, A.A. (1993) A conserved parasite serine-protease processes the *Plasmodium falciparum* merozoite surface protein-1. Molecular and Biochemical Parasitology 62, 103-114.

Blood, C. H., Sasse, J., Brodt, P.A. and Zetter, B.R. (1988) Identification of a tumor cell receptor for VGVAPG, a elastin derived chemotactic peptide. Journal of Cell Biology 107, 1987-1992.

Bloom, B. R. (1979) Games parasites play: how parasites evade the immune surveillance. Nature 279, 21-26.

Borst, P. (1991) Molecular genetics and antigenic variation. Immunology Today 12, A29-A33.

Boulter, N., Knight, P.A., Hunt, P.D., Hennessey, E.S., Katzer, F., Williamson, S., Brown, D., Baylis, H.A. and Hall, F.R. (1994) *Theileria annulata* sporozoite surface antigen (SPAG-1) contains neutralizing determinants in the C terminus. Parasite Immunology 16, 97-104.

Boulter, N.R., Glass, E.J., Knight, P.A., Bell-Sakyi, L., Brown, C.G.D. and Hall, R. (In press) *Theileria annulata* sporozoite antigen fused to hepatitis B core antigen used in vaccination trial. Vaccine.

Bradford, M. (1976) A rapid and sensitive method for the quantification of microgram quantities of protein utilising the principle of protein-dye binding. Analytical Chemistry 72, 248-254.

Braunbreton, C., Blisnick, T., Moralesbetoulle, M.E., Barale, J.C. and Langsley, G. (1994) Malaria parasites - enzymes involved in red-blood-cell invasion. Brazilian Journal of Medical and Biological Research 27, 363-367.

Brown, C. G. D., Stagg, D.A., Purnell, R.E., Kanhai, G.K. and Payne, R.C. (1973) Infection and transformation of bovine lymphoid cells in vitro by infective particles of *Theileria parva*. Nature 245, 101-103.

Brown, C. G. D. and Gray, M.A. (1981) Infection of bovine fibroblastic cell lines with *Theileria parva* and *Theileria annulata*. Transaction of the Royal Society of Tropical Medicine and Hygiene 75, 323-324.

Brown, C. G. D. (1987) *Theileridae*. In: *In vitro* methods for parasite culture. (Taylor, A.E.R. and Baker, J.R., eds.) Academic Press London. pages 230-250.

Brown, C. G. D. (1990a) Control of Tropical Theileriosis (*Theileria annulata* infection) of cattle. Parasitologia 32, 23-31.

Brown, S. D. and Van der Ploeg, L.H.T. (1994) Single-stranded DNA-protein binding in the procyclic acidic repetitive protein (PARP) promoter of *Trypanosoma brucei*. Molecular and Biochemical Parasitology 56, 109-122.

Brown, S. J. and Askenase, P.W. (1983) Immune rejection of ectoparasitic (ticks) by T cell and IgG1 antibody recruitment of basophils and eosinophils. Federation Proceedings 42, 1744-1749.

Brown, W. C., Sugimoto, C., Conrad, P.A. and Grab, D.J. (1989) Differential response of bovine T-cell lines to membrane and soluble antigens of *Theileria parva* schizont infected cells. Parasite Immunology 11, 567-583.

Brown, W. C., Lonsdale-Eccles, J.D., deMartini, J.C. and Grab, D.J. (1990b) Recognition of soluble *Theileria parva* antigen by bovine helper T cell clones: characterization and partialpurification of the antigen. The Journal of Immunology 144, 271-277.

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

Campbell, J. D., Brown, D.J., Glass, E.J., Hall, F.R. and Spooner, R.L. (1994) *Theileria annulata* sporozoite targets. Parasite Immunology 16, 501-505.

Carey, M., Lin, Y.-S., Green, M.R. and Ptashne, M. (1990) A mechanism for synergistic activation of a mammalian gene by GAL4 derivatives. Nature 345, 361-364.

Cerami, C., Frevert, U., Sinnis, P., Takacs, B., Clavijo, P., Santos, M.J. and Nussenzweig, V. (1992) The basolateral domain of the hepatocyt plasma membrane bears receptors for the circumsporozoite protein of *Plasmodium falciparum* sporozoites. Cell 70, 1021-1033.

Cerami, C., Frevert, U., Sinnis, P., Takacs, B. and Nussenzweig, V. (1994) Rapid clearance of malaria circumsporozoite protein (CS) by hepatocytes. Journal of Experimental Medicine 179, 695-701.

Cesbron-Delauw, M.F., Guy, B., Tropier, G., Pierce, R.J., Lenzen, G., Cesbron, J.Y., Charif, H., Lepage, P., Darcy, F., Lecocq, J.P. and Capron, A. (1989) Molecular characterization of a 23-kilodalton major antigen secreted by *Toxoplasma gondii*. Procedings of the National Academy of Science of the USA 86, 7537-7541.

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 the merozoite surface protein-1. Molecular and Biochemical Parasitology 60, 303-312.

Chaudhri, S. S., and Subramanian, G. (1992a) In vitro infectivity of sporozoites of *Theileria annulata* to viable lymphocytes of cattle and buffalo. Indian Journal of Animal Sciences 62, 225-226.

Chaudhri, S. S., and Subramanian, G. (1992b) Cell-mediated immune responses to sporozoites and macroschizonts of *Theileria annulata*. Veterinary Parasitology 41, 23-34.

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

Clarke, L. E., Tomley, F.M., Wisner, M.H., Foulas, I.J. and Boursell, M.E.G. (1990) Regions of an Eimeria teneila antigen contain sequences which are observed in circumsporozoite proteins from *Plasmodium* spp and which are related to the thrombospondin gene family. Molecular and Biochemical Parasitology 41, 269-280.

Clarke, R. M., 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.

Clayton, C. E. and Mowatt, M. (1989) The procyclic acid repetitive proteins of *Trypanosoma brucei*. Journal of Biological Cemistry 264, 15088-15093.

Conrad, P. A., Baldwin, C.L., Brown, W.C. Sohanpal, B., Dolan, T.T., Goddeeris, B.M., Demartini, J.C. and Ole-Moiyoi, O.K. (1989) Infection of bovine T cell clones with genotypically distinct *Theileria parva* parasites and analysis of their cell surface phenotype. Parasitology 99, 205-213.

Cooper, J. A., Cooper, L.T. and Saul, A.J. (1992) Mapping of the region predominantly recognized by antibodies to the *Plasmodium falciparum* merozoite surface antigen MSA-1. Molecular and Biochemical Parasitology 51, 301-312.

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

Courey, A. J. and Tjian R. (1988) Analysis of Sp1 in vivo reveals multiple transcriptional domains including a novel glutamine-rich activation motif. Cell 55, 887-898.

Cowdry, E. V. and Ham, A.W. (1932) Studies on East Coast Fever. I. The life cycle of the parasites in the ticks. Parasitology 24, 1-43.

Demelo, E.J.T., Decarvalho, T.U. and Desouza, W. (1992) Penetration of *Toxoplasma gondii* into host cells induces changes in the distribution of the mitochondria and the endoplasmic-reticulum. Cell Structure and Function 17, 311-317

Devereux, J., (1989) Genetics computer group sequence analysis package. Proceedings of the National Academy of Science, USA 74, 5463-5467.

Dickson, J. and Shiels, B.R. (1993) Antigenic diversity of a major merozoite surface molecule in *Theileria annulata*. Molecular and Biochemical Parasitology 57, 55-64.

Dobbelaere, D. A. E., Spooner, P.R., Barry, W.C. and Irvin, A.D. (1984) Monoclonal antibody neutralizes the sporozoite stage of different *Theileria parva* stocks. Parasite Immunology 6, 361-370.

Dobbelaere, D. A. E., Leitch, B.L., Voigt, W.P. and Irvin, A.D. (1985) *Theileria parva* : Expression of a sporozoite surface coat antigen. Experimental Parasitology 60, 90-100.

Dolan, T. T. (1989) Theileriasis: a comrehensive review. Revue Scientifique et Technique (Paris) 8, 11-36.

Dolan, S.A., Miller, L.H. and Wellems, T.E. (1990) Evidence for a switching mechanism in the invasion of erythrocytes of *Plasmodium falciparum*. Journal of Clinical Investigation 86, 618-624.

Dolan, S.A., Proctor, J.L., Alling, D.W., Okubo, Y., Wellems, T.E. and Miller, L.H. (1994) Glycophorin-B as an EBA-175 independent *Plasmodium falciparum* receptor of human erythrocytes. Molecular and Biochemical Parasitology 64, 55-63.

Dschunkowsky, E. and Luhs, J. (1904) Die *Piroplasmosen* der Rinder Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene 35, 486-492.

Dubremetz, J.F. and Schwartzman, J.D. (1993) Subcellular organelles of *Toxoplasma gondii* and host cell invasion. Research in Immunology 144, 31-33.

Dyer, M. and Tait A. (1987) Control of lymphoproliferation by *Theileria annulata*. Parasitology Today 3, 309-311.

Dyer, M., Shiels, B. and Tait, A. (1992) *Theileria annulata* : Alterations in phosphoprotein and protein kinase activity profiles in infected leukocytes of the bovine host, *Bos taurus*. Experimental Parasitology 74, 216-227.

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

Eichhorn, M. and Dobbelaere, D.A.E. (1994) Induction of Signal Transduction pathways in lymphocytes infected by *Theileria parva*. Parasitology Today 10, 469-472.

Ellis, J., Ozaki, L.S., Gwadz, R.W., Cochrane, A.H., Nussenzweig, V. and Nussenzweig, R.S. (1983) Cloning and expression in *E. coli* of the malarial sporozoite surface antigen gene from *Plasmodium falciparum*. Nature 302, 536-538.

Elmoudni, B., Philippe, M., Monsigny, M. and Schrevel, J. (1993) N-acetylglucosamine-binding proteins on *Plasmodium falciparum* merozoite surface. Glycobiology 3, 305-312.

Elsholtz, V., Albert, V.R., Treacy, M.N. and Rosenfeld, M.G. (1990) A two-base change in a POU factor binding site sitches pituitary-specific to lymphoid specific gene expression. Genes and Development 4, 43-51.

Emery, D. L., MacHugh, N.D. and Morrison, W.I. (1988) *Theileria parva* (Muguga) infects bovine T-lymphocytes in vivo and induces expression of BoT4 and BoT8. Parasite Immunology 10, 379-391.

Evans, R. M. (1988) The steroid and thyroid hormone receptor gene superfamily. Science 240, 889-895.

Falkner, F. G., Moickat, R. and Zachau H.G. (1986) Sequences closely related to an immunoglobulin promoter/enhancer element occur also upstream of other eukaryotic and prokaryotic genes. Nucleic Acid Research 13, 7847-7863.

Fawcett, D. W. and Doxsey S. (1982) The entry of sporozoites of *Theileria parva* into bovine lymphocytes *in vitro*. Electron microscopic observations. European Journal of Cell Biology 27, 10-21.

Fawcett, D. W., Buscher, G. and Doxsey, S. (1982) Salivery gland of the tick vector of East Coast Fever. III. The ultrastructure of sporogeny in *Theileria parva*. Tissue and Cell 14, 183-206.

Fawcett, D., Musoke, A. and Voigt, W. (1984) Interaction of sporozoites of *Theileria parva* with bovine lymphocytes *in vitro* . I. Early events after invasion. Tissue and Cell 16, 873-884.

Fenton, B., Clark, J.T., Wilson, C.F., McBride, J.S. and Walliker, D. (1989) Polymorphism of a 35-48 kDa *Plasmodium falciparum* merozoite surface antigen. Molecular and Biochemical Parasitology 34, 79-86.

Fidock, D.A., Bottius, E., Brahimi, K., Moelans, I.I., Aikawa, M., Konings, R.N., Certa, U., Olafsson, P., Kaidoh, T. and Asavanich, A. (1994a) Cloning and characterization of a novel *Plasmodium falciparum* sporozoite surface antigen, STARP. Molecular and Biochemical Parasitology 64, 219-232.

Fidock, D.A., Sallenave-Sales, S., Sherwood, J.A., Gachihi, G.S., Ferreira-da-Cruz, M.F., Thomas, A.W. and Druilhe, P. (1994b) Conservation of the *Plasmodium falciparum* sporozoite surface protein gene, STARP, in field isolates and distinct species of *Plasmodium*. Molecular and Biochemical Parasitology 67, 255-267.

Florent, I., Lebonniec, S., Carcy, B., Grellier, P., Mercereaupuijalon, O., Bonnefoy, S., Dhermy, D., Monsigny, M., Mayer, R. and Schrevel, J. (1994) *Plasmodium falciparum* proteases- cloning of a putative gene coding for the merozoite proteinase for erythrocyte invasion (MPEI) and determination of hydolysis sites of spectrin by PF37 proteinase. Memorias do Instituo Oswaldo Cruz 89, S2, 47-49.

Flores, O., Lu, H. and Reinberg, D. (1992) Factors involved in specific transcription by mammalian RNA polymerase II. Journal of Biological Chemistry 267, 2786-2793.

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

Frevert, U. (1994) Malaria sporozitehepatocyte interactions. Experimental Parasitology 79, 206-210.

Foussard, F., Leriche, M.A. and Dubremetz, J.F. (1991) Characterization of the lipid content of *Toxoplasma gondii* rhoptries. Parasitology 102, 367-370.

Furtado, G.C., Cao, Y. and Joiner, K.A. (1992) Laminin on *Toxoplasma gondii* mediates parasite binding to β1 integrin receptor α6β1 on human foreskin fibroblasts and chinese hamster ovary cells. Infection and Immunity 60, 4925-4931.

Gajadhar, A. A., Maequardt, W.C., Hall, R., Gunderson, J., Ariztia-Carmona, E.V. and Sogin, M.L. (1991) Ribosomal RNA sequences of *Sarcocystis muris*, *Theileria annulata* and *Crypthecodinium cohnii* reveal evolutionary relatioships among apicomplexans, dinoflagellates, and ciliates. Molecular and Biochemical Parasitology 45, 147-154.

Gaston, K. and Fried, M. (1992) The isolation of transcription factors from λgt11 cDNA expression libraries: human steroid 5a-reductase 1 has sequence-specific DNA binding activity. Nucleic Acid Research 20, 6297-6301.

Gibson, W. C., Schwinkels, B.W. and Borst, P. (1988) Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in *Trypanosoma brucei*. Journal of Molecular Biology 201, 315-325.

Gill, B. S., Bhattacharyulu, Y. and Kaur, D. (1976) Immunization against bovine tropical theileriosis (*Theileria annulata* infection). Research in Veterinary Science 21, 146-149.

Gill, B. S., Bansal, G.C., Bhattacharyulu, Y., Kaur, D. and Singh, A. (1980) Immunological relationship between strains of I Dschunkowsky and Luhs 1904. Research in Veterinary Science 29, 93-97.

Gilmore, T. D. (1990) NF-kappa B, KBF1, dorsal and related matters. Cell 62, 841-843.

Glascodine, J., Tetley, L., Tait, A., Brown, D. and Shiels, B. (1990) Developmental expression of a *Theileria annulata* merozoite surface antigen. Molecular and Biochemical Parasitology 40, 105-112.

Glass, E. J., Innes, E.A., Spooner, R.L. and Brown, C.G.D. (1989) Infection of bovine monocyte/macrophage populations with *Theileria annulata* and *Theileria parva*. Veterinary Immunology and Immunopathology 22, 355-368.

Glass, E. J., Spooner, R.L. (1990) Parasite-accessory cell interactions in Theileriosis. Antigen presentation by *Theileria annulata* -infected macrophages and production of continuously growing antigen-presenting cell lines. The European Journal for Immunology 20, 2491-2497.

Godowski, P. J., Picard, D. and Yamaoto, K. (1988) Signal transduction and transcriptional regulation by glococorticoid receptor: regulation by glucocorticoid receptor-Lex A fusion proteins. Science 241, 812-816.

Goebl, M. G. (1990) The PU-1 transcription factor is the product of the putative oncogene Spi-1. Cell 61,

Good, M. F., Pombo, D., Quakyi, I., Riley, E.M., Houghten, R.A., Menon, A., Alling, D.W., Berzofsky, J.K. and Miller, L.H. (1988b) Human T-cell recognition of the circumsporozoite protein of *Palsmodium falciparum* : Immunodominant T-cell domains map to the polymorphic regions of the molecule. Proceedings of the National Academy of Science, USA 85, 1199-1203.

Good, M. F., Pombo, D., Maloy, W.L. de la Cruz, V.F., Miller, L.H. and Berofsky, J.A. (1988c) Parasite polymorphism present within minimal T-cell epitopes of *Plasmodium falciparum* circumsporozoite protein. Journal of Immunology 1645-1650.

Goundis, D. and Reid, K.B.M. (1988) Properdin, the terminal complement components, thrombospondin and the circumsporozoite protein of malaria parasites contain simmilar sequence motifs. Nature 335, 82-85.

Gray, M. A. and Brown, C.G.D. (1981) In vitro neutralisation of theilerial sporozoite infectivity with immune serum. In "Advances in the Control of Theileriosis. Eds. A.D. Irvin, M.P. Cunningham and A.S. Young. Martinus Nijhoff Publishers, The Hague. pp 127-129.

Grellier, P., Precigout, E., Valentin, A., Carc, B. and Schrevel, J. (1995) Characterization of a new 60-kDa apical protein of *Plasmodium falciparum* merozoite expressed in late schizogeny. Biology of the Cell 82, 129-138

Grimwood, J. and Smith, J.E. (1992) *Toxoplasma gondii*: The role of a 30-kDa surface protein in host cell invasion. Experimental Parasitology 74, 106-111.

Gunderson, J. H., Sogin, M.L., Wollett, G., Hollingdale, M., de la Cruz, V.F., Waters, A.P. and McCutchan, T.F. (1987) Structurally distinct, stage-specific ribosomes occur in *Plasmodium*. Science 238, 933-937.

Guttinger, M., Caspers, P., Takacs, B., Trzeciak, A., Gillessen, D., Pink, J.R.L. and Sinigaglia, F. (1988) Human T cells recognize polymorphic and non-polymorphic regions of the *Plasmodium falciparum* circumsporozoite protein. EMBO 7, 2555-2558.

Hadani, A., Pipano, E. and Senf, Z. (1963) Studies on the transmissoin of *Theileria annulata* by ticks (Ixodoides, Ixodidae), I. *Hyalomma excavatum*. Journal of Protozoology 10, Supplement, 35.

Hadley, T. J., Klotz, F.W. and Miller, L.H. (1986) Invasion of erythrocytes by malaria parasites: A cellular and molecular overview. Annual Reviews in Microbiology 40, 451-477.

Hall, R., Osland, A., Hyde, J.E., Simmons, D.L., Hope, I.A. and Scaife, J.G. (1984) Processing, polymorphism, and biological significance of P190, a major surface antigen of the erythrocitic forms of *Plasmodium falciparum*. Molecular and Biochemical Parasitology 11, 61-80.

Hall, R., Hyde, J.E., Goman, M., Simmons, D.L., Hope, I.A., MacKay, M., Scaife, J., Merkli, B., Richle, R. and Stocker, J. (1984) Major surface-antigen gene of a human malaria parasite cloned and expressed in bacteria. Nature 311, 379-382.

Hall, F. R., (1988) Antigens and immunity in *Theieleria annulata*. Parasitology Today 4, 257-261.

Hall, R., Coggins, L., McKellar, S., Shields, B. and Tait, A. (1990) Characterisation of an extrachromosomal DNA element from *Theileria annulata*. Molecular and Biochemical Parasitology 38, 253-260.

Hall, R., Hunt, P.D., Carrington, M., Simmons, D., Williamson, S., Mecham, R.P. and Tait, A. (1992a) Mimicry of the elastin repetitive motifs by *Theileria annulata* sporozoite surface antigen. Molecular and Biochemical Parasitology 53, 105-112.

Hall, R. (1994) Molecular mimicry. Advances in Parasitology 34, 81-132.

Hanes, S. D. and Brent, R. (1989) DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix 9 Cell 57, 1275-1283.

Haynes, J. D., Dalton, J.P., Klotz, F.W., McGinnis, M.H., Hadley, T.J., Hudson, D.E. and Miller, L.H. (1988) Receptor-like specificity of a *Plasmodium knowlesi* malaria protein that binds to Duffy antigen ligands on erythrocytes. Journal of Experimental Medicine 167, 1873-1881.

Headstorm, R. C., Campbell, J.R. and Leef, M.L. (1990) A malaria sporozoite surface antigen distinct from the circumsporozoite protein. World Health Organization Bulletin 68, 152-157.

Hesselgesser, J., Chitnis, C.E., Miller, L.H., Yansura, D.G., Simmons, L.C., Fairbrother, W.J., Kotts, C., Wirth, C., Gillececastro, B.L. and Horuk, R. (1995) A mutant of melanoma growth-stimmulating activity does not activate neutrophils but blocks erythrocyte invasion by malaria. Journal of Biological Chemistry 270, 11472-11476.

Hilder, A. A., Sanghu, 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 antigen of *Plasmodium falciparum*. Parasitology 94, 199-208.

Hinek, A., Wrenn, D.S., Mecham, R.P. and Barondes, S.H. (1988) The elastin receptor: A galactoside-binding protein. Science 239, 15391541.

Holder, A. A. (1994) Proteins on the surface of the malaria parasite and cell invasion. Parasitology 108, S5-S18.

