CHARACTERISATION OF TWO MEMBERS OF A MACROSCHIZONT GENE FAMILY, *TASHAT1* AND *TASHAT3*, FROM *THEILERIA ANNULATA*

by

Rowena F. Stern, B.Sc. (Hons).

A thesis submitted for the degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF GLASGOW

MARCH 2003

Department of Veterinary Parasitology,
University of Glasgow

Author's declaration

The experimental work described in this thesis was carried out in the Department of Veterinary Parasitology, between January 1997 to June 1999 and also in June and October 2001. Except where stated otherwise, the results are my own.

Abstract

Theileria annulata is a protozoan parasite of cattle, that causes the disease tropical theileriosis throughout sub-tropical regions of the Old World. *Theileria* parasites have the ability to immortalise the host leukocyte they infect causing clonal expansion and dissemination of infected leukocytes throughout the host. This property has allowed the development of an *in vitro* system for the culture of bovine cells infected by the macroschizont stage of the parasite. In addition, differentiation of the parasite towards the next life cycle stage, the merozoite, can be induced in culture. The signals that cause the macroschizont to differentiate into merozoites are not fully understood, although it is known that this event is associated with a major elevation in merozoite gene expression (Shiels *et al.*, 1994).

Recently a small family of parasite genes that are negatively regulated early during differentiation to the merozoite were identified. One member, known as *TashAT2* contained predicted AT hook DNA binding motifs and was shown to be localised to the host cell nucleus. It has been postulated that the TashAT2 polypeptide may play a role in the regulation of macroschizont or modulation of host cell gene expression (Swan *et al.*, 1999). The focus of this project was to characterise *TashAT1*, a second member of the *TashAT* gene family. To this end, the *TashAT1* gene was sub-cloned and sequenced and mapped to a region of the genome containing *TashAT2* and a third *TashAT* gene, *TashAT3*. The 1.4kb open reading frame of *TashAT1* was virtually identical to the five prime end of *TashAT3*, indicating that *TashAT1* or *TashAT3* (*TashAT1/3*) were derived from a recent duplication event. The predicted amino acid sequence of *TashAT1/3* contained four AT hook motifs, a nuclear localisation signal and a signal sequence.

Northern blot analysis revealed that TashAT1, TashAT2 and TashAT3 mRNA were down regulated early, during differentiation to the merozoite *in vitro*. However, no down regulation was observed for any of the TashAT transcripts in a cell line that was severely attenuated with respect to parasite differentiation. Sequence analysis of the upstream regions of *TashAT1/3* identified a motif element (TashUM) located 43bp upstream of the putative transcription start site of *TashAT1/3* that was highly related to a sequence upstream of *TashAT2* and another, unrelated macroschizont gene, *Tash1*. Preliminary electromobility band shift analysis of TashUM revealed that it bound to a factor found in host and parasite enriched nuclear extract, which appeared to decrease in abundance as the parasite differentiated towards merogony.

Antisera generated against a region of *TashAT1* failed to recognise a TashAT1 polypeptide by Western blot analysis. However, a 180kDa polypeptide that was down regulated with respect to merogony and co-localised to the host nucleus was specifically recognised. The detected polypeptide was identified as TashAT3 on the basis of size, sequence identity and predicted expression profile. Immunofluorescence analysis showed that the anti-TashAT1 antisera reacted against both the host nucleus and parasite. This reactivity was lost as the parasite differentiated to the merozoite. The host reactivity was probably due to recognition of TashAT3, while it could not be concluded that the parasite reactivity was directed against TashAT1.

Taken together, the results indicated that TashAT3 and possibly TashAT1 are additional candidates for parasite encoded factors that are translocated to the host nucleus, bind to DNA and alter host cell gene expression. This modulation of gene expression could directly or indirectly alter the phenotype of the host cell and be involved in parasite dependent regulation of leukocyte cell division.

Acknowledgement

I would like to thank all the members of the Veterinary Parasitology lab, University of Glasgow, including Colin Chapman and Andy Tait for preparation of antisera and Sue McKellar, Nazreen Aslam, Kirsten Phillips, Aiobhin McDonell, Frank Katzer for their advice and technical assistance. My appreciation also goes to David Swan, Laura Stadler, Sue McKellar and Brian Shiels for access to unpublished data and use of their antisera. I am grateful to all the members of the Wellcome Unit of Molecular Parasitology for their advice and use of sequencing facilities and to Keith Vass of the CRC Beatson Institute (Glasgow) for assistance and use of the densitometric scanner. Thanks also to Allan May of the University of Glasgow, for preparation of many of the photographs presented in this thesis and the BBSRC for providing funding for part of this project.

On a personal level, I would like to thank Caroline Dando, Alistair McGregor, Julie Nybo, Colin Chapman, Chris Oura and Mark Fox, Afshan Fairley for their help, generosity and support. A special thanks goes to my sister, Lorna, for her continued support and guidance, especially during my write up. Finally, the last people I would like to thank are my assigned supervisors Brian Shiels and Andy Tait and my unofficial supervisor David Swan.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abbreviations

aa	amino acid
approx.	approximately
bp	base pair
cDNA	complementary DNA
chg	charge
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
DNA	Deoxyribonucleic acid
EMSA	Electromobility Band Shift Assay
GST	glutathione S-transferase
Init	initial
kb	kilo bases
kDa	kilo Dalton
mAb	monoclonal antibody
NLS	Nuclear Localisation Signal
nuc	nucleotide
pGex	Pgex2TK
pos	positive
RNA	ribonucleic acid
sbjct	subject
Seq	sequence
sn rRNA	small nuclear ribosomal RNA
spp.	species
TashAT1/3	TashAT1 or TashAT3
transcr.	transcription

1. General Introduction

1.1 Introduction

Tropical Theileriosis or Mediterranean Coast Fever is a debilitating and frequently fatal disease of cattle and Asiatic buffalo caused by *Theileria annulata*, a tick borne, protozoan parasite. Originally discovered by Dschunkowsky and Luhs (1904, cited by Norval *et al.*, 1992), *Theileria annulata* is a member of the apicomplexan phylum, or subphylum in some classifications, which includes *Plasmodium, Toxoplasma, Eimeria* and *Babesia* (Levine, 1988), so called because all of its members possess an "apical complex". There are thirty-four species of *Theileria* which mainly infect cattle and ruminants, but are also known to infect camels and horses. The most pathogenic species are *Theileria annulata* and to a greater extent, *Theileria parva*, the causative agent of East Coast Fever. Both *T. annulata* and *T. parva* pose a severe economic burden by constraint of livestock productivity in countries where these species are endemic.

Since its discovery, the classification of *Theileria annulata* has often been changed by taxonomists because of the initial poor understanding of the intracellular stages of this parasite. Dschunkowsky and Luhs initially named *T. annulata Piroplasma annulatum*, from the pear shaped bodies observed, known as piroplasms, which were similar in general form to *Babesia* piroplasms (Norval *et al.*, 1992). Later electron micrograph studies by Friedhoff and Schlotyseck (1968), Büttner (1967, cited by Norval *et al.*, 1992) and Jarrett and Brocklesby (1966) revealed that *Theileria spp* belonged to the phylum Apicomplexa because they contained some organelles that formed an apical complex. A typical apical complex consists of a polar ring and rhoptry, but may also include a conoid complex, subpellicular tubules and micronemes and are formed during the budding process of the parasite at some stage in its life cycle (Jura *et al.*, 1983; Shaw and Tilney, 1992; Schein *et al.*, 1978). Levine (1973) discovered that *Theileria* contained a reduced apical complex, restricted to rhoptries and the polar ring. *Piroplasma annulata* was placed into the *Theileria* genus with the discovery that this organism contained a schizont stage (Wenyon, 1926, cited by Norval *et al.*, 1992).

The similarity between *Babesia* and *Theileria* within the sub-phylum Apicomplexa led to suggestions that *Theileria* and *Babesia* were of the same genus. However, Levine (1973) observed that *Theileria* invades leukocytes, where it transforms into an intracellular

schizont stage, whereas *Babesia* exclusively infects erythrocytes (Mehlhorn and Schein, 1984). Levine (1973) also showed there were structural differences between the apical complex of the two species. Nevertheless, controversy over the speciation of some *Babesia spp* remains since two species, namely *Babesia microti* and *Babesia Equi*, both have a macroschizont stage in equine and rodent lymphoid cells respectively (Schein *et al.*,1981; Moltmann *et al.*, 1983; Mehlhorn and Schein, 1984). Moreover, neither of these parasites are trans-ovarially transmitted, a common feature of *Babesia* (Norval *et al.*, 1992). Current classification of *B. microti* and *B. equi* based on β-tubulin and rRNA analysis have not resolved this issue and it is possible that these species could form a new genus. A recent classification of the *Theileria* species by Irvin (1987) is described below (see Table 1.1):

Sub Kingdom	Protozoa, single cell eukaryotes
Phylum	Apicomplexa: apical complex present at least in some stages; reproduce sexually by syngamy
Class	Sporozoea; sporogonic stage producing sporozoites
Sub Class	Piroplasmia; piroform, rod shaped or amoeboid; parasites in erythrocytes and some other cells
Order	Piroplasmida; asexual and sexual reproduction; ticks are vectors.
Family	Theileriidae; schizont stages in lymphocytes
Genus	*Theileria*; piroplasm stages in erythrocytes lacks pigment
Species	*Theileria annulata*

Table 1.1 Classification of *Theileria annulata* (taken from Irvin, 1987).

Theileria annulata is transmitted by the three main species of the ixodid tick *Hyalomma (H. anatolicum anatolicum, H. excavatum* and *H. detritum). T. annulata* causes the syndrome known as tropical or Mediterranean theileriosis; infections occur in the asiatic buffalo (*Bos taurus, Bos Indicus*) and the water buffalo (*Bubalis bubalis*) but the parasite is more pathogenic in exotic, taurine cattle (Uilenberg, 1981; Robinson, 1982). Five other species of *Theileria* infect cattle, moreover, there are many more *Theileria spp* that infect bovines and bovids (reviewed by Mehlhorn and Schein, 1984). *Theileria parva* is the most pathogenic of the *Theileria spp* and is transmitted by ticks of the genus *Rhipicephalus. T. parva* causes East Coast Fever (ECF) and infects the Cape buffalo, Waterbuck, Asiatic buffalo and cattle, but is particularly pathogenic to the latter (Maloo *et al.*, 2001; Lawrence *et al.*, 1983; Stagg *et al.*, 1983). Also of significance is *Theileria sergenti,* the major infective *Theileria spp* of cattle in East Asia, although this species usually causes milder infections than *T. parva* and *T. annulata*. The remaining three, less pathogenic or benign *Theileria spp* are *Theileria mutans* (Young *et al.*, 1977); *Theileria taurotragi* and *Theileria velifera* (Reviewed by Uilenberg, 1981), which exist mainly in Africa. However, the focus of this study will be on *T. annulata,* with other *Theileria* species mentioned as appropriate.

1.2 Distribution and Economic Impact of *Theileria annulata*

Theileria annulata is distributed across the tropical and subtropical climatic zones of the old world (see Fig. 1.1). In the west it occurs in Portugal, Spain and Morocco and the range extends eastwards along the Mediterranean coasts of Europe and North Africa. This parasite is also found in South-East Europe, and is spread across the near East, middle East, southern Russia , Siberia, India, China and the Far East (reviewed by Robinson, 1982).

The distribution and seasonal occurrence of tropical theileriosis is closely related to the distribution of the transmitting *Hyalomma* vector and its life cycle (reviewed by Robinson, 1982). These ticks undergo a 2-host cycle, whereby the nymph and larvae may feed on the first host and the adult on a second host, or a 3-host cycle, where all three stages feed on a fresh host. Studies by Flach and Ouhelli (1992) have shown that in Morocco, disease transmission occurs entirely by ticks feeding on cattle in the summer; in the winter months the parasite maintains itself as piroplasms in cattle or as zygotes in overwintered ticks. Further infection of cattle then occurs when infected nymphs develop into adults and feed on the cattle in the spring Flach and Ouhelli (1992). Pipano (1989b) also showed that sporadic cases occur throughout the year in Israel and North Africa. *Hyalomma* ticks hide in cattle barns where they feed and develop on cattle (Pipano, 1976): the warm and humid conditions have been shown to increase infectivity (Samish, 1977). However, extremes of temperature have a detrimental effect on the development and infectivity of *T. parva* and *Babesia* and is likely to affect *T. annulata* in a similar manner (Friedhoff, 1988; Lewis, 1950; Lewis and Fotheringham, 1941; Young and Leitch, 1981). This may explain why this disease is limited to the tropical and sub-tropical zones (Robinson, 1982). Alternatively, the geographical distribution of this disease may be linked to tick survival since the related *Babesia* species is found in Northern Europe (Homer *et al.*, 2000). Other factors which influence infectivity rates are male to female tick ratio, as females have more "e" type salivary cells than males, which are targeted by *Theileria* kinetes (Young *et al.*, 1980). Finally, the susceptibility of cattle breeds to the disease is a major factor in the spread of tropical theileriosis. In Africa, farming practices have increased the distribution of the disease, by the introduction of exotic, non resistant cattle breeds to endemic areas while limiting calf exposure to ticks, thereby preventing the development of resistance to subsequent parasite challenge (Robinson, 1982; Fivaz *et al.*, 1989; Jahnke, 1982, cited by Norval *et al.*, 1992).

The true economic impact of disease caused by *T. annulata* is difficult to assess but it has been estimated that approximately 200-250 million cattle are at risk from tropical theileriosis, mainly in developing countries (Dyer and Tait, 1987). However, the number of cattle that have been infected by *T. annulata* is probably underestimated, due to lack of accurate data (Dyer and Tait, 1987). Affected cattle suffer from lymph node hyperplasia, anaemia and death while chronic disease generates weight loss, infertility and low milk yields. The disease thus results in reduced productivity, which poses a severe economic constraint in developing countries where cattle are an important source of nutrition, fertiliser and tractive power. In addition tropical theileriosis halts the introduction of high milk yield livestock, particularly in crossbreed, exotic or imported taurine breeds where mortality rates are between 40-60%, hindering the improvement of local breeds (Brown, 1990). By contrast, the indigenous low milk yield cattle have a mortality rate of approximately 5%, and these cases are mainly restricted to calves (reviewed by Dyer and Tait, 1987). In India, where crossbreeding programmes have been established to increase milk productivity these losses were estimated to be U.S$800 million or 10% of the gross national product (Brown, 1990; Devandra, 1995).

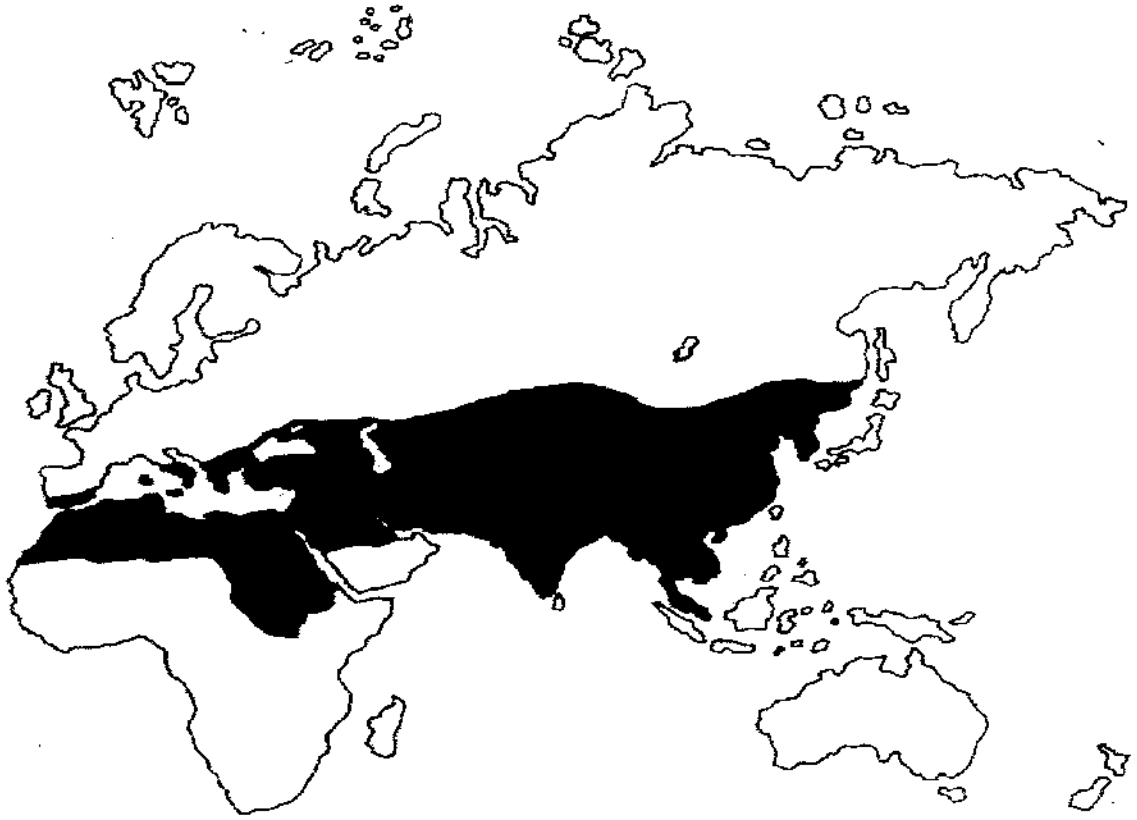


Fig. 1.1: Geographical Distribution of *Theileria* (obtained from Dyer and Tait, 1987).

1.3 The Life cycle of *Theileria*

Theileria annulata, like other protozoan parasites, has evolved a life cycle that involves differentiation through a number of different life cycle forms that occur either in the bovine host or the tick vector. Differentiation enables parasites to enter, survive and reproduce within the host cell and is fundamental to the long-term maintenance of the parasite populations and their transmission. During the life cycle of *T. annulata* (Fig. 1.2), the parasite undergoes sexual reproduction within the tick gut, and travels to the tick salivary glands, where its undergoes asexual reproduction, known as sporogony. In the bovine host, two further stages of asexual reproduction occur in the white blood cells and are known as schizogony and merogony. Liberated merozoites invade erythrocytes resulting in the production of piroplasms. Infected erythrocytes are taken up by feeding ticks, thus completing the life cycle of the parasite in the bovine. The individual stages of the life cycle of *T. annulata* are described in more detail below.

1.3.1 Bovine Host

In the bovine, the developmental cycle of *T. annulata* is initiated when, an adult tick (usually) takes its first bloodmeal and inoculates sporozoites derived from the salivary glands into the bloodstream of the host. Shaw *et al.*, (1991) reported that sporozoites were ovoid shaped, approximately 0.9μm long and 0.8μm wide, and surrounded by a unit membrane, which enclosed an apical complex consisting of several rhoptries and a polar ring. Micronemes, containing enzymes were also observed in sporozoites. *In vitro* experiments demonstrated that *T. annulata* sporozoites target Major Histocompatibility complex (MHC) class II positive cells such as monocytes, macrophages or cells of a B type lineage (Glass *et al.*, 1989; Spooner *et al.*, 1989; Campbell *et al.*, 1994; Forsyth *et al.*, 1997, 1999). However, there was less evidence of B cell infection *in vivo* as certain B cell markers were lost upon infection with *T. annulata* (Baldwin *et al.*, 1988). By contrast *T. parva* infects MIIC I positive α/β T cells (Baldwin *et al.*, 1988; Morrison *et al.*, 1996).

Sporozoites contact and penetrate the monocyte in as little as 5 minutes, but usually within 15 minutes, post infection, and infect on average between 10 and 20 % of lymphocytes: within a 60 minute interval, up to 15 sporozoites could infect each lymphocyte by *in vitro* studies (Jura *et al.*, 1983; Shaw *et al.*, 1991). Two subpopulations of sporozoites were found that were internalised by receptor mediated endocytosis either at the basal end or evenly distributed around the lymphocyte (Jura *et al.*, 1983). Curiously, Jura *et al.* (1983)

found that *T. annulata* sporozoites do not attach to the host cell membrane by their apical complex, unlike other apicomplexans such as *Plasmodium, Toxoplasma, Sarcocystis* and *Eimeria* (Mehlhorn and Schein, 1984). In addition, *in vitro* studies with *T. parva* sporozoites showed that attachment involved the progressive binding of ligands on the parasite to receptors on the host plasma membrane to form a close attachment. This "zippering" action continues until the entire sporozoite is enclosed and internalised by the host plasma membrane (Fawcett *et al.*, 1984; Jura *et al.*, 1983). Sporozoite entry was found to be passive, supported by the fact that internalisation can occur at 2°C (Fawcett *et al.*, 1986) and the host cell fails to develop pseudopodia (Fawcett *et al.*, 1984). In the case of *T. parva*, sporozoite internalisation has been shown to be mediated by MHC class I receptors on the host cell surface (Shaw *et al.*, 1995), however, it is not known if the same mechanism applies to *T. annulata* sporozoites.

Within thirty minutes, the sporozoites undergo dedifferentiation: the micronemes and rhoptries discharge their contents, a 10-15nm thick "fuzzy" layer appears on the sporozoite surface and the host cell membrane is dissolved. In contrast to *Plasmodium* and *Toxoplasma*, the use of the apicomplex to destroy the host cell membrane after internalisation means that *Theileria* is not encapsulated inside a parasitophorous vacuole (Mehlhorn and Schein, 1984). Thus, host lysozymes are unable to fuse with the vacuol membrane and discharge their contents, so enabling the parasite to escape destruction in this way (Fawcett *et al.*, 1984). The final invasive steps involve the formation of an orderly, hammock-like network of host cell derived microtubules around the outer layer of the former, dissolved host membrane that previously surrounded the sporozoite. (Shaw *et al.*, 1991; Fawcett *et al.*, 1982; Jura *et al.*, 1983; Williams and Dobbelaere, 1993). Experiments by Shaw *et al.* (1991) revealed that events subsequent to the entry of the sporozoite into the host cell are energy dependent, requiring the participation of live sporozoites and host cells.

Once the sporozoite has successfully established in the host, it transforms into a transitory uninucleate, motile growing stage known as the trophozoite within the next two hours (Jura *et al.*, 1983). Within three hours some parasites have developed a cytosome and by 18-24 hours, bundles of intranuclear, microtubules and spindle pole bodies have appeared, forming an acentric microtubular spindle. Soon after, binary fission occurs to form a binucleate schizont, the first asexual multiplication to occur in the bovine host. By 72 hours the trophozoite has grown to 2μm in diameter and undergone more nuclear divisions to form a multinucleate schizont or macroschizont (Irvin *et al.*, 1982; Jura *et al.*, 1983;

Fawcett *et al.*, 1982, 1984). Multiple infection by sporozoites results in the development of multiple macroschizonts, but these cells die out early (Stagg *et al.*, 1981).

Shaw and Tilney (1992) observed that the average multinucleate macroschizont contains 15-20 nuclei within its syncitium and is separated from the host cytoplasm by a cell membrane. The nuclei were surrounded by a nuclear membrane with pores slightly smaller than the nuclear pores of the host. Very few organelles were visible in the macroschizont other than numerous, free ribosomes and some polysomes, randomly scattered mitochondria and some membrane bound vesicles. The smooth and rough endoplasmic reticulum and golgi apparatus appeared to be absent in the schizont cytoplasm by electron microscopy (Shaw and Tilney, 1992). The ribosomes observed in the schizont were smaller than those of the host although during the onset of merogony, both host and parasite ribosomes were comparable in size. Shaw and Tilney (1992) had postulated that the ribosomes of the schizont may not be fully assembled and, if so, the parasite may be dependent on the host for all its metabolic and synthetic requirements.

Amongst the apicomplexan parasites, *Theileria* macroschizonts have the unique ability to transform the host cell, inducing lymphoblastogenesis and clonal expansion of infected cells. Experiments have shown this process is reversible by treating infected cells with the anti-theileriacidal drug, buparvaquone, demonstrating that the parasite is responsible for this transformation (Dobbelaere *et al.*, 1988). Studies by Tsur and Adler (1963) and Brown *et al.*, (1973) showed that *Theileria* infected cells can be cultured *in vitro* indefinitely. Once transformed, the cell becomes enlarged with cytoplasm and the growing macroschizont begins nuclear division. Host cell division occurs at rapid, regular intervals resulting in up to a 10 fold increase in infected cells over a period of three days or less *in vivo* (Jarrett *et al.*, 1969; Radley *et al.*, 1974; Irvin *et al.*, 1982). During the early stages of infection and within infected cultured cells, the schizonts have been shown to divide in synchrony with host cell division *in vitro* due to the foreshortened or absent G_2 phase of the parasite (Hulliger *et al.*, 1964; Irvin *et al.*, 1982). Studies by Hulliger *et al.* (1964), Stagg *et al.* (1980), Vickerman and Irvin (1981) confirmed that synchronous host and parasite cellular division was achieved by the macroschizont attaching to the host mitotic spindle apparatus in prophase and metaphase. After host division, each daughter cell receives at least one schizont (Dyer and Tait, 1987) and recent studies by Kinnaird *et al.* (1996, 2001) have identified two cyclin dependent kinase genes TaCRK2 and TaCRK3 in *T. annulata* and *T. parva* that are likely to play a central role in all stages of parasite nuclear division.

In vivo, the macroschizont continues to enlarge (up to 6-10μm in diameter) within the host cell until some undefined signal begins the process of merozoite formation which occurs in a proportion of cells as observed by Shaw and Tilney, (1992), shown in Fig. 1.3. Here, the enlarged macroschizont differentiates into a transitory form known as a microschizont, 8-10 days post infection. During this process, several ultrastructural changes were detected: these included an extensive elaboration of the nuclear envelope and the formation of an external coat on the surface of the schizont plasma membrane, that persists to cover the mature merozoite. Other observed features are the reappearance of the rough and smooth endoplasmic reticulum and golgi apparatus, the micronemes also become visible and the free mitochondria start to associate with each schizont nuclei (Fig. 1.3A, stage 2). Initially, the DNA becomes condensed in the schizont nuclei, which are arranged at the periphery, so that the schizont becomes rosette shaped (Fig. 1.3A, stages 2 and 3) . A polar ring and rhoptries, thought to play a role in nuclear division, form in small clusters at the apical pole of the schizont nuclei and the nucleus becomes attached to the schizont plasma membranes (Fig. 1.3A, stage 3). Beneath the outer plasma membrane of the schizont, tubular structures emerge and connect the rhoptries and the nuclear envelope with the inner schizont plasma membrane in a inwardly projecting peg-like structure (Fig.1.3A and B, stage 4). These structures are thought to be implicated in merozoite budding. Soon after, merozoites bud synchronously from the syncytial schizont, and are liberated by the breakdown of the host cell plasma membrane (Fig. 1.3A: stages 5 and 6). It is not known whether this breakdown is due to a specific parasite induced lysis or is a physical disruption caused by the large number of merozoites within the host cell (Shaw and Tilney, 1992).

The merozoites released from the schizont as described by Shaw and Tilney (1992) are pear shaped bodies of approximately 1-2μm in length and 0.6μm in diameter; consisting of one eccentric nucleus, approximately 3 to 6 rhoptries, 1 or 2 mitochondria, microspheres and free ribosomes contained in the cytoplasm. *Theileria* merozoites lack a fully formed apical complex, in which a conoid or similar apical structure is absent. Upon release into the bloodstream, merozoites quickly infect the host erythrocytes. (Shaw and Tilney, 1992; Shaw *et al.* 1995) found the mechanism of entry by the merozoites into the erythrocytes was similar to sporozoite invasion of the lymphocytes. Merozoites entered the erythrocyte in any orientation by forming a continual close junction after initial attachment to the erythrocyte surface. This is followed by the progressive "zippering" of the two membranes until the merozoite is completely internalised by endocytosis; a process which has been postulated to occur by receptor mediated endocytosis as in *Plasmodium* spp. (Kawamoto *et al.*,1990). The encapsulated merozoite escapes from the surrounding erythrocyte plasma

membrane by discharging the contents of the rhoptries.

The *Theileria* parasite, now known as a piroplasm lies free within the erythrocyte cytoplasm but is not attached to host microtubules, unlike the schizont stage. The piroplasms are variable in size and may be spherical or comma shaped. The different forms and sizes occur with different frequencies across the *Theileria* genera. In *T. annulata* both forms occur in approximately equal frequencies, with the comma shaped forms reaching up to 2.5μm in size. However, in *T. parva*, 80% of piroplasms are comma shaped and are typically 1.0-1.5μm in size. The comma-shaped forms have a simple cell membrane, a small amount of endoplasmic reticulum and are characterised by a double walled vacuole and an ovoid shaped nucleus at opposite poles. The spherical forms are bounded by a single cell membrane and contain mitochondria-like organelles. Differences between *Theileria* and other related apicomplexan intraerythrocytic forms have been described. For example, it is thought that the metabolic processes of *Theileria spp* are different to that of *Plasmodium*, as no pigmented residual bodies are seen when *Theileria* piroplasms feed on the erythrocytic cytoplasm. In certain *Theileria spp* the host cytoplasm is crystallised, a process thought to be due to partial digestion of haemoglobin (Young *et al.*, 1978; van Vostenbosch *et al.*, 1978; Fawcett *et al.*, 1987).

There is disagreement on how piroplasms divide: Mehlhorn and Schein (1984) claim that comma-shaped piroplasms never divide by schizogony but by binary division. By contrast, Conrad *et al.* (1985, 1986) and Fawcett *et al.* (1987) claim that division occurs by schizogony from the spherical to a maltese-cross form. It was postulated that the spherical forms could be precursors of gametes (Mehlhorn and Schein, 1984). The production of intra-erythrocytic merozoites, which are identical to those released from the schizont stage (Conrad *et al.*, 1985) leads to the destruction of the host cell. There is uncertainty whether intra-erythrocytic merozoites from *T. parva* re-infect erythrocytes, but this is thought to happen in other *Theileria spp* (Conrad *et al.*, 1985). Transmission of the parasite from bovine host to invertebrate vector occurs when piroplasms in the bloodstream and infected erythrocytes are taken up by feeding *hyalomma* ticks. In analogy to both *Plasmodium* and *Babesia* it is likely that the infective form for the tick vector are the intra-erythrocytic gametocytes (Mehlhorn and Schein, 1984).

1.3.2 The Invertebrate vector

Ingestion of parasitised erythrocytes by ticks of the *Hyalomma* species usually occur at the larval and nymph stages on the first host before engorging on a second host in the adult stage (2 -host cycle), alternatively they may feed on a fresh host at all three stages (3-host cycle). Following ingestion, lysis of the infected erythrocytes occurs in the tick gut. Some of the free piroplasm "gametocytes" proceed to the sexual stages of development and soon develop ray bodies or "strahlenkörper" two to four days post feeding by electron micrograph studies where they were thought to be part of the sexual phase of the life cycle of *Theileria* spp. (Mehlhorn and Schein, 1976). Ray bodies are 8-12 µm in length and 0.8µ m in diameter, with thin tubular projections, an electron dense thorn-like structure and a slender posterior pole bound by a unit membrane (Mehlhorn and Schein, 1984). They are formed from the developing ovoid or spherical intra-erythrocytic stages and are considered to be microgametes because after the fifth day after tick feeding, the ray bodies contain four nuclei and thorn-like structures, which eventually lead to the formation of a uninucleated gamete stage. Within the tick gut, larger spherical stages of 4-5µm are also observed: these are the macrogametes (Mehlhorn and Schein, 1984; Young *et al.*, 1980).

After six days or less, depending on the temperature, the gametes undergo syngamy, ultimately forming a spherical, diploid zygote and enter the gut epithelium where they grow in clusters (Young and Leitch, 1981). The zygote is not enclosed in a parasitophorous vacuole and it is likely that the zygotes enter the gut epithelium through receptor-mediated endocytosis (Walker, 1990; Mehlhorn and Schein, 1984). Eventually a motile, kinete containing an apical complex develops from the zygote that is surrounded by an inner pellicular complex apart from the apical pole, where it forms a modified apical ring (Mehlhorn and Schein, 1984). The liberation of kinetes appears to be synchronous and may be linked to the moulting stage of the developing tick (Young *et al.*, 1980). Once the tick has formed salivary glands, the kinetes travel to the salivary glands via the haemolymph where they undergo sporogony, stimulated by tick feeding (Bhattacharyulu *et al.*, 1975; Singh *et al.*, 1979; Samish and Pipano; 1978). The developing parasite differentiates into a sporont that becomes multinucleate, the micronemes disappear, while vacuoles and mitochondria increase in number during sporogony (Fawcett *et al.*, 1982, Fawcett *et al.*, 1985). The cytoplasmic and nuclear volume of the sporont grows when the tick feeds, aided by labyrinth structures to increase the surface area of the syncytium, now known as a sporoblast (Fawcett *et al.*, 1982, 1985; Young *et al.*, 1983). Three or four days after tick attachment to the bovine, sporozoites are released from the sporoblast into the feeding

lesion they associate with their target cells of the blood and lymphatic systems (Fawcett *et al.*, 1985; Walker, 1990), completing the life cycle of *Theileria*.