Hollenberg, S. M. and Evans, R.M. (1988) Multiple and cooperativetrans-acctivation domains of the human glucocorticoid receptor. Cell 55, 899-906.

Holloway, S. P., Sims, P.F.G., Delves, C.J., Scaif, J.G. and Hyde, J.E. (1989) Isolation of a-tubulin genes from the human malaria parasite *Plasmodium falciparum* : sequence analysis of α-tubulin I. Molecular Microbiology 3, 1501-1510.

Holloway, S. P., Gerousis, M., Delves, C.J., Sima, P.F.G., Scaife, J.G. and Hyde, J.E. (1990) The tubulin genes of the human malaria parasite *Plasmodium falciparum*, their chromosomal location and sequence analysis of the α-tubulin II gene. Molecular and Biochemical Parasitology 43, 257-270.

Hooshmand-Rad, P. and Hawa, N.J. (1973a) Malignenet theileriosis of sheep and goats. Tropical Animal Health and Production 5, 97-102.

Hooshmand-Rad, P. and Hawa, N.J. (1973b) Transmission of *Theileria hericl* in sheep by *Hyalomma anatolicum anatolicum*. Tropical Animal Health and Production 5, 103-109.

Hope, I. A. and Struhl, K. (1986) Functional dissection of a eukaryotic transcriptional activator GCN4 of yeast. Cell 46, 885-894.

Horuk, R., Chitnis, C.E., Darbonne, W.C., Colby, T.J., Rybicki, A., Hadley, T.J. and Miller, L.H. (1993) Science 261, 1182-1184.

Howard, C. J., Sopp, P., Preston, P.M., Jackson, L.A. and Brown, C.G.D. (1993) Phenotypic analysis of bovine leukocyte cell lines infected with *Theileria annulata*. Veterinary Immunology and Immunopathology 39, 275-282.

Huard, T. K., Malinoff, H.F. and Wicha, M.S. (1986) Macrophages express a plasma membrane receptor for basement membrane laminin. American Journal of Pathology 123, 365-370.

Hudson-Taylor, D.E., Dolan, S.A., Klotz, F.W., Fujioka, H., Aikawa, M., Koonin, E.V. and Miller, L.H. (1995) *Plasmodium falciparum* protein associated with the invasion junction contains conserved oxidoreductase domain. Molecular Microbiology 15, 463-471.

Hulliger, L., Wilde, J.K.H., Brown, C.G.D. and Turner, L. (1964) Mode of multiplication of *Theileria* in cultures of bovine lymphocyte cells. Nature 203, 728-730.

Hunt, P. D. and Hall, R. (1993) Immunoblotting & Enzyme linked immunosorbent assay in: "Protocols in Molecular Parasitology" Chapter 30. (Editor Hyde, D.J.E.) Human Press, New Jersey

Hyde, J. E. and Sims, P.F.G. (1987) Anomalous dinucleotide frequencies in both coding and non-coding regions from the genome of the human malaria parasite *Plasmodium falciparum*. Gene 61, 177-187.

Iams, K. P., Hall, R., Webster, P. and Musoke, A.J. (1990) Identification of λgt11 clones encoding the major antigenic determinants expressed by *Theileria parva* sporozoites. Infection and Immunity 58, 1828-1834.

Imagawa, M., Chiu, R. and Karin, M. (1987) Transcription factor AP-2 mediates induction by two different signal-transduction pathways: Protein kinase C and cAMP. Cell 51, 251-260.

Ingraham, H. A., Chen, R., Mangalam, H.J., Elsholtz, H.P., Flynn, S.C., Linn, C.R., Simmons, D.M., Swanson, L. and Rosenfeld, M.G. (1988) A tissue specific factor containing a homeo domain specifies a pituitary phenotype. Cell 55, 519-529.

Innes, E. A., Millar, P. and Glass, E.J. (1989a) In vitro infection of bovine alloreactive cytotoxic T cell lines with sporozoites of *Theileria annulata* and *T. parva*. Research in Veterinary Science 46, 367-374.

Innes, E. A., Millar, P., Brown, C.G.D. and Spooner, R.L. (1989b) The development and specificity of cytotoxic cells in cattle immunized with autologous or allogeneic *Theileria annulata* -infected lymphoblastoid cell lines. Parasite Immunology 11, 57-68.

Irvin, A. D. and Boarer, C.D.H. (1980) Some implication of a sexual cycle in *Theileria*. Parasitology 80, 571-579.

Irvin, A. D. (1987a) Characterization of species and strains of *Theileria*. Advances in Parasitology 26, 145-197.

Irvin, A. D. and Morrison, W.I. (1987b) Immunopathology, immunology, and immunoprphylaxis of *Theileria* infections. Immune responses in parasitic infection: immunology, immunopathology, immunoprophylaxis. Volume III, Protozoa. Soulsby, E.J.L. (Editor), CRC Press Inc., Boca Raton, Florida., Pages 223-274.

Ivanov, V., Stein, B., Bauman, I., Dobbelaere, D.A.E., Herrlich, P. and Williams, R.O. (1989) Infection with the intracellular protozoan parasite *Theileria parva* induces constitutively high levels of NF-KB in bovine T lymphocytes. Molecular and Cellular Biology 9, 4677-4686.

Ivanov, V. N. and Williams, R.O. (1991) The activation of trans-acting transcription factors in T-lymphocytes infected by intracellular protozoan parasite *Theileria parva*. Genetika 27, 197-209.

Jaynes, J. B. and O'Farrell, P.H. (1988) Activation and repression of transcription by homeodomain-containing proteins that bind a common site Nature 336, 744-749.

Johnson, P. F. and McKnight, S.L. (1989) Eukaryotic transcriptional regulatory proteins. Annual Review of Biochemistry 58, 799-839.

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

Joiner, K.A. and Dubremetz, J.F. (1993) Infection and Immunity 61, 1169-1172.

Jura, W. G. Z. O., Brown, C.G.D. and Kelly, B. (1983) Fine structure and invasive behaviour of the early developmental stages of *Theileria annulata in vitro*. Veterinary Parasitology 12, 31-44.

Jura, W. G. Z. O. (1984) Factors affecting the capacity of *Theileria annulata* sporozoites to invade bovine peripheral blood lymphocytes. Veterinary Parasitology 16, 215-223.

Kadonga, J. T., Carner, K.R., Masiarz, F.R. and Thian, R. (1987) Isolation of cDNA encoding the transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 51, 1079-1090.

Kain, K.C., Orlandi, P.A., Haynes, J.D., Sim, B.K.L. and Lanar, D.E. (1993) Evidence for 2-stage binding by the 175-kD erythrocyte binding antigen of *Plasmodium falciparum*. Journal of Experimental Medicine 178, 1497-1505.

Katende, J. M., Goddeeris, B.M., Morzaria, S.P., Nkonge, J.M. and Musoke, A.J. (1990) Identification of *Theileria mutans*-specific antigen for use in an antibody and antigen detection ELISA. Parasite Immunology 12, 419-433.

Katzer, F., Carrington, M., Knight, P., Williamson, S., Tait, A., Morrison, I.W. and Hall, R. (1994) Polymorphism of SPAG-1, a candidate antigen for inclusion in a sub-unit vaccine against *Theileria annulata*. Molecular and Biochemical Parasitology 67, 1-10.

Kasper, L.H., Mineo, J.R. (1994) Attachment and invasion of host cells by*Toxoplasma gondii*. Parasitology Today 10, 184-188.

Kawamoto, S., Takahashi, K., Onuma, M., Kurosawa, T. and Sonoda, M. (1990) Invasion of Bovine erythrocytes by *Theileria sergenti* piroplasms *in vitro*. Japanese Journal for Veterinary Science 52, 1261-1263.

Kimber, C. D., Purnell, R.E. and Sellwood, S.A. (1973) The use of fluorescent antibody techniques to detect *Theileria parva* in salivary glands of the tick *Rhipicephalus appendiculatus*. Research in Veterinary Sciences 14, 126-127.

Klotz, F. W., Orlandi, P.A., Reuter, G., Cohen, S.J., Haynes, J.D., Schauer, R., Howard, R.J., Palese, P. and Miller, L.H. (1992) Binding of *Plasmodium falciparum* 175-kilodalton erythrocyte binding antigen and invasion of murine erythrocytes requires N-acetylneuraminic acid but not its O-acetylated form. Molecular and Biochemical Parasitology 51, 49-54.

Knight, P. A. (1993) The characterisation of *Theileria annulata* sporozoite surface antigens. Ph.D. Thesis, University of Glasgow.

Koch, R. (1906) Beitraege zur Entwicklungsgeschichte der *Piroplasmosen*. Zeitschrift fuer Hygiene und Infektionskrankheiten 54, 1-9.

Kumar, V. and Chambon, P. (1988) The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. Cell 55, 145-156.

Kurtti, T. J., Munderloh, U.G., Irvin, A.D. and Buscher, G. (1981) *Theileria parva* : Early events in the development of bovine lymphoblastoid cell lines persistently infected with macroschizonts. Experimental Parasitology 52, 280-290.

Laemmli, U. K. (1970) Cleavage of the structural proteins during the assembly of the head of the bacteriophage T4. Narure 227, 680-682.

Laiblin, C. (1978) Klinische Untersuchungen zur *Theileria annulata* - Infektion des Rindes. Berliner und Munchener Tierarztliche Wochenschrift 91, 48-50.

Lalor, P. A., Morrison, W.I. and Black, S.J. (1986) Monoclonal antibodies to bovine leukocytes define heterogeneity of target cells for *in vitro* patrasitosis by *Theileria parva* . Edited by Morrison W.I. In: The ruminant system in health and disease. Cambridge University Press. pp 72-87.

Landschultz, W. H., Johnson, P.F. and McKnight, S.L. (1989) The DNA binding domain of the rat liver nuclear protein C/EBP is bipartite. Science 243, 1681-1688.

Langsley, G. and Roth, C. (1987) Antigenic variation in parasitic protozoa. Microbial Sciences 4, 280-285.

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

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

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

Latchman, D. S. (1990) Gene Regulation, A eukaryotic perspective. Unwin Hyman, London

Latchman, D. S. (1991) Eukaryotic Transcription Factors. Academic Press Limited, London

Latchman, D. S. (1993) Transcripton Factors: A practical Approach. (Editor Latchman, D.S., Series editors: Rickwood, D. and Hames, B.D.). IRL Press, Oxford University Press.

Lecordier, L., Mercier, C., Torpier, G., Tourvieille, B., Darcy, F., Liu, J.L., Maes, P., Tartar, A., Capron, A. and Cesbron_Delauw. (1993) Molecular structure of a *Toxoplasma gondii* dense granule antigen (GRA-5) associated with the parasitophorous vacuole membrane. Molecular and Biochemical Parasitology 59, 143-154.

Lee, W., Mitchell, P. and Tjian, R. (1987) Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell 49, 741-752.

Lee, M. G.-S. and Van der Ploeg, L.H.T. (1990) Homologous recombination and stable transfection in parasitic protozoan *Trypanosoma brucei*. Science 250, 1583-1587.

Lee, M. G.-S. (1995) A foreign transcriptioon unit in the inactivated VSG gene expression site of the procyclic form of *Trypanosoma brucei* and formation of large episomes in stably transformed trypanosomes. Molecular and Biochemical Parasitology 69, 223-238.

Lenardo, M. J. and Baltimore, D. (1989) NF-kappa B: a pleotropic mediator of inducible and tissue-specific gene control. Cell 58, 227-229.

Lewis, A. P. (1989) Cloning and analysis of a gene encoding the 230-kilodalton merozoite surface antigen of *Plasmodium yoelii*. Molecular and Biochemical Parasitlogy 36, 271-282.

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

Lockyer, M. J., Marsh, K. and Newbold, C.I. (1989) Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T-cell epitopes of the circumsporozoite protein. Molecular and Biochemical Parasitology 37, 275-280.

Maniatis, T., Goodbourn, S. and Fischer, J.A. (1987) Regulation of inducible and tissue specific gene expression. Science 236, 1237-1245.

Mash, K. and Howard, R.J. (1986) Antigens induced erythrocytes by *Plasmodium falciparum* , expression of diverse and conserved domains. Science 231, 150-153.

Mason, P. J., Shiels, B.R., Tait, A., Beck, P. and Hall, R. (1989) Sequence and expression of a gene from *Theileria annulata* coding for a 70-kilodalton heat-shock protein. Molecular and Biochemical Parasitology 37, 27-36.

Mathews, G. V. and Vanderberg, J.P. (1994) *Plasmodium* sporozoites possess both positively and negatively charged sites accessible on their surface. Journal for Parasitology 80, 526-532.

McAndrew, M. B., Read, M., Sims, P.F.G. and Hyde, J.E. (1993) Characterisation of the gene encoding an unususally 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 from an antigenic complex. Molecular and Biochemical Parasitology 23, 71-84.

McCutchan, T. F., Lal, A.A., do Rosario, V. and Waters, A.P. (1992) Two types of sequence polymorphism in the circumsporosoite gene of *Plasmodium falciparum*. Molecular and Biochemical Parasitology 50, 37-46.

McGhee, R. B. (1953) The infection by *Plasmodium lophurae* of duck erythrocytes in the chick embryo. Journal of Experimental Medicine 97, 773-779.

McHardy, N. (1978) *In vitro* studies on the action of menoctone and other compounds on *Theileria parva* and *Theileria annulata*. Annals of Tropical Medicine and Parasitology 72, 501-511.

McHardy, N., Wekesa, L.S., Hudson, A.T. and Randall, A.W. (1985) Antitheilerial activity of BW720C (buparvaquone): a comparison with parvaquone. Research in Veterinary Sciences 39, 29-33.

McHardy, N., (1991) Butalex (Buparvaquone) - a new therapeutic for theileriosis: with comments on its value in an integrated control strategy. In "Proceedings of the Second EEC Workshop on Orientation and Co-ordination of Research on Tropical Theileriosis". Eds. D.K. Singh and B.C. Varshney. Anand Press, Anand. pp 133-140

McKean, P. G., O'Dea, K. and Brown, K.N. (1993) A single amino acid determines the specificity of a monoclonal antibody which inhibits *Plasmodium chaudi* AS in vivo. Molecular and Biochemical Parasitology 62, 211-222.

Mecham, R. P., Hinek, A., Entwistle, R., Wrenn, D.S., Griffin, G.L. and Senior, R.M. (1989) Elastin binds to a multifunctional 67 kilodalton membrane protein. Biochemistry 28, 3716-3722.

Melhorn, H. and Schein, E. (1984) The Piroplasm: Life cycle and sexual stages. Advances in Parasitology 23, 37-103.

Melrose, T. M, Brown, C.G.D., Morzaria, S.P., Ocama, J.G.R. and Irvin, A.D. (1984) Glucose Phosphate Isomerase polymorphism in *Theileria annulata* and *T. parva*. Tropical Animal Health Production 16, 239-245.

Mendis, K. N., David, P.H., Hommel, M., Carter, R. and Miller, L.H. (1983) Immunity to malaria antigens on the surface of *Plasmodium falciparum* infected erythrocytes. American Journal of Tropical Medicine and Hygiene 32, 926-930.

Mendis, K. N., Ihalamulla, R.I. and David, P.H. (1988) Diversity of *Plasmodium vivax* induced antigens on the surface of infected human erythrocytes. American Journal of Tropical Medicine and Hygiene 38, 42-46.

Mendis, K. N., David, P.H. and Carter, R. (1991) Antigenic polymorphism in malaria: is it an important mechanism for immune evasion? Immunology Today 12, A34-A37.

Mercier, C., Lecordier, L., Darcy, F., Deslee, D., Murray, A., Tourvieille, B., Maes, P., Capron, A. and Cesbron-Delauw, M.F. (1993) Molecular characterization of a dense granuole antigen (GRA2) associated with the network of the parasitophorous vacuole in *Toxoplasma gondii*. Molecular and Biochemical Parasitology 58, 71-82.

Mermond, N. O., Kelly, T.J. and Tjian, R. (1989) The prolin-rich transcriptional activator of CTF/NF-1 is distinct from the replication and DNA binding domain. Cell 58, 741-753.

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

Miller, L. H., Mason, S.J., Dvorak, J.A., McGinnis, M.H. and Rothman, I.K. (1975) Erythrocyte receptors for (*Plasmodium knowlesi*) malaria: Duffy blood group determinants. Science 189, 561-563.

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

Miller, L. H., Roberts, T., Shahauddin, M. and McCutchan, T. (1993) Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Molecular and Biochemical Parasitology 59, 1-14.

Minami, T., Spooner, P.R., Irvin, A.D., Ocama, J.G.R., Dobbelaere, D.A.E. and Fujinaga, T. (1983) Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. Research in Veterinary Science 35, 334-340.

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

Miller, L.H., Good, M.F. and Milon, G. (1994) Malaria pathogenesis. Science 264, 1878-1883.

Mitchell, G. H., Hadley, T.J., Klotz, F.W., McGinnis, M.H. and Miller, L.H. (1986) Invasion of erythrocytes by *Plasmodium falciparum* malaria parasites: evidence for receptor heterogeniety and two receptors. Blood 67, 1519-1521.

Moelans, I.I., Meis, J.F., Kocken, C., Konings, R.N. and Schoenmakers, J.G. (1991) A novel protein antigen of the malaria parasite *Plasmodium falciparum*, located on the surface of gametes and sporozoites. Molecular and Biochemical Parasitology 45, 193-204.

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

Morrison, W. I., Lalor, P.A., Goddeeris, B.M. and Teale, A.J. (1986) Theileriosis: Antigens and host-parasite interactions. In: Parasite antigens: Towards new stratagies for vaccines (ed. by T.W. Pearson) 167-213, Marcel Dekker, New York.

Morrison, W. I., Goddeeris, B.M., Brown, W.C., Baldwin, C.L. and Teale, A.J. (1989a) *Theileria parva* in cattle: Characterization of infected lymphocytes and the immune responses they provoke. Veterinary Immunology and Immunopathology 20, 213-237.

Morrison, W. I. and Goddeeris B.M. (1989b) The role of the MHC in the immunity of cattle to the protozoan parasite *Theileria parva*. Ann. Rech. Vet. 20, 385-588.

Mozaria, S. P., Spooner, P.R., Bishop, R.P., Musoke, A.J. and Young J.R. (1990) Sfi I and Not I polymorphisms in *Theileria* stocks detected by pulse field gel elecrophoresis. Molecular and Biochemical Parasitology 40, 203-212.

Mozaria, S. P. and Nene, V. (1990) Bovine theileriosis: progress in immunisation methods. International Journal on Animal Science 5, 1-14.

Muller, H.M., Reckmann, I., Hollingdale, M.R., Bujard, H. and Robson, K.J.H. (1993) Thrombospondin related anonymous protein (TRAP) of *Plasmodium falciparum* binds specifically to sulfated glycoconjugAtes and to HEPG2 hepatoma-cells suggesting a role for this molecule in sporozoite invasion of hepatocytes. EMBO Journal 12, 2881-2889.

Murre, C., McCaw, P.S. and Baltimore, D. (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD and myc proteins. Cell 56, 777-783.

Musisi, F. L. (1990) Methods currently used for the control of East Coast fever: their validity and proposals for future control strategies. Parasitologia 32, 15-22.

Musoke, A. J., Nantulya, V.M., Buscher, G., Masake, R.A. and Otim, B. (1982) Bovine immune responses to *Theileria parva* : neutralizing antibodies to sporozoites. Immunology 45, 663-668.

Musoke, A. J., Nantulya, V.M., Rurangirwa, F.R. and Buscher, G. (1984) Evidence for a common protective antigen determinant on sporozoites of several *Theileria parva* strains. Immunology 52, 231-238.

Musoke, A., Moraria, S., Nkonge, C., Jones, E. and Nene, V. (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. Proceedings of the National Academy of Sciences, U.S.A. 89, 514-518.

Musoke, A., Nene, V. and Morzaria, S.P. (1993) A sporozoite-based vaccine for *Theileria parva*. Parasitology Today 9, 385-388.

Myler, P. J. (1989) Transcriptional analysis of the major merozoite surface antigen precursor (GP195) gene of *Plasmodium falciparum*. In "Parasites: Molecular Biology, Drugs, Vaccine Design" (N.Agabian and A Cerami, Eds.), pp. 123-137, Wiley-Liss, New York.

Nardelli, J., Gibson, T.J., Vesque, C. and Charnay, P. (1991) Base spequence discrimination by zinc-finger DNA binding domains. Nature 349, 175-178.

Ndengele, M.M., Messineo, D.G., Samyellowe, T. and Harwalkar, J.A. (1995) *Plasmodium falciparum* - effects of membrane modulating agents on direct binding of rhoptry proteins to human erythrocytes. Experimental Parasitology 81, 191-201.

Neitz, W. O. (1957) *Theileriosis, gonderioses* and *cytauxzoonoses.* The Onderstepoort Journal of Veterinary Research 27, 275-430.

Nene, V., Gobright, E., Musoke, A.J. and Lonsdale-Eccles, J.D (1990) A single exon codes for the enzyme domain of a protozoan cystein protease. Journal of Biological Chemistry 265, 18047-18050.

Nene, V., Iams, K.P., Gobright, E. and Musoke, A.J. (1992) Characterisation of a gene encoding a candidate vaccine antigen of *Theileria parva* sporozoites. Molecular and Biochemical Parasitology 51, 17-28.