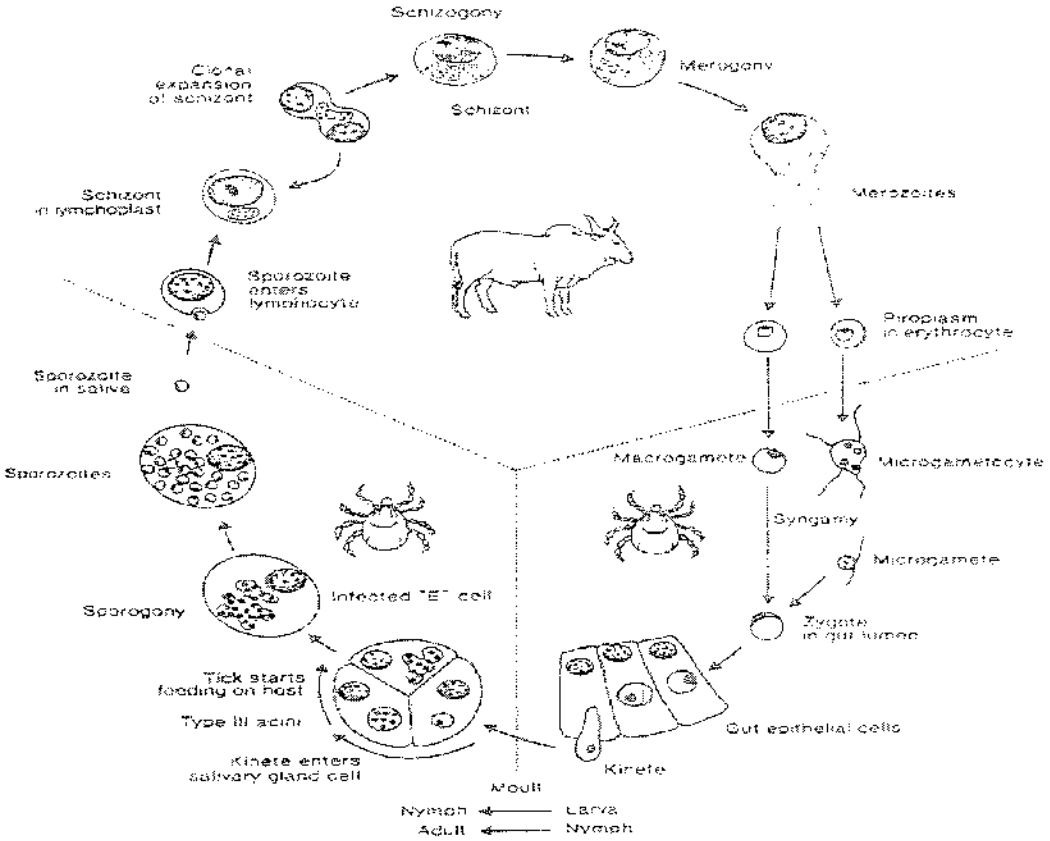


Fig. 1.2: Life cycle of *Theileria* (taken from A.S. Young in Norval *et al.*, 1992)

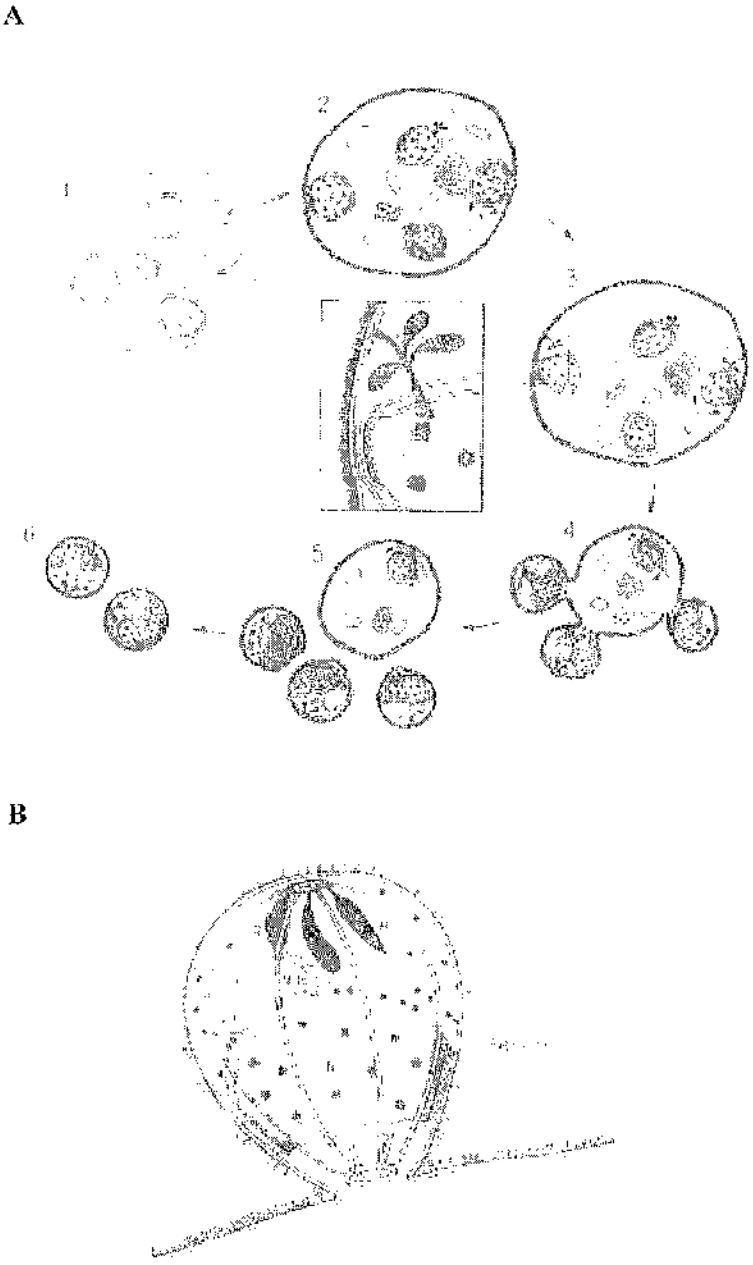


Fig. 1.3: Diagrammatic summary of the stages involved in merozoite formation. A: Stages (1-6) of merogony. B: detail of budding merozoite from stage 4 (panel A), where R: Rhoptry complex; Mito: mitochondrion; and N: nucleus (obtained from Shaw and Tilney, 1992).

1.4 Pathogenesis of Tropical Theileriosis

The pathology of tropical theileriosis is variable within different bovine species, and is particularly severe in exotic cattle breeds (reviewed by Robinson, 1982). Mortality rates vary from 90% in introduced exotic breeds to 5% in indigenous breeds (Neitz, 1957, cited by Dyer and Tait, 1987; Rafyi *et al.*, 1965). However, the severity of the disease is also directly proportional to the number of sporozoites injected into the host by the tick and the virulence of the parasitic stock (Uilenberg, 1981; Preston *et al.*, 1992a). The disease affects the lymphoid tissue and the erythrocytes of the host, caused by the intracellular schizont and the intraerythrocytic piroplasm stages, respectively. Subsequent to tick inoculation of sporozoites, the lymph nodes swell, draining the site of tick attachment and become large and hyperplastic (Srivastava and Sharma, 1981). Schizont infected cells are then disseminated throughout the lymphoid system, including the spleen and thymus forming tumour like masses (Forsyth *et al.*, 1999; Fell *et al.*, 1990). By day 7, schizont infected cells have spread to the non-lymphoid tissue such as the liver, kidneys, lung, abomasum, adrenal and pituitary glands, finally reaching the brain and heart by day 12 and day 14, respectively, post infection (Forsyth *et al.*, 1999; Fell *et al.*, 1990). During the early stages of infection, the animal suffers from a persistent fever of 41°C until death or recovery. The late stages of the disease are accompanied by merozoites and piroplasm infected cells, which are removed by the liver and spleen, causing the severe haemolytic anaemia often observed in severe cases of tropical theileriosis (Hooshmand-Rad, 1976; Uilenberg, 1981; Barnett, 1977). The animal may also suffer oedema of the lungs at the terminal stages, causing severe respiratory distress (Neitz, 1957). Preston *et al.* (1992b) has also reported rapid and severe leucopenia accompanied by lymphocytopenia. Other symptoms include, malaise and dysponea, jaundice, anorexia, diarrhoea often accompanied by blood and mucus, swelling of the eyelids and discharge from the eyes and nose. In severe cases ulceration of the abomasum is observed by post-mortem examination (Robinson, 1982; Grootenhuis *et al.*, 1980; Barnett, 1977). In acute cases death usually occurs within 20 days post infection, although in some cases pre-acute episodes may lead to death within 3-4 days (Barnett, 1977; Robinson, 1982). Cattle that survive recover eventually, but in severe cases, recovery is often incomplete and the animal remains debilitated, anorexic and unproductive. These cattle remain as carriers of piroplasms (Irvin and Morrison, 1987).

Hall and co-workers (Baylis *et al.*, 1992, 1995; Somerville *et al.*, 1998a) discovered that matrix metalloproteases (MMPs) are expressed in *T. annulata* infected cells *in vitro* and *in*

vivo and may play a role in metastasis of the lymphoid cells, leading to disease symptoms such as ulceration of the abomasum and digestion of connective tissue. However, recent studies have challenged the view that the clinical symptoms from an infection with *T. annulata* are primarily the result of uncontrolled proliferation of schizont infected lymphoid cells, in a similar fashion to *T. parva* (Forsyth *et al.*, 1999). Further studies show that *ex-vivo T. annulata* infected macrophage cells predominantly, produce a number of cytokines in a cascade including tumour necrosis factor alpha (TNF-α) and interferon alpha-1 (IFN-α_1) that trigger nitric oxide producing natural killer cells and macrophages (Preston *et al.*, 1999). TNF-α was found to be activated by MMPs (Adamson and Hall, 1996), and when administered to cattle, produced similar symptoms to those of a *T. annulata* infection, such as high fever, leucopenia, loss of weight and condition (Beutler and Cerami, 1986; Ulich *et al.*, 1987). TNF-α is also known to suppress red blood cell production, contributing to anaemia. Nitric oxide is thought to cause pathological lesions (Visser *et al.*, 1995). Thus, it is now thought that tropical theileriosis is caused by multiplying, metastasising schizont infected cells (Forsyth *et al.*, 1999) which produce cytokines that contribute to dissemination of the parasite and the clinical symptoms of the disease. (Forsyth *et al.*, 1999; Preston *et al.*, 1999).

1.5 Diagnosis

Tropical theileriosis can be detected by the clinical symptoms of infected bovines and by Giemsa stained blood smears, which detect the macroschizont and piroplasm stages. The Indirect Fluorescence Antibody Test (IFAT) utilised specific antibodies generated against *Theileria* and has been used to distinguish different *Theileria spp* (Burridge, 1971; Morzaria *et al.*, 1977; Burridge and Kimber, 1972, cited by Young and Leitch, 1981). IFAT, using monoclonal antibodies can also be used to identify population diversity within a species and a range of monoclonal antibodies raised against the macroschizont stage have also been used for this purpose (Shiels *et al.*, 1986; Ben-Miled *et al.*, 1994). However this technique is labour intensive and inconvenient for mass screening.

A more convenient and sensitive test is the enzyme linked immunosorbent assay (ELISA) developed by Voller *et al.* (1976) to detect antibodies to *Theileria* infection. ELISA tests have shown that piroplasm and schizont antigens could be used to detect *T. annulata* specific antibodies (Manuja *et al.*, 2000). Gubbels *et al.* (2000) developed an ELISA test based on recombinant antigens derived from two *Tams-1* alleles that allowed the specific

detection of *T. annulata* and not any other *Theileria* or *Babesia* species, aside from *T. parva* which is geographically distinct from *T. annulata*.

DNA based techniques have also been used both in the diagnosis of clinical cases and to discriminate different *Theileria* spp and stocks. Polymerase Chain Reaction (PCR) based techniques have been successfully used to detect parasite DNA sequences at low concentrations in infected and carrier animals. For example, PCR techniques specifically detect the *T. annulata* gene encoding the major merozoite surface antigen, *Tams-1*, from infected cattle blood (d'Oliviera *et al.*, 1995; Kirvar *et al.*, 2000) and in vector ticks (Kirvar *et al.*, 2000). Gubbels *et al.* (1999) used reverse line blotting (RLB) to simultaneously detect different species of bovine tick borne parasites from *Babesia* and *Theileria* in infected and carrier bovines. Meanwhile, DNA probes based on the large sub-unit ribosomal RNA (LSU rRNA) revealed a variable region which could discriminate two closely related *Theileria* species (Bishop *et al.*, 1995). Ben-Miled *et al.* (1994) has been able to distinguish five variants from 53 different Tunisian stocks of *T. annulata*, using two DNA probes.

1.6 Bovine Immune Response to *T. annulata* infection

Each stage of the *T. annulata* life cycle presents the bovine immune system with new antigens, to which the host responds by cell–mediated and/or humoral responses, which are both thought to play important roles in protection against tropical theileriosis. These responses have been reviewed extensively (Brown, 1990; Tait and Hall, 1990; Hall, 1988; Campbell and Spooner, 1999; Preston *et al.*, 1999).

1.6.1 Humoral Response

When sporozoites are introduced into the bovine host, they are briefly exposed to the immune system, before invading the lymphocytes, so the sporozoite is the most likely stage to be targeted by protective antibodies. The first indication that antibodies were generated against *T. annulata* sporozoites came from studies by Gray and Brown (1981). This work showed that sera taken from animals previously infected with *T. annulata* sporozoites could neutralise sporozoite infectivity of peripheral blood mononuclear cells (PBMs) *in vitro*. Further evidence that antibodies were responsible for neutralising sporozoite invasion of PBMs was produced when monoclonal antibodies, such as 1A7 blocked sporozoite invasion successfully. Moreover, IFAT tests with sporozoites, but not

macroschizonts and piroplasms using mAB 1A7 were also positive (Williamson, 1988; Williamson *et al.*, 1989). In the bovine, it appears that disease severity is correlated with sporozoite dose as repeated challenge with live sporozoites increased the neutralising activity (Preston *et al.*, 1992a). This would support the postulation that viable sporozoites are briefly exposed to the immune system and repeated exposure would enhance the humoral response. Preston and Brown (1985) suggested that antibodies derived from calf serum infected with *T. annulata* could retard the growth of PBMs infected with trophozoites. However, these findings have been revised and it has now been shown that a number of cytokines are likely to be responsible for growth inhibition of trophozoite infected cells by infected calf serum (Preston *et al.*, 1992b). The discovery that mAB 1A7, which recognises the sporozoite surface antigen (SPAG-1) could inhibit sporozoite invasion has given rise to a potential sub-unit vaccine. Indeed, a partial protective response has been achieved in cattle inoculated with recombinant SPAG-1 antigen and with its orthologues, p67, in *T. parva* (Boulter *et al.*, 1995, 1998; Boulter and Hall, 1999; Musoke, 1992).

Studies have shown that protective immunity against the intracellular macroschizont is cell–mediated, and that this response is a major contribution to the pathology of the disease (reviewed by Tait and Hall, 1990). Antibodies to the schizont stage are detected in *T. parva* and *T. annulata* infected animals (Wagner *et al.*, 1974; Kachani and Spooner, 1992), but these antibodies are not thought to contribute to protective immunity. Thus in the study of Creemers (1982), sera of infected and recovered cattle did not recognise cell membrane antigens of *T .parva* infected cell lines, nor was there evidence of antibodies directed against parasite specific antigens on the surface of infected cells. These results appear to be supported by Pipano *et al.* (1981), who found that antibodies from recovering animals were directed against the parasite and not the macroschizont infected cell. Furthermore, no correlation was found between antibody titre and the degree of protection in this study. Similarly, sera of infected bovines did not illicit a positive response to the surface of *T. annulata* infected cells by IFAT (Shiels *et al.*, 1989) and serum transfer studies from immune to naïve cattle did not afford protective immunity in the recipients (Muhammed *et al.*, 1975).

The merozoite stage of the life cycle, like the sporozoite stage is extracellular and could therefore be a target for a protective humoral response. Certainly, piroplasm and merozoite specific antibodies were detected in recovering cattle (Ahmed *et al.*, 1988). Studies on differentiating macroschizont infected cells showed that there was a shift in the antigen

profile from macroschizont to the merozoite/piroplasm, whereby some macroschizont antigens disappeared while other merozoite antigens showed significant upregulated production following the differentiation event (Glascodine *et al.*, 1990). One such antigen was a surface polypeptide,Tams-1, for which there were initially two antigenic types described: Tams-1 and Tams-2, with molecular weights of 30kDa and 32kDa, respectively (Glascodine *et al.*, 1990; Dickson and Shiels, 1993). This antigen was shown to be recognised by infected cattle sera and by a monoclonal antibody, mAb 5E1, and was located to the surface of merozoites and piroplasms (Glascodine *et al.*, 1990). Interestingly, this molecule has been shown to be polymorphic within different stocks of *T. annulata* and the predicted amino acid sequences of different allelic types consist of divergent N-linked glycosylation sites that could play a role in evasion of the humoral immune response (Shiels *et al.*, 1995; Katzer *et al.*, 1998). Ahmed *et al.* (1988) has also shown that cattle immune sera specifically opsonise free merozoites and that complement induced lysis of merozoites can also occur. Therefore, merozoites may be targeted directly by the host antibodies. Richardson *et al.* (1998) and Preston *et al.* (1999) have postulated that nitric oxide produced by macrophages in response to schizont infection may also be responsible for merozoite destruction. In contrast, studies on piroplasms revealed that that immune sera from cattle did not recognise antigens on the surface of erythrocytes (Hall, 1988). It can be postulated that there is no specific protective humoral response to the piroplasm in the bovine, probably because this stage of the parasite is intracellular and proliferation and re-invasion of erythrocytes by the parasite occurs at a low level (Persing and Conrad, 1995).

1.6.2 Cell Mediated Response

Previous studies (Pipano *et al.*,1981; Dhar and Gautam,1978; Creemers,1982; Emery, 1981; Emery *et al.*, 1981, 1982) had all concluded that the humoral response was not the primary mechanism involved in protective immunity to *T. annulata* and *T. parva* infection. It is now believed that cell mediated immunity plays the major role in protection and in pathology, in particular cytotoxic T-cells, natural killer cells, helper T-cells and macrophages. One of the first findings of a cell mediated response to *Theileria* infection occurred when Emery (1981) transferred T-lymphocytes of the thoracic duct from a calf immunised against *T. parva* to its naïve chimeric twin, resulting in protective immunity to the recipient. Direct evidence of cell mediated immunity was generated by Emery *et al.* (1981), who found that cytotoxic cells capable of killing allogeneic infected lymphocytes were present in the peripheral blood of infected cattle at the latter stages of infection. Other studies proposed that the immune response to *T. annulata* infection appeared to be similar.

Preston *et al.* (1983) showed the recovery from tropical theileriosis was accompanied by the disappearance of macrophages from the lymph nodes and the appearance of cytotoxic cells in the blood and lymphatic system. By contrast, in acute, fatal cases, no cytotoxic cells were found, suggesting that the cytotoxic T cells involved had a protective role.

Data showed that the PBMs of naïve cattle infected with *T. annulata* showed cytotoxic activity in two peaks: the cells of the first peak occurred one to two weeks post infection and were bovine leukocyte antigen (BoLA) (MHC class I) restricted, similar to cytotoxic T-cells. The second peak occurred three to four weeks post infection and did not show genetic restriction in all animals (Preston *et al.*, 1983). The presence of cytotoxic cells in the blood was concomitant with the disappearance of schizont infected cells. This study also indicated that these cytotoxic responses were protective as they were always associated with recovering cattle and almost never present in calves that died as a result of the disease. Similar results were also obtained in a later study by Chaudhri and Subramanian (1992) who found that cattle which recovered from a inoculation with a virulent *Theileria* stock of sporozoites responded with a single wave of cytotoxic cells, that had the ability to kill autologous schizont infected cells. Calves that did not survive, showed a weak cytotoxic response or none at all.

Innes *et al.* (1989) demonstrated that different T cell responses were elicited depending on whether the animal was infected with autologous or allogeneic infected cells. Here, cattle inoculated with allogeneic *T. annulata* infected cells only showed mild clinical symptoms, whereas animals infected with autologous infected cells had severe symptoms. The animals infected with the allogeneic infected cell line developed a strong cytotoxic response to the allogeneic MHC antigens of the inoculated cell line by day 9. A second response was directed against the recipients own parasite infected cells after three weeks. However, the autologous group showed little cytotoxic response that was only MHC restricted after nearly three weeks. Both groups showed BoLA restricted and non-restricted responses against parasite antigens. When challenged with a heterologous sporozoite stock, both groups of cattle were immune and developed a cytotoxic response. This suggests that MHC molecules are important in eliciting protective cell mediated immunity to different parasite types.

The non-BoLA restricted responses observed by Innes *et al.* (1989) and Preston *et al.* (1993) were thought to be natural killer (NK) cells, which have been implicated in other protozoan infections. In sublethal infections, NK cells have been shown to produce a

number of cytokines that produce a synergistic cascade that leads to the activation of NK cells and macrophages to combat the infection (Preston *et al.*, 1993, 1999). NK cells trigger the release of cytokines such as IFN γ, to stimulate macrophages that in turn produce TNF-α and IFN-α. This in turn, stimulates NK cells and induces production of nitric oxide (NO) by macrophages which destroy schizont-infected cells (Preston *et al.*, 1993; 1999; Visser *et al.*, 1995; Richardson *et al.*, 1998). The adaptive immune system is also triggered by schizont infected cells and is thought to play a co-operative role by activating macrophage anti-microbial activity, via $CD4^+$ cells and generating cytotoxic $CD8^+$ T cells (Preston *et al.*, 1999). In fatally infected cattle, it is thought that excessive amounts of TNF-α, which could account for most of the disease symptoms in *T. annulata* and in other related protozoan infections, results in death. Indeed, abnormally high levels of IFN-γ and matrix metalloproteinases, which enhance levels of TNF-α, were found in fatally infected animals (Campbell *et al.*, 1997, 1998; Adamson and Hall, 1997).

Evidence exists to show that schizont infected cells also promote an inappropriate immune response in animals that fail to recover from a primary infection. For example, IL-2, thought to be stimulated by *T. parva* infection, stimulates the proliferation of macroschizont infected cells (Campbell and Spooner, 1999). Other studies found that IL-2 receptors and MHC class II molecules were expressed on the surface of $CD4^+$ and $CD8^+$ T cells from naïve cattle infected with *T. annulata*, and was shown to induce the proliferation of autologous, resting T cells *in vitro*, resulting in a failure to mount a proper, protective T-cell response (Campbell *et al.*, 1995, 1997; Campbell and Spooner, 1999). Furthermore, *in vivo*, it has been postulated that infected macrophages present antigens to $CD4^+$ T cells, which are inappropriately activated in large numbers within the medulla and not the paracortex- the normal region for priming T cells. This would result in the proliferation of large numbers of polyclonal IL-$2R^+$ T cells, which leave the lymph node and abolish an effective immune response against macroschizont infected cells (Campbell and Spooner, 1999).

1.7 Control Measures

Currently, there are three control measures against tropical theileriosis and involve tick control, chemotherapy and host vaccination. Different measures that are currently being used or under development are discussed below:

1.7.1 Tick Control

Conventional tick control measures have centred on the use of anti-tick agents known as acaricides, such as amitraz and butocarb, in the form of dips, sprays, impregnated ear tags, slow release rumen boluses and "pour-ons" in cattle (Chizyuka and Mulilo, 1990; Musisi, 1990; Urquhart *et al.*, 1987; de Castro and Newson, 1993). However acaricides are expensive, cause environmental damage and can result in residual contamination of milk and meat (Drummond *et al.*, 1976). Furthermore, continuous use leads acaracide resistance in ticks and loss of tick-immunity in cattle, making them more susceptible to infection. Improvements have been made to cattle housing by the design of new livestock sheds and barns which discourage or eliminate tick settlement and reduce tick infestation levels (Pipano, 1989a). The maintenance of exotic, high yielding cattle have led to the practice of confinement with regular acaracide treatment and tick free feed. However, this solution is very labour intensive, expensive and often fails leading to infection of some, many or all cattle (Lawrence, 1990).

1.7.2 Chemotherapy

This method of treatment is more widely used in the control of *T. parva* and *T. sergenti* infections compared with *T. annulata* infections. The most effective anti-theileriacidal drugs are the hydroxynapthoquinones, parvaquone and buparvaquone (McHardy and Morgan, 1985; Tait and Hall, 1990; Hagiwara *et al.*, 1993). These drugs are thought to disrupt the cytochrome bc_1 complex of the electron transport chain of the parasite (Hall and Baylis, 1993). Buparvaquone has been found to have greater anti-theileriacidal activity than parvaquone (Hashemi-Fesharki, 1991), but parvaquone is active against all *Theileria* stages whereas buparvaquone only acts against the schizont and piroplasm. Other drugs that have had some success are the coccidiostat Halofuginone, which, like buparvaquone, acts on the schizont stage of the life cycle, and is cheap (Schein and Voigt, 1979). However, this drug is no longer used because it has a narrow therapeutic range. Antibiotics have been found to reduce the level of parasitosis, such as the ionophorous antibiotic monensin, but this drug has not been used due to adverse side effects (McHardy and Rae, 1982). Oxytetracyclines has been shown to reduce macroschizont development *in vitro* (Spooner 1990) and has been used with some success in "Infection and treatment" vaccination programmes (see section 1.7.3) but was found to be unsuccessful in acute infections (Singh *et al.*, 1993). The major drawback to chemotherapy is that it can prevent the development of immunity in the animal. Moreover, drug treatment programmes are expensive, and may eventually lead to

parasite resistance.

1.7.3 Host Vaccination

Chemotherapeutic agents have been used in "Infection and treatment" programmes to vaccinate cattle against *Theileria*. Long-acting oxytetracyclines have been used for *T. parva*, whereas buparvaquone is preferred for *T. annulata*. Infection and treatment involves the deliberate inoculation of sporozoites into cattle followed by drug treatment at a predetermined and critical time so that bovine immunity is activated after the sporozoites are established in the host's lymphocytes, but before clinical symptoms appear (Mozaria and Nene, 1990). Such treatment results in solid immunity to homogeneous and sometimes heterogeneous challenge, but have not been used widely, due to the large expenses involved. Moreover, piroplasms remain in infective carrier animals, which assist in the spread of the disease (Tait and Hall, 1990).

The most common and effective form of vaccination against *T. annulata* is the attenuated live vaccine. Attenuation occurs through long term, *in vitro* culture of macroschizont infected lymphocytes, and results in a loss of pathogenicity of the parasite (Brown, 1990). Initially, the parasite loses its ability to produce merozoites and thus piroplasms in most cases (Pipano, 1989a), but after prolonged culture, also loses its virulence or infectivity (Pipano and Tsur, 1966).

The mechanisms of attenuation are unclear: research on the related apicomplexan parasite *Plasmodium berghei* have shown that loss of virulence is associated with genome rearrangements (Janse *et al.*, 1992). Studies by Hall *et al.* (1999) have not been able to demonstrate large, genomic rearrangements in attenuated *T. annulata* cell lines, although Preston *et al.* (2001) found that alterations to host cell surface antigens were linked to permanent changes in the parasite genome. A second possibility for attenuation is the selection of avirulent subpopulations: this has been demonstrated in *Babesia bovis* (Cowman *et al.*, 1984; Carson *et al.*, 1990) and in *T. annulata*. In the latter case, Melrose (1984) demonstrated that only single isotypes of the enzyme glucose phosphate isomerase (GPI) were found after continuous *in vitro* culture of infected cell lines that originally contained a number of GPI isotypes. However, this theory does not completely account for attenuation of *Theileria* infected cells lines, since studies by Preston *et al.* (2001) have shown that cell lines continue to decline in virulence over long periods of *in vitro* culture, which was not caused by preferential growth of particular host cell types. This has led to

the suggestion that attenuation in *T. annulata* is achieved by alteration of gene expression, resulting in specific alterations to both parasite and host polypeptide production (Sutherland *et al.*, 1996; Shiels *et al.*, 1998; Hall *et al.*, 1999; Oura *et al.*, 2001). This theory has been supported by studies that show that long term *in vitro* culture is accompanied by the loss of parasite induced expression of host metalloproteinases (MMPs), previously shown to be associated with the pathogenesis of tropical theileriosis (Adamson *et al.*, 2000ab; Hall *et al.*,1999; Somerville *et al.*,1998b). Hall *et al.* (1999) also demonstrated that loss of MMP expression is a stable transferable trait, implying that this process occurs at the genetic level of the parasite. It has been suggested that virulence is caused by the interaction of a number of parasite sub-populations, which, under long term culture conditions becomes simplified, reducing the complexity of virulence factors below a threshold value (Hall *et al.*, 1999).

Vaccination with attenuated macroschizont infected cells have been shown to be 95-100% efficient at providing immunity to heterologous challenge in cattle (Brown, 1990). Immunity is tested by challenge from live, infected ticks or with sporozoite stabilates, which have given a range of results from mild symptoms (Gill *et al.*, 1976; Ouhelli *et al.*, 1989) to death in some cases (Ozkok and Pipano, 1981; Shukla and Sharma, 1991; Adalar *et al.*, 1992). Complete attenuation is achieved when cultured schizonts no longer cause clinical symptoms in the animal, usually after 60-250 passages in culture, with 10^6-10^7 infected cells per animal, depending on the isolate (Pipano, 1995). So far, there is no evidence of a reversion to virulence in attenuated cell lines. Furthermore, vaccines can be preserved in liquid nitrogen for a considerable time period without significant loss in viability (Wathanga *et al.*, 1986). However, protection from subsequent infection after initial immunisation has been reported to decline with time (Tsur *et al.*, 1964; Sergent *et al.*, 1945). There have been a number of wide ranging estimates for the length of protection from 3.5 years (Zablotsky, 1983, cited by Pipano, 1995) to 6 months (Beniwal *et al.*, 2000). These conflicting estimates may be explained by immunogenic differences between different *T. annulata* isolates (Barnet, 1963; Adler and Ellenbogen, 1935, cited by Pipano, 1995). In contrast to *T. annulata*, it has not been possible to generate a live attenuated vaccine for control of *T. parva* infections (Dolan *et al.*, 1984; Morrison *et al.*, 1981). It is thought that this is due to an inability of *T .parva* schizonts to transfer to cells of recipient animals, thus, failing to remove the histocompatiability barrier against recognition of infected cells by cytotoxic $CD8^+$ cells (Musoke *et al.*, 1996; Uilenberg, 1999; Boulter and Hall, 1999). Although this method of vaccination is effective and cheap the main limitations of live attenuated vaccines are storage and delivery: the vaccine has a shelf-life

of approximately 1 week at 20°C or 1 month at 4°C and there are difficulties in transporting frozen, vaccine to remote regions.

To counteract the difficulties associated with live attenuated vaccination, efforts have also been made to develop a sub-unit vaccine against *Theileria* parasites. Recent identification of candidate *Theileria* antigens, such as Tams-1 and SPAG-1 have generated some promising results. The sporozoite and merozoite stages are the only extracellular stages of the parasite and are, therefore, logical targets for selection of antigen sub-units to develop vaccines. In the sporozoite, studies have concentrated on SPAG-1, because the monoclonal antibody, 1A7, generated against this molecule was shown to be able to abrogate sporozoite infectivity *in vitro* (Williamson, 1988). Further characterisation of this molecule identified several immunodominant sites, at the N and C terminus, including the epitope recognised by 1A7, which were found on this antigen when fragments of SPAG-1 were reacted against a range of bovine immune sera (Boulter, 1996; Knight *et al.*, 1996). In addition, SPAG-1 contains C- terminal epitopes, distinct from the mapped 1A7 epitope, capable of neutralising sporozoite infectivity (Williamson, 1988, Williamson *et al.*, 1989; Hall and Baylis, 1993; Boulter, 1996; Boulter *et al.*, 1994, 1995). Partial protection has been achieved *in vivo* using recombinant SPAG-1 (Boulter, *et al.*, 1995, 1998; Boulter and Hall, 1999). In *T. parva*, a SPAG-1 homologue called p67 has been found and has shown to be cross-reactive with SPAG-1. Parts of the C- terminal domain of SPAG-1 and p67 are similar enough to form a common epitope. Musoke *et al.* (1992) found that vaccination with recombinant p67 provided protection in 6 out of 9 calves against *T. parva* sporozoites. However, there is evidence that SPAG-1 and p67 could be used to provide immunity against *T. parva* and *T. annulata* sporozoites: mAb 1A7 has been shown to neutralise *T. parva* sporozoite infectivity with 100% efficiency (Knight *et al.*, 1996; Katzer *et al.*, 1994). Other vaccination trials have shown cross protection against heterologous challenge when cattle immunised with either SPAG-1 or p67 are challenged with *T. parva* or *T. annulata* sporozoites, respectively (Boulter *et al.*, 1998; Boulter and Hall, 1999; Hall *et al.*, 2000). Interestingly, a strong T-cell response to the N-terminus of SPAG-1 has also been observed in the presence of IL-2 in immune animals suggesting that sub-unit vaccines elicit both cell-mediated and humoral responses (Boulter and Hall, 1999).

Studies on the antigenic determinants of the merozoite stage of *T. annulata* initially identified a 30kDa surface molecule, Tams-1 (Glascodine *et al.*, 1990), which has been shown to exhibit extensive amino acid and antigenic diversity (Dickson and Shiels, 1993; Katzer *et al.*, 1998; Gubbels *et al.*, 2000). This molecule has been shown to display

significant antigenic diversity particularly at putative glycosylation sites located in regions of amino acid hypervariability (Shiels *et al.*, 1995), suggestive of an immune evasion strategy. Over 40 predicted amino acid variants of Tams-1 have been identified and evidence for selection of novel antigenic types following passage through ticks has recently been reported (Gubbels *et al.*, 2001). The level of diversity makes Tams-1 a difficult candidate for a sub- unit vaccine. However, studies by Ahmed *et al.* (1988) demonstrated that bovine antibodies could specifically lyse merozoites. It is likely, however, that the best candidates for inclusion in a sub-unit vaccine are antigens encoded by the macroschizont. Theoretically, these antigens are probably processed and presented as peptides on the surface of the infected leukocyte, in association with the class I MHC molecules. Such antigens could include macroschizont surface polypeptides or proteins secreted by the parasite into the host cell environment. Although the polymorphic immunodominant macroschizont antigen (PIM) of *T. parva* (Katende *et al.*, 1998) has been used successfully to identify antibodies against *T. parva*, a protective schizont antigen has not been identified to date. Future efforts to improve vaccines should focus on an effective antigen delivery system, and these are likely to be based on inoculation of recombinant organisms capable of intracytoplasmic growth that express and produce *Theileria* antigen genes. Such a system has been successful in inducing a cytotoxic T- cell response against *P. falciparum* using recombinant *Salmonella typhamurium* (Aggarwal *et al.*, 1990). To date development of DNA vaccines for *Theileria* has been limited. Trials with a recombinant Tams-1 DNA vaccine have showed that two thirds of cattle were protected using this technique, although no antibodies were detected (d'Oliviera *et al.*, 1997).

In summary, the main priorities for recombinant vaccine development are to find protective antigens expressed by macroschizont infected cells and development of an effective antigen delivery system that, possibly, mimics the presentation of the native molecule to the immune system.

1.8 Stage Differentiation in *T. annulata*

The development of stage differentiation in apicomplexan parasites such as *Theileria* enables the parasite to establish and multiply within hosts, and transmit itself between hosts. Stage differentiation is accompanied by significant changes to the parasite polypeptide profile, brought about by altered control of gene expression (Shiels *et al.*, 1992, 1994, 1997; Carrington *et al.*, 1995). Ultimately, this results in the shut down of genes encoding the structural and metabolic polypeptides specific to the macroschizont

stage while initiating production of merozoite polypeptides. Most of the research on differentiation in *Theileria* has centred on the production of merozoites (merogony) as macroschizont infected cells can be cultured and induced to undergo differentiation to the merozoite *in vitro* (Brown, 1990; Shiels *et al.*, 1992). Much effort has been applied to defining agents that can trigger differentiation from one life cycle to another in protozoans, in general, and this research has identified a number of signals. These include alterations in temperature (Hulliger, 1965; Soete *et al.*, 1994; Van der Ploeg, 1985), pH (Soete *et al.*, 1994; Zilberstein *et al.*, 1991), agents that act on the signalling pathways, such as cAMP, (Heath *et al.*, 1990) nitric oxide (Bohne *et al.*, 1994), and intermediates of the TCA cycle (Brun and Schonenberger, 1981, cited by Fox, 1997). However, the most common inducer of differentiation in protozoan parasites is an alteration of temperature *in vitro* that may mimic temperature fluctuations that may occur *in vivo* as the parasites are transmitted to and from their warm blooded hosts.

It has been proposed that the induction of parasite differentiation at an elevated temperature is directly linked to elevated levels of heat shock proteins (hsps), which could switch on genes expressed by the next life cycle stage (Van der Ploeg *et al.*, 1985; Polla, 1991; Wiesgigl and Clos, 2001; Weiss *et al.*, 1998). Hsps are common in parasites that transfer from poikilothermic vector to a homeothermic mammal, such as *Trypanosoma brucei* (Van der Ploeg *et al.*, 1985) and are regulated by alterations in temperature. Mason *et al.* (1989) isolated the *T. annulata* heat shock protein 70 (hsp70) polypeptide and demonstrated up-regulation at the mRNA level when infected cells were placed at 41°C. However, in a subsequent study, there was no detectable difference in hsp70 mRNA levels at 41°C between a cell line that differentiates well compared to a cell line that is severely attenuated for the differentiation process. It was concluded that hsp70 gene expression was not directly involved in regulating differentiation (Shiels *et al.*, 1998). This is supported by studies in other protozoan parasites, that show no link between hsp expression and the ability of a cell to differentiate (Shapira *et al.*, 1988; Zilbertstein *et al.*, 1991).

When particular *Theileria* infected cell lines are maintained at an elevated temperature of 41°C, the schizont fails to divide in synchrony with the host cell (asynchronous division), and, after approximately 5 days, merozoite production occurs in a sub-population of cells until the majority of cells produce merozoites and the culture is no longer viable (Hulliger *et al.*, 1966; Shiels *et al.*, 1992). Since *Theileria* infection induces a fever of 41°C in infected cattle, it was thought that an elevation of temperature directly initiates parasite differentiation. However, Jarret *et al.* (1969) found no correlation between the onset of

fever and merogony. An alternative hypothesis of differentiation based on a mitotic clock was proposed by Temple and Raff (1986) and proposed that differentiation to the merozoite and the appearance of piroplasms *in vivo* occurred after a set number of mitotic divisions undergone by the infected leukocyte.

Studies by Shiels *et al.* (1992) showed that when infected cells were placed at 41°C and then replaced at 37°C for variable time periods differentiation was reversible during the early stages of the process (up to 4 days), but was irreversible after a certain time period at 41°C. Shiels *et al.* (1992) also measured the rates of infected cell growth and differentiation. These parameters were compared in two cell lines, one with a diminished ability to differentiate (diminished) and the other with an enhanced ability to differentiate (enhanced). Differentiation in the enhanced cell line was characterised by an increase in schizont size and nuclear number after two days at 41°C and was followed by a decrease in host cell division, after day 2 until it eventually stopped. However the parasite continued to undergo nuclear division, and, as the parasite is dependent on host mitosis to undergo cellular division, the parasite size and nuclear number increased significantly, resulting in an enlarged macroschizont within the host (Hulliger *et al.*, 1966; Shiels *et al.*, 1992). Thus the ability to differentiate did not appear to be linked to a fixed number of mitotic divisions but was associated with an increase in parasite size or condition relative to host cell division. In contrast, in cells with a diminished ability to differentiate, the level of host cell proliferation was elevated, whilst that of the parasite was lower at all time points tested (at 37°C) and the disruption between parasite and host cell division was markedly reduced when placed at 41°C (Shiels *et al.*, 1992). In this case it would take longer for the parasite to reach a predetermined state for differentiation, resulting in its diminished phenotype. It was concluded that disruption in the synchrony between parasite and host cell division was a factor that predetermined differentiation.

Further studies by Shiels *et al.* (1994) investigating the molecular changes during differentiation to the merozoite, revealed that down regulation of a macroschizont polypeptides was temporally linked to the up-regulation of merozoite polypeptides, including the merozoite surface polypeptide, Tams-1. The expression of Tams-1 polypeptide was first detected very early at day 2, and its corresponding mRNA at day 0 and its signal increased to day 8. This indicated that gene products present at high levels in the merozoite were also expressed at low levels in the preceding macroschizont stage. The expression of Tams-1, as detected by the mAb 5E1 by IFAT was found to be reversible in the majority of infected cells placed alternately at 37°C and 41°C during the initial

(reversible) phase of differentiation. However, some cells stained intensely with 5E1, and these were postulated to be cells that had become committed to differentiate. Thus, the asynchronous nature of merogony in *T. annulata* has led to the suggestion that differentiation is a stochastic process which depended upon merozoite factors reaching a critical level before the parasite commits to differentiation to the merozoite.

Further evidence of a stoichiometric model for differentiation involving changes to parasite growth and cellular division was provided when infected cells were treated with agents to disrupt parasite protein synthesis (a measure of growth) and host DNA synthesis (a measure of host cell division) (Shiels *et al.*, 1997). The drugs respectively caused changes to the rate of parasite growth compared division, and this in turn, altered the timing of macroschizont differentiation to the merozoite. Shiels *et al.* (1997) specifically found that pre-treatment with a DNA synthesis inhibitor resulted in increased levels of parasite differentiation in the cell population after commitment. Conversely, pre-treatment of infected cells with a parasite protein synthesis inhibitor postponed the onset of differentiation in a quantitative manner. Levels of Tams-1 were found to be directly proportional to the drug-altered differentiation events. These observation led to the proposal that increased levels of polypeptide synthesis relative to DNA synthesis provides an initial signal for the parasite to undergo merogony (Shiels *et al.*, 1998). These observations led to the postulation of a model that proposes that the commitment to differentiate is brought about by a quantitative increase in levels of polypeptide growth factors present in the preceding macroschizont that regulate merozoite gene expression relative to the levels of the nucleic acid templates they bind to (Shiels *et al.*, 2000a). Further investigations by Shiels *et al.* (2000a), of *Tams-1* regulation during differentiation seemed to confirm this hypothesis: Tam-1 expression was shown to be controlled at least, in part, at the transcription level. Electromobility Band Shift Assay (EMSA) studies revealed two complexes that bound to an upstream region of *Tams-1* in the macroschizont that were found to be elevated during the initial phase of differentiation to the merozoite, whilst a third complex was detected during the reversible phase of differentiation associated with high level Tams-1 expression.

Similarities exist between the characteristics of a number of protozoan differentiation systems; for example reduced proliferation coupled to an inherent asynchrony, led Shiels *et al.* (1998) to hypothesise that a stoichiometric model could be applicable to a range of parasites, such as bradyzoite formation in *Toxoplasma gondii* (Soete *et al.*, 1994) and gametocyte production in *Plasmodium* (Carter *et al.*, 1979). However it should be pointed

out that alternative mechanisms could be involved. For example, asynchronous division may be explained if there was a requirement for a cell to be at a particular cell-cycle position or state in order to undergo differentiation (reviewed by Shiels *et al.*, 1998).

Studies by Swan *et al.* (2001b) showed a possible link between macroschizont and merozoite gene regulation by investigating the expression of two macroschizont genes, *Tash1* and *Tash2* in relation to *Tams-1*, whose expression is associated with commitment to merogony. *Tash1* and *Tash2* were shown to be down regulated as the Tams-1 transcript became up regulated. Moreover the expression of Tash1 and Tash2 polypeptides was found to be reversible during the initial reversible phase of differentiation when infected cells were alternately pulsed between 37°C and 41°C. This study demonstrated that that regulation of macroschizont and merozoite gene expression was temporally and perhaps, mechanistically linked (Swan *et al.*, 2001b). It has been postulated that parasite factors involved in host cell division may be down regulated during differentiation to the merozoite, resulting in a reduction or cessation of host cell division (Carrington *et al.*, 1995). Recently a small parasite encoded gene family, *TashAT*, were found to be down regulated early on during differentiation to the merozoite, coincident with the decrease in host cell proliferation (Swan *et al.*, 1999). These genes could be potential parasite regulatory proteins that modulate host cell genes that control host cell proliferation (Swan *et al.*, 1999).

1.9 *T. annulata* Induced Alterations of Gene Expression

Transformation of host cells by *Theileria* is characterised by the continuous proliferation of infected cells in culture, without additional growth factors or cytokines (Dobbelaere, 1988). Immortalisation of *Theileria* infected cells in culture was shown to be accompanied by changes in surface antigens detected by mAbs (Baldwin *et al.*, 1988), pleiomorphism (Naessens *et al.*, 1985) and short generation times. In addition Irvin *et al.* (1975) and Fell *et al.* (1990) showed that *Theileria* infected cells cause tumour-like masses that metastasise throughout the organs of irradiated, athymic or SCID (severe combined immunodeficient) mice, implying that the cells undergo transformation to become cancerous. However, experiments with buparvaquone have shown that parasite induced host transformation is entirely reversible and is, therefore, not due genetic changes to the host, but is more likely brought about by parasite induced alteration to the control of host gene expression (Dobbelaere *et al.*, 2000). Higher eukaryotic cellular transformation is often accompanied by alterations to the cellular environment and involves alterations to protein kinase activity,

cytokine mediated signal transduction changes to the levels and activity of transcription factors and metalloproteinase gene expression. These mechanisms have also been defined for immortalised/transformed *Theileria* infected cells and are detailed below:

1.9.1 Alterations in Protein Kinase Activity

Dyer *et al.* (1992) reported alterations to the profile of protein kinases in infected cells compared to their uninfected counterparts, two of which were found to be unique to infected cells. These authors noted the significance of protein kinases as key indicators of cellular transformation in eukaryotic cells that when expressed abnormally can lead directly to a transformed phenotype (Hunter and Sefton, 1980; Hunter *et al.*, 1985; Hanafusa, 1986; Seldin and Leder, 1995). More recently, increased levels of host Casein Kinase II (CKII), a ubiquitous, conserved serine-threonine specific protein kinase was observed in *T. parva* infected cells *in vitro* (ole-Moi Yoi, 1995). CKII is involved in receptor mediated signalling pathways and is associated with cellular proliferation and transformation (Pinna, 1990; Tuazon and Traugh, 1991; Meisner and Czech, 1991). Experiments showed that CKII levels were found to increase when B cells were stimulated to divide using mitogens (DeBenedette and Snow, 1991). Also, some of the clinical symptoms seen in lymphocytes of transgenic mice with dysregulated CKII levels (Seldin and Leder, 1995) are similar to the pathology of *T. parva* infections, such as tissue infiltration by affected lymphocytes. However there are also significant differences between the *Theileria* transformed cells and the transgenic mouse model; namely the failure of CKII overexpressing mice to develop B cell lymphomas, the time taken for tumour development and the percentage of T cells that become transformed (Chaussepied and Langsley, 1996). Later studies by Shayan and Ahmed *et al.* (1997) showed that CKII expression was down regulated when the parasite was eliminated with buparvaquone treatment, confirming that CKII expression was parasite induced. Interestingly, *Theileria* is known to possess a gene encoding a molecule with significant identity to the catalytic α-subunit of CKII. This parasite molecule contains a sequence motif with significant identity to a signal peptide, and might be inserted into the parasite plasma membrane or transported to the host cell cytoplasm. Here, it could alter the cell cycle regulation of the host by phosphorylation of host molecules or by increasing the levels of host CKII (ole-Moi Yoi *et al.*, 1992; Dobbelaere and Heussler, 1999). Thus host or parasite CKII has the potential to play a major role in *Theileria* induced host cell immortalisation.

Further evidence for the involvement of a signalling pathway was found by Fich *et al.* (1998) who showed that transformation by *T. parva* influences the expression of at least one major Src kinase member, $p60^{fyn}$. Src kinases direct early membrane signalling in T and B cells and many are considered to be proto-oncogenes (Fich *et al.*, 1998). Using tyrosine kinase inhibitors, Fich *et al.* (1998) showed that proliferation of *Theileria* infected cells could be blocked, suggesting that tyrosine phosphorylation is essential for the maintenance of the transformed host cell. $p60^{fyn}$ levels are elevated in *T. parva* infected cells, but were found to be reduced with buparvaquone treatment. Furthermore, $p60^{fyn}$ co-precipitated with two weakly phosphorylated polypeptides. As these phosphoproteins were absent in non-infected cell lines, it was postulated by Fich *et al.* (1998) that *T. parva* proteins either directly or indirectly activate $p60^{fyn}$. However, studies in other eukaryotic systems have shown that Src activation by cellular receptors are only transient and are insufficient to stimulate cell transformation alone (Thomas and Brugge, 1997).

1.9.2 Cytokine Mediated Signal Transduction

Work performed by Dobbelaere *et al.* (1999) has demonstrated that the parasite may interfere with the normal T cell antigen receptor pathway to induce continuous proliferation. Parasite induced disruption of the signalling pathway was suspected due to the fact that some *T. parva* infected cells express high affinity IL-2 receptors that could be involved in an autocrine loop with IL-2. Moreover, these infected cells did not require antigenic stimulation to divide. Further studies showed that *T. parva* bypassed the T cell receptor (TCR) pathway, as there was a lack of phosphorylation of receptors typical to that pathway. In addition proliferation was not inhibited by immunosuppressive drugs and the IL-2 receptor was found to be constitutively activated (reviewed by Dobbelaere *et al.*, 2000). Nonetheless it should be noted that the IL-2R/IL-2 autocrine loop is only observed in some cell lines and is therefore not thought to be a primary cause of host cell proliferation (Chaussepied and Langsley, 1996). The only other cytokine shown to enhance host cell proliferation is TNFα, possibly through the parasite induced expression of TNFα receptors, as lymphocytes do not express this cytokine normally (Preston *et al.*, 1992b; Chaussepied and Langsley, 1996).

A key finding to understanding parasite controlled proliferation of the host cell was the detection of constitutively high levels of NF-κB in infected cells (Ivanov *et al.*, 1989). Interestingly it was found that expression of active nuclear NF-κB protects the infected cell from apoptosis, as treatment with inhibitors or transfection with dominant-negative

mutants resulted in cell death (Heussler *et al.*, 1999). NF-κB is a transcription factor complex composed of members of the Rel family that include p50 or p52 sub-units. These members all contain an N-terminal 300 amino acid Rel homology domain that encodes a leucine zipper region for dimerisation, a DNA binding domain and a nuclear localisation signal (Joyce *et al.*, 2001). NF-κB has an important role in cell growth as it directly controls the expression of cyclin D and c-myc, which regulate the G0/G1-S transition phase of the cell cycle (reviewed by Hinz *et al.*, 1999). This molecule is activated by a number of cytokines, including TNF-α, and cellular stress and is regulated by IκB proteins, most notably IκBα and IκBβ which, bind to NF-κB in the cytoplasm, masking the nuclear localisation signals and preventing the molecule entering the nucleus (Joyce *et al.*, 2001).

In *T. parva* infected cells, it is thought that parasite associated activation of NF-κB occurs via continuous degradation of the IκBs, by phosphorylation via IκB kinases (IKKs), allowing NF-κB to enter the nucleus and activate genes involved in the protection of apoptosis (Dobbelaere *et al.*, 2000). Recently, it has been postulated that continuous degradation of IκB occurs through the association and activation of the multi-protein IKK complex with the macroschizont itself (Dobbelaere and Heussler, 1999). It was envisaged that this could occur by direct activation of the IKK complex by upstream parasite activators, via IKK-γ, required for stimulation of IKK by NFκB inducing kinase (NIK), or the MAPK/ERK kinase, involved in NFκB activation pathways (Dobbelaere and Heussler, 1999). Alternatively, parasite molecules could interact with the IKK-complex-associated protein, a structural protein that directly interacts with IKK and the upstream kinase NIK (Dobbelaere and Heussler, 1999). Apoptosis induced by NF-κB inhibition has also been demonstrated in other transformed cell lines, including a lymphoma cell line. Indirectly, NF-κB induces IL-10 expression which has been observed in patients with human T-lymphotrophic virus type-1 induced T-cell leukaemia (Höllsberg, 1999). IL-10 was the only cytokine which showed universal expression in a study performed on a number of *T. parva* infected cell lines (McKeever *et al.*, 1997).

The mitogen-activated protein kinase (MAPK) pathway and the phosphoinositide-3-kinase (PI-3K) pathways have also been implicated with parasite infection. The JNKs (jun-NH_2-terminal kinases), a class of MAPK kinases that respond to cellular stress were found to be solely responsible for the activation of the transcription factors AP-1 and ATF-2 via the phosphorylation of c-Jun (Chaussepied *et al.*, 1998; Botteron and Dobbelaere, 1998). It has been postulated that the JNK pathway and NF-κB are regulated by MAPK/ERK kinase

kinase I (MEKK1), which is known to induce both JNKs and NF-κB via the IKK complex (reviewed by Mercurio and Manning, 1999). Ultimately, AP-1 and ATF-1 together with NF-κB are thought to activate a number of genes involved in protection from apoptosis (Dobbelaere *et al.*, 2000).

The second pathway thought to be involved in parasite induced proliferation is the PI-3K pathway, a group of signal transducers that activate growth factors, immune receptors and also interfere with apoptotic signalling (reviewed by Stambolic *et al.*, 1999). These molecules are thought to play a role in *T. parva* induced proliferation, because inhibiting agents of this pathway block host cell proliferation (Dobbelaere *et al.*, 1999). However, it is unclear if the PI-3K pathway protects against cell death as inhibitors of the PI-3K pathway did not cause the cells to apoptose, unlike the inhibition of NF-κB (Dobbelaere and Heussler 1999; Heussler *et al.*, 2001). Indeed, work by Heussler *et al.*, (2001) demonstrated that a downstream target of PI-3K pathway, Akt/PKB, appears to act independently of NF-κB activation and that NF-κB dependent protection against apoptosis does not involve the PI-3K-Akt/PKB pathway. Studies by Baumgartner *et al.* (2000) supported the findings that PI3-K activity is necessary for proliferation, and revealed that it was mediated by the induction of granulocyte-monocyte colony stimulating factor. However, in contradiction to Heussler *et al.* (2001), it was found that PI3-K is involved in the constitutive activation of NF-κB and AP-1 (Baumgartner *et al.*, 2000). A third possibility is that PI3-K induced proliferation operates in more than one pathway . As the PI3-K pathway is also implicated in cell motility and cytoskeletal changes, this pathway may also be sequestered to aid host cell proliferation in *Theileria* infected cells by moving components of the signal transduction pathways via the cytoskeleton to and from the parasite surface. This is supported by the discovery that the IKK complex has been found near to the schizont (see above) (Dobbelaere *et al.*, 2000).

It has been postulated that NF-κB targets cyclin dependent kinases which regulate NF-κB binding to specific cell cycle associated factors (Dobbelaere *et al.*, 2000). The identification of two *Theileria* specific cyclin dependent kinases, TaCRK3 and TaCRK2, (Kinnaird *et al.*, 1996, 2001), or other related *Theileria* polypeptides may be two such cyclin dependent kinases that are involved in the specific activation or repression of NF-κ B, brought to the parasite via the cytoskeleton. It is possible that TaCRK2 could be a candidate for NF-κB interaction as it is similar in sequence to the higher eukaryotic CDK1/2 family, some members of which are known to regulate NF-κB (Perkins *et al.*,

1997). However, no data has been generated to date suggesting that TaCRK2 or TaCRK3 have a functional signal sequence or are transported to the macroschizont surface or host cell cytoplasm/nucleus. Analysis of these genes suggests that TaCRK2 could control parasite nuclear division and this is likely to be their primary function (Kinnaird *et al.*, 2001).

1.9.3 Metalloproteinases

Prolonged culture of *Theileria* infected cells is known to lead to host cell proliferation and the loss of differentiation potential and virulence. Shiels *et al.* (1997) provided a link between reduced differentiation capacity and increased levels of host cell proliferation in certain infected cell lines, and an inability to up regulate merozoite gene expression in a cloned cell line severely attenuated for the differentiation process. Other studies investigating attenuation of schizont infected cells suggested that it is associated by alteration of parasite and host cell gene expression (Sutherland *et al.*, 1996; Hall *et al.*, 1999). The host gene studied in greatest detail encodes a member of a group of enzymes known as metalloproteinases (MMPs). Initially, the MMP9 metalloproteinase was found to show elevated expression in low passage *Theileria* infected cells but not in their uninfected counterparts (Adamson and Hall, 1996, Adamson *et al.*, 2000ab). However, expression was significantly reduced in infected cells that had undergone significant passage *in vitro*. Further data suggested that MMPs also may be associated with parasite virulence and metastasis of *Theileria* infected cells in SCID mice (Somerville *et al.*, 1998a).

MMP9 was shown to be regulated at the transcriptional level at least in part, by the AP-1 transcription factor (Adamson *et al.*, 2000b). There were several lines of evidence to support this: firstly the bovine MMP9 gene contained an upstream consensus AP-1 binding site; secondly, promoter activity was observed in constructs containing an AP-1 binding site by transient transfection. Thirdly, EMSA studies showed one of the specific AP-1 binding mobility complexes was lost in high passage cell lines compared to low passages cells (Adamson *et al.*, 2000b). Interestingly, a potential NF-κB binding site was also found upstream of the MMP9 gene, which was shown by deletion studies to actively contribute towards MMP9 activity in the low passage cell lines (Adamson *et al.*, 2000b).

It was revealed that both *T. annulata* and *T. parva* infection also induces the expression of AP-1 (Baylis *et al.*, 1995; Chaussipied and Langsley, 1996). Moreover, the promoter region of IL-2 contains AP-1 binding sites (see review by Rao 1994). Previous studies have shown

that the parasite is responsible for inducing expression of c-Fos, c-Jun and JunD, components of the AP-1 complex, with JunD forming a major part of the transcriptionally active AP-1 complex (Baylis *et al.*, 1995; Chaussipied and Langsley, 1996). Taken together, it has been suggested that after prolonged culture, changes in parasite gene expression cause attenuation, in part, by modulating the expression of bovine host genes such as cytokines (e.g. IL2) and virulence factors (e.g. MMP9) (Adamson *et al.*, 2000b), possibly via the induction of AP-1 and NF-κB (Chaussipied and Langsley, 1996). However, it is likely that these events are not directly linked to proliferation, although they might influence proliferation ability, because there is no absolute difference between low and high passage cells in their ability to divide.

1.9.4 TashAT Genes

To date, most studies have only identified secondary events that lead to proliferation after *Theileria* infects the host cell, and little is known of how the parasite directly induces and modulates host cell growth and division. Host cell proliferation has been shown to be reduced during differentiation to the merozoite (Hulliger *et al.*, 1966; Shiels *et al.*, 1994, 1997) and it has been suggested that during this process, parasite factors controlling host cell division are down regulated (Carrington *et al.*, 1995). Cloning genes that are down regulated during merogony could identify these parasite factors which induce host cell proliferation. One group of parasite candidates that might modulate host cell gene expression and hence proliferation could be the *TashAT* gene family, which were shown to be down regulated during merogony at the mRNA level (Swan *et al.*, 1999). Northern blot analysis indicated there were likely to be three members of the *TashAT* gene family. However, *TashAT2*, was the only member to be fully isolated and characterised at the time of this study. *TashAT1*, the first *TashAT* gene to be identified, was only was partially characterised.

Sequence analysis of *TashAT1* and *TashAT2* showed they both encoded AT hook motifs. The AT hook motif was originally found in HMG proteins but has since been identified in a wide range of transcription factors with other DNA binding function (see reviews by Aravind and Landsman, 1998). This motif is comprised of mainly basic amino acids, with a core consensus sequence of Arginine-Glycine-Arginine-Proline (R-G-R-P), flanked by positively charged lysine and arginine residues. These amino acids are known to bind preferentially to A and T residues on the minor groove track of DNA (Bustin *et al.*, 1990; Reeves, 2001). In solution, the AT hook has very little α helical structure or β sheets, and

is mainly composed of random coiled structures (Reeves and Nissen,1990; Slama-Schwok *et al.*, 2000; Schwanbeck *et al.*, 2001). However, the AT hook forms an arc when bound to DNA by virtue of the Pro residues at close proximity to each other, whilst the Arginine turn or hook at the amino terminus holds the molecule rigid (Reeves and Nissen, 1990). HMGI molecules have been found to bind to distorted DNA structures (e.g. cruciform or bent DNA) and also induce conformational changes to the DNA structure such as DNA bending (reviewed by Reeves, 2001). Thus, it was postulated that the presence of AT hook domains within predicted TashAT1 and TashAT2 polypeptides suggested a DNA binding role for this gene family.

Previous sequence comparisons between the predicted polypeptide of TashAT2 with the SWISS-PROT database showed the AT hooks of TashAT2 were most similar to those of the HMGI(Y) group of proteins (Swan *et al.*, 1999). These genes have important roles in chromatin structure and also activate gene transcription, in association with other factors (Reeves, 2001). HMGI(Y) binding sites are present in genes whose products have been implicated in the transcriptional activation of mostly positively regulated genes, with the exception of interleukin-4 (Chuvpilo *et al.*, 1993). Many of these genes are associated with the immune system and cell growth, such as lymphotoxin and TNF-β (Fashena *et al.*, 1992), human papovavirus JC genes (Leger, *et al.*, 1995), the mixed lineage leukaemia (MLL) gene (Ernst, *et al.*, 2001), the α subunit of the IL-2 receptor (IL-2Rα) (Reeves *et al.*, 2000), the human insulin receptor (Brunetti *et al.*, 2001); IFN-β (Thanos and Maniatis, 1992) and c-Fos (Chin *et al.*, 1998). Studies have shown that HMGI(Y) proteins also bind directly to a number of transcription factors, such as the leucine zipper region of activating transcription factor 2, which enables the complex to bind to the IFN- β promoter (Du and Maniatis, 1994). A number of studies have linked HMGI(Y) proteins with cell division: HMGI(Y) proteins are known to stimulate adipocyte cell growth and pre-adipocytic cell differentiation (Melillo *et al.*, 2001). By contrast, HMGI(Y) levels were found to be low or undetectable in other fully differentiated cells (Lundberg *et al*, 1989; Bustin and Reeves, 1996). Abnormally high levels of HMGI(Y) have consistently been found to be associated with neoplastic cellular transformation and tumour formation in humans and rats and is thought to be due to the inappropriate activation of genes involved in cell growth (Tallini and Dal Cin, 1999; Beckerbauer *et al.*, 2000; Reeves, 2001; Berlingieri *et al.*, 1995). The similarity between the AT hook domains in HMGI(Y) and TashAT2 predicted polypeptide led to the postulation that TashAT2 and possibly TashAT1 might have a similar function to that of the HMGI(Y) family in modulating host cell gene expression or even proliferation.

HMG I(Y) proteins themselves are known to be regulated by environmental stimuli, chemical stimuli, and by a large number of transcription factors involved in cell division, including AP-1, c-Myc, Epidermal growth factor (EGF), transforming growth factor α (TGF-α) and platelet-derived growth factor (PDGF). Some HMGI(Y) proteins are also strongly activated by members of the PI-3 kinase, the Ras/MAP kinase and the oxidative stress signalling pathway (Ayoubi *et al.*, 1999; Zentner *et al.*, 2001). Over expression of c-Myc or HMGI(Y) has been shown to lead to cancerous phenotypes. Phorbol esters, which stimulate both AP-1 and HMGI(Y), induced neoplastic transformation in some mouse epithelial cell lines (Cmarik *et al.*, 1998). As some transcriptional activators of HMGI(Y) are also associated with *Theileria* infections (such as AP-1), it could be conceivable that the *TashAT* genes may also be regulated by similar transcription factors in the infected host cell.

Sequence analysis of the predicted polypeptide sequence of TashAT2 suggested that it acts on host DNA to modulate host gene expression. The potential TashAT2 polypeptide also contains nuclear localisation signals which are capable of transporting proteins to the nucleus (reviewed by Whiteside and Goodbourn, 1993). In addition, TashAT2 has a potential transcriptional activation domain, which have been found to assist in transcription through protein-protein interactions (Triezenberg, 1995). This suggestion was supported by evidence that TashAT2 was located to the host nucleus of infected cells using antisera generated against the AT- hook encoding region and another distinct region of TashAT2 by IFAT (Swan *et al.*, 1999). Moreover, transient transfection of a recombinant TashAT2 construct into COS7 cells resulted in host nuclear fluorescence, demonstrating that TashAT2 has the structural information required for transport into the host nucleus. This, together with protein translation inhibition and immunoprecipitation studies provided strong evidence that TashAT2 was translocated to the host nucleus.

The aims of this project were to identify and characterise the full length *TashAT1* gene, and compare it to *TashAT2* sequence and determine any similarity. This project will also attempt to characterise the expression of TashAT1, and other TashAT gene products in cells that are enhanced (D7) or attenuated (D7B12) with respect to differentiation. Since attenuated infected cells are associated with a lack of differentiation and increased proliferation in the host cell (Shiels *et al.*, 1992), an increase in *TashAT* gene expression levels might suggest these genes modulate host cell proliferation or parasite differentiation. To elucidate the potential function of the TashAT1 polypeptide, IFAT will be conducted to locate this gene product in the infected host cell. A host location might support the theory

that TashAT1 modulates host gene expression. The expression profile of TashAT1 and TashAT2 will also be compared in differentiating cells to determine *TashAT* gene regulation: similar expression profiles might suggest some form of common regulation. Finally, any potential regulatory motifs upstream of the *TashAT* genes will be analysed for DNA binding under differentiation conditions. Ultimately this might lead to the identification of factor(s) that regulate macroschizont differentiation or host cell proliferation.

2. Materials and Methods

2.1 Materials

2.1.1 General buffers and reagents

- EDTA: Ethylenediaminetetra-acetic acid disodium salt (BDH), in distilled water and adjusted to pH 8.0

- Tris-cl: Tris(hydroxymethyl)aminomethane (ICN Biochemicals) in distilled water adjusted to the appropriate pH with concentrated HCl (Sigma)

- 50x TAE stock: 242g Tris base, 57.1ml glacial acetic acid (Sigma), 100ml in EDTA (pH 8.0) made up to 1l with distilled water.

- 10x PBS (Phosphate buffered saline): 80g NaCl (Sigma), 2.01g KCl (Sigma), 6.1g Na_2HPO_4 (Sigma), 2g KH_2PO_4 (Sigma) adjusted to pH 7.1 and made up to 1l with distilled water.

2.1.2 Cell culture

- Supplemented RPMI-1640 media: RPMI-1640 media (with 25mM Hepes, L-glutamine) (Gibco, BRL), 0.05% $NaHCO_3$ (BDH), 8 $\mu g\ ml^{-1}$ ampicillin (Sigma), 8 units μl^{-1} fungizone (Gibco) with 10% Foetal Calf Serum (Sigma) or 20% myoclone super plus foetal bovine serum FCS (Gibco, BRL), filter sterilised.

2.1.3 Cell Protein Extraction

- 2 x Protein Sample buffer: 100mM Tris-Cl, pH 6.8, 4% SDS, 20% glycerol, 0.2% bromophenol blue, 200mM dithiothreitol in distilled water

- Solution A: 20mM PIPES (Piperazine-N,N'-bis-2-ethanesulphonic acid) (BDH), pH 7.5, 15mM NaCl, 60mM KCl, 14mM β mercaptoethanol (BDH), 0.5mM Ethylene glycol-bis(2-aminoethylether)-N,N,N',N'-tetraacetic acid (EGTA) (BDH), 4mM EDTA, 0.15mM Spermine (Sigma), 0.5mM Spermidine (Sigma) in distilled water.

- Solution C: 5mM HEPES (N-[2-hydroxyethyl] piperazine-N'-[2-ethanesulphonic acid) (Sigma), 1.5mM $MgCl_2$ (Sigma), 0.2mM EDTA, 0.5mM dithiothreitol (Gibco,BRL), 0.5mM PMSF (Sigma), 26% glycerol (BDH) in distilled water.

2.1.4 Bacterial and λ phage Culture

- LB medium: 1% bacto-tryptone (Difco), 0.5% bacto-yeast extract (Difco), 1% NaCl in distilled water.

- LB Top agarose: LB with 1.5% agarose (Gibco,BRL) in distilled water.

- LB Agar: LB with 1% bactoagar (Difco) in distilled water.

- Supplemented (λ phage) LB agar: LB Agar plus 0.3% glucose (Sigma), 0.075mM $CaCl_2$, 0.004mM $FeCl_2$ and 2mM $MgSO_4$ in distilled water.

- SM buffer: 50mM Tris-Cl, pH 7.5, 10mM $MgSO_4$ (Sigma), 100mM NaCl, 0.01% gelatin (Sigma) in distilled water.

2.1.5 Bacterial Cell Transformation

- TYM medium: 2% tryptone, 0.5% bacto-yeast extract, 0.1M NaCl, 10mM $MgSO_4.7H_2O$ in distilled water.

- Transformation buffer I: 30mM potassium acetate (Sigma), 50mM $MnCl_2$ (BDH), 100mM KCl, 10 mM $CaCl_2$ (Sigma), 15% glycerol in distilled water.

- Transformation buffer II: 10mM MOPS, 75mM $CaCl_2$, 10mM KCl and 15% glycerol in distilled water.

- X-gal/IPTG solution: 5mg ml^{-1} of X-gal (5-bromo-4 chloro 3 indolyl-β-D galactosidase) (Sigma), 25mM IPTG (isopropythio-β-D galactosidase) (Sigma) in Dimethylformamide (BDH) stored in glass container covered with foil.

2.1.6 DNA Extraction

- TE buffer: 10mM Tris-Cl, pH 7.5,1mM EDTA in distilled water.
- 1 x SSC solution: 0.15M NaCl, 0.015M trisodium citrate (Sigma) in distilled water.
- Genomic lysis solution: 100mM Tris-Cl (pH 7.5), 100mM NaCl, 10mM EDTA and 1% sodium lauryl sarkosyl (Sigma) in distilled water.
- Resuspension solution: 50mM Tris-Cl, pH 7.5, 10mM EDTA, 100 $\mu g\ ml^{-1}$ of RNAase A (Sigma) in distilled water.
- Cell Lysis solution: 0.2M NaOH and 1% SDS (Sodium dodecyl lauryl sulphate) (Sigma) in distilled water.
- Neutralisation solution: 2.55M potassium acetate, pH 4.8 in distilled water.

2.1.7 Polymerase Chain Reaction (PCR) and DNA Modification

- PCR buffer: 10mM Tris-Cl (pH 8.3), 50mM KCl in distilled water.
- 10x Alkaline Phosphatase buffer: 10mM $ZnCl_2$ (Sigma), 10mM $MgCl_2$, 100mM Tris-Cl (pH 8.3).

2.1.8 DNA Electrophoresis and Southern Blot Analysis

- Loading buffer: 0.25 % bromophenol blue (Electron), 0.25% xylene cyanol (Sigma), 40% (w/v) sucrose (Sigma) in distilled water.
- 10mg ml^{-1} Ethidium Bromide (Sigma) solution, dissolved in distilled water, kept in darkness.
- Denaturation solution: 1.5M NaCl, 0.5M NaOH (Sigma) in distilled water.
- Neutralisation solution: 2M NaCl, 0.5M Tris-Cl, pH 7.4 in distilled water.

2.1.9 Radioactive Labelling

- STE buffer: 0.1M NaCl, 10mM Tris-Cl (pH 8.0), 1mM EDTA (pH 8.0) in distilled water.

2.1.10 Nucleic Acid Hybridisation

- Hybridisation solution: 250mM sodium phosphate buffer adjusted to pH 7.2 with 85% orthophosphoric acid (BDH), 1mM EDTA, 7% SDS in distilled water.

2.1.11 RNA Extraction, Electrophoresis and Northern Blot Analysis

- Phosphate Transfer buffer: 25mM Na_2HPO_4, adjusted to pH 5.5 with 25mM NaH_2PO_4 in distilled water.

- 5x MOPS solution: 0.2M MOPS (3-Morpholinopropanesulfonic acid) (Sigma), 50mM sodium acetate, 5mM EDTA, pH 7.0 in distilled water.

- RNA loading dye: 95% formamide (BDH), 20mM EDTA, pH 7.6, 0.05%. bromophenol blue, 0.05% xylene cyanol FF in distilled water.

- 10mg ml^{-1} Ethidium Bromide in distilled water.

2.1.12 DNA Sequencing

- Sequencing stop solution: 98% formamide, 10mM EDTA, pH 8.0, 0.025% bromophenol blue, 0.025% xylene cyanol FF in distilled water.

- 10X TBE stock: 162g Tris base stock, 27.5g boric acid (Sigma), 9.3g of EDTA, made up to 1l with distilled water.

- 10% APS: 0.1g (ammonium persulphate) (Sigma) in 1ml of distilled water.

- TEMED (tetra-methyl-1,2-diaminoethane) (Sigma)

2.1.13 IFAT and Western Blot Analysis

- DABCO/glycerol: 50% glycerol with 2.5% (w/v) DABCO (1,4 diazabicyclo 2.2.2 octane) (Sigma), pH 8.0 in distilled water.