Neuberg, M., Adamkiewicz, J., Hunter, J.B. and Muller, R. (1989) A Fos protein containing the Jun leucine zipper forms a homeodimer which binds to the AP1 binding site. Nature 341, 243-245.

Norman, C., Runswick, M., Pollock, R. and Treisman, R. (1988) Isolation and properties of cDNA clones encoding SRF, a transcription factor that binds to the c-fos serum response element. Cell 55, 989-1003.

Nussenzweig, V. and Nussenzweig, R.S. (1989) Rationale for the development of an engineered sporozoite malaria vaccine. Advances in Immunology 45, 283-224.

Odea, K.P., McKean, P.G., Harris, A. and Brown, K.N. (1995) Processing of the *Plasmodium chabaudi chabaudi* asmerozoite surface protein-1 *in vitro* and *in vivo.* Molecular and Biochemical Parasitology 72, 111-119.

Orlandi, P. A., Sim, B.K., Chulay, J.D. and Haynes, J.D. (1990) Characterization of the 175 kDa erythrocyte binding antigen of *Plasmodium falciparum.* Molecular and Biochemical Parasitology 40, 285-294.

Orlandi, P. A., Klotz, F.W. and Haynes, J.D. (1992) A malaria invasion receptor, the 175 kDa erythrocyte binding antigen of *Plasmodium falciparum* recognizes the terminal Neu5Ac(alpha2-3)Gal-sequence of glycophorin A. Journal of Cell Biology 116, 901-909.

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

Ossorio, P.N., Dubremetz, J.F. and Joiner, K.A. (1994) A soluble secretory protein of the intracellular parasite *Toxoplasma gondii* associates with the parasitophorous vacuole membrane through hydrophobic interactions. Journal of Biological Chemistry 269, 15350-15357.

Pancake, S.J., Holt, G.D., Mellouk, S. and Hoffman, S.L. (Malaria sporozoite and circumsporozoite proteins bind specifically to sulfated glycoconjugates. Journal of Cell Biology 117, 1351-1357.

Parsons, M. (1990) Current concepts in stage-regulated gene expression in kinetoplastida. Parasitology Today 6, 244-245.

Pays, E., Tebabi, A., Pays, A., Coquelet, H., Revelard, P., Salmon, D. and Steinert, M. (1989) *Trypanosoma brucei* : post-transcriptional control of the variable surface glycoprotein gene expression site. Cell 57, 835-845.

Perkins, M. (1988) Stage-dependent processing and localization of a *Plasmodium falciparum* protein of 130000 molecular weight. Experimental Parasitology 65, 61-68.

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. (1989) Erythrocyte invasion by the malarial merozoite: Recent advances. Experimental Parasitology 69, 94-99.

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

Peterson, M. G., Nguyendinh, P., Marshal, V.M., Elliott, J.F., Collins, W.E. and Anders, R.F. (1990) Apical membrane antigen of *Plasmodium fragile*. Molecular and Biochemical Parasitology 39, 279-284.

Peterson, D.S., Miller, L.H. and Wellems, T.E. (1995) Isolation of multiple sequences from the *Plasmodium falciparum* genome that encode conserved domains homologous to those in erythrocyte-binding proteins. Proceedings of the National Academy of Sciences of the USA 92, 7100-7104.

Petterson, U., Wigzell, H., Perlman, P. (1987) Malaria diagnosis. Science 236, 11-12.

Pinder, M., Withey, K.S. and Roelants, G.E. (1981) *Theileria parva* parasites transform a subpopulation of T lymphocytes. The Journal of Immunology 127, 389-390.

Pipano, E., Goldman, M., Samish, M. and Friedhoff, K.T. (1977) Immunisation of cattle against *Theileria annulata* using killed schizont vaccine. Veterinary Parasitology 3, 11-22.

Pipano, E. (1981) Schizonts and tick stages in immunisation against *Theileria annulata* infection. In "Advances in the Control of Theileriosis". Eds. A.D. Irvin, M.P. Cunningham and A.S. Young. Martinus Nijhoff, The Hague. pp 242-252

Pipano, E. (1989) Bovine theileriosis in Israel. Revue Scientifique et Technique (Paris) 8, 79-87.

Pologe, L. G. and Ravetech, J.V. (1986) A chromosomal rerangement in a *P. falciparum* histidine-rich gene is associated with the knobless pheneotype. Nature 322, 474-477.

Potocnjak, P., Yoshida, N., Nussenzweig, R.S. and Nussenzweig, V. (1980) Monovalent fragments (Fab) of monoclonal antibodies to a sporozoite surface antigen (Pb44) protect mice against malaria infections. Journal of Experimental Medicine 151, 1504-1513.

Prater, C. A., Plotkin, J., Jaye, D. and Frazier, W.J. (1991) The properdin-like type I repeats of the human thrombospondin contain a cell attachment site. Journal of Cell Biology 112, 1031-1040.

Premawansa, S., Peiris, J.S.M., Perera, K.L.R.L., Ariyarante, G., Carter, R. and Mendis, K.N. (1990) Target antigens of transmission blocking immunity of *Plasmodium vivax* malaria. Characterization and polymorphism in natural parasite isolates. journal of Immunology 144, 4376-4383.

Preston, P. M., Brown, C.G.D. and Sponner, R.L. (1983) Cell-mediated cytotoxicity in *Theileria annulata* infection of cattle with evidence for BoLA restriction. Clinical and Experimental Immunology 53, 88-100.

Preston, P. M. and Brown, C.G.D. (1985) Inhibition of lymphocyte invasion by sporozoites and the transformation of trophozoite infected lymphocytes *in vitro* by serum from immune cattle. Parasite Immunology 7, 301-314.

Preston, P. M., McDougall, C., Wilkie, G., Shiels, B.R., Tait, A. and Brown, C.G.D. (1986) Specific lysis of *Theileria annulata* -infected lymphoblastoid cells by a monoclonal antibody recognizing and infection-associated antigen. Parasite Immunology 8, 369-380.

Preston, P. M. and Brown, C.G.D. (1988a) Macrophage-mediated cytostasis and lymphocyte cytotoxicity in cattle immunized with Theileria annulata sporozoites or macroschizont-infected cell lines. Parasite Immunology 10, 631-647.

Preston, C. M., Frame, M.C. and Campbell, M.E.M. (1988b) A complex formed between cell components and a herpes simplex virus structural polypeptide binds to a viral immidiate early gene regulatory DNA sequence. Cell 52, 425-434.

Prickett, M.D., Smarz, T.R. and Adams, J.H. (1994) Dimorphism and intergenic recombination within the microneme protein (MP-1) gene family of *Plasmodium knowlesi*. Molecular and Biochemical Parasitology 63, 37-48.

Prince, J.B., Araujo, F.G., Remington, J.S., Burg, J.L., Boothroyd, J.C. and Sharma, S.D. (1989) Cloning of cDNAs encoding a 28 kilodalton antigen of *Toxoplasma gondii*. Molecular and Biochemical Parasitology 34, 3-13.

Pugh, B. F. and Tjian R. (1991) Transcription from TATA-less promoter requires a multisubunit TFIID complex. Genes in Development 5, 1935-1945.

Pugh, B. F. and Tjian R. (1992) Diverse transcriptional functions of the multisubunit eukaryotic TFIID complex. Journal of Biological Chemistry 267, 679-682.

Qazi, M.H., Hoessli, D.C. and Walkemasir, E. (1994) The role of cell-surface carbohydrates in malaria. Pure and Applied Chemistry 66, 2259-2262.

Raju, K. and Anwar, R.A. (1987) Primary structures of bovine elastin a,b, and c deduced from sequences of cDNA clones. Journal for Biological Chemistry 262, 5755-5762.

Ravetch, J. V., Feder, R., Pavlovec, A. and Blobel, G. (1984) Primary structure and genomic organization of the histidine-rich protein of the malaria parasite *Plasmodium lophurae*. Nature 312, 616-620.

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.

Rawlings, D. J., Fujioka, H., Fried, M., Keister, D.B., Aikawa, M. and Kaslow, D.C. (1992) α-tubulin II is a male-specific protein in *Plasmodium falciparum*. Molecular and Biochemical Parasitology 56, 239-250.

Robert, R., et al. (1991) Infection and Immunity 59, 4670-4673.

Robson, K. J. H., Hall, J.R.S. and Jennings, M.W. (1988) A highly conserved amino acid sequence in thrombospondin, properdin and in proteins from sporozoites and blood stages of a human malaria parasite. Nature 335, 79-82.

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

Rogers, W.O., Rogers, M.D., Hedstrom, R.C. and Hoffman, S.L. (1992a) Characterisation of the gene encoding sporozoite surface protein-2, a protective *Plasmodium-yoelii* sporozoite antigen. Molecular and Biochemical Parasitology 53, 45-51.

Rogers, W.O., Malik, A., Mellouk, S., Nakamura, K., Szarfman, A., Gordon, D.M., Nussler, A.K., Aikawa, M. and Hoffman, S.L. (1992b) Characterization of *Plasmodium falciparum* sporozoite surface protein 2. Procedings of the National Academy of Science of the USA 89, 9176-9180.

Rosenber, R., Wirtz, R.A., Lanar, D.E., Sattabongkot, J., Hall, T. and Waters, A.P. (1989) Circumsporozoite protein heterogeneity in the human malaria parasite *Plasmodium vivax*. Science 245, 973-976.

Rudenko, G., Bishop, D., Gottesdiener, K. and Van der Ploeg, L.H.T. (1989) Alpha-amanitin resistant transcription of protein coding gene in insect bloodstream from *Trypanosoma brucei*. EMBO Journal 13, 4236-4259.

Rudenko, G., Le Blancq, S., Smith, J., Lee, M.G.-S., Rattray, A. and Van der Ploeg, L.H.T. (1990) Procyclic Acidic repetitive protein (PARP) genes located in an unsusually smal alpha-amanitin-resistant transcription unit: PARP promoter activity assayed by transient DNA transfection of *Trypanosoma brucei*. Molecular and Cellualr Biology 10, 3492-3504.

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

Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual. 2nd Edition. Cold Spring Harbor Laboratory Press, Cold Pring Habor, NY.

Smayellowe, T.Y., Fujioka, H., Aikawa, M. and Messinieo, D.G. (1995) *Plasmodium falciparum* rhoptry proteins of 140/130/110 kDa (RhopH) are located in an electron lucent compartment in the neck of the rhoptries. Journal of Eukaryotic Microbiology 42, 224-231.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. Proceedings of the National Academy of Science of the USA 74, 5463-5467.

Santoro, C., Mermond, N., Andrews, P.C. and Tjian, R. (1988) A family of human CCAAT box binding proteins active in transcription and DNA replication: cloning and expression of multiple cDNAs. Nature 334, 218-224.

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

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

Senior, R. M., Griffin, G.L., Mecham, R.P., Wrenn, D.S., Prasad, K.U. and Urry, D.W. (1984) Val-Gly-Val-Ala-Pro-Gly, a repeating peptide in elastin, is chemotactic for fibroblasts and monocytes. The Journal of Cell Biology 99, 870-874.

Shapiro, S. Z., Fujisaki, K., Mozaria, S.P., Webster, P., Fujinaga, T., Spooner, P.R. and Irvin, A.D. (1987) A life-cycle stage-specific antigen of *Theileria parva* recognized by anti-machroschizont monoclonal antibodies. Parasitology 94, 29-37.

Shaw, M. K., Tilney, L.G. and Musoke, A.J. (1991) The entry of *Theileria parva* sporozoites into bovine lymphocytes: Evidence for MHC class I involvement. The Journal of Cell Biology 113, 87-101.

Shaw, M. K., Tilney, L.G., Musoke, A.J. and Teale, A.J. (1995) MHC class I molecules are an essential cell surface component involved in *Theileria parva* sporozoite binding to bovine lymphocytes. Journal of Cell Science 108, 1587-1596.

Shiels, B. R., McDougall, C., Tait, A. and Brown, C.G.D. (1986) Identification of infection-associated antigens in *Theileria annulata* transformed cells. Parasite Immunology 8, 69-77.

Shiels, B., Hall, R., Glascodine, J., McDougall, C., Harrison, C., Taracha, E., Brown, D. and Tait, A. (1989) Characterization of surface polypeptides on different life-cycle stages of *Theileria annulata*. Molecular and Biochemical Parasitology 34, 209-220.

Sibley, L.D., Niesman, I.R., Parmley, S.F. and Cesbron-Delauw, M.F. (1995) Regulated secretion of multi-lamellar vesicles leads to formation of a tubulo-vesicular network in host-cell vacuoles occupied by *Toxoplasma gondii*. Journal of Cell Science 108, 1669-1677.

Sim, B.K.L., Carter, J.M., Deal, C.D., Holland, C., Haynes, J.D. and Gross, M. (1994) *Plamodium falciparum* - Further characterisation of a functionally active-region of the merozoite invasion ligand EBA-175. Experimental Parasitology 78, 259-268.

Sim, B.K., Chitnis, C.E., Wasniowska, K., Hadley, T.J. and Miller, L.H. (1994) Receptor and ligand domAins for invasion of erythrocytes by *Plasmodium falciparum*. Science 264, 1941-1944.

Sim., B.K.L. (1995) EBA-175: An erythrocyte-binding ligand of *Plasmodium falciparum*. Parasitology Today 11, 213-217.

Sina, B. J., Wright, C., Atkinson, C.T., Ballou, R., Aitawa, M. and Hollingdale, M. (1995) Characterization of a sporozoite antigen common to *Plasmodium falciparum* and *Plasmodium berghei*. Molecular and Biochemical Parasitology 69, 239-246.

Sinden, R. E. (1985) A cell biologist's view of host cell recognition and invasion by malarial parasites. Transactions of the Royal Society of Tropical Medicine and Hygiene 79, 598-605.

Singh, D. K., Ouhelli, H., Suramanian, G., Sayin, F. and Bhattacharyulu, Y. (1986) Orientation and coordinadion of research on tropical theileriosis (Report of an EEC sponsored workshop). 24-76.

Singh, D. K. (1990) Methods currently used for the control of *Theileria annulata* : Their validity and proposals for future control strategies. Parasitologia 32, 33-40.

Singh, H., LeBowitz, J.H., Baldwin, Jr. A.S. and Sharp, P.A. (1988) Molecular cloning of an enhancer binding protein: Isolation by screening of an expression library with a recognition site DNA. Cell 52, 415-423.

Smith, D. B. and Johnson, S. (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. Gene 67, 31-40.

Sopp, P., Howard, C.J. and Parsons, K.R. (1991) Investigating monoclonal-antibodies binding to bovine null-cell antigens using 2-color immunofluorescence. Veterinary Immunology and immunopathology 27, 163-168.

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

Spooner, R. L., Innes, E.A., Glass, E.J., Millar, P. and Brown, C.G.D. (1988) Bovine mononuclear cell lines transformed by *Theileria annulta* express different subpopulation markers. Parasite Immunology 10, 619-629.

Spooner, R. L., Innes, E.A., Glass, E.J. and Brown, C.G.D. (1989) *Theileria annulata* and *T. parva* infect and transform different bovine mononuclear cells. Immunology 66, 284-288.

Stagg, D. A., Dolan, T.T., Leitch, B.L. and Young, A.S. (1981) The initial stages of infection of cattle cells with *Theileira parva* sporozoites *in vitro*. Parasitology 83, 191-197.

Stagg, D. A., Young, A.S., Leitch, B.L., Gootenhuis, J.G. and Dolan, T.T. (1983) Infection of mammalian cells with *Theileria* species. Parasitology 86, 243-254.

Steuber, S., Frevert, U., Ahmed, J.S., Hauschild, S. and Schein, E. (1986) *In vitro* susceptibility of different mammalian lympocytes to sporozoites of *Theileria annulata*. Zeitschrift fur Parasitenkunde 72, 831-834.

Sugimoto, C., Mutharia, L.M., Conrad, P.A., Dolan, T.T., Brown, W.C., Goddeeris, B.M. Pearson, T.W. (1989) Protein changes in bovine lymphoblastoid cells induced by infection with the intracellular parasite *Theileria parva*. Molecular and Biochemical Parasitology 37, 159-170.

Su, S.D., Sanadi, A.R., Ifon, E. and Davidson, E.A. (1993) A monoclonal antibody capable of blocking the binding of PF200 (MSA-1) to human erythrocytes and inhibiting the invasion of *Plasmodium falciparum* merozoites into human erythrocytes. Journal of Immunology151, 2309-2317.

Suss, G., Matile, H., Meloen, R.H., Takacs, B. and Pink, J.R.L. (1993) Identifying polymorphic regions of the p190 protein from different *Plasmodium falciparum* strains by using T cells. Parasite Immunology 15, 127-134.

Tait, A. and Hall, F.R. (1990) *Theileria annulata* : control measures, diagnosis and the potential use of subunit vaccines. Revue Scientifique et Technique (Paris) 9, 387-403.

Tanabe, K., MacKay, M., Goman, M. and Scaife, J.G. (1987) Allelic dimorphism in a surfacce gene of the malaria parasite *Plasmodium falciparum* . Journal of Molecular Biology 195, 273-287.

Theilen, G. H., Rush, J.D. Nelson-Rees, W.A., Dungworth, D.L., Munn, R.J., and Switzer, J.W. (1968) Bovine leukaemia: establishment and morphological characteristion of continuous cell suspension culture BL-1 Journal of the National Cancer Institute 40, 737-749.

Theiler, A. (1906) *Piroplasma mutans* (n.spec.) of South African cattle. Journal of Comparative Pathology and Therapy 19, 292-438.

Theiler, A. (1904) East coast fever. Transvaal Agriculture Journal 2, 421-438.

Tomley, F. M., Clarke, L.E., Kawazoe, U., Dijkema, R. and Kok, J.J. (1991) Sequence of the gene encoding an immunodominant microneme protein of *Eimeria tenella*. Molecular and Biochemical Parasitology 49, 277-288.

Towbin, H., Staehelin, T. and Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: Procedure and some applications. Procedings of the National Academy of Sciences, USA 76, 4350-4354.

Toye, P. G., Goddeeris, B.M., Iams, K., Musoke; A.J. and Morrison, W.I. (1991) Characterization of a polymorphic immunodominant molecule in sporozoites and schizonts of *Theileria parva*. Parasite Immunology 13, 49-62.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1990) A single amino acid can determin the DNA binding specificity of homeodomain proteins. Cell 59, 553-562.

Tsuboi, T., Kappe, S.H.I., Alyaman, F., Prickett, M.D., Alpers, M. and Adams, J.H. (1994) Natural variation within the principal adhesion domain of the *Plasmodium vivax* Duffy binding-protein. Infection and Immunity 62, 5581-5586.

Tuszynski, G. P., Karczewski, J., Smith, L., Murphy, A., Rothman, V.L. and Knuasen, K.A. (1989) The GPIIb-IIa-like complex may function as a human cell adhesion receptor for thrombospondin. Experimantal Cell Research 182, 473-481.

Uilenberg, G. (1964) *Haematoxenus veliferus* , n.g., n. sp., parasite inceratae sedis du sang de bovins a Madagascar. Revue d'Elevage et de Medecine Veterinaire des Pays Tropicaux 17, 655-662.

Uilenberg, G. (1981) Theileria infections other than East Coast fever. In "Diseases of Cattle in the Tropics". Eds. M. Ristic and I. McIntyre. Martinus Nijhoff, The Hague. pp 411-427

Uilenberg, G., Camus, E. and Barre, N. (1983) Existence in Guadeloupe (French West Iindies) of *Theileria mutans* and *Theileria velifera* (Sporozoa, Theileridae). Revue d'Elevage et de Medecine Veterinaire des Pays Tropicaux 36, 261-264.

Umesono, K. and Evans, R.M. (1989) Determinants of target gene specificity for steroid/thyroid hormone receptors. Cell 57, 1139-1146.

Urquart, G. M., Duncan, J.L., Dunn, A.M., and Jennings, F.W. (1987) Veterinary Parasitology. Longman U.K.

Vanderberg J. P. (1968) Further studies on the *Plasmodium berghei-Anopheles stephens*i-rodent system of mmalian malaria, Journal of Parasitology 54, 1009-1016.

Vanderberg, J. P. and Stewart, M.J. (1990) *Plasmodium* sporozoite-host cell interactions during sporozoite invasion. Bulletin of the World Health Organization 68, 74-79.

Verdini, A. S., Chiappinelli, L. and Zanobi,A. (1991) Towards the elucidation of the mechanism of the attachment and the entry of malaria sporozoites into cells: Synthetic polypeptides from the circumsporozoite protein of *Plasmodium falciparum* bind Ca2+ and interact with model phospholipid membranes. Biopolymers 31, 587-594.

Vijayasarathy, S., Ernest, I., Itzhaki, J.E., Serman, D., Mowatt, M.R., Michals, P.A.M. and Clayton, C.E. (1990) The genes encoding fructose biphosphate aldolase in *Trypanosoma brucei* are interspersed with unrelated genes. Nucleic Acid Research 18, 2967-2975.

Wakelin, D. (1984) Immunity to Parasites: How Animals Control Parasite Infections. Edward Arnold.

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

Warnecke, M., Schein, E., Voigt, W.P. and Uilenberg, G. (1979) On the life cycle of *Theileria velifera* (Uilenberg, 1964) in the gut and the haemolymph of the tick vector *Amblyomma variegatum* (Fabricius, 1794). Zeitschrift fur Tropenmedizin and Parasiologie 30, 318-322.