- DAPI stain solution: DABCO/glycerol with 1mg ml $^{-1}$ DAPI (4,6-diamidino-2-phenylindole) (Sigma) and 1mg ml^{-1} p-phenydiamine (Sigma) in distilled water.

- SDS-PAGE electrophoresis tank buffer: 50mM Tris, 384mM glycine, 2mM EDTA and 1% SDS in distilled water.

- Coomassie Blue staining solution: 0.001% Coomassie Brilliant Blue R-250 (Sigma), 10% acetic acid, 40% methanol (BWR international) in distilled water.

- Destaining solution: 10% acetic acid, 40% methanol in distilled water.

- Transfer buffer: 25mM Tris, 192mM glycine (ICN biochemicals), 20% methanol in distilled water.

- Protein block buffer: 10mM Tris-Cl, pH 7.4, 150mM NaCl, 0.1% Tween-20 (Sigma), 5% non-fat milk powder, 10% horse serum (Sigma) in distilled water.

- Protein wash buffer: 10mM Tris-Cl, pH 7.4, 150mM NaCl, 0.1% Tween-20 in distilled water.

2.1.14 Electromobility Shift Assay (EMSA)

- EMSA binding buffer: 10mM HEPES, pH 7.9, 1mM EDTA, 5% Ficoll (Sigma) in distilled water.

- 25% APS: 0.25g APS in 1ml distilled water.

2.2 Methods

2.2.1 Cell Culture

The cloned cell lines, D7 and D7B12 (Shiels *et al.*, 1992), derived from *T. annulata* (Ankara) (TaΛ2) macroschizont infected cell lines (Shiels *et al.*, 1992) were cultured at 37° C or 41°C in the presence of 5% CO_2. Cell counts were measured using a haemocytometer and diluted to a cell density of 1.4 x 10^5 cells ml^{-1} every 48 hours. Cells were grown in supplemented RPMI-1640 media and 10% heat inactivated foetal calf serum (FCS) (Sigma) or 20% myoclone super plus foetal bovine serum FCS (Gibco, BRL) in the case of D7B12 cells. Uninfected Bovine Lymphosarcoma cells (BL20) (Morzaria *et al.*, 1982), Ta46Λ cells (Baylis *et al.*, 1992) and TBL20 cells, a *Theileria* infected cell line from the same lineage as BL20 cells (Shiels *et al.*, 1986) were cultured as for D7B12 cells at 37°C.

2.2.2 Cryopreservation and Recovery of Cells

To preserve cells in liquid N_2, 5 x 10^6 cells ml^{-1} were centrifuged at 400xg for 5 minutes and gently resuspended in 3ml of medium at 4°C containing 10% Dimethylsulphoxide (Sigma). The cells were divided into 1.5ml aliquots in cryotubes, wrapped in cotton wool and placed in a polystyrene box and stored at -70°C for 24 hours, after which they were transferred to liquid N_2 storage. Cryopreserved cells were recovered by rapidly thawing the vial at 37°C and then added to 10ml of pre-warmed medium. The cells were centrifuged at 400xg for 5 mins, given a further wash in 10ml of medium and resuspended in 5ml of supplemented medium. After being in culture at 37°C for 24 hours, the cells were checked for viability and the culture volume brought up to 10ml with supplemented medium.

2.2.3 Staining of Cells with Giemsa Reagent

100 µl of cells at an approximate density of 5 x 10^5 cells ml^{-1} were centrifuged in a Shandon cytospin 2 centrifuge at 240xg for 5 minutes onto superfrost glass slides (BDH). The cells were air dried at 37°C for 5 minutes, fixed with methanol for 30 minutes and then stained in 4% Giemsa's stain (Gurr's improved R66, BDH) diluted in distilled water for 40 minutes. The stained cells were rinsed with distilled water and viewed under oil under x 100 objective after drying.

2.2.4 Preparation of Cell and Nuclear Extracts

Total cell extracts were prepared by pelleting approximately 10^7 cells by centrifugation at 400xg for 5 minutes. The cells were washed three times in an equal volume of PBS, resuspended in 0.5ml of PBS and lysed with 0.5 volumes of 2x protein sample buffer. The lysate was sheared by a 23 gauge syringe to break down the genomic DNA.

Preparation of host and parasite nuclei was modified from extraction methods developed for *Plasmodium* by Lanzer *et al.* (1992). 10^7-10^8 cells were pelleted by centrifugation at 400xg, rinsed in 1xPBS, and resuspended in 6ml nuclear extract solution A, to which 0.125mM PMSF (Phenylmethylsuphonyl fluoride) was freshly added. The cells were homogenised in a dounce homogeniser with a tight pestle for 6 strokes on ice and centrifuged at 400xg for 5 minutes at 4°C to pellet the host nuclei. The supernatant was decanted and re-centrifuged as before to remove any remaining host nuclei. The final supernatant was centrifuged at 2000xg for 10 minutes at 4°C to pellet the parasite nuclei. The host and parasite pellets were resuspended in 300 µl and 160µl respectively of nuclear extract solution C, to which NaCl was added to a total concentration of 300mM. The extract was incubated on ice for 30 minutes and then centrifuged at 15800xg for 10 minutes to remove any insoluble material. The supernatant containing the nuclear extracts were either snap frozen in dry ice and stored in 10µl aliquots in liquid N_2 or SDS protein loading buffer was added for SDS-PAGE analysis (see section 2.2.27). All samples were stored at -20°C. The Protein concentration of diluted nuclear extract was estimated by the Warburg-Christian Method (1941/1942 cited by Seidman and Moore, 2000), by measuring the sample at $O.D_{280}$ and $O.D_{260}$ in a quartz cuvette and accounting for nucleic acid contamination using the following formula:

[Protein] $\cong 1.55A_{280} - 0.757\ A_{260}$ (in units of mg/ml).

2.2.5 Culturing and Storing Bacteria

The *Escherichia coli* strains used were JM109 (Yanisch-Perron *et al.*,1985) and XL1-Blue (Bullock *et al.*, 1987), obtained from Promega and Stratagene, respectively. These strains were grown in LB medium at 37°C in an orbital incubator or streaked out on LB agar plates. These strains and derivatives containing recombinant plasmids were stored at -70°C in LB medium containing 25% glycerol. Cells were recovered from glycerol stocks by streaking out onto an LB agar plate, and grown overnight at 37°C. Single colonies were

then grown up in LB medium.

2.2.6 Transformation of E.coli Competent Cells

To prepare competent cells for transformation, 1ml of JM109 or XL1-Blue *E. coli* cells, previously grown overnight at 37°C were used to inoculate a 2l flask containing 100ml of TYM medium. This culture was grown for approximately 2 hours until mid-log phase (optical density reading at 600nm ($O.D._{600}$) of 0.5-0.9). At this point, TYM broth was added to the culture to a final volume of 500ml and the cells were grown until the culture reached an $O.D._{600}$ value of 0.6, whereupon the culture was cooled in ice water. The cells were pelleted by centrifugation at 4200xg for 15 minutes (at 4°C), resuspended in 100ml of cooled transformation buffer I and centrifuged again at 4200xg for 8 minutes. The pelleted competent cells were resuspended in 20ml of cold transformation buffer II and dispensed in 0.1-0.5ml aliquots in pre-chilled microfuge tubes, frozen in liquid N_2 and stored at -70°C.

For the transformation process, competent cells were thawed at room temperature until they started to melt and then placed on ice. 5-10 µl of ligated DNA was added to 100µl of competent cells and incubated for 30 minutes. The cells were then subjected to heat shock at 42°C for 1 minute and returned to ice immediately. After cooling, 600µl of LB medium was added to the transformed cells and incubated for 90 minutes in a 37°C waterbath. Transformed cells were plated onto LB agar plates supplemented with 50µg ml^{-1} of ampicillin, which were previously spread with 150µl of X-gal/IPTG solution. The plates were covered in plastic wrap, to prevent moisture loss, and incubated at 37°C over night. Recombinant colonies containing the sub-clones of interest were chosen on the basis of blue/white colony selection, according to the manufacturer's protocol for XL1-Blue or JM109 cell transformation (Stratagene or Promega). White colonies were picked with a sterile metal loop and re-streaked onto fresh ampicillin plates for maintenance and stored at 4°C and/or kept as glycerol stocks (see section 2.2.5).

2.2.7 Expression of GST-Fusion Proteins in E.coli Cells

Induction of GST-fusion protein expression in XL1-Blue *E.coli* cells were performed according to the suppliers instructions (Pharmacia) either at 37°C or 30°C for up to 4 hours.

2.2.8 Purification of the TashAT1-N Fusion Protein

The induced bacterial cell culture was pelleted by centrifugation at 7700xg for 10 minutes at 4°C, the supernatant discarded and the cell pellet placed on ice. The pellet was resuspended in 10ml of 1x PBS (see section 2.1.1) and the cells lysed by 6 cycles of sonication on ice (Soniprep 150, MSE/Sanyo) at amplitude 7; Lysis was deemed successful by a partial clearance of the cell suspension. 20% Triton X-100 (in 1x PBS) was then added to a final concentration of 1% and the cell debris was pelleted by centrifugation at 12000xg for 10 minutes at 4°C where the supernatant, containing the fusion protein was transferred to a fresh tube. Purification of the fusion protein was performed on a prepared Glutathione Sepharose 4B column according to the manufacturer's instructions (Pharmacia). The purified fusion protein eluates were pooled together and dialysed against 4l of 1 x PBS in dialysis tubing (Medical International ltd), previously sterilised by boiling in 2% sodium bicarbonate, 1mM EDTA solution and rinsed. The Protein concentration was estimated by a method adapted from Bradford (1976), using the BCA Protein Assay Reagent kit (Pierce), according to the supplier's instructions. The concentration of the protein sample was determined by interpolation from the standard curve, generated from Bovine Serum Albumin (BSA) samples of known concentration versus their absorbance at 562nm (A_{562}).

2.2.9 Preparing High Titre λ dash II phage Stocks

Bacteriophage were grown according to the protocol developed by Promega for Wizard Lambda Preps DNA purification system. Preparation of high titre stocks of λ bacteriophage were carried out using plate lysis method, using XL1-Blue cells as a host culture using supplemented (λ phage) LB agar. The plate was incubated over night at 37°C and then inspected for confluent lysis. To determine the concentration of λ phage stock, plate lysates were prepared from serial dilutions of the λ phage stock (prepared in SM buffer). The number of plaque forming units (pfu) were counted and calculated according to the method described by Sambrook *et al.* (1989).

2.2.10 Preparation of λ dash II

Preparation of λ bacteriophage DNA was based on a method by Sambrook *et al.* (1989), using high titre stocks of λ bacteriophage clones (see section 2.2.9). 1ml of an overnight culture of XL1-blue cells were used to inoculated two 1l flasks, each containing 100ml of

LB medium supplemented with 10mM $MgSO_4$. The flasks were grown at 37°C in an orbital shaker until the cells were at a density of 2-3 x 10^8 cells ml^{-1} (in the growth phase). This density corresponded to an $O.D_{600}$ reading of 0.45-0.6, using plastic cuvettes. The XL1-Blue cells were infected with a total of 2-3 x 10^{10} pfu of bacteriophage stock and the flasks were shaken vigorously at 37°C in the orbital shaker at 250 revolutions per minute (rpm). Incubation continued until the cell suspension became clear. At this point 200µl of chloroform was added to the flasks and shaken for a further 15 minutes to release the remaining bacteriophage particles. DNAase I (Sigma) and RNAase A (Sigma) were added to a final concentration of 10µg ml^{-1} in the lysate and incubated for a further 30 minutes at 37°C. Phage particles were precipitated by the addition of NaCl to 4%, and PEG (6000) (Sigma) to 10% of the total volume at 4°C for 12 hours. The λ bacteriophage were pelleted by centrifugation at 5000xg for 15 minutes, the supernatant was removed and the pellet was resuspended in 0.5ml of SM buffer. λ bacteriophage DNA was purified using the Wizard Lambda Preps DNA purification system kit (Promega) according to the manufacturers instructions, using the vacuum manifold method.

2.2.11 Preparation of Genomic DNA

Approximately 2 x 10^8 *Theileria* infected cells cultured at 37°C were harvested by centrifugation at 400xg for 5 minutes and resuspended in 5ml of 1xSSC solution. The cells were then lysed for 2-3 minutes in an equal volume of lysis solution. To remove all protein, proteinase K (sigma) was added to a concentration of 100 µg ml^{-1} and incubated at 55°C for 2 hours. DNA was progressively extracted by Phenol (Sigma) and then by phenol/chloroform. This procedure involved the addition of phenol (equilibrated to pH 7.8-8.0) saturated with TE to the cell lysis solution in a 1:1 ratio of TE saturated phenol: cell lysis solution. The mixture was agitated vigorously, centrifuged for 10 seconds and the aqueous phase removed. The same procedure was repeated with phenol:chloroform (1:1) extraction but this time the aqueous phase of the phenol/chloroform extraction was mixed with chloroform (Sigma) alone to remove any excess phenol that could interfere with downstream enzymatic reactions. The top, aqueous layer of the chloroform extraction was removed and DNA was precipitated by the addition of 2.5x volume of 100% ethanol (pre-cooled at -20°C) and 0.1x volume of sodium acetate. The DNA was left to precipitate for 12 hours at -20°C and then centrifuged at 20800xg for 30 minutes for collection. The DNA pellet was washed in pre-cooled 70% ethanol, centrifuged at 20800xg for 10 minutes, and resuspended in a suitable volume of TE after drying under vacuum pressure and stored at 4

°C. The concentration was determined by measuring a diluted sample at $O.D_{260}$ in a quartz cuvette as described by Sambrook *et al.* (1989).

2.2.12 Preparation of Plasmid DNA

For small scale preparations of *E.coli* cells bearing the desired plasmid construct, 3ml cultures of LB medium, containing the antibiotic ampicillin (100mg ml^{-1} in distilled water), were inoculated with a single colony lifted using a wire loop and grown overnight at 37°C. The cells were pelleted by centrifugation at 20800xg for 30 seconds, resuspended in Resuspension solution, then lysed by gentle mixing with cell lysis solution. Lastly, the lysate was neutralised with 0.2ml of neutralisation solution followed by centrifugation at 20800xg for 20 minutes. The supernatant was decanted into a fresh tube and the DNA precipitated by the addition of 2 volumes of 100% ethanol. The precipitated DNA was pelleted by centrifugation at 15800xg for 30 minutes, washed in 70% Ethanol and re-pelleted as before. Excess ethanol was drained from the tube and then removed by drying at room temperature for approximately 5 minutes. The pellet was resuspended in 50 to 100µl distilled water or TE buffer and stored at -20°C. Larger quantities of plasmid DNA (400µg-1.00mg) were collected using the maxiprep kit (Qiagen) from 100-500ml of *E.coli* cell culture according to the manufacturer's instructions.

2.2.13 Restriction Digestion of DNA

DNA was digested with restriction enzymes using the buffers provided by the suppliers (Gibco, BRL or Promega). Typically, 0.2-1.0 µg plasmid or λ dash II DNA or 45µg of genomic DNA were used for each reaction. In general, 5 units of enzyme were used to digest 1 µg of DNA for 1 hour at the temperature recommended by the supplier.

2.2.14 Sub-cloning DNA

Following electrophoresis of PCR amplified (see section 2.2.15) or restriction digested DNA (see section 2.2.13), the DNA fragments of interest were isolated from low melting point agarose gel slices using GeneClean kit (qbiogen), according to the manufacturer's instructions. PCR products were ligated into the pGEM-T easy vector according to the manufacturer's instructions. Restriction digested DNA was ligated into pGEM7zf (Promega) or pGEX-2TK (Pharmacia) vectors, previously digested with the appropriate restriction enzymes and purified by GeneClean kit as described above. Ligation reactions

were typically performed by adding the following to a 20µl reaction mix and incubating for 12 hours at 16°C: 50ng of DNA insert, 50ng of digested vector, 2 µl of 10x ligation buffer (Gibco, BRL) and 100 units of T4 DNA ligase (Gibco, BRL).

2.2.15 Amplification of DNA products by the Polymerase Chain Reaction (PCR)

DNA fragments were amplified from the DNA template in a standard 50µl reaction mixture containing 1.25 units of high fidelity Taq DNA polymerase (amplitaq); DNA template:0.1-0.5µg for plasmid/λ dash II DNA or 0.5- 1µg for genomic DNA: 10pmol of 5' and 3' primers (synthesised commercially by Cruachem); 200 µM each dNTP (amplitaq); 2mM $MgCl_2$ and PCR buffer (amplitaq) at the appropriate dilution. The DNA amplification reactions were performed in a Perkin-Elmer 480 thermocycler. A typical reaction involved an initial denaturation step of 95°C for 5 minutes, followed by 30 thermocycling steps and finally a 10 minute extension step at 70°C. The parameters used for each cycle were typically: denaturation: 95°C for 1 minute, anneal: 55°C for 1 minute, extend: 72°C for 1 minute. However the annealing temperature (Ta) varied according to the melting temperature (T_m) of the primers used, and were calculated using the following formula:

$Tm = 4(G + C) + 2(A + T)^{o}C.$

The Ta was set at 5°C below the lowest Tm of the pair of primers to be used (Innis and Gelfand, 1990).

2.2.16 5' RACE (Rapid Amplification of cDNA Ends)

This method was employed to determine the transcription start site of *TashAT1* using the 5'RACE kit, version 2 (Gibco, BRL) and was performed according to the manufacturer's instructions. The 3' gene specific nested primers GSP1, 2 and 3 were named rsp1, 2 and 3 (see Appendix A). The first PCR reaction was performed in a Perkin Elmer thermocycler as follows: initial denaturation at 94°C for 1 minutes then 30 cycles of denaturation at 94°C for 1 minute, annealing at 55°C for 1 minute, extension at 72°C for 1 minute followed by a final extension at 70°C for 10 minutes. The second PCR amplification reaction was performed as described above, but in this case the annealing temperature was 45°C. The

product was electrophoresed and purified from an agarose gel using the GeneClean® kit and ligated into pGem T Easy vector (Promega) as described in section 2.2.14, and transformed into XL1-Blue cells (see section 2.2.6). DNA was purified from recombinant colonies using the GeneClean kit according the manufacturer's instructions and sequenced (see section 2.2.24).

2.2.17 Electrophoresis of DNA

DNA separation was carried out as described by Sambrook *et al.* (1989). The agarose gel (Gibco, BRL) was prepared by melting agarose powder in 1xTAE to a percentage determined by the size of DNA used, normally 0.7%. Once the agarose had cooled to 55°C, Ethidium bromide (Sigma) was added to a final concentration of 0.5 $\mu g\ ml^{-1}$, poured into a casting frame and left to solidify. Low melting point agarose (Sigma) was used when the DNA was to be isolated as a gel slice. In this case the low melting point gel was cast at 4°C and DNA electrophoresis was performed at 4°C to prevent the gel from melting. DNA samples were loaded into wells with an equal proportion of loading buffer. The gel was placed in an electrophoresis apparatus (Pharmacia or Gibco, BRL), covered in 1x TAE buffer and arranged so that the negatively charged DNA would run from the wells, nearest the negative cathode, towards the positive anode. Samples were electrophoresed until the dye front was one third from the end of the gel. DNA fragments were visualised as fluorescent bands under u.v. light. The DNA size was estimated by a comparison with the 1kb DNA marker ladder (Gibco, BRL), run alongside the sample. The size of the DNA restriction fragments were determined by interpolation from a calibration curve of $\log_{10}$ DNA molecular weight markers against distance migrated.

2.2.18 Southern Blotting

After electrophoresis, DNA was treated and transferred to Hybond-N (Amersham) by the following method. The gel was placed in denaturation solution for 45 minutes and then in neutralisation solution for 30 minutes. Finally, the gel was equilibrated in Phosphate Transfer buffer (see section 2.1.11) for another 30 minutes. DNA was transferred onto a nylon membrane in transfer buffer for at least 18 hours by capillary action as described by Southern (1975). The DNA was fixed to the membrane by exposure to 150 Joules of U.V. light radiation in a GS Gene Linker box (BioRad). The membrane was either stored in aluminium foil at -20°C or prepared for hybridisation with a radioactively labelled probe. Hybridisation was performed at 65°C over night with a probe radiolabelled by the random

priming method (see section 2.2.22) using the methods and reagents developed by Church and Gilbert (1984). After hybridisation, excess radioactive DNA was removed by washing the filter three times in 0.2 x SSC, 0.1% SDS solution every 20 minutes; the filter was exposed to autoradiographic film (Kodak X-Omat AR) in a cassette box at -70°C for an adequate time period. To visualise the hybridised bands, the film was submerged in developing solution for 3 minutes (Kodak) followed by 30 seconds in wash solution (3% acetic acid), and then 3 minutes in Ilford Hypam fixing solution (Kodak) in darkness. Afterwards, the film was rinsed in water for 5 minutes and dried.

2.2.19 Preparation of RNA

To purify total RNA, approximately 10^8 -10^9 cells were pelleted at 100xg, and washed twice in 10ml of 1x PBS buffer. After the final wash, RNA was purified from the cells using the Tri-Reagent (Sigma) according to the manufacturer's protocol, using equipment and reagents pre-treated with DEPC (di-ethyl pyrocarbonate) (Sigma). Once purified, the RNA was resuspended in 200µl of 0.5% SDS solution by incubation at 55°C for 10 minutes and used immediately or stored at -70°C. Poly $(A)^+$ RNA was donated by D.G. Swan. Quantification and purity of the RNA sample was determined by spectrophotometric methods as described by Sambrook *et al.* (1989).

2.2.20 Electrophoresis of RNA

RNA samples were electrophoresed on 1.2% agarose-formaldehyde gels using electrophoresis apparatus previously treated with DEPC in 1x MOPS buffer. The RNA gel was prepared by melting agarose (low EEO) (Sigma) in distilled water to a concentration of 1.2%. When the solution had cooled to 55°C, 0.2 volumes of 5x MOPS solution and 0.02 volumes of 37% formaldehyde (BDH) were added and the gel was cast. RNA samples plus the RNA marker ladder (Gibco, BRL) were prepared in the following solution: 10µg total RNA, 0.1 volumes of 5x MOPS, 0.18 volumes of 37% formaldehyde, 0.5 volumes of formamide. The RNA was denatured by incubation at 55°C for 15 minutes and rapidly cooled on ice. 0.2 volumes of RNA loading dye was added to the samples, which were then loaded onto the gel and electrophoresed at 30-40V for 16 hours with a peristaltic pump attached to recirculate the buffer. Electrophoresis was stopped when the first dye front had reached three quarters of the length of the gel. The RNA marker lane was removed and stained with 60µl ethidium bromide in 250ml of water for viewing under u.v. illumination.

2.2.21 Northern Blotting

Northern blotting of the remaining RNA was performed according to Sambrook *et al.* (1989) with Hybond-N membrane using 25mM phosphate transfer buffer, pH 5.5, for 18 hours. The RNA was fixed to the membrane as described in section 2.2.18. Hybridisation was carried out as described by Church and Gilbert (1984) at 55°C. The membrane was washed twice in either 1 x SSC (see section 2.1.6) or 0.1 x SSC with 0.1% SDS every 20 minutes and the filter exposed to autoradiographic film at -70°C which was developed as described in section 2.2.18. The size of the RNA species detected were determined by interpolation from a calibration curve of $\log_{10}$ RNA molecular weight markers against distance migrated.

To re-use radioactively labelled membranes, 1l of boiling 0.1% SDS solution was poured over the membrane, left for 30 minutes, rinsed and checked for radioactivity. The process was repeated until no radioactivity was detected on the membrane. Before being re-labelled with a second radioactive probe, the stripped membrane was exposed to autoradiographic film for a time period equivalent to that used for it's original exposure (or longer in the case of abundant RNA species) to check no residual label remained on the membrane.

2.2.22 Radioactive Labelling of DNA using the Random Primed Method

This method was used to prepare ^{32}P radioactively labelled probes for hybridisation with membranes containing DNA or RNA. The probe was prepared using the Random Primed DNA labelling kit (edition 7, Boehringer Mannheim) according to the manufacturer's instructions, using 50μCi of [α-^{32} P]CTP (ICN biochemicals). The desired DNA fragment was excised from a low melting point agarose gel, sterile, distilled water was added (3ml/g of gel) and the solution was heated for 7 minutes at 100°C to melt the gel and denature the DNA and was either cooled to 37°C for immediate use, or stored at -20°C. The DNA sample was denatured by heating the sample to 100°C for 5 minutes and then cooling the sample rapidly on ice before it was added to the reaction mix. After incubation at 37°C for 1 hour, 80μl of STE buffer was added to stop the reaction and the probe was denatured by boiling for 5 minutes.

2.2.23 Radioactive Labelling of DNA using the End Labelling Method

In order to label the 5` end of oligonucleotides with ^{32}P-γ-ATP, it was necessary to dephosphorylate the end termini of the oligonucleotide, using 1 unit of Alkaline Phosphatase (Sigma) for every µmol of oligonucleotide, plus 1µl of 10x alkaline phosphatase buffer in a total volume of 10µl. The reaction was incubated for 30 minutes at 37°C and the enzyme denatured by heat inactivation at 65°C for 1 hour in the presence of 3mM EDTA (pH 8.0). The oligonuclotide was then purified using GeneClean kit, according to the manufacturer's instructions. For the end labelling reaction, 10pmol of oligonucleotide was mixed with 1µl of 160 µCi/µl^{-1} of [γ-^{32}P]ATP (ICN biochemicals), 1µl of 10x polynucleotide kinase buffer (Promega), 1µl of T4 polynucleotide kinase (10 units µl^{-1}) (Promega) made up with distilled water to 10µl. The mixture was incubated at 37°C for 30 minutes, after which 40µl of STE was added and the probe purified on a NucTrap® column (Stratagene), inside a Beta Shield device (Stratagene), according to the manufacturer's instructions.

2.2.24 Densitometric Analysis

Densitometric scanning of autoradiographs were performed using a Quantity One scanner (pdI) at the CRC Beatson Institute (Glasgow), according to the manufacturer's instructions.

2.2.25 Automated DNA Sequencing

DNA samples were prepared by PCR amplification using infrared fluorophore (laser dye IRD41) labelled primers for detection by a LI-COR model 4000(L) DNA sequencer as described by Middendorf *et al.* (1992). The LI-COR sequencer contains a laser diode emitting at 785nm which causes excitation of the fluorophore labelled DNA fragments passing through the laser diode during electrophoresis, which are detected by a microscope/detector. The DNA sequence images were collated into one image directly on computer using ImagIR™ software.

The sequencing reactions were performed using the Fluoro Sequenase kit (Epicentre Technologies) in the following reaction mix: 50ng DNA, 2pmol IRD41 labelled primer (T7, T3 or SP6), 2.5µl of 10 x sequitherm Excel™II sequencing buffer (Epicentre

Technologies), 1µl of biopro thermostable DNA polymerase (Bioline) made up to 17 µl with distilled water. 4µl of reaction mix was added to each of four thermocycler tubes, each containing 2 µl of sequitherm Excel™II dideoxy-ATP, -CTP,-GTP,-TTP termination mix (Epicentre Technologies), respectively. 30 µl of mineral oil were added to each tube, which were placed in a Perkin Elmer thermocycler for a total of 30 cycles starting with a 95°C denaturation step for 5 minutes and then 30 seconds at 95°C, 30 seconds at 60°C (annealing), 1 minute at 70°C (elongation). After PCR amplification, the reactions were stopped with 4µl of sequencing stop solution and stored at 4°C in darkness. Just prior to loading, samples were denatured at 95°C for 5 minutes.

The acrylamide sequencing gel was prepared by adding 21g of urea to 6ml of 50% LongRanger™ mix (FMC) and 6ml of 10x TBE made up to total volume of 50ml with distilled water and degassed to remove any air bubbles. Polymerisation of the acrylamide gel was achieved with the addition of 25µl TEMED and 250 µl of 10% APS and cast using LI-COR plates. When the acrylamide was set the plates were assembled into the LI-COR sequencer as described in the manufacturer's instructions with 1 x TBE buffer used as electrophoresis buffer. Before the samples were loaded, the gel was pre-run for 30 minutes until the gel temperature reached 50°C, at which point the sequencing gel was loaded with 2µl of sample and electrophoresed at 1500V (35mA) for approximately 6.5 hours. The sequencing image and sequence was retrieved and manually edited on computer using the ImagIR™ data collection software set designed for the LI-COR model 4000 sequencer.

2.2.26 Sequence Analysis

DNA sequence information was transferred to a database and analysed using the Genetics Computer Group (GCG) Wisconsin package software, version 10.2. Sequence comparisons were performed using the FastA programme. The predicted polypeptide and nucleotide sequences of TashAT1 was compared to other polypeptide and nucleotide sequences using the programs BLAST (for polypeptides) and ENTREZ (for nucleotides) at the ncbi website (http://www.ncbi.nlm.nih.gov/). Preliminary sequence data from the *T. parva* genome project was obtained from The Institute of Genome Research (TIGR) website at http://www.tigr.org. The predicted molecular weight of TashAT1 and the fusion protein was obtained from the Predictprotein program, GCG. Potential motifs and secondary structure predictions were analysed using the following software programs:

PSORT II	http://psort.nibb.ac.jp/
iPSORT	http://hypothesiscreator.net/iPSORT/
SMART	http://smart.embl-heidelberg.de/
PROSITE	http://www.expasy.ch/prosite/
SignalP and TargetP	http://www.cbs.dtu.dk/services/
ProDom	http://prodes.toulouse.inra.fr/prodom/doc/prodom.html
PESTfind	http://www.at.embnet.org/embnet/tools/bio/PESTfind/
SEG and PHD analysis	http://dodo.cpmc.columbia.edu/pp/predictprotein.html
MatInspector	http://transfac.gbf.de/TRANSFAC.

2.2.27 Indirect Fluorescence Antibody Test (IFAT)

Cell cultures were grown to a density of 1 x 10^6 cells/ml and prepared on slides as described in section 2.2.3. The cells were fixed onto the slides by incubation with acetone at -20°C for 30 minutes and left to equilibrate at room temperature for 5 minutes. 20 µl of the first antibody (see section 2.2.30), diluted in RPMI-1640 media or neat, was applied to the fixed cells and incubated at room temperature in a humid chamber for 30 minutes. The slides were washed three times in 1 x PBS and then dried at room temperature for 5 minutes. 20 µl of FITC (fluorescein isothiocyanate) conjugated rabbit anti-species IgG (Sigma), diluted by 1:200, was applied to the cells (which generates green fluorescence under the microscope), incubated and washed as before. The slides were placed in 0.1% of Evans Blue in 1x PBS for 2 minutes to counter stain the cells (red under microscope) and then rinsed in PBS. After drying, the cells were mounted in DABCO/glycerol solution or DAPI stain and examined under x 60 or x 100 objective for fluorescence with a Leitz ortholux II transmitted light fluorescence microscope. Photographs of the images were taken by an Olympus BX60 fluorescence camera, attached to a SPOT digital camera (Diagnostics Instruments Inc), where images were digitally loaded onto a computer and viewed using the SPOT basic program.

2.2.28 SDS-Polyacrylamide Gel Electrophoresis (PAGE)

This method was developed by Laemmli (1970) and performed in Mini Protean II gel electrophoresis apparatus (BioRad) using mini-protean II gel plates. The resolving gel varied in acrylamide concentration according to the size of protein to be detected; normally an 8% resolving gel was used. A 10ml resolving gel contained 2.7ml of 30% acrylamide/0.8% N-N Bis-methylene acrylamide (w/v) mix (Scotlab) in 2.5ml of 1M Tris-Cl, pH 8.8 and 0.1ml of 10% SDS, 4.6ml of distilled water. The gel was polymerised with the addition of 0.5ml of freshly prepared 10% APS and 0.5ml of TEMED (see section 2.1.12), and immediately poured into the casting plates to three quarters of the height of the

gel plates (approximately 6ml gel mix). To prevent uneven polymerisation of the gel at the air interface, the gel was overlayed with water saturated 2-butanol and polymerisation was complete by 30 minutes. Once the water saturated 2-butanol was poured off and the resolving gel rinsed with distilled water, the remaining quarter of the gel plates were used for the 4% acrylamide stacking gel, which was poured on top of the resolving gel. Preparation of 5ml of stack gel was made as follows: 0.67ml of 30% acrylamide/0.8% N-N Bis-methylene acrylamide mix (Scotlab), 1.25ml of 0.5M Tris-Cl, pH 6.8 and 0.05 % SDS and 3.5ml distilled water. This gel was polymerised by the addition of 0.025ml of 10% APS and 0.005ml TEMED. Approximately 2ml of stacking gel was required for casting in the mini protean II plates. The gel apparatus was transferred to the mini protean II gel electrophoresis tank, and filled with electrophoresis tank buffer. In order to prepare the samples and molecular weight markers for loading, 0.5 volumes of 2 x protein sample buffer was added to protein samples, boiled for 5 minutes and loaded immediately into the wells. Electrophoresis of polypeptide samples were carried out for 2 hours at 100V and either processed for Western blotting or stained with Coomassie Blue staining solution for 2 hours to visualise the protein bands. In the latter case, protein bands became visible when the coomassie blue was selectively removed using Destaining solution. The gel was dried under vacuum at 80°C for 2 hours in a BioRad gel dryer between clear plastic sheets for long term storage. The molecular mass of the polypeptide(s) of interest was determined by interpolation from a calibration curve of $\log_{10}$ molecular weight markers against distance migrated.

2.2.29 Western Blotting

This method was adapted from Towbin *et al.* (1979). After electrophoresis, the polypeptides were transferred from the SDS-PAGE gel onto a Protran nitrocellulose membrane (Schleicher and Schuell) in transfer buffer using a mini protean electroblotter apparatus (Bio-Rad) at 1V cm^{-2} for 90 minutes. The filter was stained with 0.2% Ponceau-S (Sigma) dissolved in 3% v/v trichloroacetic acid for 5 minutes, and then washed in distilled water, to check for transfer efficiency. The marker lane was removed and the filter was destained and incubated with block buffer for 1-3 hours. The primary antibody was diluted in a suitable volume of block buffer and incubated with the filter for 8-10 hours in a sealed bag on a rocking platform. The filter was washed of primary antibody solution twice with protein wash buffer at room temperature for 20 minutes and was incubated with protein block buffer for a further 20 minutes before incubation with the secondary antibody for a further hour. This was either alkaline-phosphatase conjugated anti-rabbit IgG (whole

molecule, Sigma) or peroxidase conjugated anti-rabbit IgG (Sigma), both used at 1: 10000 dilution in block buffer. The filters were washed as before. In the former case, detection of antibody binding was performed as described by Harlow and Lane (1988), by incubating the filter in bromochloroindoyl phosphate/tetrazolium (BCIP/NBT) which is converted to a purple substrate by alkaline phosphatase. When the polypeptides were sufficiently visible the reaction was stopped by rinsing with distilled water. Detection of antibody/antigen complexes bound to the peroxidase conjugated secondary antibody was performed by ECL (Amersham) according to the manufacturers instructions. When the substrate was added to the filter, chemiluminescence occurred at the secondary antibody binding site, which was detected on autoradiographic film at increasing time periods, depending on the strength of the signal.

2.2.30 Peptide Blocking

Peptide blocking was performed as described by Mottram *et al.* (1993). The fusion protein was added to primary antisera diluted in block buffer to an appropriate concentration required to bind all the antibodies, for 3 hours at room temperature. The blocked antisera was then incubated with the Western blot as described in section 2.2.28.