Wasylyk, B., Wasylyk, C., Flores, P., Begue, A., Leprince, D. and Staehelin, D. (1990) Thec-est proto-oncogenes encode transcription factors that co-operate with c-Fos and c-Jun for transcriptional activation. Nature 346, 191-193.

Waters, A. P., Syin, C. and McCutchan, T.F. (1989) Developmental regulation of stage-specific ribosome populations in Plasmodium. Nature 342, 438-440.

Wathanga (1984) M.Sc. Thesis, Edinburgh University.

Webster, P., Dobbelaere, D.A.E., Fawcett, D.W. (1985) The entry of sporozoites of *Theileria parva* into bovine lymphocytes in vitro. Immunoelectron microscopic observations. European Journal of Cell Biology 36, 157-162.

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

Wellems, T. E. and Howard, R.J. (1986) Homologous genes encode two distinct histidine-rich proteins in a cloned isolate of *Plasmodium falciparum*. Proceedings of the National Academy of Sciences, USA 83, 6065-6069.

Werk, R. (1985) How does *Toxoplasma gondii* enter host cells? Reviews of Infectious Diseases 7, 449-457.

Werk, R. (1989) In: Intracellular Parasitism (Ed. Moulder, J.W.) pp259-269, CRC Press.

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

Wesseling, J. G., de REE, J.M., Ponnundurai, T., Smits, M.A. and Schoenmakers, J.G.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., Snijders, P.J.F., van Someren, P., Jansen, J., Smits, M.A. and Schoenmakers, J.G.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.

Weston, K. and Bishop, J.M. (1989) Transcriptional activation by the v-myb oncogene and its cellular progenitor c-myb. Cell 58, 85-93.

White, R. J. a Jackson S.P. (1992) The TATA-binding protein: a central role in transcription by RNA polymerases I, II and III. Trends in Genetics 8, 284-288.

Wizel, B., Rogers, W.O., Houghten, R.A., Lanar, D.E., Tine, J.A. and Hoffman, S. (1994) Induction of murine cytotoxic T-lymphocytes against *Plasodium falciparum* sporozoite surface protein-2. European Journal of Immunology 24, 1487-1495.

Wilkie, G. M., Melrose, T.M., Fletcher, J.D. and Brown, C.G.D. (1986) GPI Isoenzyme identificaion of cloned populations of *Theileira annulata* (Ankara) after passages through the vertebrate and invetebrate host. Transactions of the Royal Society of Tropical Medicine and Hygiene 80, 987-988.

Williamson, S. M. (1988) A *Theileria annulata* sporozoite surface antigen as a potential vaccine against tropical theileriosis. Ph.D. Thesis, Edinburgh University.

Williamson, S. Tait,A., Brown, D., Walker, A., Beck, P., Shields, B., Fletcher, J. and Hall, R. (1989) *Theileria annulata* sporozoite surface antigen expressed in *Escherichia coli* elicits neutralizing antibodies. Proceedings of the National Academy of Sciences, U.S.A. 86, 4639-4643.

Wong, S.Y. and Remington (1993) AIDS 7, 299-316.

Wrenn, D. S., Griffin, G.L., Senior, R.M. and Mecham, R.P. (1986) Characterization of biologically active domains on elastin: identification of a monoclonal antibody to a cell recognition site. Biochemistry 25, 5172-5176.

Wrenn, D. S., Hinek, A. and Mecham, R.P. (1988) Kinetics of receptor-mediated binding of tropelastin to ligament fibroblasts. Journal of Biological Chemistry 263, 2280-2284.

Wu, Y., Sifri, D., Lei, H.-H., Su, X.-Z. and Wellems, T.E. (1995) Tranfection of *Plasmodium falciparum* within human red blood cells. Proceedings of the National Academy of Sciences, U.S.A. 92, 973-977.

Yakimoff, W. L. and Dekhtereff, N.A. (1930) Zur Frage ueber die Theileriose in Ostsibirien. Archiv fuer Protistenkunde 72, 176-189.

Yamamoto, K. K., Gonzalez, G.A., Biggs, W.H. and Montiminy, M.R. (1988) Phosphorylation-induced binding and transcriptional efficacy of nuclear factor CREB. Nature 334, 494-498.

Yoshida, N., Potocnjak, P., Nussenzweig, V. and Nussenzweig, R.S. (1981) Biosynthesis of Pb44, the protective antigen of sporozoites of *Plasmodium berghei*. Journal of Experimental Medicine 154, 1225-1236.

Young, A. S., Purnell, R.E., Payne, R.C., Brown, C.G.D. and Kanhai, G.K. (1978) Studies on the transmission and course of infection of a Kenyan strain of *Theileria mutans*. Parasitology 76, 99-115.

Young, A. S., Grootenhuis, J.G., Leitch, B.L. and Schein, E. (1980) The development of *Theileria = Cytauxzoon taurotragi* (Martin and Brocklesby, 1960) from eland in its tick vector *Rhipicephalus appendiculatus*. Parasitology 81, 129-144.

Appendix: Publications

Parasite Immunology, 1994: 16: 97–104

Theileria annulata sporozoite surface antigen (SPAG-1) contains neutralizing determinants in the C terminus

NICKY BOULTER[1], PAMELA A. KNIGHT[2], PHILIP D. HUNT[1], EMMA S. HENNESSEY[1], FRANK KATZER[1], ANDREW TAIT[2], SUSANNA WILLIAMSON[3], DUNCAN BROWN[3], HOWARD A. BAYLIS[1] & ROGER HALL[1]

[1]Department of Biology, University of York, York Y01 5DD, UK
[2]Wellcome Unit for Molecular Parasitology, Department of Veterinary Parasitology, University of Glasgow, Bearsden Road, Glasgow, G61 1QH, Scotland, UK
[3]Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, Scotland, UK

SUMMARY

SPAG-1 is a surface antigen on Theileria annulata *sporozoites that is a candidate both for inclusion in a sub-unit vaccine and as a ligand for host cell recognition. We have pinpointed major neutralizing epitopes to the C terminus. To facilitate this we expressed SPAG-1 as a series of defined fragments in the pGEX system. These constructs were validated by sequencing and by their spectrum of reactivity with monoclonal antibody (MoAb) BA4. This MoAb recognizes the elastin motif VGVAPG, that is predicted to occur three times in the N terminal half of SPAG-1. The recombinant proteins were then tested by Western blotting with a neutralizing MoAb (1A7) and two neutralizing bovine sera (10T and 34A). The results demonstrate that 1A7 and the bovine sera react with determinants unique to the C terminus. We mapped the neutralizing determinant recognized by MoAb 1A7 to a 16 residue sequence (residues 807–822) using synthetic peptides. Interestingly the bovine sera do not recognize the 1A7 epitope. The potential role of the C terminus as a ligand for host cell recognition and the implications for sub-unit vaccine production are discussed.*

Keywords Theileria annulata, *sporozoite, surface antigen, ligand, sub-unit vaccine, recombinant protein*

Correspondence: Dr Roger Hall
Received: 12 March 1993
Accepted for publication: 19 October 1993

INTRODUCTION

Control of tropical theileriosis relies principally upon vaccination with attenuated macroschizont infected cell lines (Pipano 1989, Singh 1990). Other methods such as chemotherapy and tick control are available but require strict management and surveillance and thus tend to be more expensive (Brown 1990). A significant drawback to the use of the attenuated vaccine is the requirement for a cold chain between the points of production and delivery. Other potential complications of using a live vaccine include the possibility of reversion to virulence and contamination with other pathogens (Tait & Hall 1990). These disadvantages would not exist if an effective, one shot sub-unit vaccine was developed. To this end we have been characterizing a sporozoite surface antigen, SPAG-1, which has been shown to carry neutralizing epitopes (Williamson *et al.* 1989). We have fully characterized the gene for this antigen and have observed that it contains regions homologous to repetitive structures present in bovine elastin (Hall *et al.* 1992). We have argued that SPAG-1 is both a potential ligand for host cell recognition and a sub-unit vaccine candidate (Hall *et al.* 1992). The analogous molecule in *Theileria parva* (p67) has recently been demonstrated to be partially protective (Musoke *et al.* 1992). We are interested in precisely determining the structure of the neutralizing epitopes on SPAG-1 with a view to designing synthetic vaccine components and investigating host cell recognition and invasion. In this paper we present data that locate neutralizing epitopes in this molecule towards the C terminal region.

MATERIAL AND METHODS

SPAG-1 expression constructs in pGEX

The pGEX system (Smith & Johnson 1988) was used to express various sub-fragments derived from the SPAG-1

Table 1 Summary of the derivation of the constructs used in this analysis. The relevant fragments were generated by digestion with the appropriate restriction enzymes shown. For ease of interpretation reference to Figure 1a should be made. In each case the pGEX vector chosen was that predicted to maintain the appropriate open reading frame.

Construct	Amino acids	Fragment	Bases	Vector
2·7	20–907	*Eco*RV-*Eco*RI (partial)	58–2724	*Sma*I-*Eco*RI cut pGEX-3X
HB	20–108	*Bam*HI	58–326	*Bam*HI cut pGEX-3X
0·4	20–97	*Eco*RI cut 2·7, discard fragment & religate	58–293	pGEX-3X
2·4	109–907	*Bam*HI-*Eco*RI (partial)	326–2724	*Bam*HI-*Eco*RI cut pGEX-2T
0·8	109–381	*Bam*HI	326–1145	*Bam*HI cut pGEX-2T
EA	109–262	*Acc*I-*Eco*RI[1] cut 0·8, discard fragment, endfill & religate	326–787	pGEX-2T
BN	109–169	*Bam*HI-*Nla*IV cut 0·8 (partial), discard fragment, endfill & religate	326–508	pGEX-2T
N6	382–632	*Bam*HI-*Hpa*I	1145–1897	*Bam*HI-*Sma*I cut pGEX-2T
NE	382–486	*Nco*I-*Eco*RI[1] cut N6, discard fragment, endfill & religate	1145–1459	pGEX-2T
N2[2]	726–907	*Nla*IV-*Eco*RI	2175–2724	*Sma*I-*Eco*RI cut pGEX-2T
S1	726–784 818–907	*Spe*I cut N2, discard fragment & religate	2175–2354 2453–2724	pGEX-2T
SE	726–784	*Spe*I-*Eco*RI[1] cut S1, discard fragment, endfill & religate	2175–2354	pGEX-2T
SR1	784–892	*Eco*RI fragment from λgt11-SR1[3]	2175–2679	*Eco*RI cut pGEX-1N

[1] The *Eco*RI site used in these manipulations is derived from the polylinker in the vector.
[2] Construct N2 did not produce any detectable fusion protein but it is shown in the table as it was the precursor to construct S1.
[3] This fragment is derived from the original SPAG-1 clone described by Williamson (Williamson *et al.* 1989) and the *Eco*RI sites are artificial.

cDNA (Hall *et al.* 1992). The strategy employed to make each of the 12 constructs is outlined in Table 1 and the actual configuration of each construct is shown in Figure 1a. The Table and Figure are self explanatory except for construct 2·7 in which an *Eco*RV site was introduced by site-directed mutagenesis (Kunkel, Roberts & Zakour 1987) using the oligonucleotide shown below:

```
      47                                  71
SPAG-1  CTGGAGCGGACAAGATGCCTGCGGG
OLIGO   CTGGAGCGGATATCATGCCTGCGGG
```

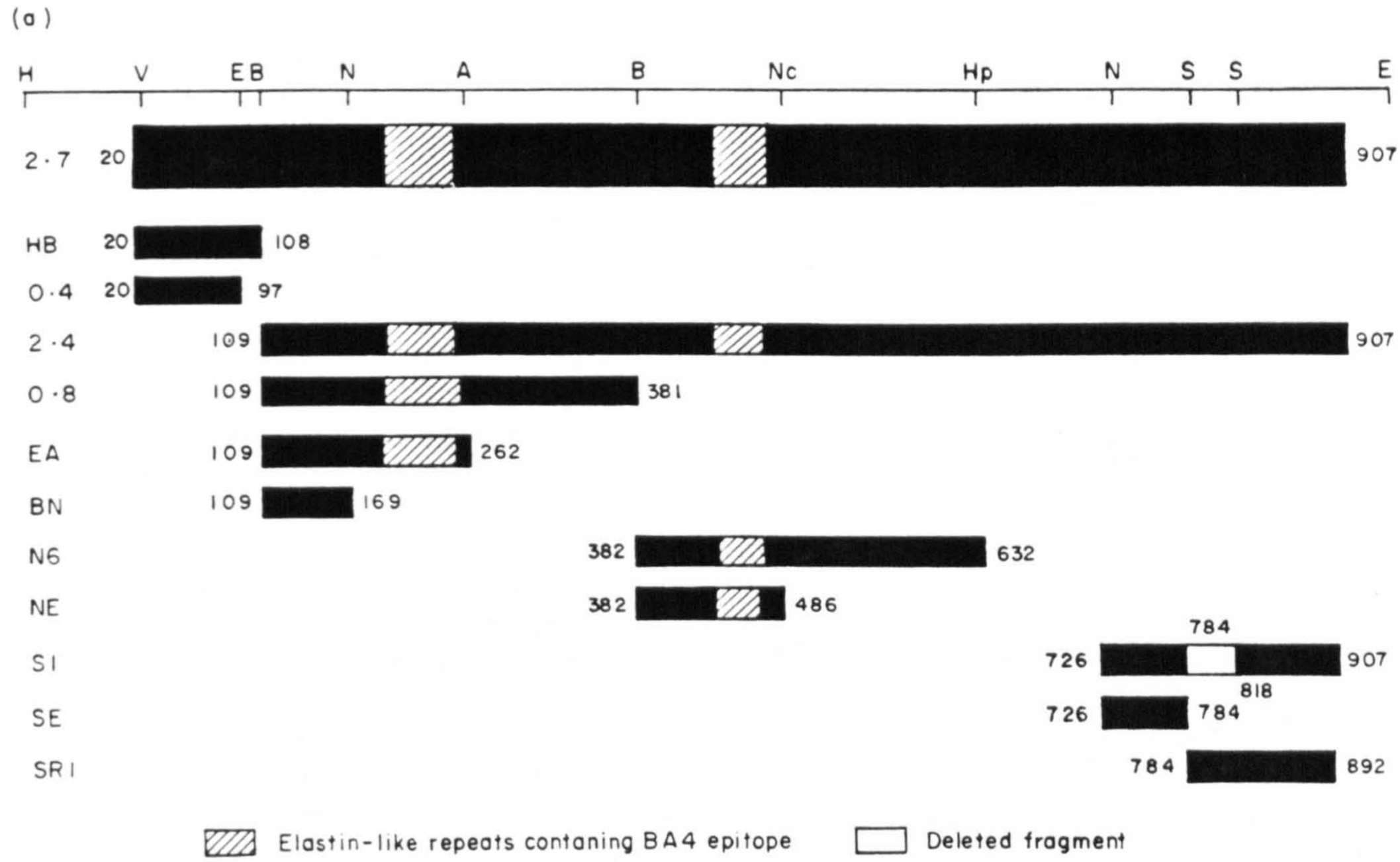
Cloning into the *Sma*I site of pGEX-3X resulted in a construct, the fusion protein product of which lacks only the first 19 amino acids of SPAG-1 and in which the twentieth has been mutated from lysine to isoleucine. The junctions of all constructs plus the internal deletion in S1 (Figure 1a) were confirmed by sequencing.

Antibodies and immunizations

MoAbs 1A7 (anti-SPAG-1) and BA4 (anti-elastin VGVAPG) have been described previously (Wrenn *et al.* 1986, Williamson *et al.* 1989, Hall *et al.* 1992). Bovine antisera (10T and 34A) were raised against recombinant SR1-β galactosidase fusion protein. The original SR1 fusion protein (corresponding to amino acids 784–892) expressed in λgt11 was purified as described (Williamson 1988, Williamson *et al.* 1989) by electroelution and inoculated intra-muscularly into calves 10T and 34A on three occasions. On days 0 and 56 the calves received 50 μg in Freunds complete adjuvant and then on day 84 they received a further 50 μg in Freunds incomplete adjuvant. The serum used in this work was taken on day 94.

Purification of fusion proteins and Western blots

Fusion proteins were purified on glutathione-Sepharose 4B (Pharmacia, Milton Keynes, Bucks, UK) from lysates of *Escherichia coli* prepared by a modified version of a published procedure (Smith & Johnson 1988). Routinely a 100 ml overnight culture was diluted into 1 litre fresh medium (LB+0·2% glucose+100 μg ml^{-1} ampicillin)



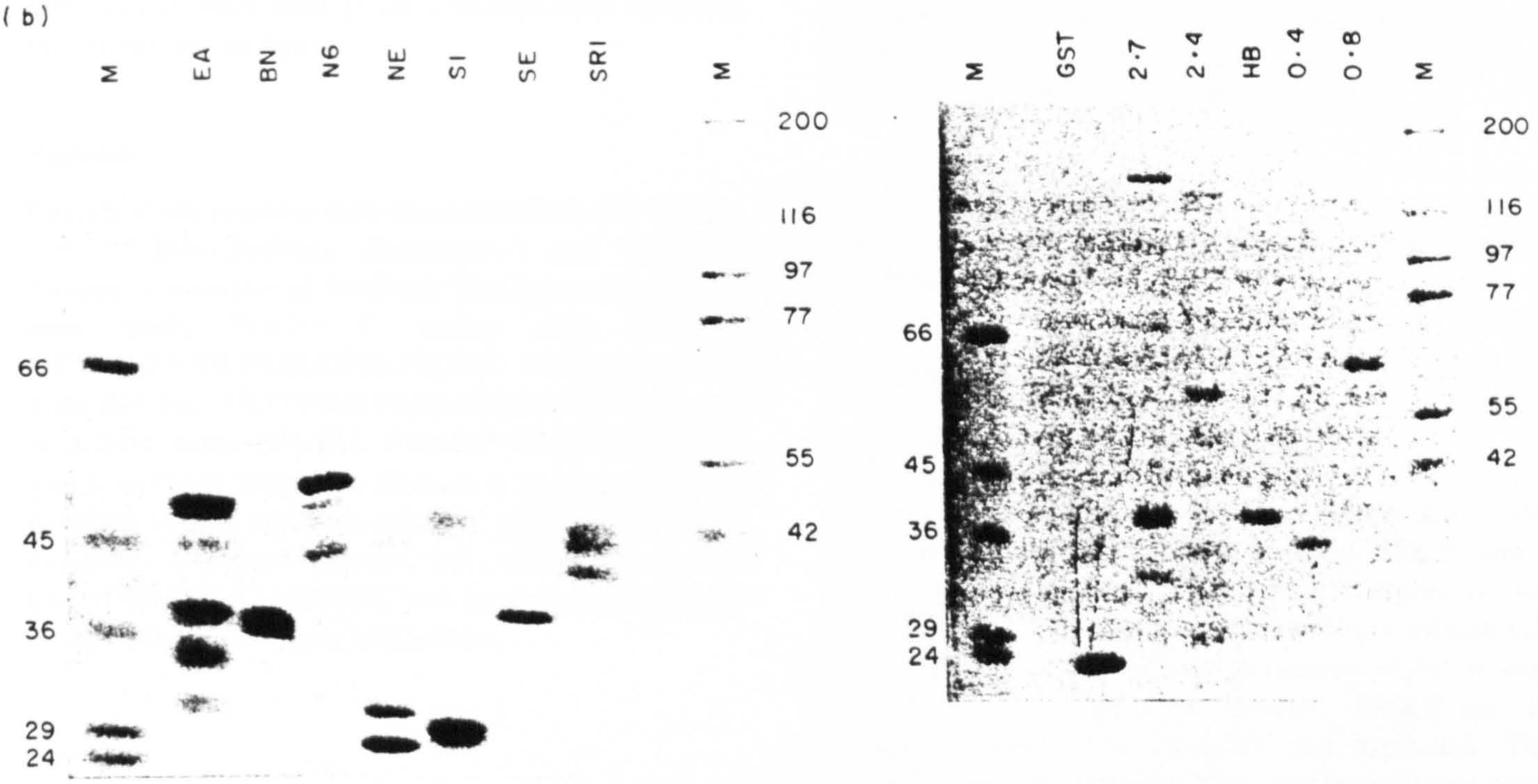


Figure 1 Details of the constructs used in this analysis. (a) Schematic representation of the pGEX-SPAG-1 constructs. The black bars represent the expressed portions of SPAG-1 for each construct whose name is given opposite down the left side. The GST partner present in all cases is not shown. The line at the top is a simplified restriction of the cDNA clone of SPAG-1 (Hall *et al.* 1992) from which all these constructs except SR1 are derived (see note 3 Table 1). The sites shown are those used to derive the constructs and are as follows: H, *Hind*III; V, an artificial *Eco*RV site; E, *Eco*RI; B, *Bam*HI; A, *Acc*I; Nc, *Nco*I; Hp, *Hpa*I; N, *Nla*IV; S, *Spe*I. (b) Coomassie stained SDS-PAGE gels (12·5%) of the proteins expressed from the constructs shown in (a). The tracks are labelled with the construct name. Size marker tracks are labelled M. The size markers are labelled with their M_r in kilodaltons.

and grown for 2–2·5 h at 37°C. Induction with 1 mM IPTG was followed by growth at 37°C for 2 h. Cultures were harvested by centrifugation (2660 *g*, 4°C). Pellets were resuspended in 10 ml ice cold MTPBS (150 mM NaCl, 16 mM Na_2HPO_4, 4 mM NaH_2PO_4) and made 1% in Triton-X100, 0·1 mM in $MgCl_2$ and the following protease inhibitors added: PMSF (0·1 mM), 1,10 phenanthroline (10 μM). The suspension was sonicated 5 × 1 min on ice at maximum amplitude (approx 7 microns) on an MSE sonicator. The lysate was clarified (10 min, 17 540 *g*, 4°C) and the supernatant applied to a 2 ml glutathione-Sepharose 4B column which was run according to the manufacturer's instructions (Pharmacia). Concentrations of eluted fusion proteins were estimated using the Bradford assay (Bradford 1976) and quality was assessed by SDS-PAGE (Anderson, Baum & Gesteland 1973). For Western blot analysis 0·5–1 μg of fusion protein were loaded per track and, following electrophoresis, the gels were processed as described (Harlow & Lane 1988). Alkaline phosphatase-IgG conjugate was used as the detection system and the substrate consisted of 0·0033% nitro blue tetrazolium (prepared in 70% dimethyl formamide) plus 0·016% bromochloroindolyl phosphate (prepared in 100% dimethyl formamide) in 10 mls 100 mM NaCl, 5 mM $MgCl_2$, 100 mM Tris pH 9·5. Sera and monoclonals were used at the concentrations shown in the legends to the figures.

Peptides

Custom made peptides were purchased from the Department of Biotechnology, Biochemistry and Molecular Biology, University of Sheffield. The following peptides were made: Peptide C, amino acids 801–822, DPSSRTPNAKPAELGPSLVIQNC; Peptide B, amino acids 804–822, SRTPNAKPAELGPSLVIQNC; Peptide A, amino acids 807–822. PNAKPAELGPSLVIQNC. The C terminal cysteine in all cases is added to facilitate coupling to BSA and does not form part of the SPAG-1 sequence. A control peptide, 191, whose sequence is LEVDDNEDERLMPSHSC was kindly provided by Dr Arthur Moir, University of Sheffield.

Dot blots

10 μg of peptide or fusion protein were spotted onto nitrocellulose filters and left to dry for at least 1 h. The filters were then blocked in TS-milk (5% Marvel in 0·15 M NaCl, 20 mM Tris-Cl pH 8·0) for 30 min at room temperature (RT). The primary MoAb was added (concentration shown in figure legend) in TS-milk (0·02% azide) and incubated overnight. The filter was washed (5 min per wash) once in TS, four times in TS-0·05% Triton X100 and incubated with goat anti-mouse peroxidase conjugate (1/1000 dilution in TS-milk; Sigma) for 2 h at room temperature. The blots were washed as before and developed using 4-chloro-1-napthol as the substrate (for details see Harlow & Lane 1988).

Table 2 Comparison of the expected size with the actual size of the longest product from each construct.

Construct	Expected size[1] (kDa)	Actual size (kDa)
2·7	112·8	145
HB	32·6	37·5
0·4	31·3	35
2·4	103·3	130
0·8	47·9	63
EA	36·8	49
BN	29·1	36
N6	48·9	50
NE	32·5	44
S1	38·7	46
SE	29·3	36
SR1	34·1	45

[1] The expected size was deduced by adding the M_r of GST in our gel system (namely 23 kDa—see Figure 1b) to the size calculated from the sequence for each expressed fragment of SPAG-1.

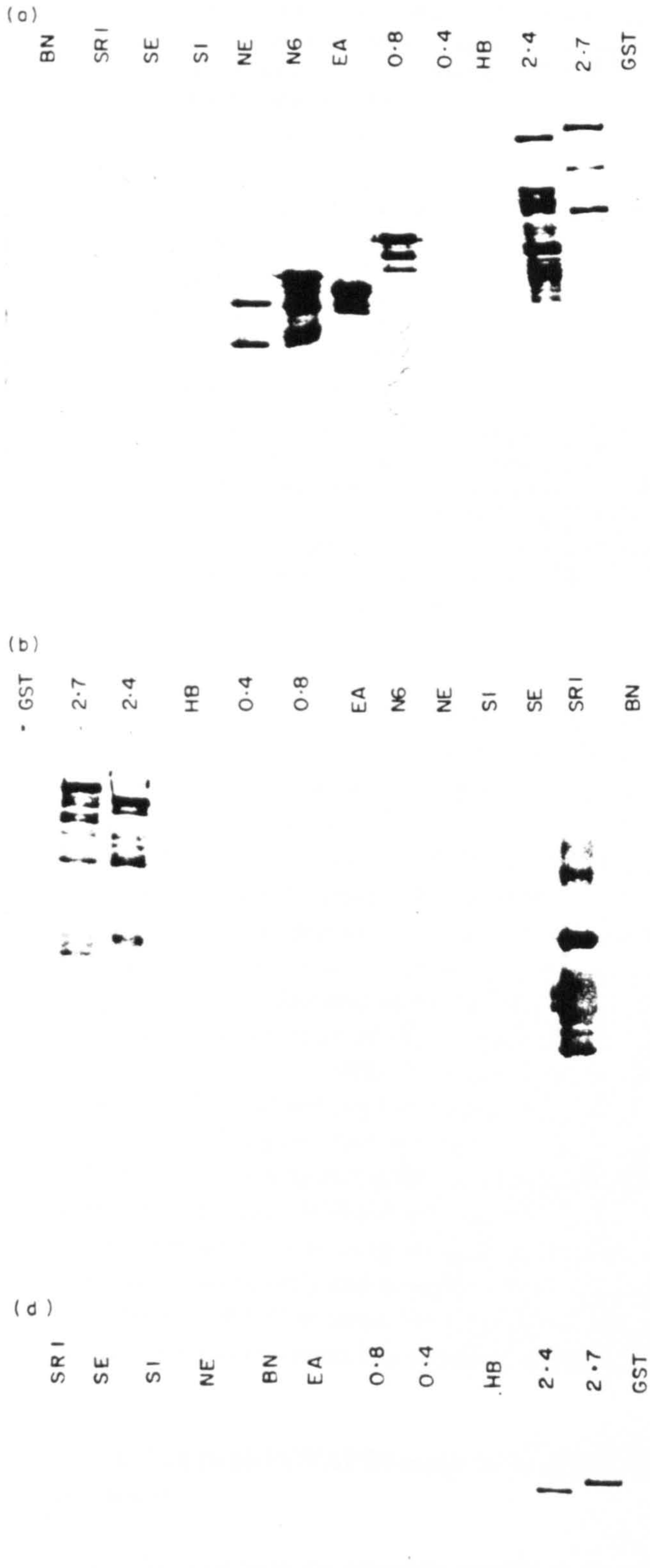
Sporozoite neutralization tests

These assays were performed as described (Williamson *et al.* 1989).

RESULTS

Expression of SPAG-1 as GST fusions

We have generated a series of constructs of SPAG-1 in the pGEX series of vectors (Smith & Johnson 1988, Pharmacia), which are shown diagramatically in Figure 1a (see also Table 1). The longest of these, 2·7, is almost full length lacking only the first 19 amino acids. It was constructed by creating an artificial *Eco*RV site as detailed (Table 1 and Materials and Methods). The remaining constructs except SR1 were all produced using conveniently located restriction sites (see Figure 1a and Table 1). SR1 was generated by inserting the *Eco*RI fragment derived from the original λgt11 clone (Williamson *et al.* 1989) into pGEX-1N. We have verified that these constructs contain the expected in frame inserts by double stranded sequencing across the junctions at both



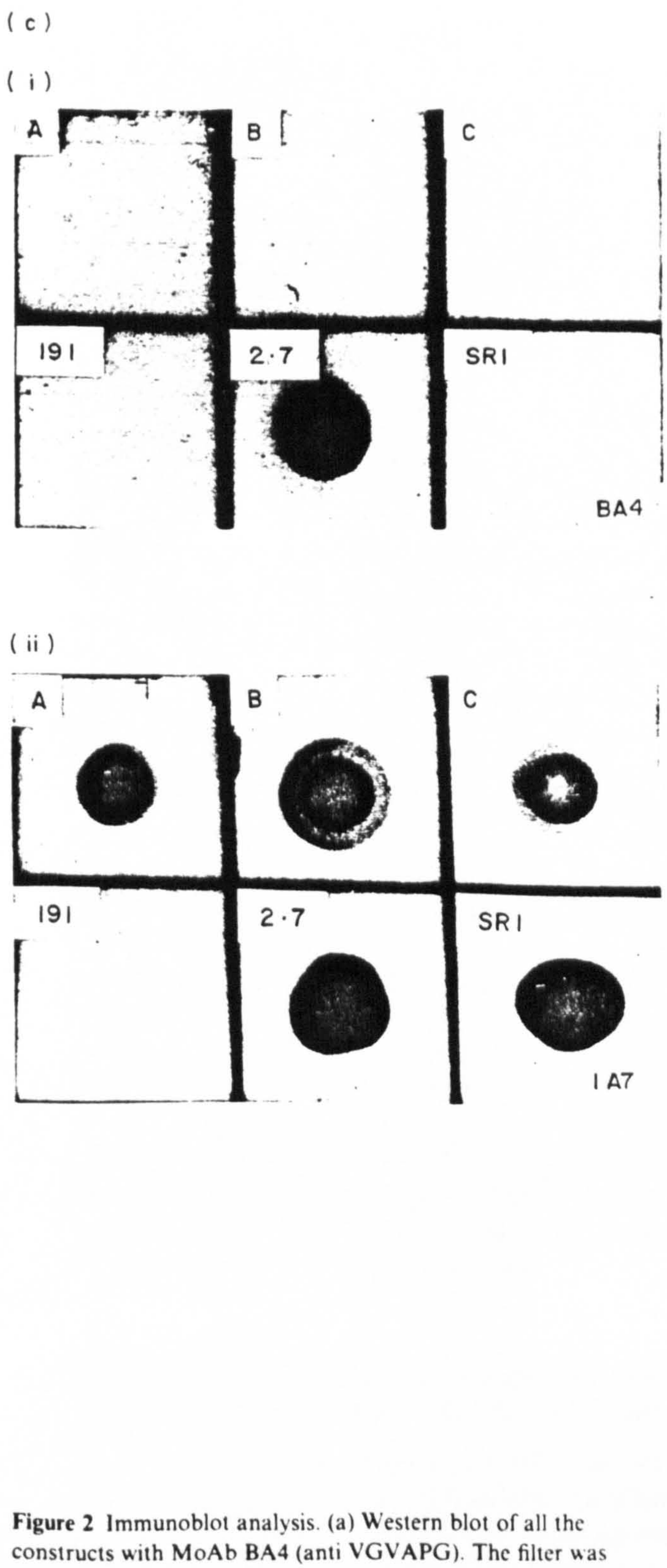


Figure 2 Immunoblot analysis. (a) Western blot of all the constructs with MoAb BA4 (anti VGVAPG). The filter was probed with ascites diluted 1/1000. The tracks are labelled with the construct name. (b) Western blot of all the constructs probed with 1A7 ascites diluted 1/1000. (c) Dot blots probed with (i) BA4 and (ii) 1A7 ascites at 1/1000 dilution. A, B and C are synthetic peptides from 807–822, 804–822 and 801–822 respectively. 191 is an unrelated control peptide. 2·7 and SR1 are protein derived from constructs of those names. (d) Western blot obtained with bovine anti-serum 10T raised against the β galactosidase-SR1 fusion protein. The blot was probed at a dilution of 1/750.

Text cut off in original

Table 3 Sporozoite neutralization assays. This table shows the results of sporozoite *in vitro* inhibition assays with selected antibodies used in this study. Sporozoites were from the Ta Hisar stock (Gill, Bhattacharyula & Kaur 1976)

Sera	Median % infection	% Inhibition
10T day 0	13·8	0
10T day 94	0	100
34T day 0	14·0	0
34T day 94	0·5	96

The sera were used undiluted. Median % schizont infected cells developing in cultures containing test and control antibodies are given from which % inhibitions for test antibodies were derived as described previously (Williamson *et al.* 1989). The differences between the control and test median % infection values were shown to be statistically significant ($P < 0·001$) using the Mann-Whitney U test for nonparametric statistics (Siegel 1956).

ends (data not shown). The affinity purified expressed products from all of these constructs were characterized by SDS-PAGE and the results are shown in Figure 1b. All of the constructs produce proteins of larger size than predicted by the sequence (see Table 2 for comparison of actual and expected values). In our earlier studies with SPAG-1 a similar phenomenon was observed and we suggested that the apparent increase in size may be due to unusual structural features resulting from the anomalous amino acid composition of the molecule (Williamson *et al.* 1989, Hall *et al.* 1992). It is noteworthy that the recloned SR1 fragment appears to produce a GST fusion protein 11 kD higher than expected, which is the same difference that was apparent for the original β galactosidase fusion protein (Williamson *et al.* 1989). In addition there is consistently a range of smaller products which vary both qualitatively and quantitatively from construct to construct. Whether these are a result of degradation and/or internal initiation has not been determined.

MoAb BA4 (anti-VGVAPG) reacts as predicted with the constructs

Figure 2a illustrates the reactivity profile of the series of constructs with anti VGVAPG monoclonal BA4 (Wrenn *et al.* 1986). Clearly those constructs which are positive (NE, N6/EA, 0·8, 2·4 and 2·7) are those that we would expect as predicted by the location of the VGVAPG motif in the sequence (Figure 1a and Hall *et al.* 1992). This reactivity profile clearly validates the nature of these constructs.

Locating the 1A7 neutralizing determinant

We also reacted the expressed constructs with MoAb 1A which we already know recognizes a determinant locate towards the C terminus present in construct SR1 (Hall *et al.* 1992). This MoAb reacted only with constructs tha span the SR1 region (2·4, 2·7, and SR1) and did not reac with any of the constructs containing upstream region (Figure 2b). This strongly suggests that the 1A7 epitope i located uniquely within the SR1 region. Interestingly, th construct S1 is not recognized by MoAb 1A7 (Figure 2b S1 spans SR1 except that it lacks residues 784–818 (Figur 1a). This immediately suggests that the 1A7 epitope lies i the stretch of residues running from 784–818. We there fore synthesized three overlapping peptides from withi this region: a 22 mer from 801–822, a 19 mer from 804 822 and a 16 mer 807–822 and tested these in dot blots These peptides reacted strongly with 1A7 but not at a with a control antibody (BA4) as shown in Figure 2c. I addition a control peptide did not react with eithe antibody. These results have been confirmed by ELISA using the same peptides coupled to BSA (not shown) Thus we conclude that the 1A7 determinant reside somewhere within the stretch of 16 amino acids compris ing residues 807–822. That this determinant is neutraliz ing has been reported previously (Williamson *et al.* 1989) Of note is the fact that MoAb 1A7 recognizes some of th smaller products of constructs 2·7 and 2·4. This i unexpected since simplistically, one would expect tha any product which bound to the affinity matrix mus contain an intact N terminal GST moiety. Therefore an smaller products should be degraded from the C termina end and lose 1A7 reactivity. However, it is formall possible that SPAG-1 oligomerizes through its C termina region and that N terminally deleted GST fusion protein could co-purify via association with intact molecules This however remains to be determined.

Bovine anti-sera against β-gal SR1 neutralizes sporozoite infectivity but do not recognize the 1A7 determinant

We wished to know whether the 1A7 epitope wa recognized by the bovine immune system. Therefore tw bovine antisera (10T & 34A) that were raised against th SR1 region (expressed as a β-gal fusion protein in λ gt11 were assayed. Both sera gave identical results. Th Western blot using serum 10T is shown in Figure 2d and i can be clearly seen that all the constructs recognized b 1A7 are also recognized by 10T. In addition there is reaction with construct S1 demonstrating that the serum reacts with determinants distinct from the 1A7 epitope Interestingly the sera 10T and 34A do not recognize th

peptides that are recognized by MoAb 1A7 and thus do not overlap in specificity at all with this monoclonal (data not shown). However these bovine sera do neutralize infectivity of sporozoites very efficiently as demonstrated in Table 3 indicating that the bovine does recognize neutralizing epitopes in the SR1 region of SPAG-1.

DISCUSSION

We have mapped the neutralizing determinant on SPAG-1 recognized by MoAb 1A7 to 16 amino acids resolution. This determinant lies within the sequence PNAK-PAELGPSLVIQN which is located towards the C terminus, from residues 807–822. In addition we have demonstrated that the bovine can recognize different neutralizing determinants in the C terminus which do not overlap with the 1A7 epitope. Within the limits of the resolution of these experiments we can say that the determinants are unique to the C terminus since the constructs that do not contain the SR1 region do not react with either 1A7 or the bovine sera. Furthermore the sequence containing the 1A7 determinant is not repeated elsewhere in the molecule.

These observations are important since they implicate the C terminus as an immunologically relevant region that can be recognized by the bovine immune response. This has implications for sub-unit vaccine design. It should be stressed however that these data do not exclude the possibility that immunologically relevant determinants may be located elsewhere in the molecule and indeed experiments are currently underway to test this. It is also possible to infer from these data that SPAG-1 carries a ligand(s) for host cell recognition in the C terminal SR1 region.

It is interesting that neutralizing epitopes reside close to the C terminus which one would imagine may be inaccessible to an immunoglobulin molecule because of its postulated proximity to the membrane. However, the answer to this may lie in the fact that SPAG-1 is processed from the N terminus (Williamson *et al.* 1989, Hall *et al.* 1992) and this might reveal functional C terminal structures such as ligands. This situation is similar to that found with the major merozoite surface protein (MSP-1) of *Plasmodium falciparum* (Blackman *et al.* 1991). The MSP-1 precursor molecule of 190 Kd is extensively processed to ultimately retain a 19 Kd C terminal fragment which is thought to be the ligand for red cell recognition since neutralizing monoclonals recognize this structure. Also relevant to this discussion are the recent results concerning the circumsporozoite protein of *P. falciparum* that demonstrate that the C terminally located conserved domain called region II carries ligands for the surface of hepatocytes (Cerami *et al.* 1992).

Logically the next phase of this work is to define the full spectrum of B and T epitopes in SPAG-1. Both these objectives can be approached using the array of constructs described in this manuscript to indicate, broadly, their location. Attempts will be made to generate sera and T cell clones directed against the full length 2·7 construct (probably with the GST cleaved off by factor Xa) and then dissecting the response obtained using the series of constructs described in the present analysis. Fine tuning will need to be achieved using synthetic peptides designed with the aid of predictive algorithms (Margalit *et al.* 1987, Rothbard & Taylor 1988). Data on the degree and nature of SPAG-1 polymorphism as well as immune response variation in the bovine will also need to be assessed. Ultimately we would hope to utilize this knowledge to define a non-variable T cell helper determinant(s) which was not influenced by host immune response gene variation. This structure, if obtained would be coupled to the C terminal B epitopic region we have defined in this paper and which we know to be relatively invariant (unpublished observations) and vaccination trials pursued. Whilst the experience with the CSP of malaria (for review see Romero 1992) indicates that we must be cautious about the proposal that SPAG-1 could alone form the basis of a synthetic sub-unit vaccine against tropical theileriosis we believe it offers potential because: a) it is a defined neutralizing antigen which can be manipulated to advance our knowledge of the parameters affecting the bovine immune response and b) it remains a serious candidate for inclusion in a multi-component, multistage sub-unit vaccine especially in the light of the encouraging results obtained recently for *Theileria parva* (Musoke *et al.* 1992). The work we describe here lays the groundwork for future investigations of both SPAG-1 and other potential components in the search for a non-living vaccine against tropical theileriosis.

ACKNOWLEDGEMENTS

This work was supported by the Agricultural and Food Research Council (P.D.H. & E.S.H.), the European Community (N.B.), the Wellcome Trust (A.T. & H.A.B.) and the Overseas Development Administration (D.B. & S.W.). P.A.K. was the recipient of a SERC studentship. We gratefully acknowledge Arthur Moir for his advice and Meg Stark for help with the figures. Thanks to Professor Bob Mecham for the BA4 MoAb. The provision of parasite material by the staff of the Protozoology Division, Centre for Tropical Veterinary Medicine, University of Edinburgh is gratefully acknowledged.

REFERENCES

Anderson C.W., Baum P.R. & Gesteland R.F. (1973) Processing of adenovirus-2 induced proteins. *Journal of Virology* **12**, 241–252

Blackman M.J., Ling I.T., Nicholls S.C. & Holder A.A. (1991) Proteolytic processing of the *Plasmodium falciparum* merozoite surface protein-1 produces a membrane bound fragment containing two epidermal growth-factor like domains. *Molecular and Biochemical Parasitology* **49**, 29–34

Bradford M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248–254

Brown C.G.D. (1990) Control of Tropical Theileriosis (*Theileria annulata* infection) of cattle. *Parassitologia* **32**, 23–31

Cerami C., Frevert U., Sinnis P., Takacs B., Clavijo P., Santos M.J. & Nussenzweig V. (1992) The basolateral domain of the hepatocyte plasma membrane bears receptors for the circumsporozoite protein of *Plasmodium falciparum* sporozoites. *Cell* **70**, 1021–1033

Gill B.S., Bhattacharyula Y. & Kaur D. (1976) Immunization against bovine tropical theileriosis (*Theileria annulata* infections). *Research in Veterinary Science* **21**, 146–149

Hall R., Hunt P.D., Carrington M., Simmons D., Williamson S., Mecham R.P. & Tait A. (1992) Mimicry of elastin repetitive motifs by *Theileria annulata* sporozoite surface antigen. *Molecular and Biochemical Parasitology* **53**, 105–112

Harlow E. & Lane D. (1988) In *Antibodies: a Laboratory Manual*, Ch 12, pp. 490–492, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, USA

Kunkel T.A., Roberts J.D. & Zakour R.A. (1987) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Methods in Enzymology* **154**, 367–383

Margalit H., Spouge J.L., Cornette J.L., Cease K.B., Delisi C. & Berzofsky J. (1987) Prediction of immunodominant helper T cell antigenic sites from the primary sequence. *Journal of Immunology* **138**, 2213–2229

Musoke A., Morzaria S., Nkonge C., Jones E. & Nene V. (1992) A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle. *Proceedings of the National Academy of Sciences, USA* **89**, 514–518

Pipano E. (1989) Bovine theileriosis in Israel. *Revue Scientifique et Technique Office International des Epizooties* **8**, 79–87

Romero P. (1992) Malaria vaccines. *Current Opinion in Immunology* **4**, 432–441

Rothbard J.B. & Taylor W.R. (1988) A sequence pattern common to T cell eptiopes. *EMBO Journal* **7**, 93–100.