2.2.31 Generation of Antisera

Antisera was prepared by A. Tait and raised against the antigen in New Zealand White Rabbits. The antisera was collected from the rabbits and dispensed into 20µl and 50µl aliquots for storage at -20°C. Antisera DE-39 and EL-24 were prepared in the same manner as described by Swan *et al.* (1999).

2.2.32 Electromobility Shift Assays (EMSA)

DNA binding reactions with host or parasite nuclei were performed in a 20µl reaction volume containing 10µg of nuclear extract, 14µl of EMSA binding buffer, 1µl of 200ng ml^{-1} poly dG.dC.. 1µl of γ -^{32}P-dATP labelled probe was added and the reaction mix was incubated for 40 minutes at 4°C. Single stranded oligonucleotide (MWG) and its reverse complement were annealed to create a double stranded oligonucleotide by incubating 400pmol of each oligonucleotide in 100µl of 0.5x polynucleotide kinase buffer (Promega) at 80°C for 2 minutes and gradually cooling to room temperature.

3. Genetic Analysis of TashAT1 and TashAT3

3.1 Introduction

Previous studies of *Theileria* infected leukocytes demonstrated that this parasite had the ability to induce a transformed phenotype in the host (Brown, 1990), by some undefined process. In addition, Shiels *et al.* (1994) found that parasite genes are expressed in a stage specific manner. These findings led to the postulation that macroschizont factors that are down regulated during differentiation to the merozoite, may be involved in host cell division (Carrington *et al.*, 1995). Thus, characterisation of genes expressed in the macroschizont that are down regulated during merogony may lead to the identification of factors that stimulate host cell division.

Swan *et al.* (1999) performed a differential immunoscreen on polypeptides derived from genomic DNA from the *T. annulata* cell line, D7 in order to identify genes that are up regulated in the macroschizont. In this experiment, antigens encoded by constitutively expressed or merozoite specific genes were blocked with antisera raised against merozoite stage of the *T. annulata* life cycle. A bovine serum, previously shown to recognise macroschizont antigens was used to identify genes specifically expressed in the macroschizont. From this screen, a 3.4kb λgt11 clone, called cl-12, was isolated (Swan *et al.*, 1999), and preliminary sequence analysis identified a partial open reading frame (called *TashAT1*) with four AT hook motifs. However, when a λ dash II D7 genomic library was screened with cl-12, the resultant isolated gene, only showed 65% identity with *TashAT1* at the DNA level and contained only three AT hook motifs. It was therefore concluded that this new gene was not *TashAT1* and was called *TashAT2* (Swan *et al.*, 1999). Therefore, the first important aim of this project was to obtain the full-length *TashAT1* gene clone for further characterisation.

Previous sequence comparisons of the *TashAT2* (Swan *et al.*, 1999) showed that its AT hook motifs were most similar to those of HMGI(Y) protein, which have an important role in chromatin structure and act as co-factors, mainly to positively activate a wide range of genes (reviewed by Reeves, 2001). The similarity between the AT hook domains in HMGI(Y) suggested that the *TashAT* gene family in *Theileria* might also bind to DNA and even similar gene targets to HMGI(Y). This was supported by evidence of TashAT2 within the host nucleus of an infected cell. IFAT studies also found the TashAT2 polypeptide in

the nucleus of COS7 cells that were transfected with a TashAT2 construct, showing that the potential NLS of TashAT2 might be able to direct the polypeptide to the host nucleus (Swan *et al.* 1999). Thus a second aim would be to identify any sequence motifs within the TashAT1 predicted peptide, such as a potential NLS to deduce the possible function of TashAT1. Similarly, analysis of the upstream region of *TashAT1* could provide information of possible transcriptional regulators of *TashAT1*.

Work by Aravind and Landsman (1998), have classified all known AT hook encoding genes into three classes, based on their extended sequence conservation outwith the core GRP region, and on their binding affinities (see Table 3.1). Type I class AT hook domains contain the 9 amino acid residue consensus sequence together with four extra basic residues C terminal to the core GRP, which form a polar network to provide additional binding. Type II class AT hook domains do not posses these extra residues and therefore have a lower affinity to DNA than the Type I class of AT hook domains. A third class of AT hook domains, Type III, has properties of both Type I and II classes. The third aim was to identify which class the *TashAT* genes belong according to the above classification and thus predict the DNA binding capacity of TashAT1 and TashAT2 polypeptides.

Southern blot analysis of genomic DNA by Swan *et al.* (1999), using cl-12 as a probe, yielded multiple DNA fragments, some of which could not be attributed to the presence of *TashAT2*. This finding and the identification of two non-identical genes that both contain AT hook motifs suggested the presence of a small gene family in *T. annulata*. Thus the fourth aim was to physically map the locus of *TashAT1*, in relation to *TashAT2* and to detect any other *TashAT* genes.

3.2 Results

3.2.1 Identification of the TashAT1 and TashAT3 genes

Previous screening of a λ dash II library derived from D7 genomic DNA with cl-12, identified λ dash 12, and another λ dash clone, λ dash 13. To isolate *TashAT1*, two primers, HMG1 and HMG2, designed to flank the AT hook encoding sequence of *TashAT2* (see Appendix A), were used to amplify *TashAT1* from both λ dash clones. This produced a 340bp product in λ dash12 (shown in Fig. 3.1) which was known to correspond to the AT hook region of *TashAT2*. However, a novel 600bp PCR product, called p600, was found in λ dash 13 and could possibly represent the AT hook encoded region of *TashAT1*. To isolate *TashAT1*, the 600bp product was amplified from λ dash 13 and cloned into the pGEM7zf vector after agarose gel extraction (see section 2.2.14).

Sequence analysis of this clone (see Fig. 3.2), revealed that the p600 predicted peptide sequence was in fact 584bp in size and contained four AT hook domains. A comparison between the p600 fragment and the preliminary *TashAT1* ORF sequence, from cl-12 showed a 92.6% identity over the first 441bp (see Fig. 3.3), which, barring sequencing errors, might indicate that the p600 product originated from *TashAT1*.

To clone the full length *TashAT1* gene, λ dash 13 DNA was subjected to restriction digestion followed by Southern blotting using the p600 fragment as a probe. This analysis (shown in Fig. 3.4) identified several DNA fragments that were distinct from the *TashAT2* gene. These were a 0.8kb and a 1.2kb fragment from the EcoRI digestion; a 3.2kb SpeI fragment; a 12b KpnI/SpeI fragment and two HindIII fragments at approximately 3.5kb and 1.6kb. Two smeared bands at approximately 1.2kb and 0.8kb were also observed in the HindIII restriction digestions, which may be due to contamination from the neighbouring EcoRI digestion given their size and the high level of background radioactivity on the autoradiograph. Double restriction digestion of λ dash 13 DNA using EcoRI and three other restriction enzymes, did not abolish the EcoRI DNA fragments, which indicated that the EcoRI fragments were internal to the SpeI and HindIII and KpnI fragments. The 0.8kb and 1.2kb EcoRI fragments were therefore considered as possible candidates for the *TashAT1* gene, cloned into a pGEM7zf vector and sequenced.

The sequence results showed that the 0.8kb EcoRI fragment (called AThook1) contained four potential AT hook encoding domains and was 96.9% identical over the first 478bp to the p600 fragment (see Fig 3.5) and 97.1% identical to the entire *TashAT1* ORF (see Fig. 3.6). This evidence suggested that the AThook1 fragment was likely to belong to *TashAT1*.

To identify the origins of the 1.2kb fragment, a comparison was made between the partial, preliminary sequence of the 1.2kb fragment (910bp total length) and the *TashAT1* ORF (see Fig.3.7). This showed 89.8% identity over the first 412bp of the 1.2kb fragment, but none after that point, in contrast to the AThook1 fragment, which showed 100% identity with *TashAT1* ORF over the entire length of the clone. Therefore, the 1.2kb fragment, although similar, was thought to be distinct from the AThook1 fragment. A third sequence comparison between the 1.2kb EcoRI fragment and the p600 showed 97.1% identity to the p600 fragment from base pairs 80-588 of the p600 fragment (Fig. 3.8), whereas the AThook1 fragment showed 96.9% identity to p600 fragment between base pairs 148 to 620 of the p600 fragment. This confirmed that the 1.2kb EcoRI fragment (now called AThook3) and fragment AThook1 were very similar over the region covered by the p600 fragment, but may be divergent outwith this region.

To determine if the AThook3 fragment belonged to *TashAT2* (3.5kb in length), a sequence comparison between AThook3 and a fragment containing *TashAT2* was performed. The results (see Fig. 3.9) showed that there was 97.5% identity between the 1.2kb EcoRI fragment (from 280 to 910bp) and *TashAT2* (between 1450bp to 2075bp) which corresponds to 1372-1997bp from the translation start site of *TashAT2*. However, the restriction map generated from the *TashAT2* sequence (Swan *et al.*, 2001a) lacked a 1.2kb internal EcoRI fragment, indicating that this 1.2kb EcoRI fragment was not part of *TashAT2*. This suggested that the 1.2kb EcoRI fragment could belong to a third AT hook encoding gene that shared identity with different regions of both *TashAT1* and *TashAT2*.

In order to isolate *TashAT1*, focus was placed on the KpnI/SpeI restriction digestion of λ dash 13 in Fig. 3.4. The SpeI and KpnI single digestions produced the 3.2kb and 12b fragments, respectively. However double digestion of λ dash 13 DNA with EcoRI/SpeI and EcoRI/KpnI digested the SpeI and KpnI fragments, leaving a 0.8kb (AThook1) and a 1.2kb (AThook3) fragment. This suggested that the AThook1 and AThook3 fragments may be contained within the 12b KpnI or the 3.2kb SpeI fragment. However, size considerations meant that the 3.2kb SpeI fragment was the more feasible to clone and sequence, and was therefore chosen for further study.

Parallel studies by Swan and Phillips (unpublished, 1997) had previously purified and cloned the 3.2kb SpeI fragment in an attempt to isolate the third putative *TashAT3* gene. A series of overlapping deletions were made to sequence this clone but work was discontinued after initial sequencing revealed that this clone might be *TashAT1*. The 3.2kb SpeI fragment was used in this project thereafter to obtain the sequence of *TashAT1*.

To confirm that the 3.2kb SpeI fragment contained *TashAT1*, the 3.2kb SpeI fragment was subjected to restriction digestion to confirm if it contained the 0.8kb EcoRI fragment. BamHI and AatII or SphI were used to linearise the subcloned DNA, which was then digested with a variety of restriction enzymes, including EcoRI. Southern blotting, using the ATHook1 fragment as a probe, revealed an 800bp fragment (see Fig. 3.10, lane 5), confirming the presence of the ATHook1 fragment. This indicated that the 3.2kb SpeI fragment contained *TashAT1* and was called λTa1.

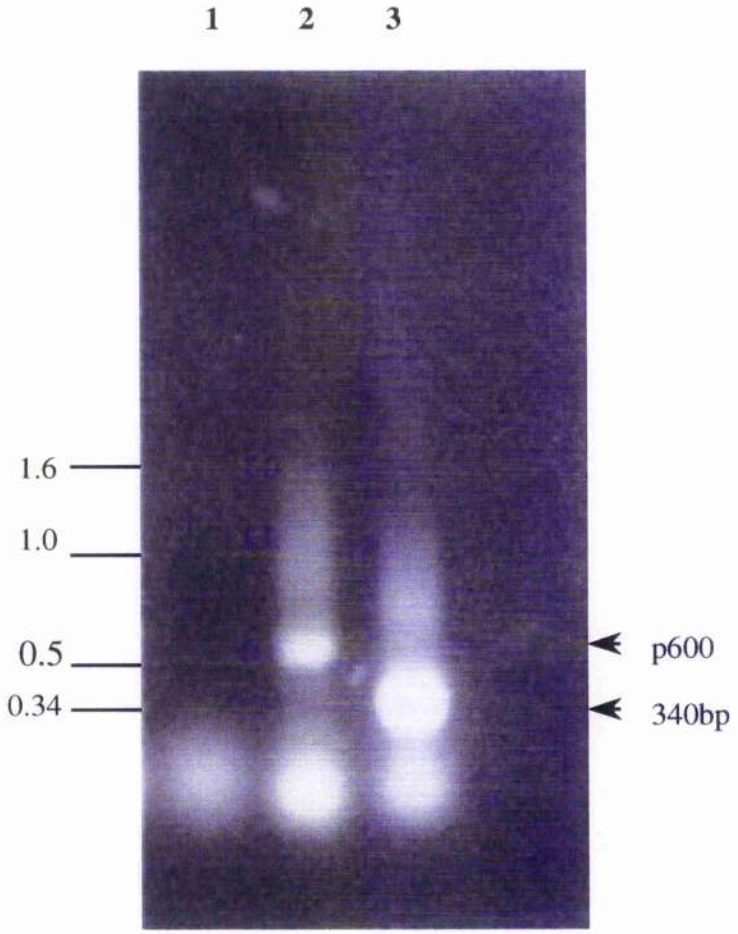
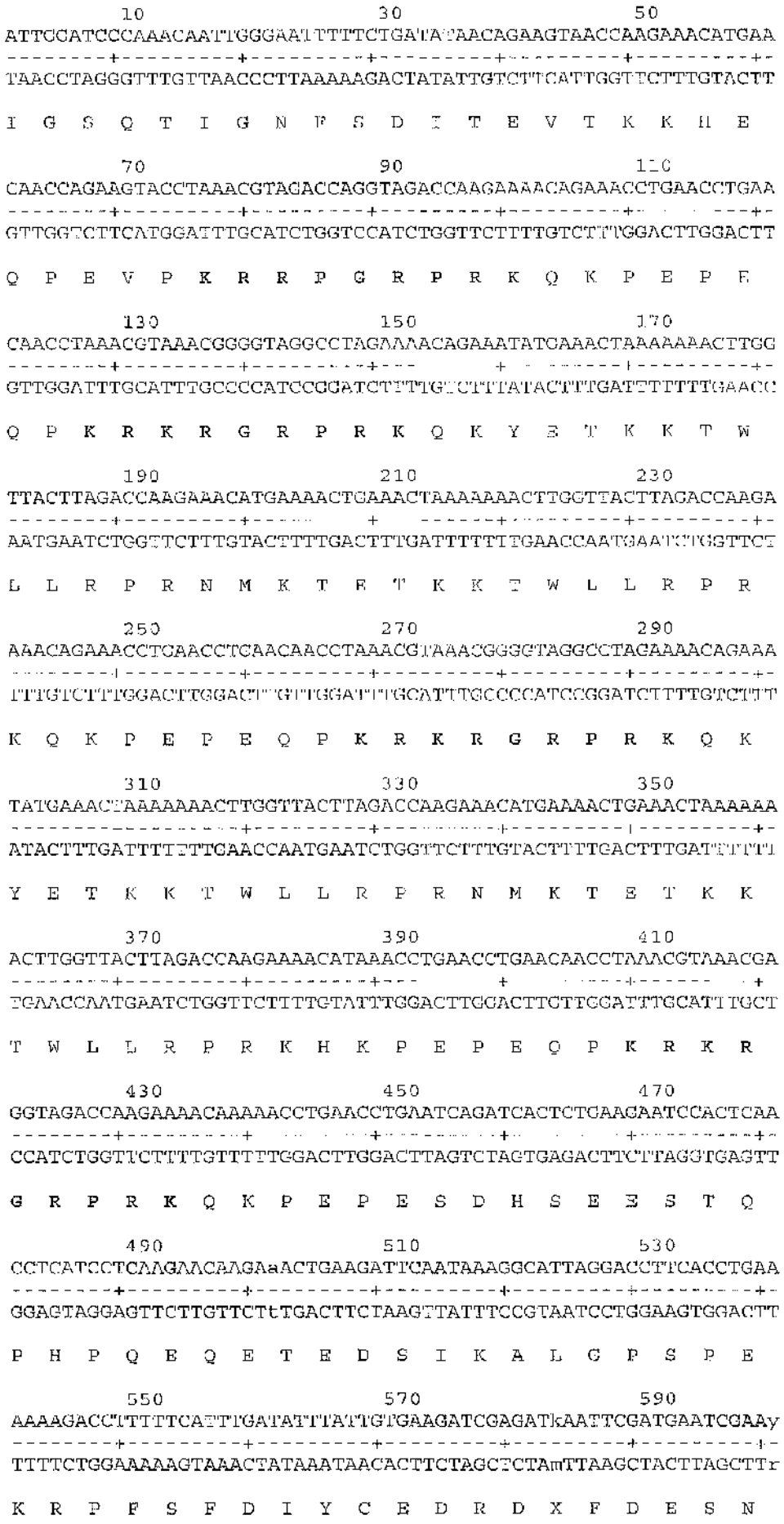


Fig. 3.1: PCR amplification of the 600bp PCR product (p600) from λ dash clone 13 using primers HMG1 and HMG2.
Lane 1: control PCR amplification with no DNA; lane 2: λ dash clone 13; lane 3: λ dash clone 12. Numbers of the left indicate the sizes of the DNA marker fragment (in Kb). Arrow heads indicate the PCR products detected.

Fig. 3.2: DNA and predicted amino acid sequence of the p600 product from λ dash 13. AT hook domains are indicated in bold type.


10 30 50
ATTCCATCCCAAACAATTGGGAATTTTTCTGATATAACAGAAGTAACCAAGAAACATGAA
TAACCTAGGGTTTGTTAACCCTTAAAAAGACTATATTGTCTTCATTGGTTCTTTGTACTT
I G S Q T I G N F S D I T E V T K K H E
70 90 110
CAACCAGAAGTACCTAAACGTAGACCAGGTAGACCAAGAAAACAGAAACCTGAACCTGAA
GTTGGTCTTCATGGATTTGCATCTGGTCCATCTGGTTCTTTTGTCTTTGGACTTGGACTT
Q P E V P K R R P G R P R K Q K P E P E
130 150 170
CAACCTAAACGTAAACGGGGTAGGCCTAGAAAACAGAAATATGAAACTAAAAAAACTTGG
GTTGGATTTGCATTTGCCCCATCCGGATCTTTTGTCTTTATACTTTGATTTTTTTGAACC
Q P K R K R G R P R K Q K Y E T K K T W
190 210 230
TTACTTAGACCAAGAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTAGACCAAGA
AATGAATCTGGTTCTTTGTACTTTTGACTTTGATTTTTTTGAACCAATGAATCTTGGTTCT
L L R P R N M K T E T K K T W L L R P R
250 270 290
AAACAGAAACCTGAACCTGAACAACCTAAACGTAAACGGGGTAGGCCTAGAAAACAGAAA
TTTGTCTTTGGACTTGGACTTGTTGGATTTGCATTTGCCCCATCCGGATCTTTTGTCTTT
K Q K P E P E Q P K R K R G R P R K Q K
310 330 350
TATGAAACTAAAAAAACTTGGTTACTTAGACCAAGAAACATGAAAACTGAAACTAAAAAA
ATACTTTGATTTTTTTGAACCAATGAATCTGGTTCTTTGTACTTTTGACTTTGATTTTTT
Y E T K K T W L L R P R N M K T E T K K
370 390 410
ACTTGGTTACTTAGACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAACGA
TGAACCAATGAATCTGGTTCTTTTGTATTTGGACTTGGACTTCTTGGATTTGCATTTGCT
T W L L R P R K H K P E P E Q P K R K R
430 450 470
GGTAGACCAAGAAAACAAAAACCTGAACCTGAATCAGATCACTCTGAAGAATCCACTCAA
CCATCTGGTTCTTTTGTTTTTGGACTTGGACTTAGTCTAGTGAGACTTCTTAGGTGAGTT
G R P R K Q K P E P E S D H S E E S T Q
490 510 530
CCTCATCCTCAAGAACAAGAaACTGAAGATTCAATAAAGGCATTAGGACCTTCACCTGAA
GGAGTAGGAGTTCTTGTTCTtTGACTTCTAAGTTATTTCCGTAATCCTGGAAGTGGACTT
P H P Q E Q E T E D S I K A L G P S P E
550 570 590
AAAAGACCtTTTTTCATTTGATATTTATTGTGAAGATCGAGATkAATTCGATGAATCGAAy
TTTTCTGGAAAAAGTAAACTATAAATAACACTTCTAGCTCTAmTTAAGCTACTTAGCTTr
K R P F S F D I Y C E D R D X F D E S N

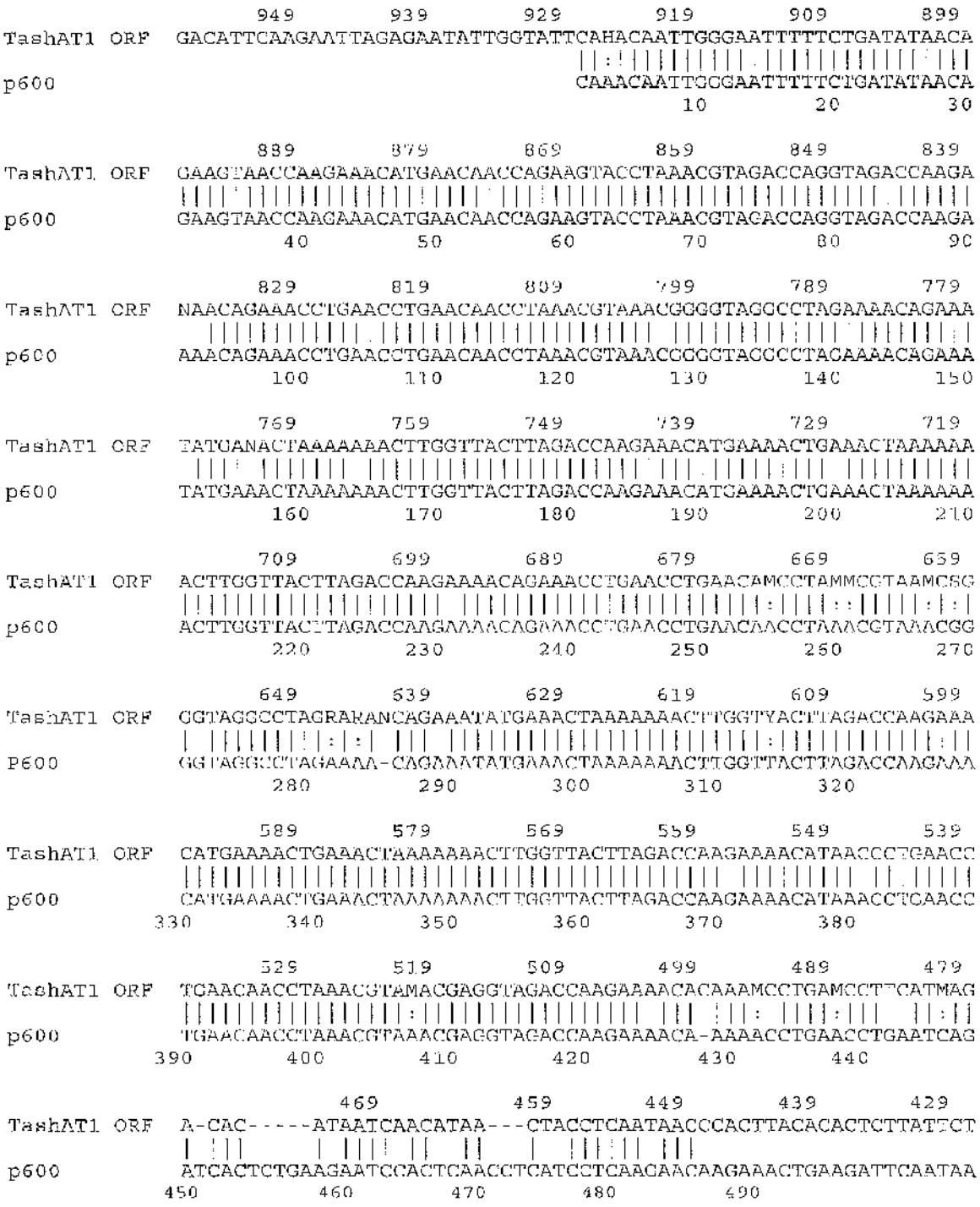


Fig. 3.3: Sequence comparison of the p600 PCR fragment with TashAT1 ORF derived from cl12: Scores 92.6% identity in a 489 nucleotide overlap

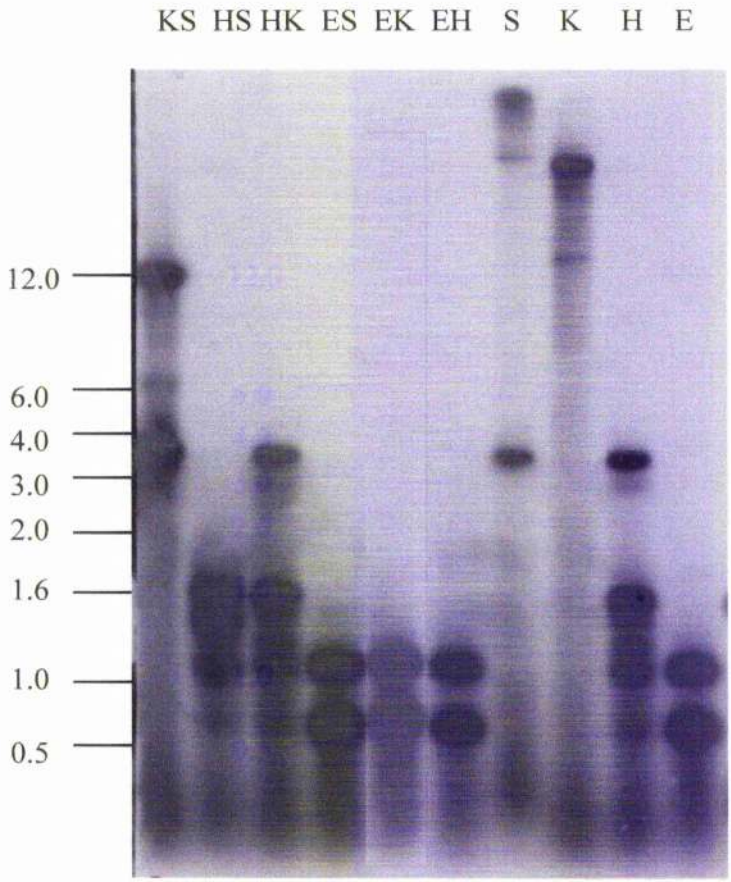


Fig. 3.4: Southern blot analysis of restriction digested λ dash 13 DNA hybridised with probe p600. E: EcoRI; H: HindIII; K: KpnI; S: SpeI. EK digestion performed on a separate blot. DNA marker sizes are indicated to the left of the figure (in kb).

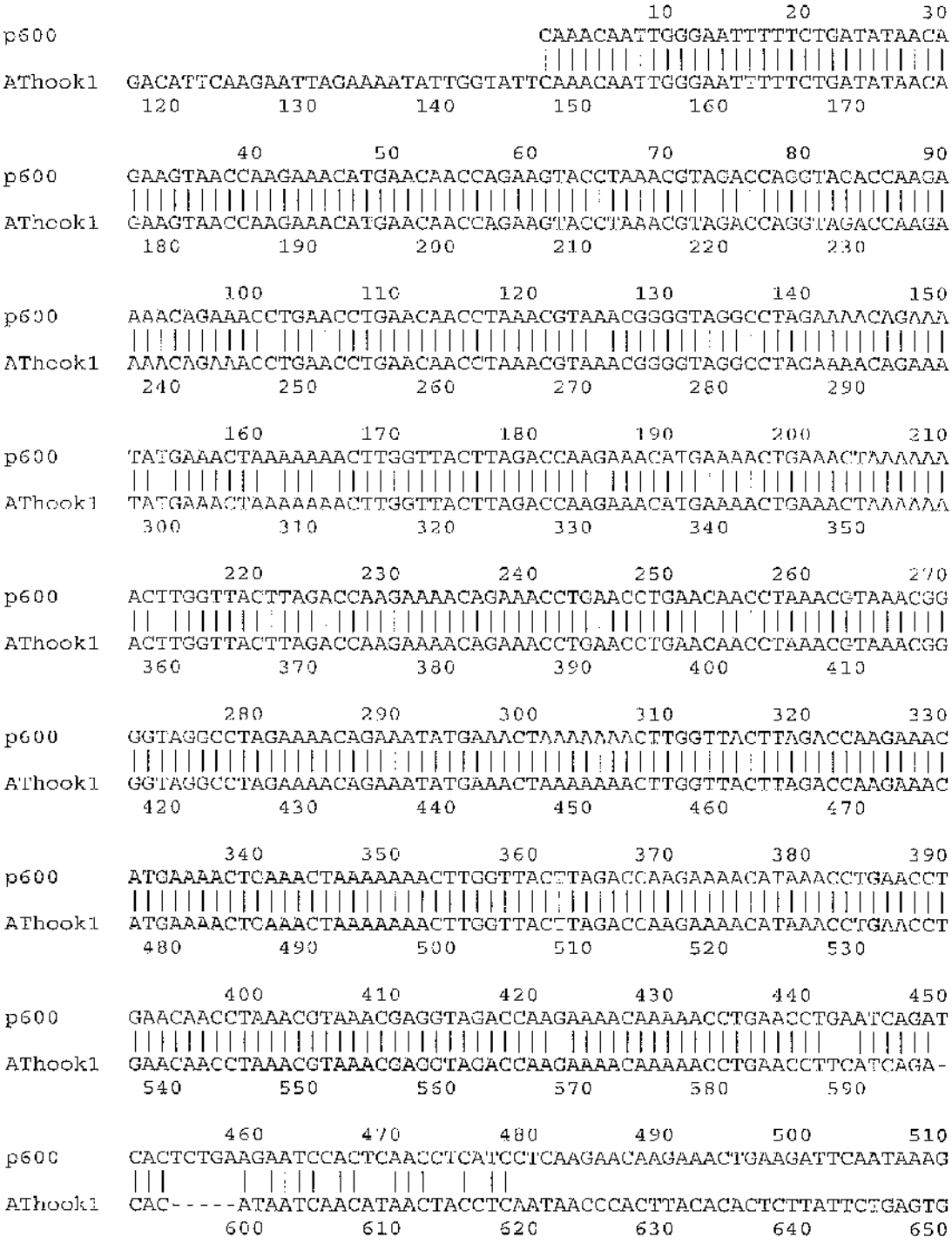
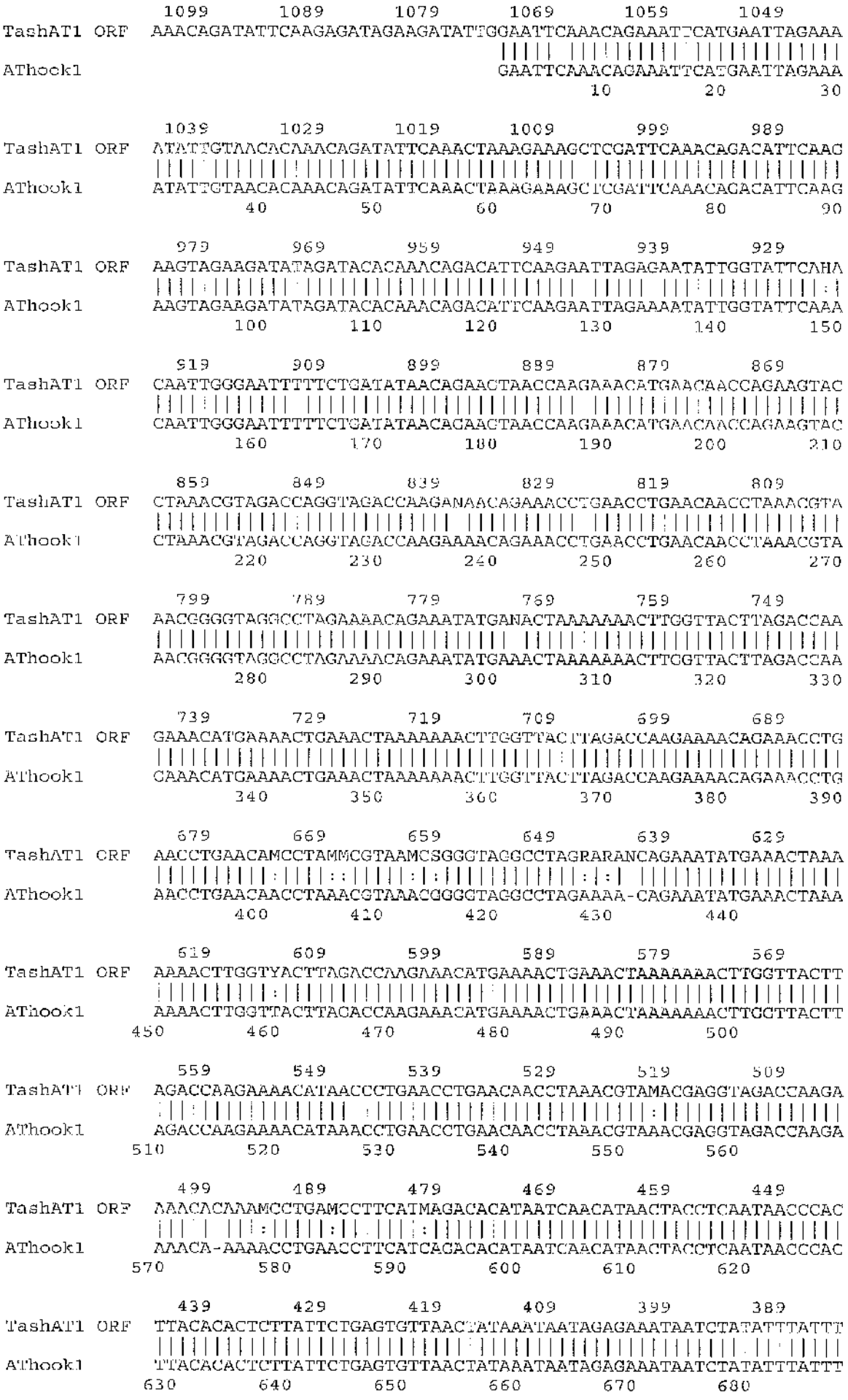
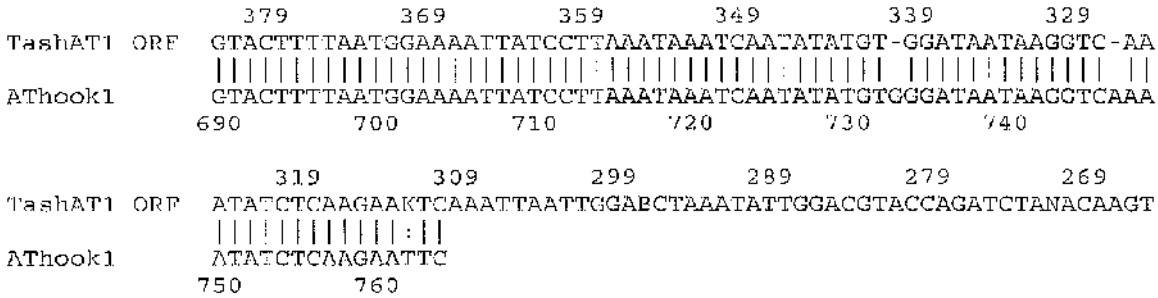


Fig. 3.5: DNA Sequence comparison of p600 PCR fragment with the 0.8kb EcoRI fragment (AThook1) from λ dash 13. 96.9% identity over a 478bp overlap.

Fig. 3.6: DNA sequence comparison of fragment ATHook1, from λ dash 13, with the TashAT1 ORF from cl12. Scores: 97.1% identity in a 765 nucleotide overlap.


1099 1089 1079 1069 1059 1049
TashAT1 ORF AAACAGATATTCAAGAGATAGAAGATATTGGAATTCAAACAGAAATTCATGAATTAGAAA
AThook1 GAATTCAAACAGAAATTCATGAATTAGAAA
10 20 30
1039 1029 1019 1009 999 989
TashAT1 ORF ATATTGTAACACAAACAGATATTCAAACTAAAGAAAGCTCGATTCAAACAGACATTCAAG
AThook1 ATATTGTAACACAAACAGATATTCAAACTAAAGAAAGCTCGATTCAAACAGACATTCAAG
40 50 60 70 80 90
979 969 959 949 939 929
TashAT1 ORF AAGTAGAAGATATAGATACACAAACAGACATTCAAGAATTAGAGAATATTGGTATTCAAA
AThook1 AAGTAGAAGATATAGATACACAAACAGACATTCAAGAATTAGAAAATATTGGTATTCAAA
100 110 120 130 140 150
919 909 899 889 879 869
TashAT1 ORF CAATTGGGAATTTTTCTGATATAACAGAACTAACCAAGAAACATGAACAACCAGAAGTAC
AThook1 CAATTGGGAATTTTTCTGATATAACAGAACTAACCAAGAAACATGAACAACCAGAAGTAC
160 170 180 190 200 210
859 849 839 829 819 809
TashAT1 ORF CTAAACGTAGACCAGGTAGACCAAGANAACAGAAACCTGAACCTGAACAACCTAAACGTA
AThook1 CTAAACGTAGACCAGGTAGACCAAGAAAACAGAAACCTGAACCTGAACAACCTAAACGTA
220 230 240 250 260 270
799 789 779 769 759 749
TashAT1 ORF AACGGGGTAGGCCTAGAAAACAGAAATATGANACTAAAAAAACTTGGTTACTTAGACCAA
AThook1 AACGGGGTAGGCCTAGAAAACAGAAATATGAAACTAAAAAAACTTGGTTACTTAGACCAA
280 290 300 310 320 330
739 729 719 709 699 689
TashAT1 ORF GAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTAGACCAAGAAAACAGAAACCTG
AThook1 CAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTAGACCAAGAAAACAGAAACCTG
340 350 360 370 380 390
679 669 659 649 639 629
TashAT1 ORF AACCTGAACAMCCTAMMCGTAAMCSGGGTAGGCCTAGRARANCAGAAATATGAAACTAAA
AThook1 AACCTGAACAACCTAAACGTAAACGGGGTAGGCCTAGAAAA-CAGAAATATGAAACTAAA
400 410 420 430 440
619 609 599 589 579 569
TashAT1 ORF AAAACTTGGTYACTTAGACCAAGAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTT
AThook1 AAAACTTGGTTACTTACACCAAGAAACATGAAAACTGAAACTAAAAAAACTTGCTTACTT
450 460 470 480 490 500
559 549 539 529 519 509
TashAT1 ORF AGACCAAGAAAACATAACCCTGAACCTGAACAACCTAAACGTAMACGAGGTAGACCAAGA
AThook1 AGACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAACGAGGTAGACCAAGA
510 520 530 540 550 560
499 489 479 469 459 449
TashAT1 ORF AAACACAAAMCCTGAMCCTTCATMAGACACATAATCAACATAACTACCTCAATAACCCAC
AThook1 AAACA-AAAACCTGAACCTTCATCAGACACATAATCAACATAACTACCTCAATAACCCAC
570 580 590 600 610 620
439 429 419 409 399 389
TashAT1 ORF TTACACACTCTTATTCTGAGTGTTAACTATAAATAATAGAGAAATAATCTATATTTATTT
AThook1 TTACACACTCTTATTCTGAGTGTTAACTATAAATAATAGAGAAATAATCTATATTTATTT
630 640 650 660 670 680


379 369 359 349 339 329
TashAT1 ORF GTACTTTTAATGGAAAATTATCCTTAAATAAATCAATATATGT-GGATAATAAGGTC-AA
AThook1 GTACTTTTAATGGAAAATTATCCTTAAATAAATCAATATATGTGGGATAATAAGGTCAAA
690 700 710 720 730 740
319 309 299 289 279 269
TashAT1 ORF ATATCTCAAGAAKTCAAATTAATTGGABCTAAATATTGGACGTACCAGATCTANACAAGT
AThook1 ATATCTCAAGAATTC
750 760

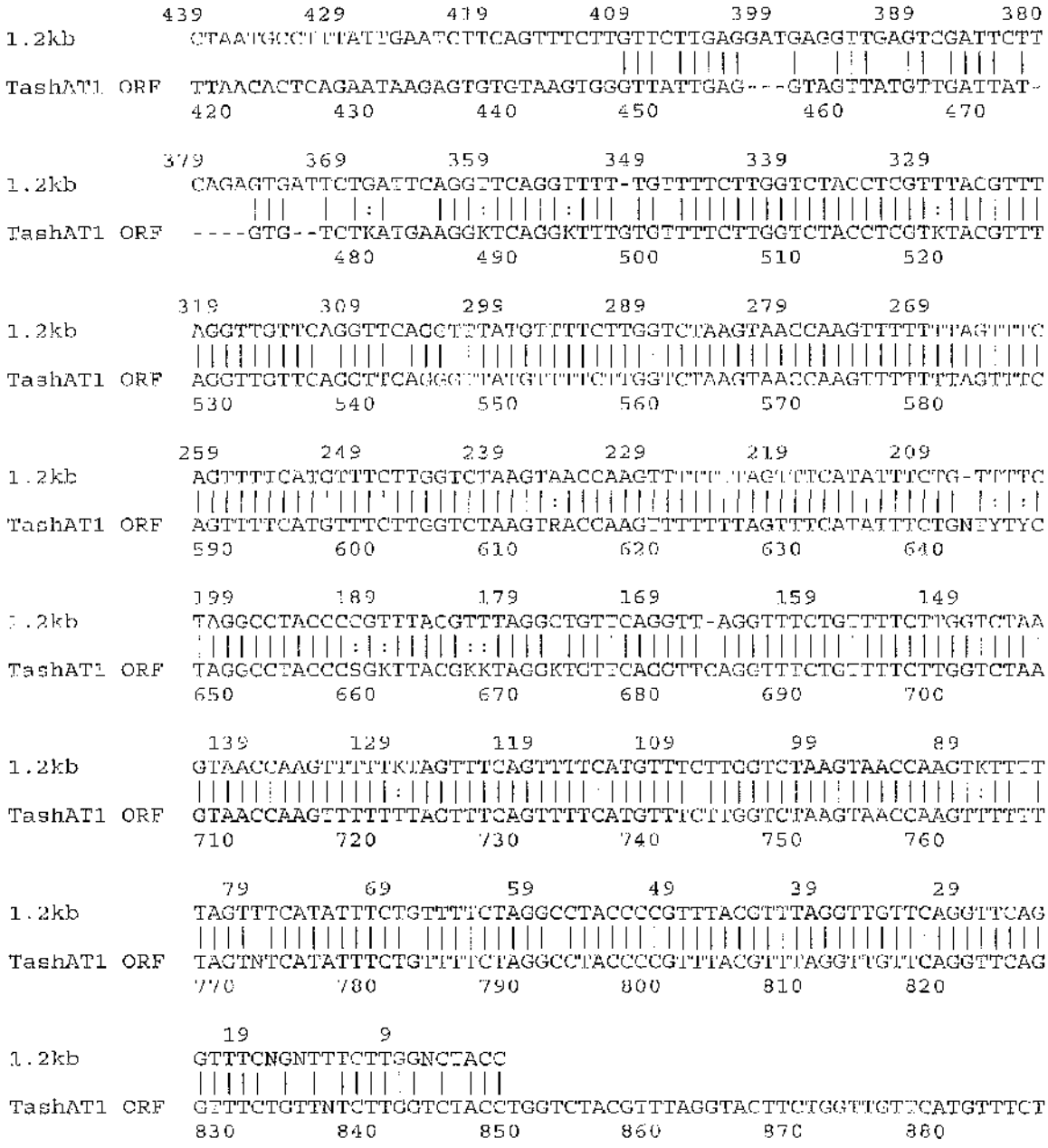


Fig. 3.7: DNA sequence comparison of the 1.2kb EcoRI fragment from λ dash 13 with the TashAT1 ORF from cl12. Score 89.8% identity in a 412 nucleotide overlap.

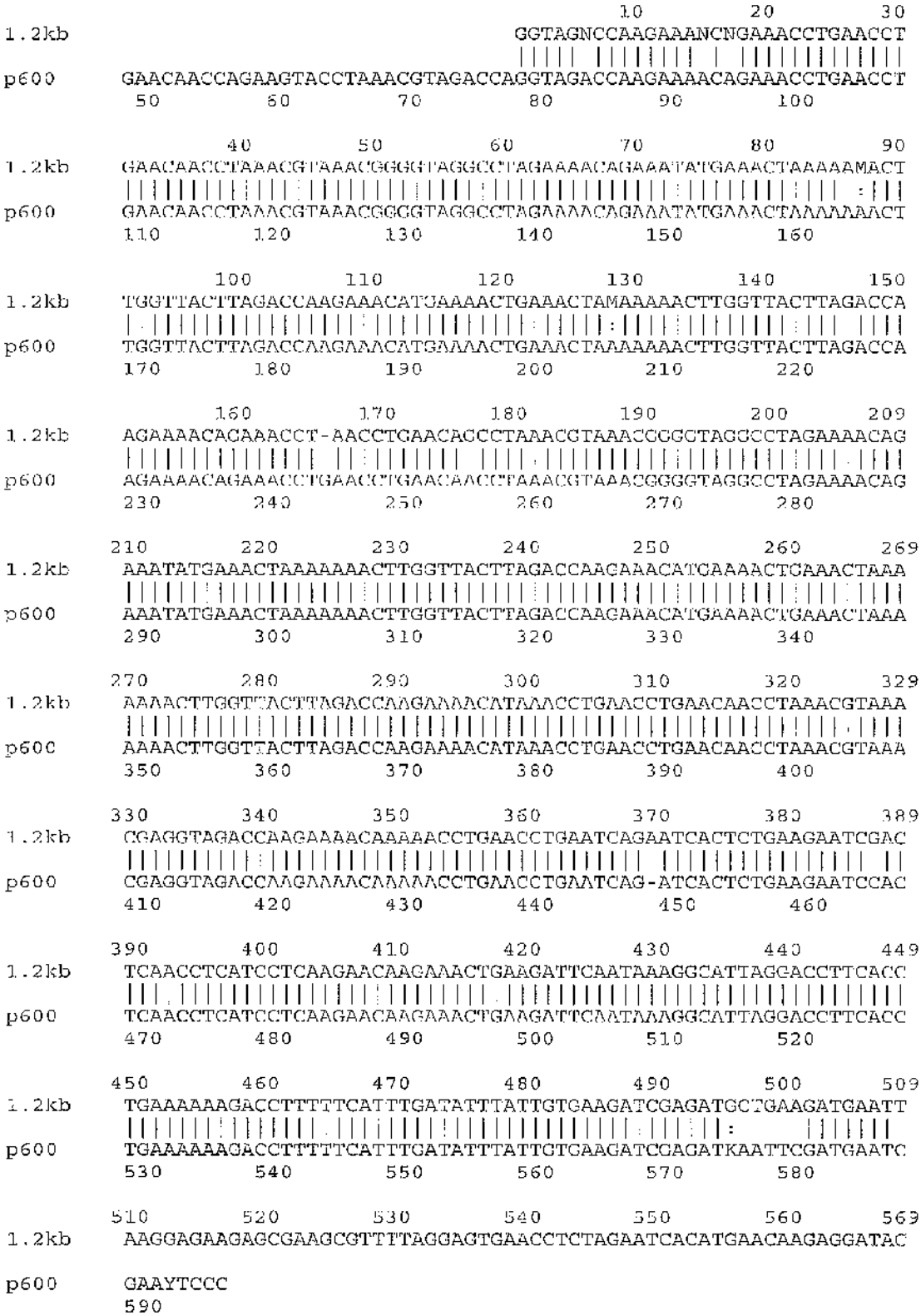
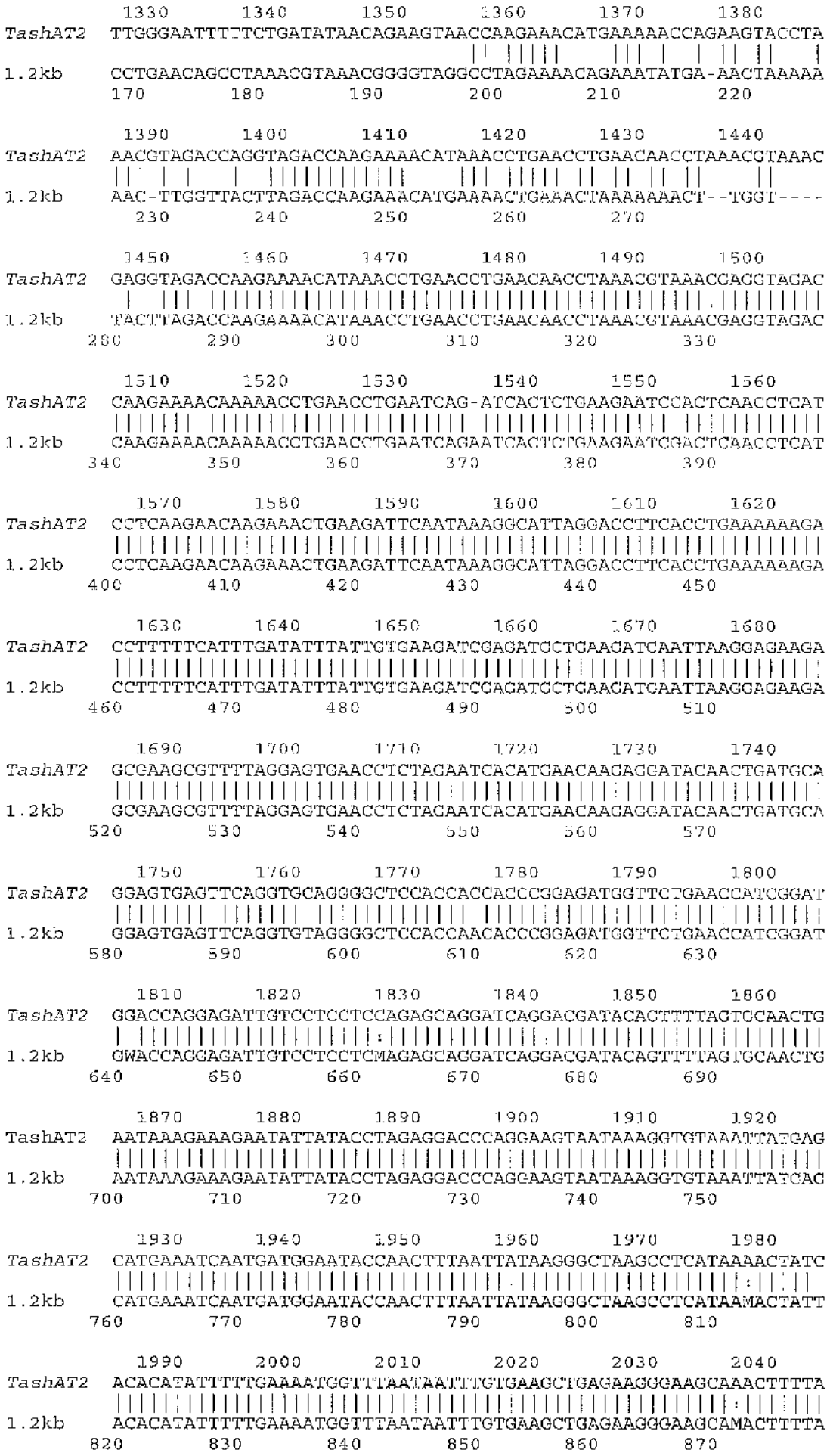
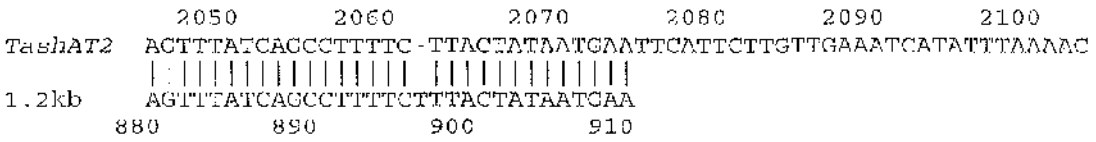


Fig. 3.8: DNA sequence comparison between the 1.2kb EcoRI fragment from λ dash 13 and the p600 PCR fragment. 97.1% identity over a 509 nucleotide overlap.

Fig. 3.9: DNA sequence comparison between the 1.2kb EcoRI fragment from λ dash 13 and *TashAT2*. 93.2% identity over a 721 nucleotide overlap.


1330 1340 1350 1360 1370 1380
TashAT2 TTGGGAATTTTTCTGATATAACAGAAGTAACCAAGAAACATGAAAAACCAGAAGTACCTA
1.2kb CCTGAACAGCCTAAACGTAAACGGGGTAGGCCTAGAAAACAGAAATATGA-AACTAAAAA
170 180 190 200 210 220
1390 1400 1410 1420 1430 1440
TashAT2 AACGTAGACCAGGTAGACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAAC
1.2kb AAC-TTGGTTACTTAGACCAAGAAACATGAAAACTGAAACTAAAAAAACT--TGGT----
230 240 250 260 270
1450 1460 1470 1480 1490 1500
TashAT2 GAGGTAGACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAACCAGGTAGAC
1.2kb TACTTAGACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAACGAGGTAGAC
280 290 300 310 320 330
1510 1520 1530 1540 1550 1560
TashAT2 CAAGAAAACAAAAACCTGAACCTGAATCAG-ATCACTCTGAAGAATCCACTCAACCTCAT
1.2kb CAAGAAAACAAAAACCTGAACCTGAATCAGAATCACTCTGAAGAATCGACTCAACCTCAT
340 350 360 370 380 390
1570 1580 1590 1600 1610 1620
TashAT2 CCTCAAGAACAAGAAACTGAAGATTCAATAAAGGCATTAGGACCTTCACCTGAAAAAAGA
1.2kb CCTCAAGAACAAGAAACTGAAGATTCAATAAAGGCATTAGGACCTTCACCTGAAAAAAGA
400 410 420 430 440 450
1630 1640 1650 1660 1670 1680
TashAT2 CCTTTTTCATTTGATATTTATTGTGAAGATCGAGATGCTGAAGATCAATTAAGGAGAAGA
1.2kb CCTTTTTCATTTGATATTTATTGTGAAGATCGAGATGCTGAACATCAATTAAGGAGAAGA
460 470 480 490 500 510
1690 1700 1710 1720 1730 1740
TashAT2 GCGAAGCGTTTTAGGAGTGAACCTCTACAATCACATGAACAAGAGGATACAACTGATGCA
1.2kb GCGAAGCGTTTTAGGAGTGAACCTCTACAATCACATGAACAAGAGGATACAACTGATGCA
520 530 540 550 560 570
1750 1760 1770 1780 1790 1800
TashAT2 GGAGTGAGTTCAGGTGCAGGGGCTCCACCACCACCCGGAGATGGTTCTGAACCATCGGAT
1.2kb GGAGTGAGTTCAGGTGTAGGGGCTCCACCAACACCCGGAGATGGTTCTGAACCATCGGAT
580 590 600 610 620 630
1810 1820 1830 1840 1850 1860
TashAT2 GGACCAGGAGATTGTCCTCCTCCAGAGCAGGATCAGGACGATACACTTTTAGTCCAACTG
1.2kb GWACCAGGAGATTGTCCTCCTCMAGAGCAGGATCAGGACGATACAGTTTTAGTGCAACTG
640 650 660 670 680 690
1870 1880 1890 1900 1910 1920
TashAT2 AATAAAGAAAGAATATTATACCTAGAGGACCCAGGAAGTAATAAAGGTGTAAATTATCAG
1.2kb AATAAAGAAAGAATATTATACCTAGAGGACCCAGGAAGTAATAAAGGTGTAAATTATCAC
700 710 720 730 740 750
1930 1940 1950 1960 1970 1980
TashAT2 CATGAAATCAATGATGGAATACCAACTTTAATTATAAGGGCTAAGCCTCATAAAACTATC
1.2kb CATGAAATCAATGATGGAATACCAACTTTAATTATAAGGGCTAAGCCTCATAAMACTATT
760 770 780 790 800 810
1990 2000 2010 2020 2030 2040
TashAT2 ACACATATTTTTGAAAATGGTTTAATAATTTGTGAAGCTGAGAAGGGAAGCAAACTTTTA
1.2kb ACACATATTTTTGAAAATGGTTTAATAATTTGTGAAGCTGAGAAGGGAAGCAMACTTTTA
820 830 840 850 860 870


2050
2060
2070
2080
2090
2100
TashAT2 ACTTTATCACCCTTTTC-TTACTATAATGAATTCATTCTTGTTGAAATCATATTTAAAAC
1.2kb AGTTTATCAGCCTTTTCTTTACTATAATGAA
880
890
900
910

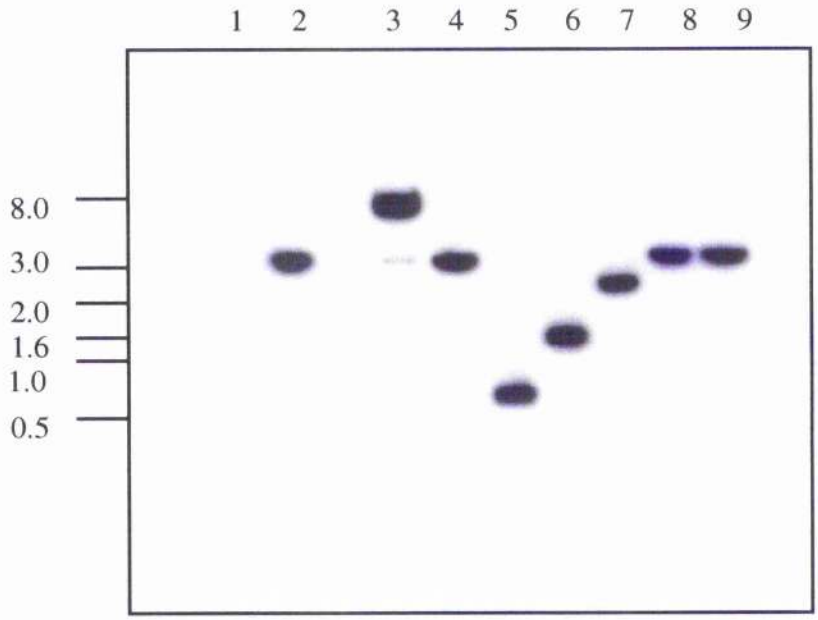


Fig. 3.10: Southern blot analysis of λTa1 restriction digested DNA, hybridised with probe AThook1. undigested PGem7Zf control (lane 1); undigested λTa1 (lane 2); λTa1 digested with BamHI and AatII (lane 3); BamHI, SphI and SpeI (lane 4); BamHI, AatII and EcoRI (lane 5); BamHI, SphI and HindIII (lane 6); BamHI, SphI and EcoRV (lane 7); BamHI, SphI and SpeI (lane 8) and BamHI, SphI and XbaI (lane 9). DNA molecular weight markers are indicated to the left of the figure (in kb).

3.2.2 Sequence analysis of TashAT1

To sequence λTa1, a total of 29 overlapping deletion clones were generated to span the entire clone, on the sense and antisense strands, (shown in Appendix B). λTa1 was sequenced on both strands and assembled into a contiguous fragment from 29 individual deletion clones combined with 5 sequenced fragments from Swan and Phillips (unpublished, 1997), using the fragment assembly program from GCG (see Appendix C). A PCR generated fragment, called rsp2 (see Appendix C), was included to span the region between 1700bp and 1980bp as the deletion clones did not span this region sufficiently on the sense strand.

The results of sequence analysis (shown in Fig. 3.11) revealed a 970bp partial ORF, and a 1.4kb continuous ORF that was located approximately 470bp downstream of the 3' end of the 970bp ORF. On inspection, the 970bp partial ORF was eliminated as a possible *TashAT1* candidate because the predicted peptide sequence did not contain any AT hook motifs, nor did it show any significant similarity to the *TashAT1* ORF, the p600 or AThook1 fragments.

The identity of the 1.4kb ORF (shown in Fig. 3.12) was confirmed to be *TashAT1* as it was 100% identical to the AThook1 fragment at the predicted amino acid level (see Fig. 3.13). The putative amino acid sequence of TashAT1 identified four AT hook motifs (1 to 4) interspersed with two RPRK sequences spanning the 120 amino acid residue domain (see Fig. 3.12). Motifs 2, 3 and 4 were identical to each other but the third amino acid in motif 1 contained an arginine (R) instead of a lysine (K) residue in motifs 2 to 4.

To determine the degree of similarity between *TashAT1* and *TashAT2,* the predicted amino acid sequence of these genes were compared (see Fig. 3.14). The results showed that *TashAT1* and *TashAT2* were 46.1% identical to each other over the 360 amino acids that corresponded to the AT hook-encoding region. However, *TashAT2* had three AT hook domains and also lacked the shorter RPRK repeat sequences seen in *TashAT1*. Motif 1 of *TashAT1* and *TashAT2* were identical to each other and motifs 2, 3 and 4 of *TashAT1* were identical to motifs 2 and 3 of TashAT2. The only difference between motif 1 and motifs 2-4 of both *TashAT1* and *TashAT2* was the third and fourth amino acids of motif 1, which contain arginine and proline, whilst the corresponding amino acids in motifs 2-4 contain lysine and arginine.

The AT hook domains of the *TashAT1, TashAT2* and the newly discovered *TashAT3* (Swan *et al.*, 2001a) genes (see Table 3.1) showed that they belonged to class II AT hook domains based on the classification of Aravind and Landsman (1998). The RPRK sequences within the AT hook regions of *TashAT1* and *TashAT3*, are not true AT hook domains but this amino acid sequence is likely to be capable of DNA binding (Aravind and Landsman, 1998)

To identify other motifs, the putative peptide sequence of *TashAT1* was analysed by the software programs ProDom (Corpet *et al.*, 1998), PSORT II (Nakai, 1996; Nakai and Horton, 1999), SMART (Schultz *et al.*, 1998) and PROSITE (Hoffman *et al.*, 1999). These motifs are shown in Fig. 3.12. SMART, PROSITE and PSORT II analysis revealed the presence of several nuclear localisation signals spanning 21 residues of the AT hook encoding region. The results of the PSORT II analysis revealed both pat 4 and pat 7 continuous nuclear localisation signals (NLSs) and bipartite NLSs within the predicted TashAT1 polypeptide (Fig. 3.12 and Appendix D). Pat 4 continuous NLSs are comprised of 4 basic amino acids (K or R) or 3 amino acids followed by an H or a P residue (Hicks and Raikhel, 1995). Pat 7 NLSs are composed of 7 amino acids starting with a P residue followed by a 3 residue spacer region and then 3 out of 4 basic K/R amino acids (Hicks and Raikhel, 1995). Bipartite NLSs contain 2 basic residues followed by a 10-residue spacer and then a basic region comprising at least 3 out of 5 basic residues (Hicks and Raikhel, 1995).

To determine if TashAT1 contains a signal sequence, the first 70 N-terminal residues of TashAT1 were analysed for a potential signal peptide using the SignalP program (Nielsen *et al.*, 1997). This gave three values (with a range between 0 and 1) of 0.644, 0.959 and 0.680 for the C (raw cleavage site) score (cut off = 0.37); the S (Signal peptide score (cut off = 0.88) and the Y (combined cleavage site) score (cut off = 0.34), with a maximum C and Y score of 25 (see Appendix E). The SignalP analysis of these scores concluded that TashAT1 contained a signal peptide, with a predicted cleavage site between residues 24 and 25 (the position before the maximal Y score). The predicted amino acid sequence of TashAT1 was also analysed using PSORTII, which did not predict a signal sequence but did identify a signal sequence cleavage site at residue 24 (see Appendix D). However, an updated version of the program PSORTII, iPSORT (Bannai *et al.*, 2002), did predict a signal peptide within the 30, N-terminal residues of the predicted amino acid sequence of TashAT1. The Target P program (Emanuelsson *et al.*, 2000) was used to determine the predicted location of TashAT1 within the cell. This analysis gave a score of 0.877 and was

within the highest predicted reliability class score (greater than 0.80) for being secreted into other subcellular compartments of the cell (see Appendix F).

The TashAT1 polypeptide was also analysed for PEST sequences, which target the polypeptide for rapid destruction, using the software program PESTfind (Rogers *et al.*, 1986; Rechsteiner and Rogers, 1996). These sequences were defined as hydrophilic, but not positively charged, stretches of amino acids, greater or equal to 12 residues; rich in Proline (P), glutamic acid (E), serine (S) and threonine (T) and flanked by lysine (K), arginine (R) and histidine (H) residues. The program produces a PEST-FIND score range between -50 and +50: a potential PEST sequence is defined by a score greater than 0, but a value greater than +5 has a higher probability of being a PEST sequence. A sequence of 30 amino acids, between residues 212 to 241, was found within the predicted polypeptide sequence of TashAT1 (see Fig.3.12 and Appendix G). It was given a PEST-FIND score of +17.11 and therefore had a higher probability of being a potential PEST sequence.

Inspection of the TashAT1 sequence also revealed a potential transcriptional activation (or transactivation) domain in *TashAT1* (see Fig. 3.12). Transactivation domains (reviewed by Triezenberg, 1995) regulate gene transcription via protein-protein interactions and a defined to be rich in acidic amino acids: in glutamine, proline or serine, threonine. A potential 21 amino acid transactivation domain of the putative TashAT1 polypeptide was found that contained a stretch of mainly acidic amino acids such as glutamine, aspartate, proline (P), serine (S), threonine (T) just upstream of the first AT hook motif.

The PROSITE search, identified numerous protein kinase C (PKC) and Casein Kinase II (CKII) potential phosphorylation sites within *TashAT1* (seen in Fig. 3.12 and Appendix H). Most of these potential phosphorylation sites were located at the N -terminus of the predicted polypeptide, and none were found within the AT hook domain region. The PROSITE analysis package also identified one myristolation and two N-Glycosylation sites within the predicted TashAT1 polypeptide (Fig. 3.12 and Appendix H) and one myristolation site from residues 72-77. By contrast, the PSORTII program failed to identify any myristolation sites.

To determine if the putative TashAT1 polypeptide was similar to other proteins, it was compared to other protein domain families using the software package ProDom. The results (see Fig 3.15) showed homology to six polypeptide families, mostly within four discrete regions of TashAT1 of approximately 30 amino acids. The only significant

homology found was over the last two highlighted regions to protein families PD184102 and PD010330, which encode the HMGs and histone proteins respectively. As expected, TashAT1 showed identity to the AT hook-encoding region of the PD184102 family and to a region rich in basic amino acids within the PD010330 protein family. Some identity was detected to the E residues between TashAT1 and PD000422 over amino acids 204 to 247. Although the PD001830 family encoded DNA binding bromodomains, and are associated with chromatin and interact with acetylated lysine (K), the homology between TashAT1 to PD001830 appears to be coincidental with the repeated E (Glu), ER or Q (Gln) residues of this protein domain family. The remaining three protein domain families (PD148762, PD000422 and PD000002) encoded protein kinases, serine threonine kinases and coil-coiled myosin repeat chain heavy filament heptad repeat muscle proteins respectively. Again the homology between TashAT1 to the protein families PD000002 and PD000422 appears to be coincidental to abundantly repeated E residues in these proteins. The identity between TashAT1 and PD148762 protein family was poor, with only three identical residues over one homologous region.

Secondary structure predictions of TashAT1 were analysed by the software programmes SEG analysis (Wootton and Federhen, 1996) and PHD (Rost, 1996). PHD analysis predicts one dimensional protein structure by profile based on a neural network model. The results of this analysis showed TashAT1 to be a mixed class polypeptide, containing 9.2% helical structures, 27.0% extended sheets and 63.7% loop structures (see Fig. 3.16). Most of the loop structures were found in the 140 amino acids at the C-terminus of TashAT1, over the AT hook encoding regions. The extended sheet structures were mainly located at the N-terminus and were interspersed with the helical domains, of which there were six in total. However, no helical domains and very few extended sheet structures were found in the last (C- terminal) 130 amino acid residues of TashAT1. Solvent accessibility (SA), a measure of folding and compactness, were low mainly in the 320 residues from the N -terminus and high over the remaining 139 amino acids.

SEG analysis is a programme that detects regions of low compositional complexity within a polypeptide, which tend to be non-globular, exposed loop structures. This programme revealed that TashAT1 contained five regions of low compositional complexity: three of which were located at the C- terminus and two in the middle of the predicted peptide sequence (see Fig. 3.17A). The majority of the C-terminal low complexity regions overlapped with the potential AT hook domains of TashAT1. Tertiary structural predictions by the program "GLOBE" showed that the predicted peptide sequence of

TashAT1 did not contain any globular structures or helix-loop-helix structures (see Fig. 3.17B), in agreement with the SEG analysis.

To identify any common regulatory motifs within the 5' upstream regions of *TashAT1* and *TashAT2*, the upstream sequences of these genes were compared using the FastA program. The results (shown in Fig. 3.18) revealed that the 5' upstream region of *TashAT2*, *TashAT1* and subsequently *TashAT3* showed poor sequence identity to each other, except for a 31bp region upstream of both genes. This motif was also found in another macroschizont encoded gene, *Tash1* (Swan *et al.*, 2001b) and was labelled *TashAT* upstream motif, or TashUM. The overall identity between the TashUM of *TashAT1* and *TashAT2* was 80.6%. However, an AT and a GC rich motif were found within the TashUM region of *TashAT1* and *TashAT2* that showed 100% identity to each other.