Seigel S. (1956) In *Nonparametric Statistics for the Behavioural Sciences*. McGraw-Hill, New York, USA

Singh D.K. (1990) Methods currently used for the control of *Theileria annulata*: their validity and proposals for future control strategies. *Parassitologia* **32**, 33–40

Smith D.B. & Johnson K.S. (1988) Single step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione-S-transferase. *Gene* **67**, 31–40

Tait A. & Hall F.R. (1990) *Theileria annulata*: control measures, diagnosis and the potential use of sub-unit vaccines. *Revue Scientifique et Technique Office International des Epizooties* **9**, 387–403

Williamson S.M. (1988) A *Theileria annulata* sporozoite surface antigen as a potential vaccine for tropical theileriosis. PhD thesis. University of Edinburgh

Williamson S., Tait A., Brown D. *et al.* (1989) *Theileria annulata* sporozoite surface antigen expressed in *E. coli* elicits neutralizing antibody. *Proceedings of the National Academy of Sciences, USA* **86**, 4639–4643

Wrenn D.S., Griffin G.L., Senior R.M. & Mecham R.P. (1986) Characterization of biologically active domains on elastin: identification of a monoclonal antibody to a cell recognition site. *Biochemistry* **25**, 5172–5176

ELSEVIER

Molecular and Biochemical Parasitology 67 (1994) 1–10

MOLECULAR AND BIOCHEMICAL PARASITOLOGY

Polymorphism of SPAG-1, a candidate antigen for inclusion in a sub-unit vaccine against *Theileria annulata*

Frank Katzer [a], Mark Carrington [b], Pamela Knight [c], Susanna Williamson [d], Andrew Tait [c], Ivan W. Morrison [e], Roger Hall [a,*]

[a] *Department of Biology, University of York, York, Y01 5DD, UK*
[b] *Department of Biochemistry, University of Cambridge, Tennis Court Rd., Cambridge, CB2 1QW, UK*
[c] *Wellcome Unit for Molecular Parasitology, University of Glasgow, Bearsden Rd., Glasgow, G61 1QH, UK*
[d] *Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, Edinburgh, EH25 9RG, UK*
[e] *AFRC Institute For Animal Health, Compton, Nr. Reading, Berkshire, RG16 0NN, UK*

Received 19 January 1994; accepted 18 April 1994

Abstract

SPAG-1, a *Theileria annulata* sporozoite surface antigen, is a vaccine candidate. Data is presented, based on the clonal segregation of SPAG-1 associated RFLPs, showing that this antigen is encoded by a single copy gene. We have cloned and sequenced a full-length genomic copy of the SPAG-1 gene and a comparison of this with a previously published SPAG-1 cDNA sequence demonstrates a high degree of polymorphism. We infer that these sequences represent two distinct allelic SPAG-1 variants. The deduced polypeptides show an overall identity of 92% with the most variable stretch (60% identity) occurring towards the middle of the molecule. The N and C termini are more conserved with identities of 92% and 97% respectively. The elastin receptor ligand, VGVAPG, present 3 times in the protein sequence derived from the cDNA is not found in that deduced from the genomic copy. Evidence for 2 further SPAG-1 alleles was obtained from PCR based sequences using macroschizont clones containing different SPAG-1 associated RFLPs. In summary we have shown the existence of at least 4 highly polymorphic SPAG-1 alleles. The implications of such polymorphism between and within distinct geographical isolates for the development of a SPAG-1 based subunit vaccine is discussed.

Keywords: *Theileria*; Sporozoite; Antigen; Polymorphism; Vaccine

1. Introduction

Antigenic polymorphism is a common theme amongst infectious agents and has serious implications for vaccine development. It is essential when considering molecules as candidates for inclusion in a sub-unit vaccine that an understanding of the nature and degree of variation in their structure is achieved. Ideally the aim is to define an immunolog-

Abbreviations: RFLP, restriction fragment length polymorphism.;

Note: Nucleotide sequence data reported in this paper have been submitted to the GenBank™ database with the accession numbers X78186–X78194.

* Corresponding author. Tel: 0904-432864; Fax: 0904-432860; E-mail: frh1@vax.york.ac.uk

ically relevant region which is not polymorphic. However, in many cases, the immunodominant protective epitopes exhibit variation, presumably as a result of immunological selection. For example, in the malaria parasite, two of the major vaccine candidates, MSP-1 and CSP, show sequence variations which occur in regions recognised by T cells [1–6]. This constrains the use of these regions in a vaccine formulation since all the different variants may be required in the final preparation. The problem is further compounded by immune response variation determined by MHC polymorphism in the host. This means that the ideal epitope must be non-polymorphic and universally recognised.

There are several reports of antigenic polymorphism in *Theileria*. Thus, Shiels et al. [7] generated a panel of monoclonal antibodies against *T. annulata* macroschizonts which revealed diversity between and within isolates. A similar approach has been employed for *T. parva* by Minami et al. [8]. The structural bases for the observed polymorphisms has not been fully established in either case. However a panel described by Minami et al. has been shown to be directed against a single antigen designated as polymorphic immunodominant macroschizont (PIM) antigen which is also expressed in the sporozoite [9,10]. Variant forms of this molecule migrate with different apparent M_rs and, since the molecule is not glycosylated then it is presumed that the size differences are reflected in the amino acid sequence. Diversity has also been described in an immunodominant 30–32-kDa *T. annulata* merozoite/piroplasm antigen, the underlying basis for which is probably differential glycosylation [11].

We have been studying a sporozoite antigen, SPAG-1, defined by a neutralising monoclonal antibody (1A7) as a vaccine candidate [12]. To date we have sequenced and expressed a cDNA clone encoding the complete polypeptide and, we have located the neutralising epitope defined by 1A7 within a linear sequence of 16 amino acids at the C terminus [13,14]. Other neutralising epitopes recognised by immune bovine sera are distinct, but also locate to a nearby position in the C terminus. In addition to attempting to localise potentially protective B and T cell epitopes it is important to define if they are encoded by conserved regions of the molecule, and therefore we examine variation in the SPAG-1 gene in this paper. Previously, we identified 3 RFLPs of SPAG-1, but the underlying genetic basis for this variation was unknown [12]. The present study demonstrates that the RFLPs segregate independently in a variety of cloned lines, derived from 2 geographically distinct isolates, and thus behave as allelic variants. To define in detail the level and nature of the polymorphism, marked by these RFLPs, we have analysed the complete sequence of 2 alleles and partial sequences of 2 further alleles. These comparisons reveal a highly polymorphic region towards the middle of the molecule (amino acid residues 284–369) and a much more conserved stretch towards the C terminus (residues 806–902).

Table 1
Origin of cloned macroschizont infected cell lines and the SPAG-1 alleles and *Eco*RI RFLPs they contain

Cloned macroschizont cell line	Size of *Eco*RI RFLP	Allele	Origin
TaA 46A	3.4 kb	A3.4	Wathanga (M.Sc. Thesis, Edinburgh, 1984)
TaA 139D4	3.4 kb	A3.4	Wilkie (pers. comm.) / Williamson (Ph.D. Thesis, Edinburgh, 1988)
TaA 139D7	3.4 kb	A3.4	Wilkie (pers. comm.)
TaA 139E3	3.4 kb	A3.4	Wilkie (pers. comm.)
TaA 139E5	3.4 kb	A3.4	Wilkie (pers. comm.) / Williamson (Ph.D Thesis, Edinburg, 1988)
TaA 46.2	3.4 kb	A3.4	Wathanga (M.Sc. Thesis,Edinburgh, 1988)
TaH 46.2	3.4 kb	H3.4	Wathanga (M.Sc. Thesis, Edinburgh, 1984)
TaH 46.3	3.4 kb	H3.4	Wathanga (M.Sc. Thesis, Edinburgh, 1984)
TaH 46.4	3.4 kb	H3.4	Wathanga (M.Sc. Thesis, Edinburgh, 1984)
TaA 139D6	4.8 kb	A4.8	Wilkie (pers. comm.) / Williamson (Ph.D. Thesis, Edinburgh, 1984)
TaA 46.3	4.8 kb	A4.8	Wathanga (M.Sc. Thesis, Edinburgh, 1984)
TaHBL3b	6.0 kb	H6.0	Katzer

TaA and TaH denote *T. annulata* Ankara and Hissar, respectively.

The implications of these findings are discussed particularly with respect to vaccine design.

2. Materials and methods

Parasite material, DNA extraction and Southern blotting. Piroplasms of the *T. annulata* Hissar (Ta Hissar) [15] and *T. annulata* Ankara (Ta Ankara) [16] stocks were harvested from infected calf blood as described previously [12]. The macroschizont-infected cloned lines used in this study are listed in Table 1 and were cultured as described by Brown et al. [17]. The Ta Hissar series (except TaHBL3b) were cloned by limiting dilution from the parental cell line TaH46 to give clones TaH46.2, TaH46.3 and TaH46.4 and have been described previously by Wathanga (M.Sc. Thesis, University of Edinburgh, 1984). Clone TaHBL3b was similarly derived by cloning the Ta Hissar infected lymphosarcoma BL3 [18] cell line TaHBL3 [19]. The Ta Ankara clones TaA139-D4, TaA139-D6, TaA139-D7, TaA139-E3, TaA139-E5 were produced by end point titration of sporozoites into peripheral blood mononuclear cells of calf 139 (Wilkie, personal communication; Williamson, S.M., Ph.D. Thesis, University of Edinburgh, 1988). Clones TaA46.A, TaA46.2 and TaA46.3 are derived from calf 46 by limiting dilution from infected cell lines as described by Wathanga (M.Sc. Thesis, University of Edinburgh, 1984). DNA extraction and Southern blotting were performed by standard methods [20].

Construction of genomic library and sequencing of selected clones. A genomic library of Ta Hissar piroplasm DNA was created in λEMBL3 essentially as described [21]. Briefly 300 μg of piroplasm DNA were partially digested with the restriction enzyme *Sau*3A, size selected (10–20 kb) on a sodium chloride gradient (25 to 5%) and inserted into the *Bam*HI site of the replacement vector λEMBL3. The phage were packaged and propagated on Q359 to yield a primary library of 200000 clones. 25000 clones were screened with the original SPAG-1 λgt11-SR1 insert [12] and several positives isolated. One clone with an insert of 13 kb contains an internal 3.4-kb *Eco*RI fragment that hybridised to the SR1 probe. This was sub-cloned into pUC18 and the plasmid sonicated and random fragments (average size 300 bp) were size-selected and recloned into m13mp18 for shotgun sequencing using the dideoxy method [22]. To complete the sequence at the 5′ end of the gene an adjacent *Eco*RI fragment of 1.4 kb was sub-cloned from the λEMBL3 clone and sequenced by using specially designed oligonucleotides. The *Eco*RI site between the two fragments was sequenced across using a PCR fragment derived from a Ta Hissar macroschizont clone segregating for the 3.4-kb RFLP as a template. The complete sequence was further confirmed by using a series of oligonucleotides as directional primers spanning the whole open reading frame.

Table 2
Sequence and location of the oligonucleotide primers

Name of primer	Sequence of primer 5′ to 3′	Position according to cDNA Sequence
420	ctgccaattcttccggttg	1919–1900
647	cttgctgaggatcctcctcc	1158–1139
711	gaagaagttttgaaagttttgg	2667–2688
815	ggtctacagggaccaggagg	784– 804
932	gtggactatgctgaatatgatt	2722–2700
Y2	gtgtaccaggaggaaaggc	945– 963
Y3	gcggacaagatgcctgcggg	53– 83

PCR, sub-cloning and double stranded sequencing. The primers used in this study are listed in Table 2 and their location is shown diagrammatically in Fig. 3A. Two series of PCR products containing the central polymorphic region and the constant C terminal region were amplified and are referred to as series 1 and series 2 respectively. The template DNA for each of the alleles was prepared from the following clones: Ta Ankara 139-D4, allele A3.4; Ta Ankara 139-D6, allele A4.8; Ta Hissar 46.2. allele H3.4; Ta Hissar TBL3b, allele H6.0. Primers Y3 and 420 were used to amplify series 1 and the primers Y2 and 932 were used to amplify series 2 (Fig. 3A). In all cases the PCR reaction mix was made up to 50 μl containing 50 μM of each primer, 100 ng genomic DNA, 10 μM dNTP, 3 units of Taq XL and 1 × Taq Buffer supplied by Northumbria Biologicals Ltd. The reaction mix was incubated at 96°C for 5 min and this was followed by 30 cycles of 1 min at 94°C, 1 min at 60°C and 2 min at 70°C after which

the reaction was left at 70°C for 10 min. Then the PCR products were purified from an agarose gel using a Gene Clean Kit (Bio 101) according to the manufacturers instructions.

The PCR products were ligated into pGEM-T (Promega) and transformed into *E. coli* strain TG1 rec O. Plasmid DNA for double stranded sequencing was prepared for the different clones using the Wizard Miniprep Kit (Promega) following the manufacturers' instructions. Double stranded sequencing reactions were performed using a T7 sequencing kit (Pharmacia) according to the manufacturers' instructions. Primers 815, Y2 and 647 were used for sequencing the polymorphic region in the series 1 clones and 711 and 932 were used for sequencing the constant region in the series 2 clones (Table 2, Fig. 3A). To minimise errors arising from Taq polymerase infidelity the sequence of 3 independent cloned PCR products was obtained for each RFLP type. We observed an error rate of 0.22% and in those cases where there were discrepancies we have presented the consensus sequence based on a two out of three match. For three of the RFLP markers other cloned macroschizont lines exist (five for 3.4-kb Ankara stock, two for 3.4-kb Hissar stock and one for 4.8-kb Ankara stock, see Table 1) and sequence analysis of a cloned PCR product for each was also obtained over the polymorphic region only. Primers were purchased from the Department of Biotechnology, Biochemistry and Molecular Biology, University of Sheffield. All sequence analysis was performed using the UWGCG [23] package on the SERC facility at the Daresbury Laboratory.

3. Results and discussion

Previously we had identified three *Eco*RI fragments of 3.4, 4.8 and 6.0 kb released from piroplasm DNA that hybridise with the 330 bp insert derived from the original SPAG-1 clone λgt11-SR1 [12]. These data are confirmed by our current results except that the original faint band of 4.8 kb in the Ta Hissar stock is no longer present (Fig. 1, lanes 1 and 2). This observation alone suggests that the multiple bands result from mixtures of at least three parasite populations. Previous data from glucose phosphate isomerase isoenzyme studies also supports the contention that the stocks used here are mixtures [24]. To resolve whether our data are the result of multiple genes per genome or mixtures containing independently segregating alleles, we studied DNA extracted from the macroschizont-infected cloned cell lines shown in Table 1. We observed that each clone contains only one of the three RFLP types as summarised in Table 1 and illustrated by Southern blot data for each RFLP in Fig. 1. Each of the 3 RFLPs behave as alternative forms of a gene at a single locus and thus we have adopted the term allele. We consider that the above data provide strong evidence in favour of there being one SPAG-1 gene per haploid genome.

We have sequenced a complete genomic copy of SPAG-1 and the derived amino acid sequence is presented in Fig. 2A labelled gH3.4 (see below). This information is obtained from a clone isolated from a partial *Sau*3A genomic DNA (Ta Hissar) library constructed in λEMBL3. This clone was

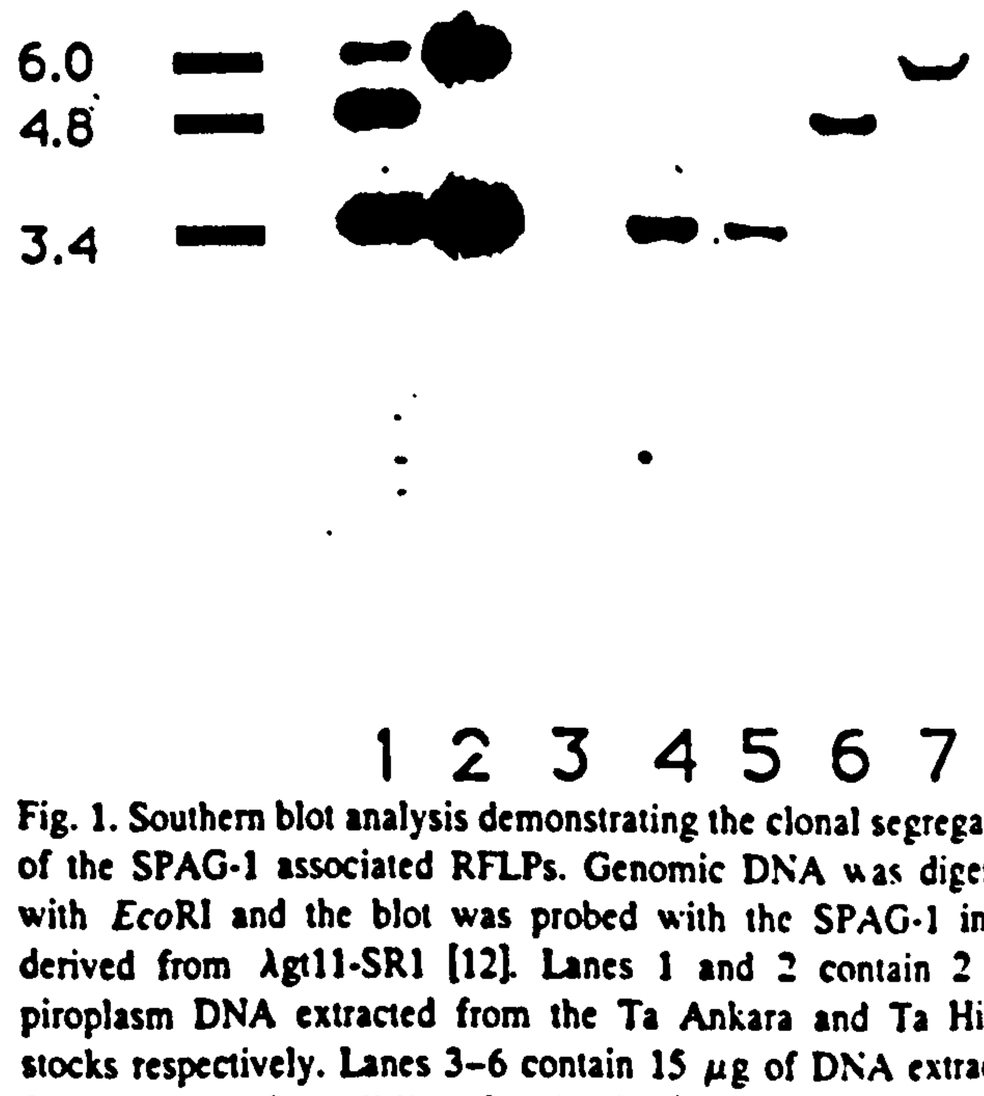
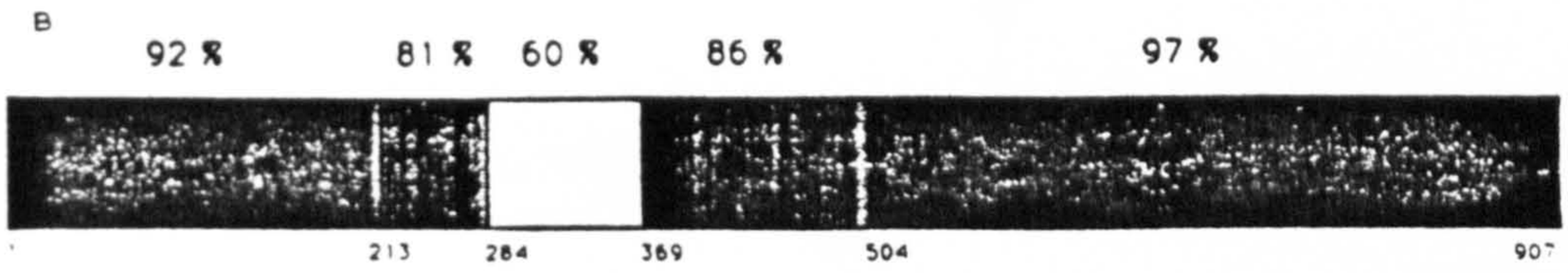


Fig. 1. Southern blot analysis demonstrating the clonal segregation of the SPAG-1 associated RFLPs. Genomic DNA was digested with *Eco*RI and the blot was probed with the SPAG-1 insert derived from λgt11-SR1 [12]. Lanes 1 and 2 contain 2 μg piroplasm DNA extracted from the Ta Ankara and Ta Hissar stocks respectively. Lanes 3–6 contain 15 μg of DNA extracted from the following cell lines (see Table 1): Uninfected BL3 (lane 3), clone TaA 139-D4 (lane 4), clone TaH 46.2 (lane 5), clone TaA 139-D6 (lane 6) and clone TaHBL3b (lane 7). The bars mark the segregating RFLPs and give their sizes in kb.

A

```
1.4   MNILHFLLTIPVIFVSGADKMPAGESSRTSKPSPLVTLESAVTQPSKDPFKTISALSKATKVWKSAVSVSGDSKTVPTPVSEPIITESFQ
      ||| ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
    1 MNIIHFLLTIPAIFVSGADKMPAGESSRTSKPSPLVTLESAVTQPSKDPFKTISALSKATKVWKSAVSVSGDSKTVPTPVSEPMITESFQ   90

      EPVSQELEFQSDTEINESGSGSDED.DDDDEEEEEEDDGSGSSKGAKGSPKAQAAVSSSSTSTASPTSPTTTSSQPGLGSSGSHGQQVPG
      |||||||||||||||||||||||| |||| ||||||| | ||| ||||||| ||||||| ||||||||| || ||| |||| || ||
   91 EPVSQELEFQSDTEINESGSGSDEDEDDDDDEEEEEDDKSTSSKNGKGSPKAQPGVSSSSTSSASPTSPTTTLSQTGLGPSGSHAQQDPG  180

      VGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVGAVPGVGGLGDDSSALPGSGGLGAGAKAGKGPGS
      |||||||||||||||||||||||||||||||    |||||||    |||||||||  ||| | | ||| ||||||||||||||||| |
  181 VGVPGVGVPGVGVPGVGVPGVGVPGVGVPGVG....GVPGVGVA....PGVGVPGVGVAPGV.GVGADSSGLPGSGGLGAGAKAGKGQGS  261

      GLQGPGGVGPGVPGVGDAASSSLPGKPPGVG.........VPGAVEPGLPGAAGVPGGKAGKSNKSSDPELDFGDESDGSGSGSAGDDD
      ||||||||| ||||| ||||| ||||||||                | | || ||||||| |   |  ||  |  ||| | | ||
  262 GLQGPGGVG.VVPGVGVAASSSSPGKPPGVGAGVMPGVGVRAQGGVIIGAPGVAGVPGGKPGQP.VSQELELKSDTEINESGSSSEGEDD  349

      DNDDEEEEDDKSTSSKGAGAKAGKGQGSVSPGGGSSASQTSSTTTSSQSGLAPSGSHPQQVPQQDPALSQPSGGGVPGVGVPGVGVPGVG
      | | |||| |||||||| |||||||||||||||||||||| || ||||| |||| || |||||| | ||||||||||||||||||||
  350 D.DEEEEEENKSTSSKGAGGKAGKGQGSVSPGGGSSASQTSPTTT.PQSGLASSGSHAQQSPQQDPAPSKPSGGGVPGVGVPGVGVPGVG  437

      VPGVGV.PGVG.VPGVGGATTSSSST..TSTTTTTTTSSSPGKPSNQGSHGTSPRKLVTRQTDSISGPIPSPGDPRAITGQMGLLYSSPS
      |||||| |||| |||||||||||||  ||| |||||  | |||| |||||||||  ||||||||||||||||||||||
  438 VPGVGVAPGVGVVPGVGGATTSSSSTTSTSTSTTTTTTTSSGKPSDQGSHGTSPRNAVTRQTDSISGPIPSPGDPRAITGQM********  519

      FVGEGERFAAQFLGDFKPKPRRYEGQETDAVKLKQFIFEEVKSLVQTLINLKLAIANDFVEISEKLKKKNQNYVPKLKLLKGEQFDTKQK
        |||||||  |||||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  520 **GEGERFAVQFLGDFKPKPRRYEGQGTDAVKLKQFIFEEVKSLVQTLINLKLAIANDFVEISEKLKKKNQNYVPKLKLLKGEQFDTKQK  607

      VANVLKGFNSLYFVFFMNLNLAKEVNKPEELAEFLWKLNTIPDKVGREFELAIEKTKGSEKKKELEEAFNSIGLGFKIAQYATNDILSSI
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  608 VANVLKGFNSLYFVFFMNLNLAKEVNKPEELAEFLWKLNTIPDKVGREFELAIEKTKGSEKKKELEEAFNSIGLGFKIAQYATNDILSSI  697

      TNSVYSLIKLKNFGDDFITEVRKSLQMVPHQKNLNGSAFIVKISEIINKKGTEDEDQTSGSGSKGTEGVSLRGQDLTEEEVLKVLDELVK
      |||||||||||||||| ||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||||||||||||
  698 TNSVYSLIKLKNFGDDFVTEVRKSLQMVPHQKNLNGSAFIVKISEIINKKGTEDQDQTSGSGSKGTEGGSLRGQDLTEEEVLKVLDELVK  787

      DVSEEQVGIGDLSDPNSRTPNAKPAELGPSLVIQNVPSDPSKVTPTQPSNLPQVPTTGPGNGTDGTTTGPGGNGEGGKDLKEGEKKEGLF
      ||||| |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  788 DVSEEHVGIGDLSDPSSRTPNAKPAELGPSLVIQNVPSDPSKVTPTQPSNLPQVPTTGPGNGTDGTTTGPGGNGEGGKDLKEGEKKEGLF  877

      QKIKNKLLGSGFEVTSIMIPMTTIIFSIVH*
      |||||||||||||| || ||||||||||||
  878 QKIKNKLLGSGFEVASIIIPMTTIIFSIVH*  907
```

B



OVERALL IDENTITY 92 %

2. Comparison of SPAG-1 protein sequences, derived from the DNA sequences of alleles gH3.4 and cH. (A) The polypeptide iences are compared using the GAP programme on the UWGCG analysis software package [23]. The VGVAPG motifs in cH are erlined and the region spanned by the putative intron is shown marked with asterisks. (B) Schematic representation of the comparison wn in (A) The figures above the blocks represent the % identity over each segment calculated by scoring a gap as one change. The bers below designate the amino acid residues at the boundaries of the sequence blocks and are based on the cH molecule as shown in

shown to contain a 3.4-kb *Eco*RI fragment that hybridised with the SPAG-1 330 bp λgt11-SR1 insert (data not shown). Thus, we define this as a 3.4-kb Ta Hissar allele and give it the designation gH3.4 (the g is to indicate that it is the sequence derived from a genomic clone). When compared to the amino acid sequence derived from the published cDNA sequence, extensive variation is apparent (Fig. 2A and B) [13]. This cDNA, which is derived from a Ta Hissar library [13], is quite clearly not the product of gH3.4 and we designate it as a separate allele, cH. Overall, the two protein sequences are 92% identical, with the C terminal half of the molecule being most homologous (residues 504–907, 97% identity). The N-terminal segment from residues 1–212 is also highly similar (92% identity, including 2 changes in the putative signal peptide). These identity values are generated by counting an insertion or a deletion as one change. The polymorphism is greater between residues 213–504 in the second quarter of the molecule and is most extensive between amino acids 284 and 369 showing an identity of only 60%. The variation is due to multiple gaps/insertions and amino acid substitutions. One apparent area of variation is probably spurious and likely to be due to the presence of a short in-frame intron of 30 bases (supporting data not shown). When decoded these specify the 10 extra amino acids GLLYSSKSPV which appear in the gH3.4 sequence starting at residue 487 with the corresponding gap in the cH sequence being denoted by asterisks in Fig. 2A. Interestingly, unlike the cDNA allele cH, the gH3.4 allele product does not contain any VGVAPG motifs which considerably weakens our previous hypothesis that this represents a ligand for host cell recognition [13]. This is supported by other data demonstrating that cells sorted with a monoclonal antibody to the elastin receptor exhibit no correlation with their susceptibility to infection by sporozoites (J. Campbell unpublished, manuscript submitted). Other features such as the elastin PGVGV repeats, the D/E and the T/S regions described previously in the cH allele product are also present in the gH3.4 polypeptide [13].

To further monitor the extent of SPAG-1 polymorphism we sequenced the most variable region

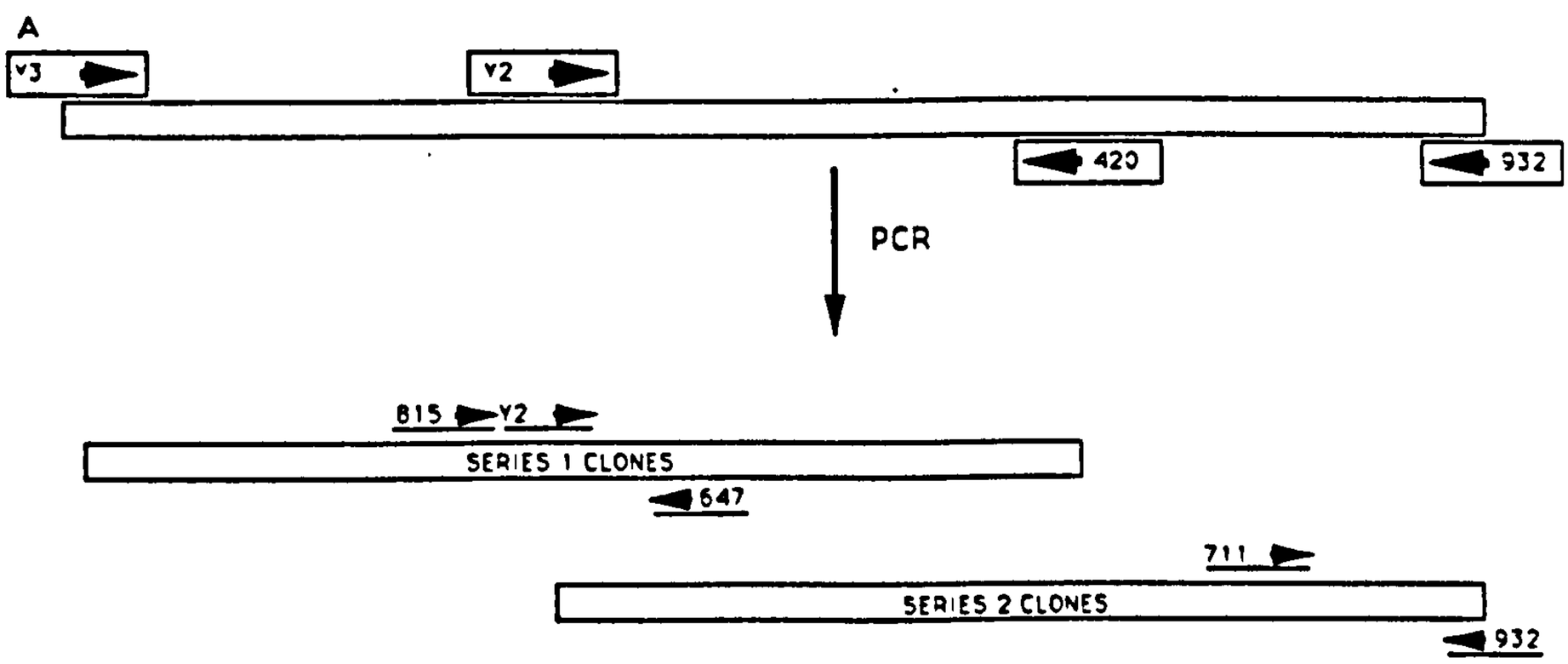


Fig. 3. Comparison of SPAG-1 alleles over the most polymorphic region and the constant C terminal region. (A) Schematic representation of the PCR based strategy used to generate the sub-clones sequenced in this analysis. The location of the primers used to generate the series 1 and series 2 fragments is shown by the boxed arrows, whilst those used for sequencing are shown by the smaller, underlined arrows. For precise details of each primer see Table 2. (B) Comparison of the SPAG-1 sequences over the most polymorphic region (bases 850–1107, amino acids 284–369). PCR derived sequences are shown for each allele from series 1. Templates were derived from the same macroschizont clones as used in Fig. 1. In addition the same region from alleles cH and gH3.4 is presented for further comparison. The data are given in (i) as the DNA sequence and in (ii) as the derived peptide sequence. (C) As in B except that comparison is between the series 2 sequences from the C terminal constant region (bases 2419–2706, amino acids 806–902). In addition the analogous sequence of the *T. parva* p67 antigen is presented for comparison. The 16 amino acids containing the 1A7 epitope [14] are underlined. All alignments are made to the cH sequence and - represents identity,. represents a gap and any substitutions are shown by the relevant letter.

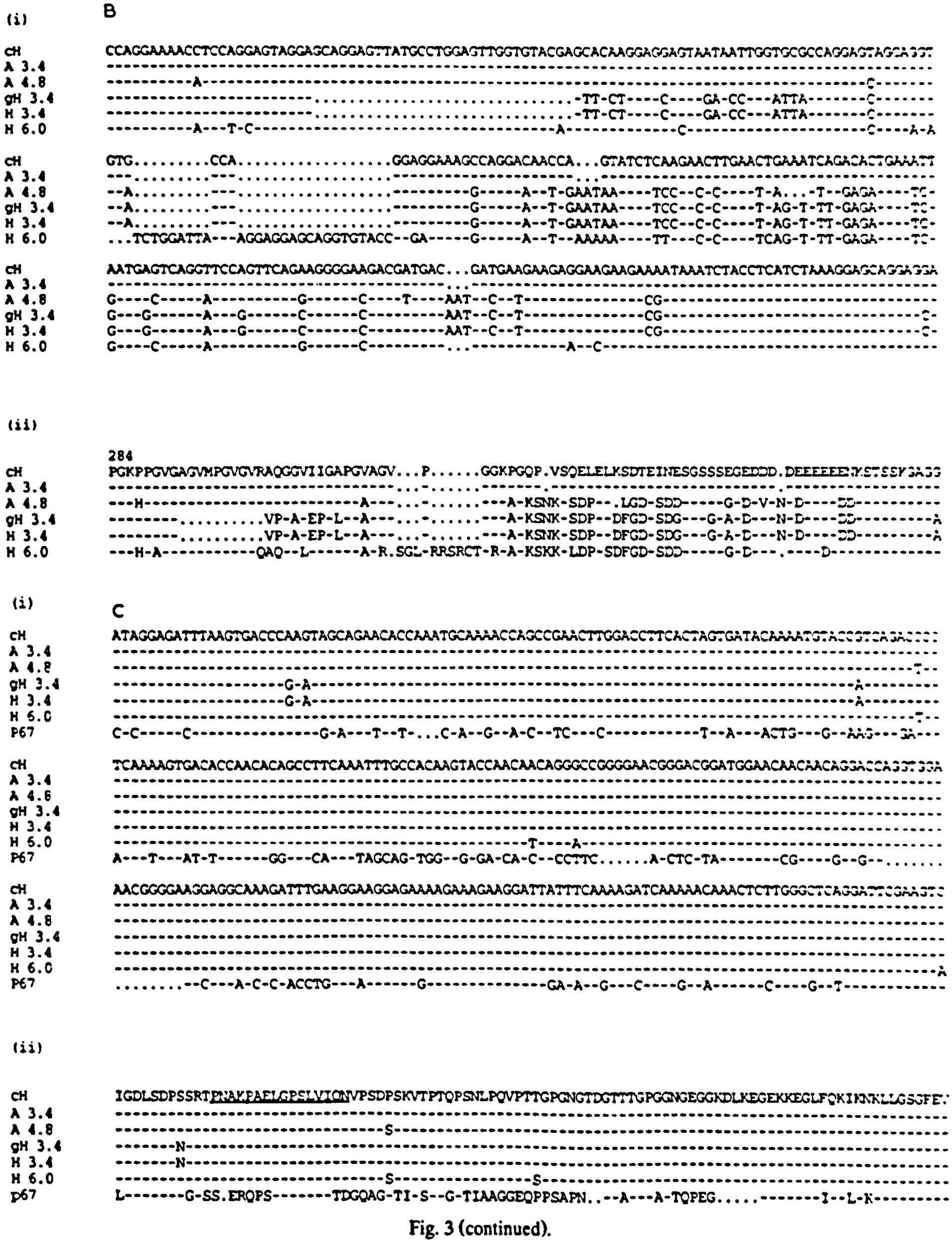


Fig. 3 (continued).

from bases 850 to 1107 corresponding to amino acids 284 to 369 (Fig. 2B). For this we used the cloned series 1 PCR genomic DNA fragments created as outlined schematically in Fig. 3A (see Materials and methods for details). Template DNA for the PCR was extracted from the cloned macroschizont

infected cell lines (TaA139-D4, TaA139-D6, TaH46.2, TaHBL3b) carrying different *Eco*RI RFLPs (see Fig. 1). The comparison of these DNA sequences along with the homologous regions of alleles cH and gH3.4 is shown in Fig. 3B (i); a comparison based on the translated sequences is shown in Fig. 3B (ii). A number of conclusions can be drawn. First, it is clear that there are 4 different sequences revealed by this analysis, namely A3.4 (which is identical to cH), H3.4 (which is identical to gH3.4), A4.8 and H6.0. Of considerable interest is the observation that the 3.4-kb RFLP is composed of two distinct sequences. One of these corresponds to the allele A3.4 carried by clone TaA139-D4 which is also found in the other 5 Ankara clones carrying the 3.4-kb RFLP as shown in Table 1 (sequence data not shown). The other corresponds to allele H3.4 found in clone TaH46.2 and another two Hissar clones (Table 1). Of note is that the allele gH3.4 is identical as expected to the PCR derived sequence for allele H3.4. Interestingly, the sequence of the Ta Hissar cH allele is identical to the Ta Ankara allele A3.4 over this variable region. It is also identical at the C terminus (see below). This strongly suggests that alleles A3.4 and cH are completely identical, although further sequence analysis of the complete A3.4 allele is required to confirm this. Since the cDNA is of Ta Hissar origin [13] this observation implies that the Ta Hissar stock also contains the A3.4 allele and tests to confirm this using specific oligonucleotides are underway. Finally, it is noteworthy that the H6.0 and A4.8 alleles are distinct over this same variable region and thus the RFLPs are markers of more extensive genetic polymorphism. The highest degree of variation over the region of sequence compared in Fig. 3B is observed between alleles A3.4 and H3.4 showing 78% and 60% identity at the DNA and protein levels respectively. We do not have any Ta Ankara macroschizont clones segregating for the A6.0 fragment and so comparison with the H6.0 cannot be made. Similarly information is also lacking for the H4.8 fragment and thus the comparison with A4.8 is not possible. Therefore we have data which demonstrates the existence of at least 4 alleles of SPAG-1.

We PCR amplified and cloned the series 2 fragments, of 288 bp from the 3′ end of SPAG-1, between bases 2419 and 2706. To do this we used DNA extracted from each of the macroschizont clones TaA139-D4, TaA139-D6, TaH46.2 and TaHBL3b (Materials and Methods and Fig. 3A). The DNA sequences corresponding to most of the original SR1 region were derived and are compared in Fig. 3C (i) whilst the corresponding amino acid sequences are displayed in Fig. 3C (ii) [12,13]. Three independent clones were analysed for each allele. This sequence was clearly highly conserved in all 4 variants with only a total of 3 differences at the protein level between the most divergent alleles H3.4 and H6.0 (Fig. 3C (ii)). Interestingly there is no variation observed in the region spanning the 16 amino acids to which the 1A7 neutralising epitope was localised (shown underlined in Fig. 3C (ii)) [14]. Also shown for comparison is the sequence for this region of the *Theileria parva* sporozoite antigen p67 [25]. A continuous stretch of 7 out of the 16 residues containing the 1A7 epitope are conserved between SPAG-1 and p67. Since monoclonal antibody 1A7 reacts weakly with recombinant p67 (Knight. P.A., Ph.D. Thesis, University of Glasgow, 1993), it can be concluded that the epitope is at least partially located in these shared residues.

The finding that the C terminal half of SPAG-1 is the most conserved region is useful information from the point of view of sub-unit vaccine design. This is enhanced by our recent epitope mapping studies which have located neutralising determinants recognised by bovine immune sera between residues 784 and 892 of the molecule [14]. It is also of considerable interest that the most conserved region between the *T. parva* p67 and SPAG-1 is the C terminal half of the molecule with 56% identity over amino acids 504–907 [25]. This raises the notion of a common vaccine to these two parasites designed on homologous regions in the C terminus. It is also interesting that the N-terminus is reasonably well conserved and that the region of most variation (including two large gaps) corresponds to the same area that exhibits most polymorphism within SPAG-1. These data may suggest that the N and C termini are conserved for functional reasons such as host cell invasion.

The definition of constant and variable regions will assume more relevance once more information about the immunodominant structures on SPAG-1 is accrued. When we locate T cell epitopes we will be able to assign some form of rating according to their

general usefulness depending upon whether they locate to polymorphic or conserved segments. Using our current knowledge we are expressing the C terminus as hepatitis antigen recombinants [26] and in *Salmonella* vectors [27] in order to see if we can obtain a delivery system which will give effective immunity. This work forms the basis for the rational design of all future vaccine constructs based on the SPAG-1 antigen.