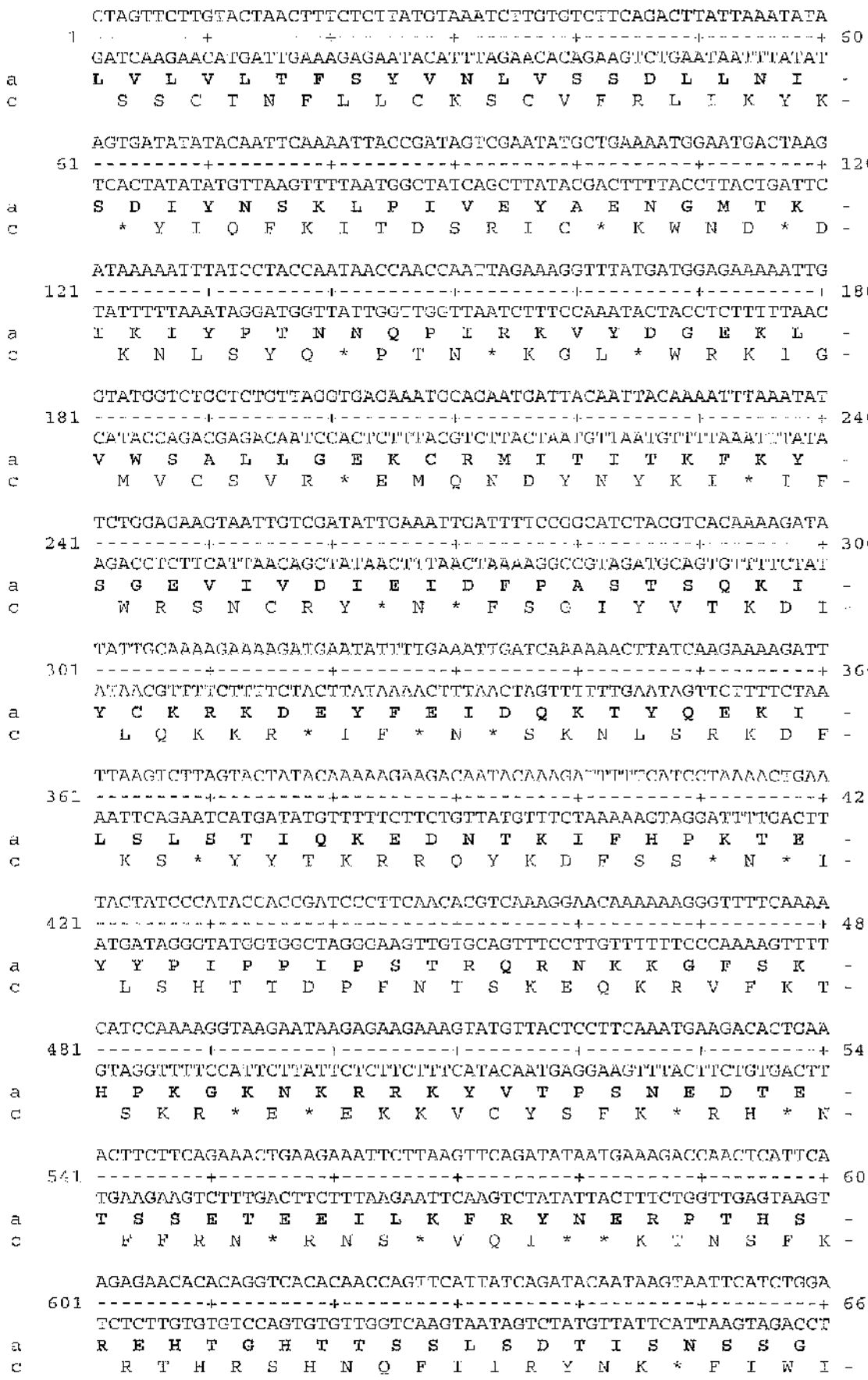
The 5' upstream region of *TashAT1* was analysed for potential transcription factor binding sites using the software program MatInspector (Quandt *et al.*, 1995), to detect any transcription factors with the same binding site as TashUM or any bovine leukocyte specific factors that might regulate *TashAT1*. Analysis with MatInspector did show 74 potential transcription factor binding sites of 5-20bp within the entire upstream region of *TashAT1* (see Fig.3.19 and Appendix I). However, this analysis failed to identify a potential transcription factor that could bind specifically to the TashUM motif or to the complete AT and GC rich motifs identified within TashUM. The most frequently found binding sites were the POU domain encoding Oct1 transcription factor binding sites (5 in total) which had good matrix similarity scores (all over 0.85). Four SATB (Special AT-rich sequence-binding protein 1) sites also were detected that had good matches with the upstream sequence of *TashAT1* (all over 0.9). Also of interest were three lymphocyte associated transcription factors: IKRS (Ikaros 1), a potential regulator of lymphocyte differentiation (matrix sim. score 0.916), PU.1- an Ets (oncogene) like factor found in macrophages and B cells (matrix sim. score 0.86). The third lymphocyte associated transcription factor was BCL6, a zinc finger encoding transcription repressor that has been associated with diffuse large cell lymphoma, when altered by translocation events.

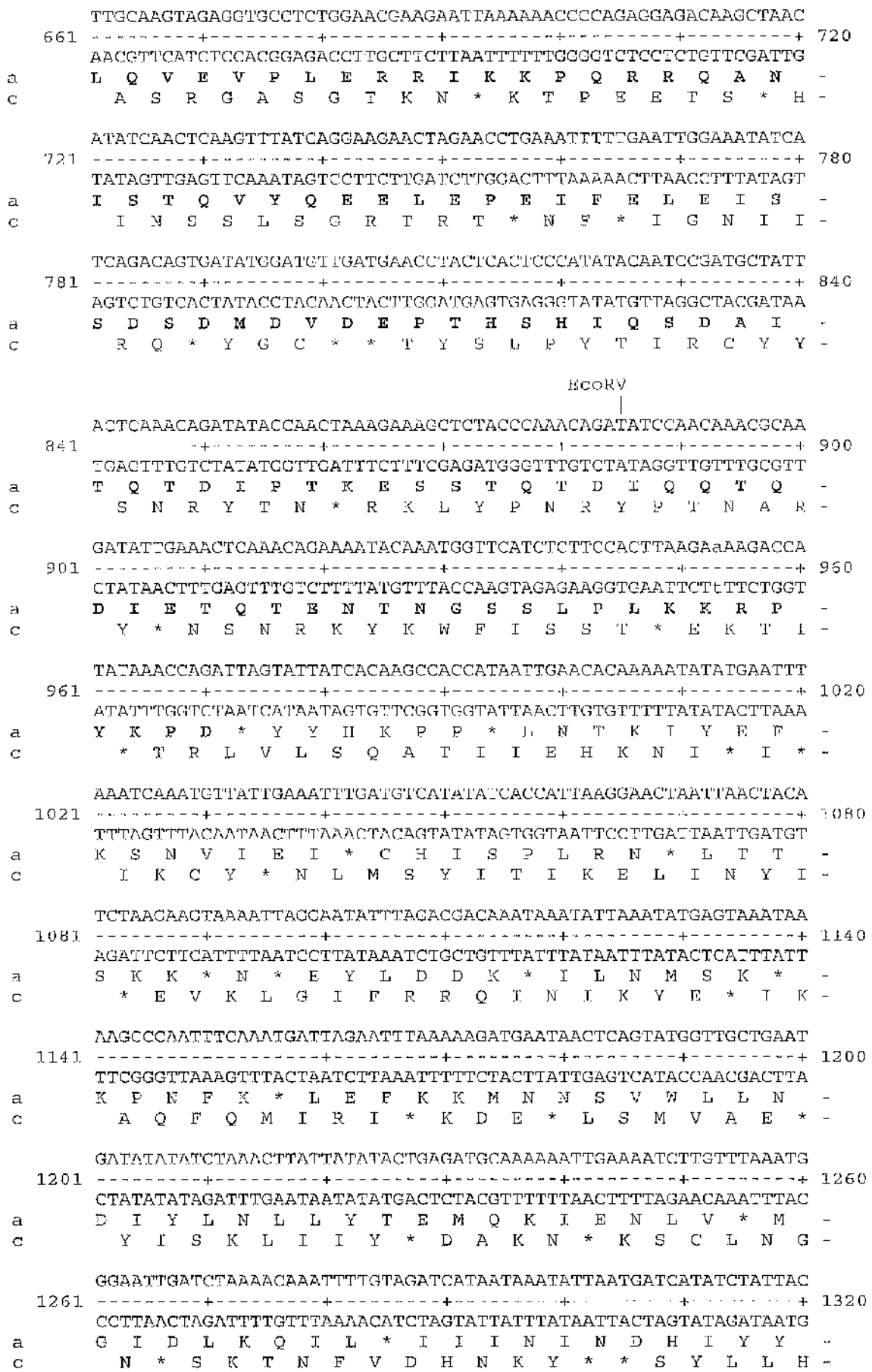
To investigate whether the *TashAT1* gene was conserved across the *Theileria* species the predicted amino acid of TashAT1 was compared with the preliminary, and, at of the time of this study, incomplete sequence data from the *T. parva* genome. TashAT1 showed similarities to 54 sequences from the *T. parva* sequence database (see Appendix J). The highest similarity to the putative TashAT1 polypeptide was from contig 443, a 3073bp

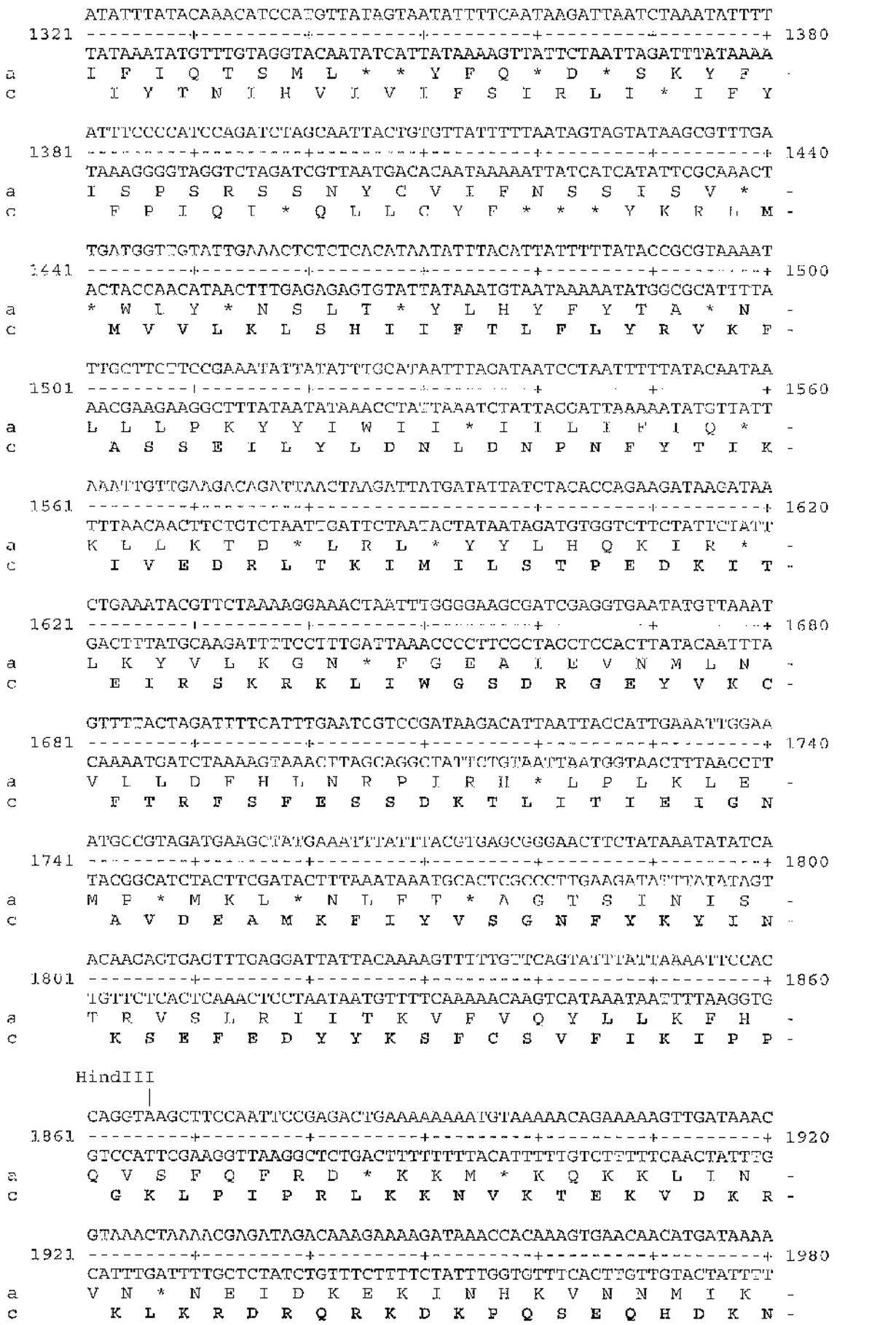
sequence from the *T. parva* database: this showed an identity score of 41% between 30658bp and 31611bp of the contig sequence (see Fig. 3.20). Most of this similarity occurred over the N-terminal region of TashAT1 with little similarity detected over the potential AT hook-encoding region. The sequence comparison between the predicted amino acid sequence of TashAT1 and the preliminary contig sequences of the *T. parva* genome (see Table 3.2) did not identify any true AT hook motifs. However, there were two AT hook-like motifs, rich in lysine (K), Arginine (R) and proline (P), which may have the potential to bind to DNA, as they are very similar to a true AT hook motif. In addition, there may be structural similarity in both sequences towards the C-terminus as they both contained repeated QT residues approximately every 10 residues.

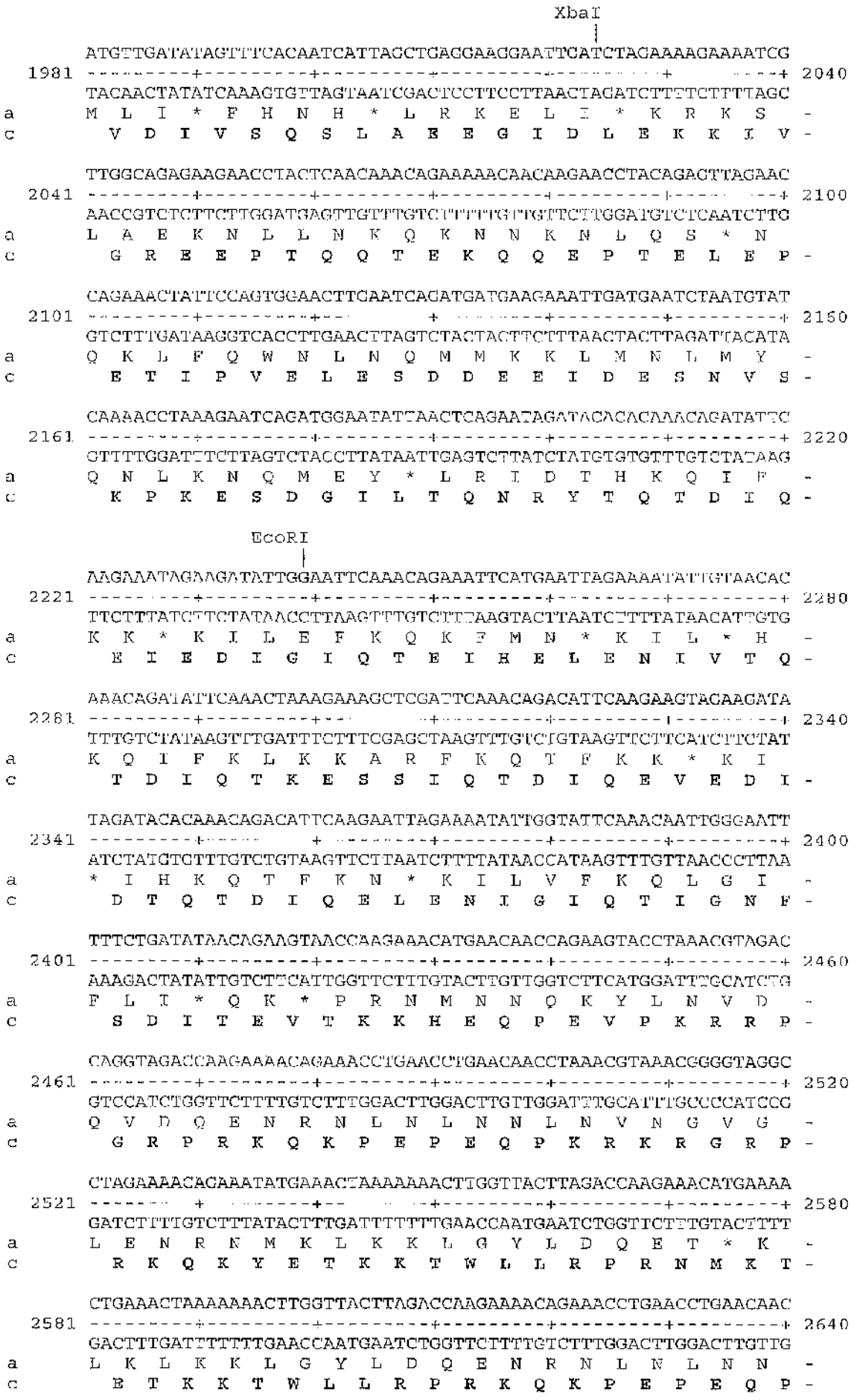
In summary, the predicted TashAT1 polypeptide sequence contains four class II AT hook domains; a NLS; a signal sequence, numerous potential phosphorylation sites and a potential transactivation domain. Secondary and Tertiary structure predictions of TashAT1, show that TashAT1 is non-globular, mixed class polypeptide mainly consisting of loops within the AT hook encoding region (at the C-terminus), and extended sheets and helices (at the N terminus). A 31bp potential regulatory motif, TashUM, was found upstream of the *TashAT1* gene, and of a third *TashAT* gene, *TashAT3*. Similar TashUM-like motifs were found upstream of *TashAT2* and an unrelated macroschizont encoded gene, *Tash1*.

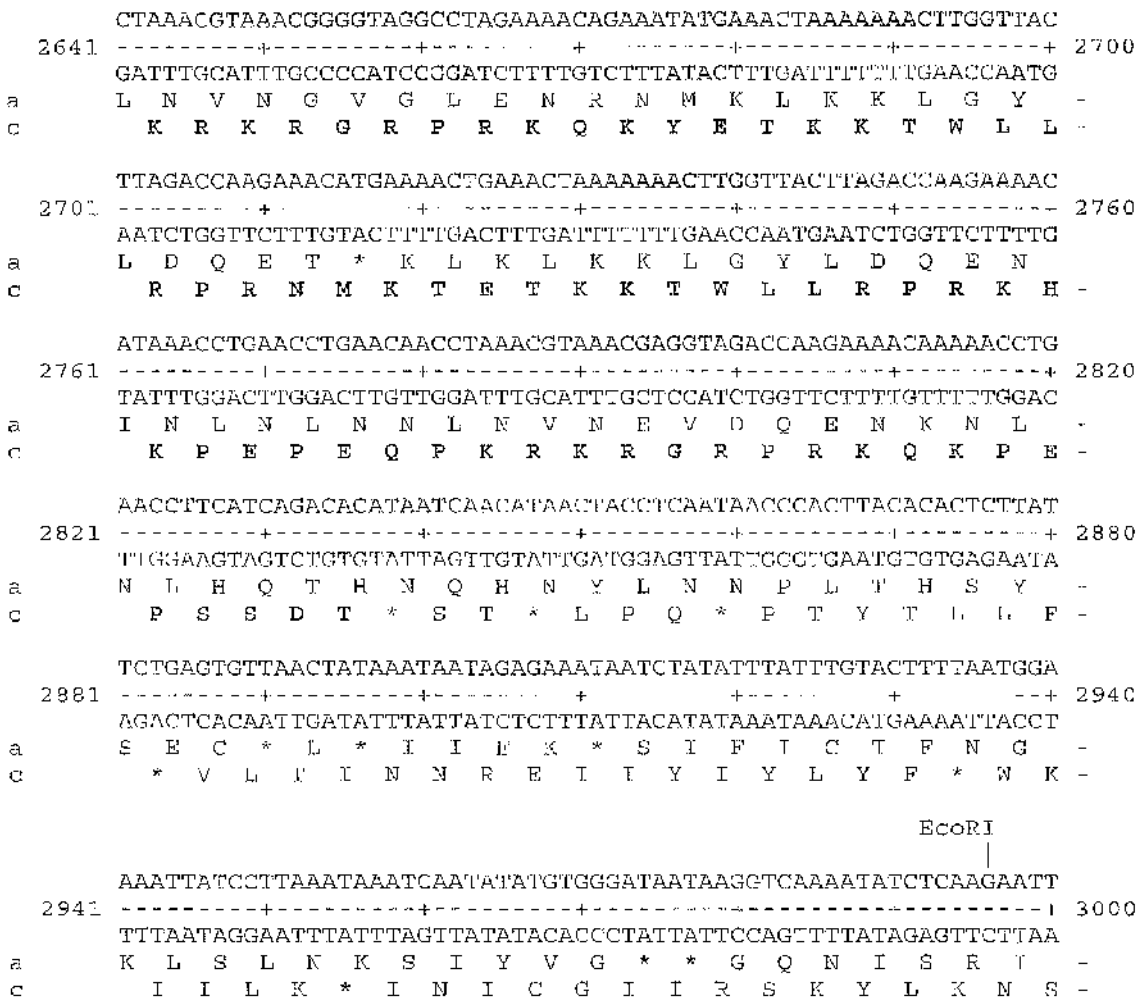
Fig. 3.11: DNA sequence and restriction map of λ Ta1. Protein translation frames a and c show the predicted amino acid sequences (in bold type) of the 970bp partial ORF and *TashAT1*, respectively.


CTAGTTCTTGTACTAACTTTCTCTTATGTAAATCTTGTGTCTTCAGACTTATTAAATATA
GATCAAGAACATGATTGAAAGAGAATACATTTAGAACACAGAAGTCTGAATAATTTATAT
a L V L V L T F S Y V N L V S S D L L N I -
c S S C T N F L L C K S C V F R L I K Y K -
AGTGATATATACAATTCAAAATTACCGATAGTCGAATATGCTGAAAATGGAATGACTAAG
TCACTATATATGTTAAGTTTTAATGGCTATCAGCTTATACGACTTTTACCTTACTGATTC
a S D I Y N S K L P I V E Y A E N G M T K -
c * Y I Q F K I T D S R I C * K W N D * D -
ATAAAAATTTATCCTACCAATAACCAACCAATTAGAAAGGTTTATGATGGAGAAAAATTG
TATTTTTAAATAGGATGGTTATTGGTTGGTTAATCTTTCCAAATACTACCTCTTTTTAAC
a I K I Y P T N N Q P I R K V Y D G E K L -
c K N L S Y Q * P T N * K G L * W R K 1 G -
GTATGGTCTCCTCTGTTAGGTGACAAATCCACAATGATTACAATTACAAAATTTAAATAT
CATACCAGACGAGACAATCCACTCTTTACGTCTTACTAATGTTAATGTTTTAAATCTATA
a V W S A L L G E K C R M I T I T K F K Y -
c M V C S V R * E M Q N D Y N Y K I * I F -
TCTGGAGAAGTAATTGTCGATATTGAAATTGATTTTCCGGCATCTACGTCACAAAAGATA
AGACCTCTTCATTAACAGCTATAACTTTAACTAAAAGGCCGTAGATGCAGTGTTTTCTAT
a S G E V I V D I E I D F P A S T S Q K I -
c W R S N C R Y * N * F S G I Y V T K D I -
TATTGCAAAAGAAAAGATGAATATTTTGAAATTGATCAAAAAACTTATCAAGAAAAGATT
ATAACGTTTTCTTTTCTACTTATAAAACTTTAACTAGTTTTTTGAATAGTTCTTTTCTAA
a Y C K R K D E Y F E I D Q K T Y Q E K I -
c L Q K K R * I F * N * S K N L S R K D F -
TTAAGTCTTAGTACTATACAAAAAGAAGACAATACAAAGATTTTTTCATCCTAAAACTGAA
AATTCAGAATCATGATATGTTTTTCTTCTGTTATGTTTCTAAAAACTAGGATTTTGACTT
a L S L S T I Q K E D N T K I F H P K T E -
c K S * Y Y T K R R Q Y K D F S S * N * I -
TACTATCCCATACCACCGATCCCTTCAACACGTCAAAGGAACAAAAAAGGGTTTTCAAAA
ATGATAGGGTATGGTGGCTAGGGAAGTTGTGCAGTTTCCTTGTTTTTTCCCAAAAGTTTT
a Y Y P I P P I P S T R Q R N K K G F S K -
c L S H T T D P F N T S K E Q K R V F K T -
CATCCAAAAGGTAAGAATAAGAGAAGAAAGTATGTTACTCCTTCAAATGAAGACACTCAA
GTAGGTTTTCCATTCTTATTCTCTTCTTTCATACAATGAGGAAGTTTACTTCTGTGACTT
a H P K G K N K R R K Y V T P S N E D T E -
c S K R * E * E K K V C Y S F K * R H * N -
ACTTCTTCAGAAACTGAAGAAATTCTTAAGTTCAGATATAATGAAAGACCAACTCATTCA
TGAAGAAGTCTTTGACTTCTTTAAGAATTCAAGTCTATATTACTTTCTGGTTGAGTAAGT
a T S S E T E E I L K F R Y N E R P T H S -
c F F R N * R N S * V Q I * * K T N S F K -
AGAGAACACACAGGTCACACAACCAGTTCATTATCAGATACAATAAGTAATTCATCTGGA
TCTCTTGTGTGTCCAGTGTGTTGGTCAAGTAATAGTCTATGTTATTCATTAAGTAGACCT
a R E H T G H T T S S L S D T I S N S S G
c R T H R S H N Q F I I R Y N K * F I W I -


EcoRV


HindIII


XbaI
EcoRI


EcoRI

A

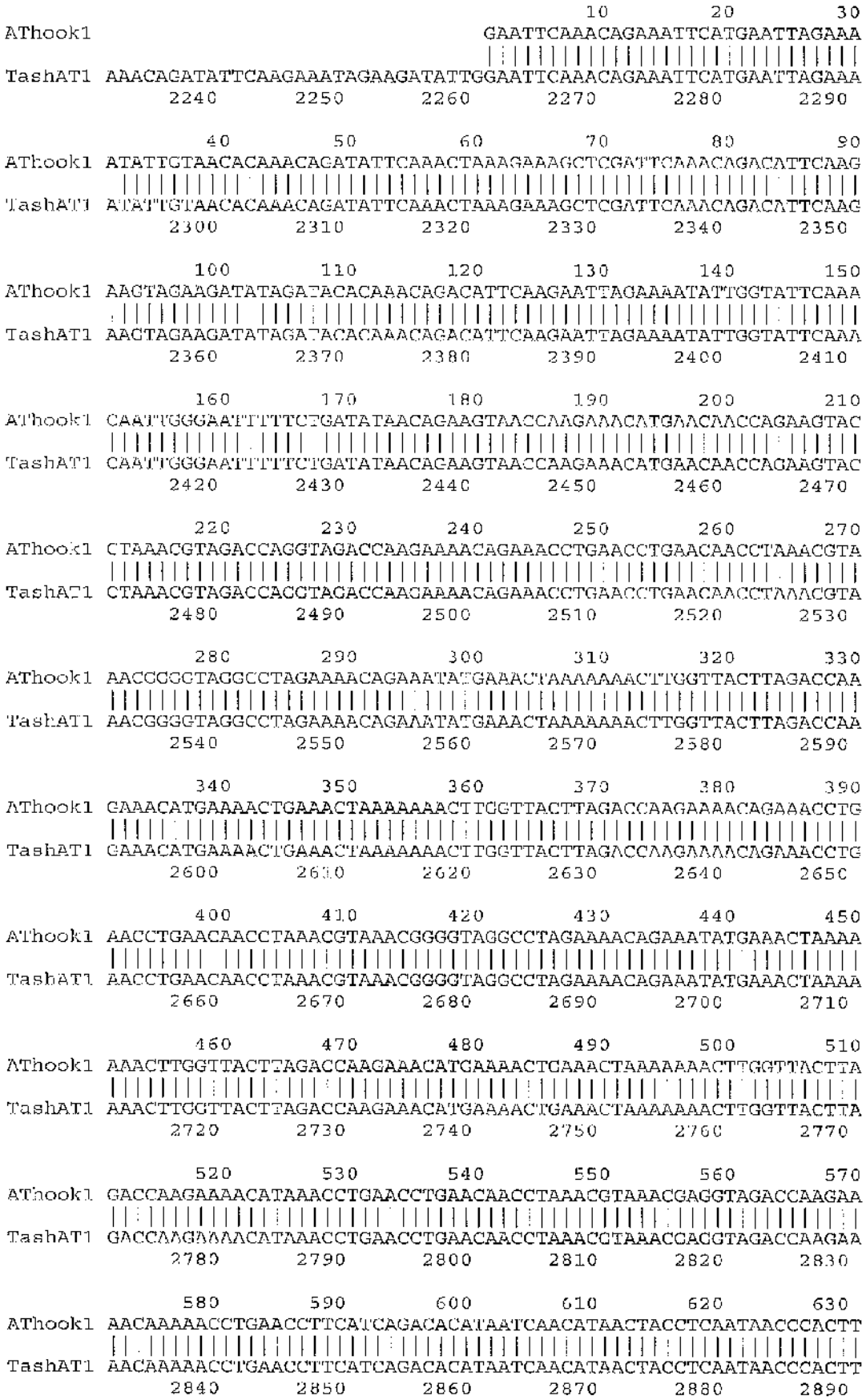
```
1    M  M  V  V  L  K  L  S  H  I  I  F  T  L  F  L
17   Y  R  V  K  F  A  S  S  E  I  L  Y  L  D  N  L  D  N  P  N
                                                    *  *
37   F  Y  T  I  K  I  V  E  D  R  L  T  K  I  M  I  L  S  T  P
57   E  D  K  I  T  E  I  R  S  K  R  K  L  I  W  G  S  D  R  G
77   E  Y  V  K  C  F  T  R  F  S  F  E  S  S  D  K  T  L  I  T
97   I  E  I  G  N  A  V  D  E  A  M  K  F  I  Y  V  S  G  N  F
                   *
117  Y  K  Y  I  N  K  S  E  F  E  D  Y  Y  K  S  F  C  S  V  F
137  I  K  I  P  P  G  K  L  P  I  P  R  L  K  K  N  V  K  T  E
157  K  V  D  K  R  K  L  K  R  D  R  Q  R  K  D  K  P  Q  S  E
                                     *
177  Q  H  D  K  N  V  D  I  V  S  Q  S  L  A  E  E  G  I  D  L
197  E  K  K  I  V  G  R  E  E  P  T  Q  Q  T  E  K  Q  Q  E  P
     *                                     *     +  +  +  +  +
217  T  E  L  E  P  E  T  I  P  V  E  L  E  S  D  D  E  E  I  D
     +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +  +
237  E  S  N  V  S  K  P  K  E  S  D  G  I  L  T  Q  N  R  Y  T
     +  +  +  +  +                                          *
257  Q  T  D  I  Q  E  I  E  D  I  G  I  Q  T  E  I  H  E  L  E
                *
277  N  I  V  T  Q  T  D  I  Q  T  K  E  S  S  I  Q  T  D  I  Q
                   *
297  E  V  E  D  I  D  T  Q  T  D  I  Q  E  L  E  N  I  G  I  Q
317  T  I  G  N  F  S  D  I  T  E  V  T  K  K  H  E  Q  P  E  V
337  P  K  R  R  P  G  R  P  R  K  Q  K  P  E  P  E  Q  P  K  R
357  K  R  G  R  P  R  K  Q  K  Y  E  T  K  K  T  W  L  L  R  P
377  R  N  M  K  T  E  T  K  K  T  W  L  L  R  P  R  K  Q  K  P
397  E  P  E  Q  P  K  R  K  R  G  R  P  R  K  Q  K  Y  E  T  K
417  K  T  W  L  L  R  P  R  N  M  K  T  E  T  K  K  T  W  L  L
437  R  P  R  K  H  K  P  E  P  E  Q  P  K  R  K  R  G  R  P  K
457  Q  K  P  E  P  S  S  D  T
```

B

1. KRRPGRPRK
2. KRKRGRPRK
3. KRKRGRPRK
4. KRKRGRPRK

Fig. 3.12: Predicted polypeptide sequence of TashAT1 (panel A) and AT hook domains of TashAT1 (panel B). Signal sequence: bold italics; Transactivation domain: underlined in blue; AT hook domains: red type; nuclear localisation signals: double underlined; RPRK potential DNA-binding motifs: bold type; N-glycosylation sites: shaded in yellow; myristolation site: shaded in green; protein kinase C sites: pink type; casein kinase II sites: asterix above residues; tyrosine kinase phosphorylation site: boxed; PEST sequences: crosses below residues.

Fig. 3.13: DNA sequence comparison of fragment AThook1, from λ dash 13, with the 1.4kb ORF (TashAT1) from λTa1. 100.0% identity over a 763 nucleotide overlap.