Acknowledgements

We are indebted to C.G.D. Brown, G. Wilkie, L. Bell-Sakyi and other support staff at the Centre for Tropical Veterinary Medicine, University of Edinburgh, for supplying parasite material. We are grateful to Nicky Boulter for helpful discussion, Howard Baylis for advice on PCR, M. Stark for photography and Phil Hunt for preliminary PCR analysis. Excellent technical assistance was provided by Anne Megson and Susan McKellar. F.K. was supported by a University of York Studentship and P.K. by a SERC-CASE Studentship. Financial support was provided by the Wellcome Trust (A.T. and R.H.), the AFRC (R.H., I.M. and F.K.) and the Overseas Development Administration (S.W.)

References

[1] Miller, L.H., Roberts, T., Shahabuddin, M. and McCutchan, T. (1993) Analysis of sequence diversity in the *Plasmodium falciparum* merozoite surface protein-1 (MSP-1). Mol. Biochem. Parasitol. 59, 1–14.

[2] Suss, G., Matile, H., Meloen, R.H., Takacs, B. and Pink, J.R.L. (1993) Identifying polymorphic regions of the p190 protein from different *Plasmodium falciparum* strains by using specific T cells. Parasite Immunol. 15, 127–134.

[3] Good, M.F., Pombo, D., Quakyi, I., Riley, E.M., Houghten, R.A., Menon, A., Alling, D.W., Berzofsky, J.A. and Miller, L.H. (1988) Human T-cell recognition of the circumsporozoite protein of *Plasmodium falciparum*: Immunodominant T-cell domains map to the polymorphic regions of the molecule. Proc. Natl. Acad. Sci. USA. 85, 1199–1203.

[4] Good, M.F., Pombo, D., Maloy, W.L., de la Cruz, V.F., Miller, L.H. and Berzofsky, J.A. (1988) Parasite polymorphism present within minimal T-cell epitopes of *Plasmodium falciparum* circumsporozoite protein. J. Immunol. 140, 1645–1650.

[5] Guttinger, M., Caspers, P., Takacs, B., Trzeciak, A., Gillessen, D., Pink, J.R.L. and Sinigaglia, F. (1988) Human T cells recognize polymorphic and non-polymorphic regions of the *Plasmodium falciparum* circumsporozoite protein. EMBO J. 7, 2555–2558.

[6] Lockyer, M.J., Marsh, K. and Newbold, C.I. (1989) Wild isolates of *Plasmodium falciparum* show extensive polymorphism in T cell epitopes of the circumsporozoite protein. Mol. Biochem. Parasitol. 37, 275–280.

[7] Shiels, B., McDougall, C., Tait, A. and Brown, C.G.D. (1986) Antigenic diversity of *Theileria annulata* macroschizonts. Vet. Parasitol. 21, 1–10.

[8] Minami, T., Spooner, P.R., Irvin, A.D., Ocama, J.G.R., Dobbelaere, D.A.E. and Fujinaga, T. (1983) Characterisation of stocks of *Theileria parva* by monoclonal antibody profiles. Res. Vet. Sci. 35, 334–340.

[9] Shapiro, S.Z., Fujisaki, K., Morzaria, S.P., Webster, P., Fujinaga, T., Spooner, P.R. and Irvin, A.D. (1987) A life-cycle stage-specific antigen of *Theileria parva* recognised by anti-macroschizont monoclonal antibodies. Parasitology 94, 29–37.

[10] Toye, P.G., Goddeeris, B.M., Iams, K.P., Musoke, A.J. and Morrison, W.I. (1991) Characterisation of a polymorphic immunodominant molecule in sporozoites and schizonts of *Theileria parva*. Parasite Immunol. 13, 49–62.

[11] Dickson, J. and Shiels, B.R. (1993) Antigenic diversity of a major merozoite surface molecule in *Theileria annulata*. Mol. Biochem. Parasitol. 57, 55–64.

[12] Williamson, S., Tait, A., Brown, D., Walker, A., Beck, P., Shiels, B., Fletcher, J. and Hall, R. (1989) *Theileria annulata* sporozoite surface antigen expressed in *Eschericia coli* elicits neutralizing antibody. Proc. Natl. Acad. Sci. USA 86, 4639–4643.

[13] Hall, R., Hunt, P.D., Carrington, M., Simmons, D., Williamson, S., Mecham, R.P. and Tait, A. (1992) Mimicry of elastin repetitive motifs by *Theileria annulata* sporozoite surface antigen. Mol. Biochem. Parasitol. 53, 105–112.

[14] Boulter, N., Knight, P.A., Hunt, P.D., Hennessey, E.S., Katzer, F., Tait, A., Williamson, S., Brown, D., Baylis, H.A. and Hall, R. (1994) *Theileria annulata* sporozoite surface antigen (SPAG-1) contains neutralizing determinants in the C terminus. Parasite Immunol. 16, 97–104.

[15] Gill, B.S., Bhattacharyulu, Y. and Kaur, D. (1976) Immunization against bovine tropical theileriosis (*Theileria annulata* infection). Res. Vet. Sci. 21, 146–149.

[16] Schein, E., Buscher, G. and Friedhoff, K.T. (1975) Lichtmikroskopie Untersuchungen über die Entwicklung von *Theileria annulata* (Dschunkowsky und Luhs) in *Hyalomma anatolicum excavatum* (Koch, 1844). Z. Parasitenkd. 48, 123–136.

[17] Brown, C.G.D. (1987) Theileriidae. In: *In Vitro* Methods for Parasite Cultivation (Taylor, A.E.R. and Baker, J.R., eds.), pp 230–250. Academic Press, London.

[18] Theilen, G.H., Rush, J.D., Nelson-Rees, W.A., Dungworth, D.L., Munn, R.J. and Switzer, J.W. (1968) Bovine leukaemia: establishment and morphological characterisation of continuous cell suspension culture BL-1. J. Natl. Cancer Inst. 40, 737–749.

[19] Baylis, H.A., Megson, A., Brown, C.G.D., Wilkie, G.F. and

Hall, F.R. (1992) *Theileria annulata*-infected cells produce abundant proteases whose activity is reduced by long-term cell culture. Parasitol. 105, 417–423.
[20] Sambrook, J., Fritsch, E. and Maniatis, T. (1989) Molecular Cloning. A Laboratory Manual. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
[21] Kaiser, K. and Murray, N.E. (1986) The use of phage *lambda* replacement vectors in the construction of representative genomic DNA libraries, In: DNA Cloning. A Practical Approach. Volume I, (Glover D.M., ed.). IRL Press, Oxford.
[22] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci USA 74, 5463–5467.
[23] Devereux, J. (1989) Genetics computer group sequence analysis package. Nucleic Acids Res. 12, 387–395.
[24] Wilkie, G.M., Melrose, T.R., Fletcher, J.D. and Brown C.G.D. (1986) GPI isoenzyme identification of cloned populations of *Theileria annulata* (Ankara) after passage through the vertebrate and invertebrate hosts. Trans. R. Soc. Trop. Med. Hyg. 80, 987–988.
[25] Nene, V., Iams, K.P., Gobright, E. and Musoke, A.J. (1992) Characterisation of a gene encoding a candidate vaccine antigen of *Theileria parva* sporozoites. Mol. Biochem. Parasitol. 51, 17–28.
[26] Brown, A.L., Francis, M.J., Hastings, G.Z., Parry, N.R., Barnett, P.V., Rowlands, D.J. and Clarke, B.E. (1991) Foreign epitopes in immunodominant regions of hepatitis B core particles are highly immunogenic and conformationally restricted. Vaccine 9, 595–601.
[27] Curtiss III, R., Kelly, S.M., Gulig, P.A. and Nakayama, K. (1989) Selective delivery of antigens by recombinant bacteria. Curr. Topics Microbiol. Immunol. 146, 35–49.

SPOROZOITE SURFACE MOLECULES OF *THEILERIA ANNULATA*: VARIATION AND FUNCTION

F. Katzer [a], P. Knight [b], S. Williamson [c], I. Morrison [d], R. Hall [a]

[a]Department of Biology, University of York, York, YO1 5DD, UK [b] Welcome Unit for Molecular Parasitology, University of Glasgow, Bearsden Rd. Glasgow, G61 1QH, UK [c]Centre for Tropical Veterinary Medicine, University of Edinburgh, Easter Bush, Roslin, Midlothian, Edinburgh, EH25 9RG, UK [d]AFRC Institute for Animal Health, Compton, Nr. Reading, Berkshire, RG16 0NN, UK

ABSTRACT :

SPAG-1 and SPAG-2, *Theileria annulata* sporozoite surface antigens, are vaccine candidates. Data is presented showing that SPAG-1 is a single copy gene. We compare a full length genomic copy of the SPAG-1 gene with a previously published SPAG-1 cDNA sequence demonstrating a high degree of polymorphism. We infer that these sequences represent two distinct allelic SPAG-1 variants. The deduced polypeptides show an overall identity of 92% with the most variable stretch (60% identity) occurring towards the middle of the molecule. The N and C termini are more conserved with identities of 92% and 97% respectively. Evidence of two further SPAG-1 alleles was obtained from PCR based sequences using macroschizont clones. In total we have shown the existence of at least 4 highly polymorphic SPAG-1 alleles. Flowcytometry studies using recombinant biotinylated SPAG-1 and SPAG-2 have been conducted. These have shown that SPAG-1 binds specifically to subpopulations of T cells, B cells and monocytes, while SPAG-2 binds specifically to a subpopulation of monocytes but not to B cells nor T cells.

INTRODUCTION :

Antigenic polymorphism is a common theme amongst infectious agents and has serious implications for vaccine development. Therefore it is essential when considering molecules as candidates for the inclusion in a sub-unit vaccine that an understanding of the nature and the degree of variation in their structure is achieved. Ideally the aim is to define an immunologically relevant region which is not polymorphic. It is also important to understand the function of the surface antigens of intracellular parasites, as this might reveal new strategies for intervention in, for example, the process of host cell recognition and/or invasion.

Three sporozoite surface antigens of *Theileria annulata* have been identified so far and their genes have been at least partly cloned and sequenced. SPAG-1 was the first of these antigens which was studied. It is defined by a neutralising monoclonal antibody (1A7) as a vaccine candidate (Williamson et al., 1989). To date we have sequenced a full length cDNA clone and a genomic clone coding for SPAG-1. The cDNA clone encoding the complete polypeptide was

expressed and, we have located the neutralising epitope defined by 1A7 within a linear sequence of 16 amino acids at the C terminus (Hall 1992 et al., Boulter et al., 1994). The C terminal 100 amino acids was expressed in the e1 loop of the Hepatitis core antigen and used in a vaccination trial (Boulter et al., accompanying manuscript). SPAG-2 and SPAG-3 were both identified via the neutralising monoclonal antibody 4B11. A cDNA coding for SPAG-2 has been isolated and sequenced. The C terminal 330 amino acids were expressed as a GST fusion protein and used in a vaccination trial (Knight, pers. comm.).

The present study investigates the extent of allelic variants of SPAG-1 between and within two geographically distinct isolates, as well as the ability of expressed recombinant SPAG-1 and SPAG-2 to bind to specific subpopulations of bovine peripheral blood mononuclear cells.

RESULTS :

Previously we had identified three *Eco*RI fragments of 3.4, 4.8 and 6.0 kb released from piroplasm DNA that hybridize with the 330 bp insert derived from the original SPAG-1 clone Tgt11-SR1 (Williamson et al., 1989). These data are confirmed by our current results except that the original faint band of 4.8kb in the Ta Hissar stock is no longer present (Fig. 1, lanes 1 and 2). This observation suggests that the multiple bands result from mixtures, of at least, three parasite populations. By Southern blot analysis of DNA extracted from 12 macroschizont-infected cloned cell lines we observed that each clone contains only one of the three RFLP types observed in piroplasm DNA (Fig. 1). We conclude that each of the 3 RFLPs behave as alternative forms of a single copy gene and thus we adopt the term allele.

A complete genomic copy of the SPAG-1 gene was sequenced and we defined this the 3.4kb Ta Hisar allele, as it contains the 3.4-kb *Eco*R1 RFLP seen in Fig. 1, and give it the designation of gH3.4. When compared to the amino acid sequence derived from the published cDNA sequence, extensive variation is apparent (Fig. 2) therefore this cDNA is clearly not the product of gH3.4 and we designate it as a separate allele, cH. Overall, the two protein sequences are 92% identical, with the C terminal half of the molecule being most homologous (residues 504-907, 97% identity). The N terminal segment from residue 1-212 is also highly similar (92%). The polymorphism is greater between residues 213-504 and is most extensive between amino acid 284 and 369 showing an identity of only 60%.

To further study the extent of SPAG-1 polymorphism we sequenced the most variable DNA region corresponding to amino acids 284 to 369 from PCR products using DNA from cloned macroschizont infected cell lines as templates. The cell lines used were the same as shown in Fig. 1. The comparison of the predicted amino acid sequences along with the same regions derived from alleles cH and gH3.4 are shown in Fig. 3a. From this comparison it can clearly be seen that there are 4 distinct SPAG-1 alleles namely A3.4 (which is identical to cH), H3.4 (which is identical to gH3.4), A4.8 and H6.0. Of considerable interest is the observation that the 3.4kb RFLP is composed of two distinct sequences.

The sequence comparison of the same macroschizont cell line clones across the C terminal region also reveals the same pattern of 4 alleles but it is also apparent that this region is highly conserved across the 4 alleles (Fig. 3b). Interestingly there is no variation observed in the region spanning the 16 amino acids to which the 1A7 epitope was located (shown underlined in Fig. 3b (Boulter et al., 1994)). a continuous stretch of 7 out of the 16 residues containing the 1A7 epitope are conserved between SPAG-1 and p67. Since monoclonal antibody 1A7 reacts weakly with recombinant p67 (Knight 1993), it can be concluded that the epitope is at least partly located in these shared residues.

The definition of constant and variable regions will assume more relevance for sub-unit vaccine development once more information about immunodominant structures of SPAG-1 is accrued.

Two colour flowcytometry using biotinylated recombinant SPAG-1, SPAG-2 and monoclonal antibodies, specific for a panel of bovine peripheral blood mononuclear cells surface antigens, was conducted. SPAG-1 binds to 76% of CD2 positive cells, 44% of gamma/delta T cells, 19% of B cells and 27% of monocytes while SPAG-2 does not bind to T cells nor B cells but binds specifically to a subpopulation of monocytes. Therefore SPAG-1 binds to a variety of cells including T cells which are not readily infected by *T. annulata* . Interestingly SPAG-1 binds to a subpopulation of B cells and monocytes (targets of *T. annulata* (Spooner et al.,1989)) as well as to CD2 positive cells. The CD2 molecule is expressed on *T. annulata* infected cells indicating that these cells might also be targets for *T. annulata* infection (Howard et al., 1993). SPAG-2 shows a high specificity to monocytes (targets for *T. annulata* infection (Spooner et al., 1989)). Neither SPAG-1 nor SPAG-2 in their own rights could explain the specificity found in the process of host cell recognition and invasion although the SPAG-2 data is very encouraging. It is likely that SPAG-1 and SPAG-2 are involved in this process to some extent but that there are other molecules which also interact with the host cell to facilitate the process of recognition and invasion.

ACKNOWLEDGEMENTS :

We are indebted to C.G.D. Brown , G. Wilkie, L. Bell-Sakyi and other support staff at the Centre for Tropical Veterinary Medicine, University of Edinburgh, for supplying parasite material. We are grateful to Phil Hunt for preliminary PCR analysis and Paul Sopp and Katherine Griffith for excellent help with the flowcytometry analysis. Figures 1, 2 and 3 have been published in Katzer et al. (1994). Financial support was provided by the University of York (F.K. and R.H.), the SERC (P.K.), the Welcome Trust (R.H.), the AFRC (R.H., I.M. and F.K.) and the Overseas Development Administration (S.W.).

REFERENCES :

Boulter, N., Knight, P.A., Hunt, P.D., Hennessey, E.S., Katzer, F., Tait, A., Williamson, S., Brown, D., Baylis, H.A. and Hall, R. (1994) Parasite Immunol. 16, 97-104.

Katzer, F., Carrington, M., Knight, P., Williamson, S., Tait, A., Morrison, I. and Hall, R.

(1994) Molecular and Biochemical Parasitology 67, 1-10.
Knight, P.A. (1993) Ph.D. Thesis, University of Glasgow. Hall, R., Hunt, P.D., Carrington, M., Simmons, D., Williamson, S., Mecham, R.P. and Tait, A. (1992) Mol. Biochem. Parasitol. 53, 105-112.
Howard, C.J., Sopp, P., Preston, P.M., Jackson, L.A. and Brown, C.G.D. (1993) Vet. Immunol. Immunopathol. 39, 275-282.
Spooner, R.L., Innes, E.A., Glass, E.J. and Brown, C.G.D. (1989) Immunology 66, 284-288.
Williamson, S., Tait, A., Brown, C.G.D., Walker, A., Beck, P., Shiels, B., Fletcher, J. and Hall, R. (1989) Proc. Natl. Acad. Sci. USA 86, 4639-4643.

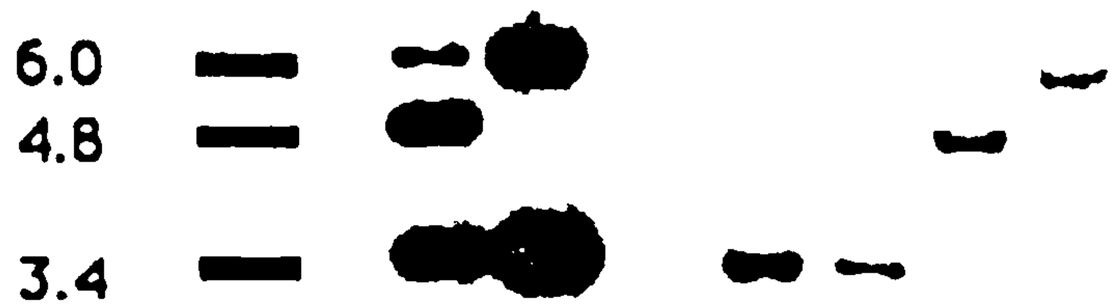


Fig. 1. Southern Blot analysis demonstrating the clonal segregation of the SPAG-1 associated RFLPs. Genomic DNA was digested with *Eco*R1 and the blot was probed with the SPAG-1 insert derived from Tgt11-SR1 (Hall et al.,1992).Lanes 1 and 2 contain piroplasm DNA extracted from Ta Ankara and Ta Hissar stocks respectively. Lanes 3-6 contain DNA of the following cell lines: Uninfected BL3 (lane 3), clone TaA 139-D4 (lane 4), clone TaH 46.2 (lane 5), clone TaA 139-D6 (lane 6) and clone TaHBL3b (lane 7). The bars mark the sizes of the RFLPs in kb.

COMPARISON OF SPAG-1 ALLELES gH 3.4 AND cH

OVERALL IDENTITY 92 %

Fig. 2. Schematic representation of the comparison of SPAG-1 protein sequences derived from the DNA sequences of alleles gH3.4 and cH. The figures above the blocks represent the % identity over each segment calculated by scoring a gap as one change. The numbers below designate the amino acid residues at the boundaries of the sequence block and are based on the cH molecule.

```
cH          PGKPPGVGAGVMPGVGVRAQGGVIIGAPGVAGV...P......GGKPGQP.VSQELELKSDTEINESGSSSEGEDDD.DEEEEEENKSTSSKGAGG
A 3.4       ------------------------------...-......-------.------------------------.----------
A 4.8       ---H-------------------------A---...-......---A-KSNK-SDP--.LGD-SDD-----G-D-V-N-D----DD----------
gH 3.4      --------.........VP-A-EP-L--A---...-......---A-KSNK-SDP--DFGD-SDG---G-A-D---N-D----DD---------A
H 3.4       --------.........VP-A-EP-L--A---...-......---A-KSNK-SDP--DFGD-SDG---G-A-D---N-D----DD---------A
H 6.0       ---H-A-----------QAQ--L------A-R.SGL-RRSRCT-R-A-KSKK-LDP-SDFGD-SDD-----G-D---.----D------------
```

Figure 3a.

```
cH          IGDLSDPSSRTPNAKPAELGPSLVIQNVPSDPSKVTPTQPSNLPQVPTTGPGNGTDGTTTGPGGNGEGGKDLKEGEKKEGLFQKIQNKLLGSGFEV
A 3.4       ----------------------------------------------------------------------------------------------------
A 4.2       --------------------------------S---------------------------------------------------------------------
gH 3.4      -------N------------------------------------------------------------------------------------------------
H 3.4       -------N------------------------------------------------------------------------------------------------
H 6.0       --------------------------------S---------------S-----------------------------------------------------
p67         L-------G-SS.ERQPS-------TDGQAG-TI-S--G-TIAAGGEQPPSAPN..--A---A-TQPEG.....-------I--L-K---------
```

Figure 3b.

ig. 3. Comparison of SPAG-1 alleles over the most polymorphic region and the constant C erminal region. (a) Comparison of the SPAG-1 predicted amino acid sequence over the most olymorphic region (residues 284-369). The PCR based sequences are derived from the same loned macroschizont infected cell lines as shown in Fig. 1. In addition the same region from lleles cH and gH3.4 is presented for further comparison. (b) As in a except that the comparison s between sequences derived from the C terminal constant region (amino acid 806-902). In ddition the analogous sequence of the *T. parva* p67 antigen is presented for comparison. The 6 amino acids containing the 1A7 epitope (Boulter et al., 1994) are underlined. All alignments re made to the cH sequence and - represents identity, . represents a gap and any substitutions re shown by the relevant letter.