AThook1 GAATTCAAACAGAAATTCATGAATTAGAAA
TashAT1 AAACAGATATTCAAGAAATAGAAGATATTGGAATTCAAACAGAAATTCATGAATTAGAAA
2240 2250 2260 2270 2280 2290
AThook1 ATATTGTAACACAAACAGATATTCAAACTAAAGAAAGCTCGATTCAAACAGACATTCAAG
TashAT1 ATATTGTAACACAAACAGATATTCAAACTAAAGAAAGCTCGATTCAAACAGACATTCAAG
2300 2310 2320 2330 2340 2350
AThook1 AAGTAGAAGATATAGATACACAAACAGACATTCAAGAATTAGAAAATATTGGTATTCAAA
TashAT1 AAGTAGAAGATATAGATACACAAACAGACATTCAAGAATTAGAAAATATTGGTATTCAAA
2360 2370 2380 2390 2400 2410
AThook1 CAATTGGGAATTTTTCTGATATAACAGAAGTAACCAAGAAACATGAACAACCAGAAGTAC
TashAT1 CAATTGGGAATTTTTCTGATATAACAGAAGTAACCAAGAAACATGAACAACCAGAAGTAC
2420 2430 2440 2450 2460 2470
AThook1 CTAAACGTAGACCAGGTAGACCAAGAAAACAGAAACCTGAACCTGAACAACCTAAACGTA
TashAT1 CTAAACGTAGACCACGTAGACCAAGAAAACAGAAACCTGAACCTGAACAACCTAAACGTA
2480 2490 2500 2510 2520 2530
AThook1 AACGGGGTAGGCCTAGAAAACAGAAATATGAAACTAAAAAAACTTGGTTACTTAGACCAA
TashAT1 AACGGGGTAGGCCTAGAAAACAGAAATATGAAACTAAAAAAACTTGGTTACTTAGACCAA
2540 2550 2560 2570 2580 2590
AThook1 GAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTAGACCAAGAAAACAGAAACCTG
TashAT1 GAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTAGACCAAGAAAACAGAAACCTG
2600 2610 2620 2630 2640 2650
AThook1 AACCTGAACAACCTAAACGTAAACGGGGTAGGCCTAGAAAACAGAAATATGAAACTAAAA
TashAT1 AACCTGAACAACCTAAACGTAAACGGGGTAGGCCTAGAAAACAGAAATATGAAACTAAAA
2660 2670 2680 2690 2700 2710
AThook1 AAACTTGGTTACTTAGACCAAGAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTA
TashAT1 AAACTTGGTTACTTAGACCAAGAAACATGAAAACTGAAACTAAAAAAACTTGGTTACTTA
2720 2730 2740 2750 2760 2770
AThook1 GACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAACGAGGTAGACCAAGAA
TashAT1 GACCAAGAAAACATAAACCTGAACCTGAACAACCTAAACGTAAACCACGTAGACCAAGAA
2780 2790 2800 2810 2820 2830
AThook1 AACAAAAACCTGAACCTTCATCAGACACATAATCAACATAACTACCTCAATAACCCACTT
TashAT1 AACAAAAACCTGAACCTTCATCAGACACATAATCAACATAACTACCTCAATAACCCACTT
2840 2850 2860 2870 2880 2890

```
                640         650         660         670         680         690
AThook1  ACACACTCTTATTCTGAGTGTTAACTATAAATAATACACAAATAATCTATATTTATTTGT
         |||||| |||||||| .|||||||||||||||||||||||| |||||||||||||||||||||
TashAT1  ACACACTCTTATTCTGAGTGTTAACTATAAATAATAGAGAAATAATCTATATTTATTTGT
               2900        2910        2920        2930        2940        2950

                700         710         720         730         740         750
AThook1  ACTTTTAATGGAAAATTATCCTTAAATAAATCAATATATGTGGGATAATAAGGTCAAAAT
         |||||||.|||||||||||||||||||||||||||||||||||||||.|||||||||||||||
TashAT1  ACTTTTAATGGAAAATTATCCTTAAATAAATCAATATATGTGGGATAATAAGGTCAAAAT
               2960        2970        2980        2990        3000        3010

                760
AThook1  ATCTCAAGAATTC
         |||||||||||| .
TashAT1  ATCTCAACAATTCAAATTAATTGGATCTAAATATTGGACGTACCAGATCTAAACAAGTAT
               3020        3030        3040        3050        3060        3070
```

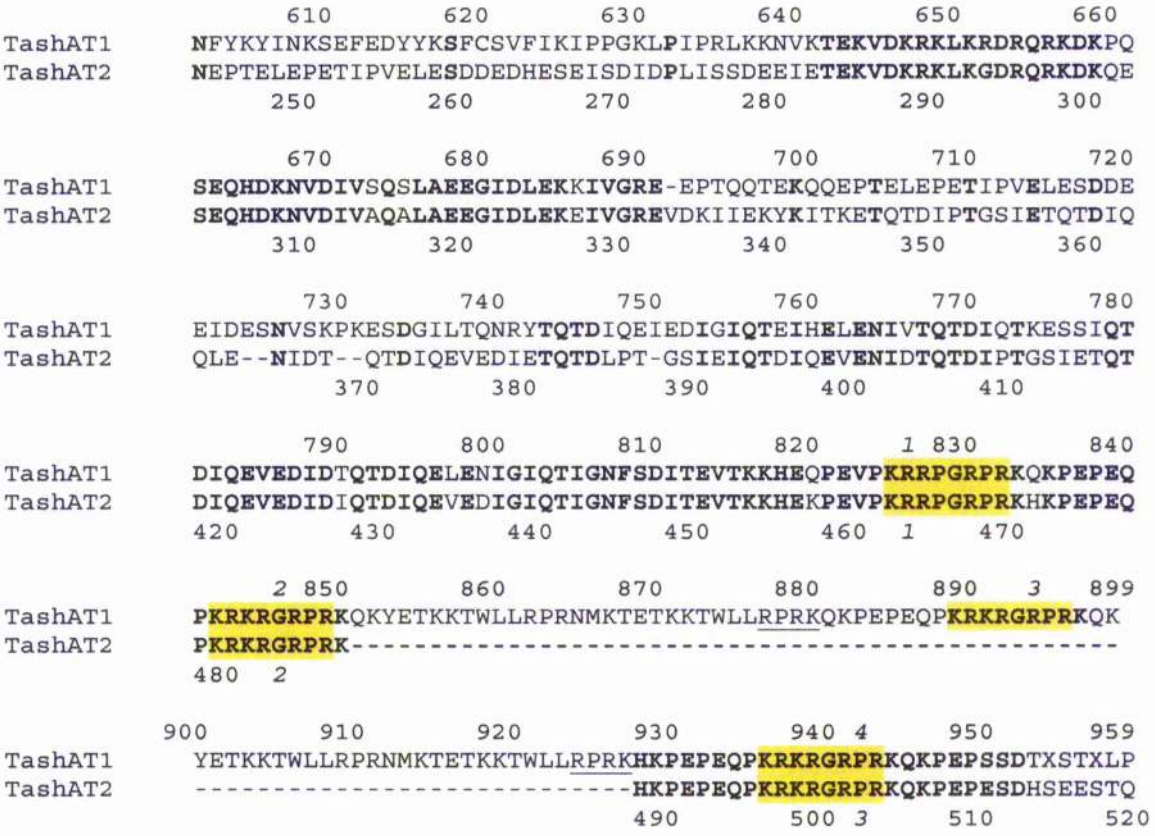
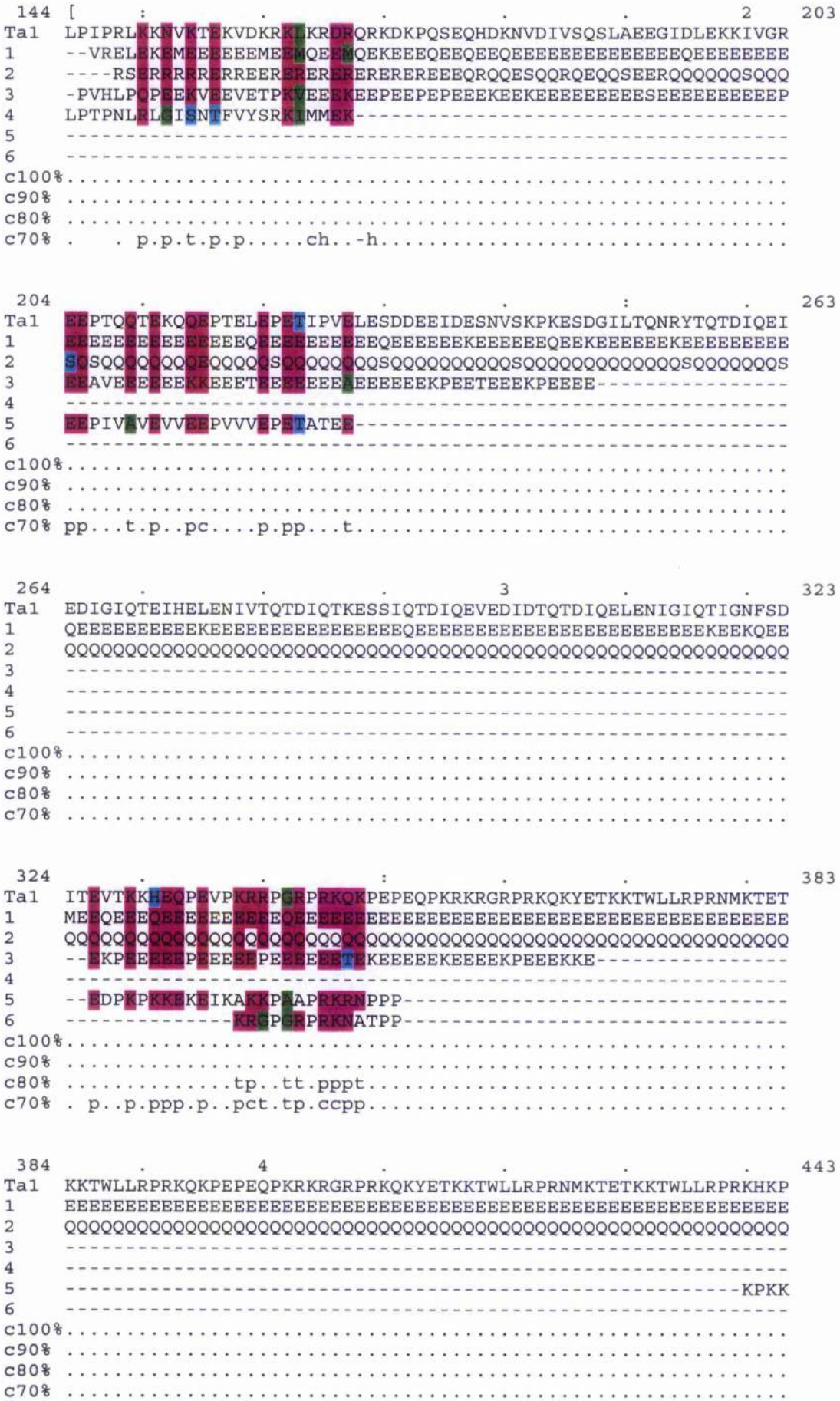


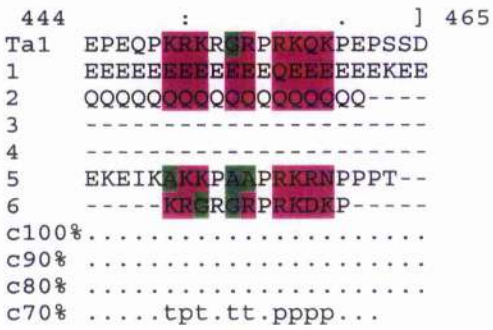
Fig. 3.14: Predicted peptide sequence comparison of TashAT1 with TashAT2.
46.1% over 360 amino acid overlap. Exact amino acid matches are represented in bold type. AT hook domains are highlighted in yellow and numbered in italics. RPRK motifs are underlined.

Name of A.T. hook sub-type	n	AT hook sequence	Motif no.
General		**KR+RGRPRK**	
CLASS II:		**KR+RGRPRK**	
TashAT1/3	4	KRRPGRPRK	1
		KRKRGRPRK	2
		KRKRGRPRK	3
		KRKRGRPRK	4
TashAT2	3	KRRPGRPRK	1
		KRKRGRPRK	2
		KRKRGRPRK	3
HMGI(Y) (human)	3	KRGRGRPRK	
		KRPRGRPRK	
		RKPRGRPRK	
CLASS I		**RPRGRPRGSKNK**	
CLASS III		**PR*RGRPKPK**	

Table. 3.1: Classification of AT hook domains from the *TashAT* gene family, based on data from Aruvand and Landsman (1998). The consensus sequence of each class of AT hook is represented in bold type. The + position has an equal probability of being a K,G,P or R. The * position denotes residues K,G or P with equal probabilities. Polar residues are underlined. The number, n, shows the frequency of AT hook motifs appearing in each gene.

Fig. 3.15: Putative domain search for TashAT1 (Ta1) using the PRODOM software. Prodom domain family 1= PD000002; 2= PD001830; 3=PD000422; 4= PD148762; 5= PD010330; 6= PD184102. Coloured regions denote 70% consensus and property of amino acids. t= polar neutral residue, h= polar positive, p and c were undefined. Purple = charged; green = hydrophobic; blue = polar.


144 [: . . . 2 203
Ta1 LPIPRLKKNVKTEKVDKRKLKRDRQRKDKPQSEQHDKNVDIVSQSLAEEGIDLEKKIVGR
1 --VRELEKEMEEEEEEMEEMQEEMQEKEEEQEEQEEQEEEEEEEEEEEEEEQEEEEEEEE
2 ----RSERRRRRERREEREREREREREREEEQRQQESQQRQEQQSEERQQQQQQSQQQ
3 -PVHLPQPEEKVEEVETPKVEEEKEEPEEPEPEEEKEEKEEEEEEEEEESEEEEEEEEEP
4 LPTPNLRLGISNTFVYSRKIMMEK----------
5 ----------
6 ----------
c100%.....
c90% .....
c80% .....
c70% . . p.p.t.p.p.....ch..-h.....
204 . . . . : . 263
Ta1 EEPTQQTEKQQEPTELEPETIPVELESDDEEIDESNVSKPKESDGILTQNRYTQTDIQEI
1 EEEEEEEEEEEEEEEQEEEEEEEEEEQEEEEEEEKEEEEEEQEEKEEEEEEKEEEEEEEE
2 SQSQQQQQQQQQEQQQQQSQQQQQQQQSQQQQQQQQQQQQSQQQQQQQQQQQQQSQQQQQQQS
3 EEAVEEEEEEKKEEETEEEEEEEAEEEEEEKPEETEEEKPEEEE----------
4 ----------
5 EEPIVAVEVVEEPVVVEPETATEE----------
6 ----------
c100%.....
c90% .....
c80% .....
c70% pp...t.p..pc....p.pp...t.....
264 . . . 3 . . 323
Ta1 EDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQELENIGIQTIGNFSD
1 QEEEEEEEEEEKEEEEEEEEEEEEEEEQEEEEEEEEEEEEEEEEEEEEEEKEEKQEE
2 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
3 ----------
4 ----------
5 ----------
6 ----------
c100%.....
c90% .....
c80% .....
c70% .....
324 . . : . . . 383
Ta1 ITEVTKKHEQPEVPKRRPGRPRKQKPEPEQPKRKRGRPRKQKYETKKTWLLRPRNMKTET
1 MEEQEEEQEEEEEEEEEEQEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
2 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
3 --EKPEEEEEEPEEEEEPEEEEEETEKEEEEEKEEEEKPEEEKKE----------
4 ----------
5 --EDPKPKKEKEIKAKKPAAPRKRNPPP----------
6 ----------KRGPGRPRKNATPP----------
c100%.....
c90% .....
c80% .....tp..tt.pppt.....
c70% . p..p.ppp.p..pct.tp.ccpp.....
384 . 4 . . . 443
Ta1 KKTWLLRPRKQKPEPEQPKRKRGRPRKQKYETKKTWLLRPRNMKTETKKTWLLRPRKHKP
1 EEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEEE
2 QQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQQ
3 ----------
4 ----------
5 ----------KPKK
6 ----------
c100%.....
c90% .....
c80% .....
c70% .....


444 : .] 465
Ta1 EPEQPKRKRGRPRKQKPEPSSD
1 EEEEEEEEEEEEQEEEEEEKEE
2 QQQQQQQQQQQQQQQQQQ----
3 ----------
4 ----------
5 EKEIKAKKPAAPRKRNPPPT--
6 -----KRGRGRPRKDKP-----
c100%
c90%
c80%
c70% .....tpt.tt.pppp...

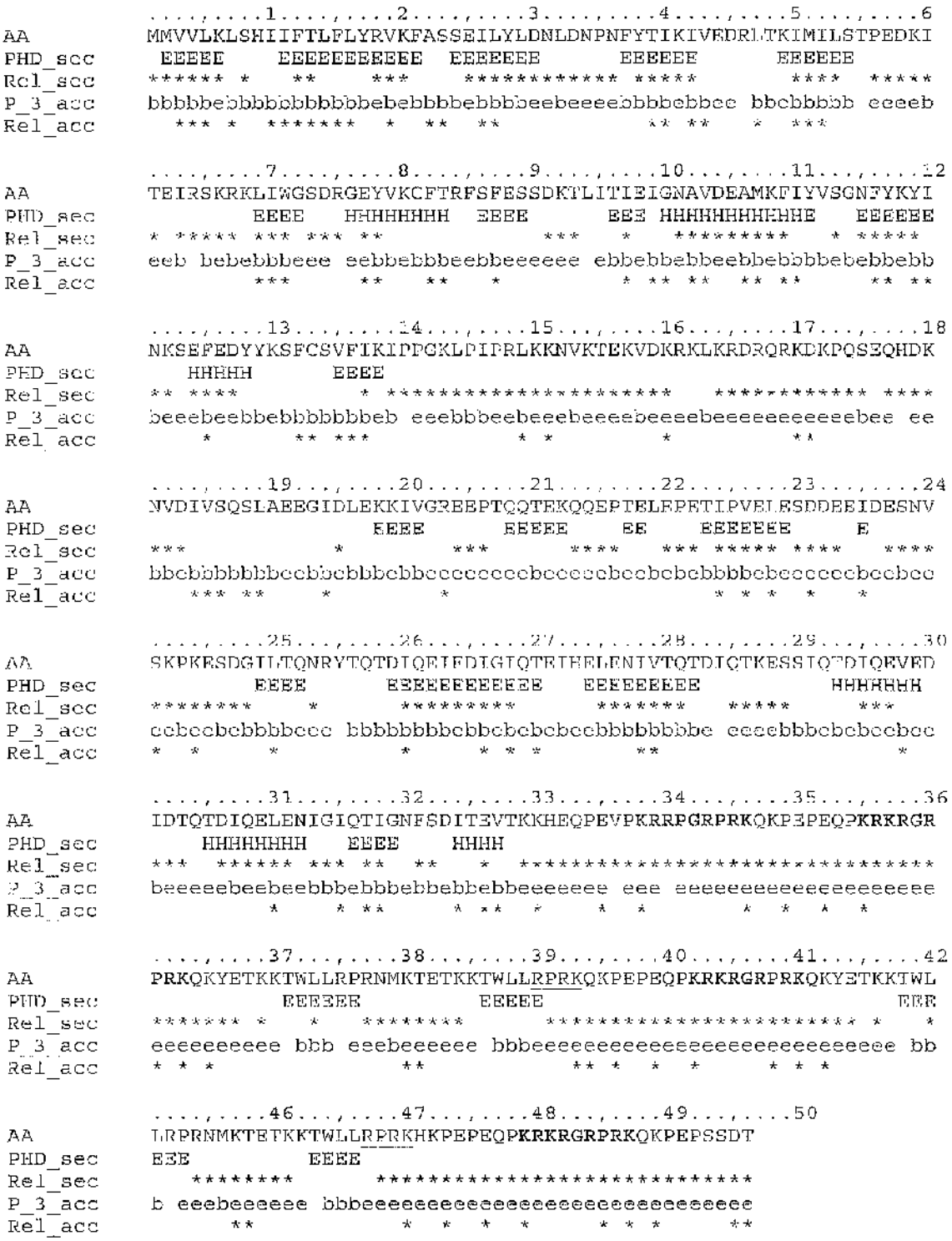


Fig. 3.16: Secondary structure predictions for the theoretical TashAT1 polypeptide using the software program "PHD". AA denotes amino acid; secondary structures (PHD_sec): E: extended sheets, H: helices and blank spaces: other (loop). P-3-acc: solvent accessibility prediction in 3 states where b: 0-9%, I: 9-36% and e is 36-100%. Rel sec and Rel acc represent the reliability index for secondary structure and solvent accessibility predictions respectively, where * denotes strong reliability predictions. AT hook domains are indicated in bold type; RPRK potential DNA-binding motifs are underlined.

A

```
  1 MMVVLKLSHI IFTLFLYRVK FASSETLYLD NLDNPNFYTI KIVEDRLTKI
 60 MILSTPEDKI TEIRSKRKLI WGSDRGEYVK CFTRFSFESS DKTLITIEIG
120 NAVDEAMKFI YVSGNFYKYI NKSEFEDYYK SFCSVFIKIP PGKLPIPRLK
180 KNVKTEKVDK RKLKRDRQRK DKPQSEQEDK NVDIVSQSLA EEGIDLEKKI
240 VGREEPTQQT EKQQEPTELE PETIPVELES DDEEIDESNV SKPKESDGIL
300 TQNRYTQTDI QEIEDICIQT EIHELENIVT QTDIQTKESS IQTDIQEVED
360 IDTQTDIQEL ENIGIQTIGN FSDITEVTKK HEQPEVPKRR PGRPRKQKPE
420 PEQPKRKRGR PRKQKYETKK TWLLRPRNMK TETKKTWLLR PRKQKPEPEQ
480 PKRKRGRPRK QKYETKKTWL LRPRNMKTET KKTWLLRPRK HKPEPEQPKR
540 KRGRPRKQKP EPSSDT
```

B

```
GLOBE prediction of protein globularity

nexp = 286     (number of predicted exposed residues)
nfit = 181     (number of expected exposed residues)
diff = 105.00  (difference nexp-nfit)

Your protein appears not to be globular
```

Fig. 3.17: Secondary (A) and Tertiary (B) structure predictions for the theoretical TashAT1 polypeptide. A: SEG analysis and B: GLOBE prediction of globularity. Regions of low-complexity are underlined; AT hook domains are represented in bold type. Numbers refer to the amino acid sequence position.

Fig. 3.18: Sequence comparison between the *TashAT* upstream motif (TashUM) of *TashAT1/3*, *TashAT2* and *Tash1*. The position of the motif relative to the possible transcription start site is indicated by numbers above each sequence. The sequences within each coloured area are 100% identical. Sequence in red and green type indicate the AT rich and GC rich regions, respectively.

Fig. 3.19: Map of possible transcription factor binding sites within the upstream region of *TashAT1*. A total of 90 matches were found over a 400bp region. Bold type denotes the five prime upstream sequence of *TashAT1*. Potential transcription factors are denoted by V$ (vertebrate) or F$ (Fungal) symbols, adjacent to their respective consensus binding sites. Numbers in brackets to the right of the transcription factor names denote the matrix score, where 1.00= perfect match and >0.80 is a good match; numbers to the left show the position of the binding site consensus sequence relative to the *TashAT1* upstream sequence. + and - denotes predicted binding to sense and antisense strands, respectively. * represents the putative translation start site of *TashAT1*. TashUM motif is underlined. Further details of these transcription factors are described in Appendix I. Figure generated from the MatInspector program (Quandt *et al.*, 1995).

```
(   33)                          +NAARYAAAYANTNNN(V$HNF3B.01(0.972))
(   34)                           +NWWTAAGTATWNN(V$NKX31.01(0.852))
(   45)                                 +NGATANGANWAGATA(VSEVI1.04(0.834))
(   50)                                    +WNWGTMAACAWWMW(V$XFD3.01(0.921))

    1  TCTAAGAAGTAAAATTAGGAATATTTAGACGACAAATAAATATTAAATAT

(    2) -NNNKRYMATAAAAYWNKNW(V$CDX2.01(0.837))
(    7)    -AAGTAAACTTTN(V$MYT1.01(0.843))
(   12)       -CAATTAWG(V$NKX25.02(0.896))
(   14)        -NNNNTGMATATTCNN(V$OCT1P.01(0.901))
(   15)        -NNRGNTATTTTTARNCMN(V$AMEF2.01(0.824))
(   31)                  -NNNTAATNNAATTANNN(VSCART1.01(0.855))
(   34)                    -NKKTAAWNATTAACC(V$HNF1.01(0.804))
(   34)                    -NNWNNNNTATTANWNNTAWNKN(VSSATB1.01(0.918))
(   38)                      -NYKNATTWNNNATGNN(V$BRN2.01(0.940))


(   55)   +STATAAAWRNNNNNNN(V$TATA.01(0.880))
(   83)                    +KNTWTAAANA(V$MEF2.05(0.970))
(   98)                           +RTGASTCAGCA(VSNFE2.01(0.862))

   51  GAGTAAATAAAAGCCCAATTTCAAATGATTAGAATTTAAAAAGATGAATA

(   64)          NYKNATTWNNNATGNN(V$BRN2.01(0.939))
(   71)             -WNGSNYCATTANNSTSWYAA(V$ISL1.01(0.818))
(   77)                -SSNNNNGNTTTAWANNN(V$MTATA.01(0.852))
(   89)                      -NKKTAAWNATTAACC(V$HNF1.01(0.793))
(   94)                        -ATGAATAAWT(V$PIT1.01(0.961))
(   95)                         -NNCATAAATCAT(V$MEIS1.01(0.781))


(  115)          +CWNAWTKWSATRYN(V$OCT1.05(0.935))
(  119)            +ATGATTTATGNN(V$MEIS1.01(0.841))
(  127)                 +TATCTM(F$NIT2.01(1.000))
(  142)                           +CWNAWTKWSATRYN(V$OCT1.05(0.906))
(  150)                              +NATGCAAATN(VSOCT1.02(0.937))

  101  ACTCAGTATGGTTGCTGAATGATATATATCTAAACTTATTATATACTGAG

(  127)                    -TATCTWNTCNTATCN(VSEV11.04(0.857))
(  129)                     -NNWATACTTAWWN(V$NKX31.01(0.899))
(  130)                     -NNAAACTTTNN(V$MYT1.02(0.890))
(  132)                      -NNNNNNYWTTTATAS(V$TATA.01(0.897))


(  156)    +SNAAAGYGAAACC(V$IRF1.01(0.891))
(  175)                 +TNTATGNTAATT(V$OCT1.06(0.862))
(  176)                  +NNNTGGGAATRCC(V$IK1.01(0.917))
(  183)                       +NATYGATSSS(V$CDPCR3HD.01(0.959))
(  184)                        +ANAGATMWWA(V$GATA3.02(0.929))
```

```
     151  ATGCAAAAAATTGAAAATCTTGTTTAAATGGGAATTGATCTAAAACAAAT

(    165)          -TTATCTTGT(V$EVI1.06(0.854))
(    168)            -NAWTGTTTATWT(V$HFH1.01(0.899))


(    203)  +TTRWSASNNTAATGRNSCNW(V$ISL1.01(0.826))
(    204)   +ANAGATMWWA(V$GATA3.02(0.935))
(    208)     +NNCATNNNWAATNMRN(V$BRN2.01(0.931))
(    209)      +NNNTAATTNN.ATTANNN(V$CART1.01(0.924))
(    210)       +CWTAATTG(V$NKX25.02(0.878))
(    211)        --WNWATAAACAWNNR(V$XFD2.01(0.903))
(    223)                +ANAGATMWWA(V$GATA3.02(0.933))
(    231)                    +TATCTM(F$NIT2.01(1.000))
(    236)                      +GTTACRTNAN(VSVBP.01(0.900))
(    243)                         +AWTTATTCAT(V$PIT1.01(0.908))
(    246)                           +TAAAYAAAYANNM(V$HFH3.01(0.982))

     201  TTTGTAGATCATAATAAATATTAATGATCATATCTATTACATATTTATAC

(    212)        -NKKTAAWNATTAACC(V$HNF1.01(0.807))
(    223)              -AATTANCATANA(VSOCT1.06(0.852))
(    225)                -NTATCTTATCT(V$EVI1.05(0.834))
(    228)                  -TWWKATCTNT(V$GATA3.02(0.950))
(    228)                  -NNWNNNNTATTANWNNTAWNKN(V$SATB1.01(0.924))
(    235)                      -NRTTACRTAAYN(V$E4BP4.01(0.874))
(    236)                       -NNNNNNYWTTTATAS(VSTATA.01(0.940))
(    238)                        -NNNNTRTTTATNTNNW(V$FREAC7.01(0.943))
(    238)                        -NNWATACTTAWWN(V$NKX31.01(0.843))
(    242)                          -TNTTTAWANM(V$MEF2.05(0.978))


(    258)       NMNWTANNWNTAATANNNNWNN(V$SATB1.01(0.913))
(    291)                          +WNMNWRTTTTATKRYMNNN(V$CDX2.01(0.922))
(    294)                            +ATAATTWATGASNN(V$BRN3.01(0.784))

     251  AAACATCCATGTTATAGTAATATTTTCAATAAGATTAATCTAAATATTTT

(    290)                          -NNNANTRTTTRYTTN(VSHNF3B.01(0.934))
(    297)                              -NNYMASTTCCTSYWNN(V$PU1.01(0.860))
(    299)                               -NNMTTKCNNMAYNY(V$CEBPB.01(0.942))


(    307)     +CCGTTCCGCTCTAGATAT(V$PAX1.01(0.675))
(    315)          +NNNNNYAATTAN...(V$S8.01(0.976))
(    329)                  +NMNWTANNWNTAATANNNNWNN(V$SATB1.01(0.925))
(    330)                   +GGTTAATNWTTAMMN(V$HNF1.01(0.792))

     301  ATTTCCCCATCCAGATCTAGCAATTACTGTGTTATTTTTAATAGTAGTAT

(    304)   -TCCCCNCN(V$MZF1.01(0.974))
(    313)         -CTTTCTAGGAATWN(V$BCL6.01(0.771))
(    317)           -NNGGTMATTAKWNTMTWAA(V$PDX1.01(0.744))
(    330)                  -RGKYWTTTTTARNSMN(VSMMEF2.01(0.920))
(    338)                       -NTAATNRSNYAATTAG(V$XVENT2.01(0.823))


                        *
     351  AAGCGTTTGATGATGGTTGT
```

```
Ta1  :MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 65
      M  LKLSHIIFTLFLY++K ASSEILYLDN+    F  IKI+E+R+T+ MI STP+ +IT
T.p  :MATLKLSHIIFTLFLYQIKIASSEILYLDNIVGSGFNIIKIIENRITRTMIYSTPDRQIT 30717

Ta1 :EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 125
     ++R  RKLIW    GE +KC T FSFESS K LITIEI N   +++KFIY+  N+++Y+
T.p :QVRQGRKLIWMGYPGESIKCLTIFSFESSSKILITIEIENPAYDSLKFIYMHRNYFRYVT 30897

Ta1 :KSEFEDYY----KSFCSVFIKIPPGKLPIPRLKKNVKTEKVDKRKLKRDRQRKDKPQSEQ 181
     K+ FE  +    K   S   K  PGKLPIPRLKK       +KRK   ++ ++ K +
T.p :KAYFETNFAMQAKPLKSPTSKPIPGKLPIPRLKKP------EKRKADAEKSKEAKKKIVG 31059

Ta1 :HDKNVDIVSQSLAEEGIDLEKKIVGREEPTQQTEKQQEPTELEPETIPVELESDDEEIDE 241
     H       +++ AE     E      ++ TQ  +KQ EP     E I VE+ SDDE  D+
T.p :HPSE----TKTEAERQTQTELYTTTSQQTTQP-QKQSEP-----EIIQVEVGSDDEGTDD 31209

Ta1 : ----SNVSK-----------------PKESD-GILTQNRYTQTDIQEIEDIGIQTEIHEL 279
         S   K                   PKE    ILTQ RYTQTD  E ED   QT+  +
T.p :  FLVVSTSQKVDLYSETVDTTTGTAIHPKEQQVKILTQIRYTQTDTHESEDTETQTDTQQS 31389

Ta1 :ENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQELENIGIQTIGNFSDITEV-TKKHEQP 338
     ++  TQT I T  +  QT I   +  DTQTD  E +    QT+   +D TE  T  HE
T.p :KDTETQTVILTDSTETQTLIP-TDSTDTQTDTHESKETETQTVIP-TDSTETQTDTHETE 31563

Ta1 :EV---P KRRPGRPRK Q 351
     ++    K R   P+KQ
T.p :DIGIQTKLRTRYPKKQ 31611
```

Fig. 3.20: Comparison of the predicted peptide sequence of TashAT1 with the *T. parva* genomic database. 41% identity over a 351 residue overlap. TashAT1 sequence is denoted by Ta1, *T. parva* contig 443 sequence is denoted by T.p. Numbers on the right adjacent to TashAT1 and contig 443 sequence refer to the amino acid number of TashAT1 and contig 443, respectively. Regions of good identity are indicated in blue type. The AT hook motif of TashAT1 is indicated in bold type.

Theileria spp.	Gene/Contig	locus	Motif no.
T.parva	443	678649	PRKRGRKPK
	443	50507	PKKRGKPRK
T.annulata	*TashAT1/2/3*		***
			KRRPGRPRK
			KRKRGRPRK

Table 3.2: AT hook like motifs found in the *T. parva* genomic database.
The two types of AT hook motifs found in all *TashAT* genes *(TashAT1/2/3)* were included for comparison with *T.parva*. * sign denotes the AT hook core.

3.2.3 Mapping TashAT1 and other TashAT genes

3.2.3.1 Southern blot analysis

Southern blot analysis was performed on restriction fragments from λ dash 13 DNA in order to map *TashAT1* and the restriction fragments possibly derived from a third *TashAT* gene (*TashAT3*) not yet cloned within λ dash 13. Following restriction digestion, λ dash 13 DNA was hybridised to two probes (see Fig.3.21): AThook1, and a 368bp PCR product, Ta369, generated from the five prime region of *TashAT1*. Probe Ta369 was designed to be specific to *TashAT1*, whilst probe AThook1 was capable of detecting all *TashAT* genes. The restriction enzymes were chosen because they were predicted to digest within the *TashAT1* sequence and some were also present within the putative *TashAT3* gene.

The Southern blot of restriction digested λ dash 13 DNA was hybridised with probe AThook1 (see Fig. 3.22). As expected, probe AThook1 also detected some of the restriction fragments (tabulated in Appendix K) hybridised to probe p600 (see Fig. 3.4). These restriction fragments were mapped relative to *TashAT1* and the putative *TashAT3* genes (see Fig. 3.23 and 3.24).

The southern blot in Fig. 3.22 indicated that there were only two copies of the AThook1 sequence, since only two fragments hybridised to probe AThook1 in all digests except those with XbaI, (which was partially digested). Moreover, the probe AThook1 did not possess any of the restriction enzyme sites used in this analysis. Given that there might be only two copies of the AThook1 sequence within λ dash 13, and one copy is derived from *TashAT1*, then the second copy is likely to be derived from *TashAT3*.

The 3.2kb HindIII fragment was mapped to *TashAT1* as it disappeared together with the 3.2kb SpeI fragment upon double digestion with HindIII and SpeI, leaving a 1.4kb fragment. By contrast, the 1.6kb HindIII fragment remained upon double digestion with SpeI, which indicated that it was not derived from *TashAT1*. In the HindIII/EcoRI double digestion, the 1.6kb HindIII fragment was removed, leaving the 0.8kb EcoRI fragment (which belonged to *TashAT1*) and the 1.2kb EcoRI fragment only. Since the 1.2kb EcoRI fragment lies within the 1.6kb HindIII fragment, and was previously deduced to be derived from *TashAT3*, then the 1.6kb HindIII fragment must therefore also be derived from *TashAT3*. The double restriction digestion with KpnI and SpeI produced a 3.2kb (SpeI) and an 8kb (KpnI/SpeI) fragment. Since the 3.2kb SpeI fragment is known to contain *TashAT1*

only, then the second AT hook-encoding gene, *TashAT3*, is likely to be derived from the 8kb KpnI/SpeI fragment.

There were a number of restriction fragments produced in the XbaI, HindIII/XbaI and SpeI/XbaI digests (see Fig.3.22) which were all feint and indistinct, and may have been partially digested. This meant that most of the XbaI sites could not be mapped with respect to *TashAT1* or the putative *TashAT3* gene. However, the 2.0kb SpeI/XbaI fragment, seen in Fig. 3.23, could be derived from *TashAT1* as an XbaI site is present 2kb from the original 5' SpeI site in the restriction map of the sequenced λTa1 (see Fig. 3.11). The bands produced in the EcoRI/XbaI digestion appeared to be more intense compared to the other bands. This might be due to the presence of two or more copies of these fragments, one of which may be derived from *TashAT3*, since the increased intensity is not seen in the equivalent restriction digests using probe Ta369.

To characterise the region surrounding the 5' end of *TashAT1*, the southern blot (shown in Fig. 3.22) was stripped and re-hybridised with the 5' *TashAT1* probe, Ta369 (see Fig. 3.23). The restriction digest with EcoRI produced a 2.5kb EcoRI fragment, which was reduced to a 2.1kb fragment upon digestion with SpeI. This indicated that the 2.5kb EcoRI fragment was derived from the 3.2kb SpeI fragment, over the region corresponding to Ta369. Two fragments at 2.8kb and 3.2kb were produced upon digestion with HindIII. The 2.8kb fragment may correspond to the Ta369 region within *TashAT1* as double restriction digestion with EcoRI and HindIII only produced a 2.1kb fragment. Given the position of the EcoRI and HindIII sites within λTa1 (see Fig. 3.11), the 2.8kb HindIII fragment is the only candidate that could produce a 2.1kb fragment upon digestion with EcoRI. Therefore the 2.1kb HindIII fragment is likely to be derived from *TashAT1*. A second copy of the Ta369 sequence may also be present within the 3.2kb HindIII fragment as a fragment of identical length was also detected by probe AThook1. However, probe AThook1 did not overlap with the region corresponding to Ta369 (1500-1862bp) in *TashAT1*, so the second Ta369 sequence may be located downstream of *TashAT1*.

Analysis of the 1.6kb HindIII/SpeI fragment suggested it could also belong to the 3.2kb HindIII fragment, as its size was inconsistent with the restriction map of *TashAT1*. The 8kb KpnI/SpeI fragment was detected with the Ta369 and also with probe AThook1 and might comprise of *TashAT3* and the second copy of the Ta369 sequence. However it was not clear if the Ta369 sequence is part of *TashAT3*.

In summary Southern blot analysis identified and mapped several fragments that belonged to *TashAT1* and another AT hook encoding gene, called *TashAT3*. Two copies of the Ta369 sequence from *TashAT1* were found within λ dash 13, one of which was mapped to an 8kb KpnI/SpeI restriction fragment that was likely to contain *TashAT3*. However, it was not possible to map the precise location of various restriction fragments with respect to each other using this technique. In order to produce a more detailed map of *TashAT3*, and its relationship with *TashAT1* and *TashAT2*, it was necessary to perform T7 or T3 mapping analysis on λ dash 13.

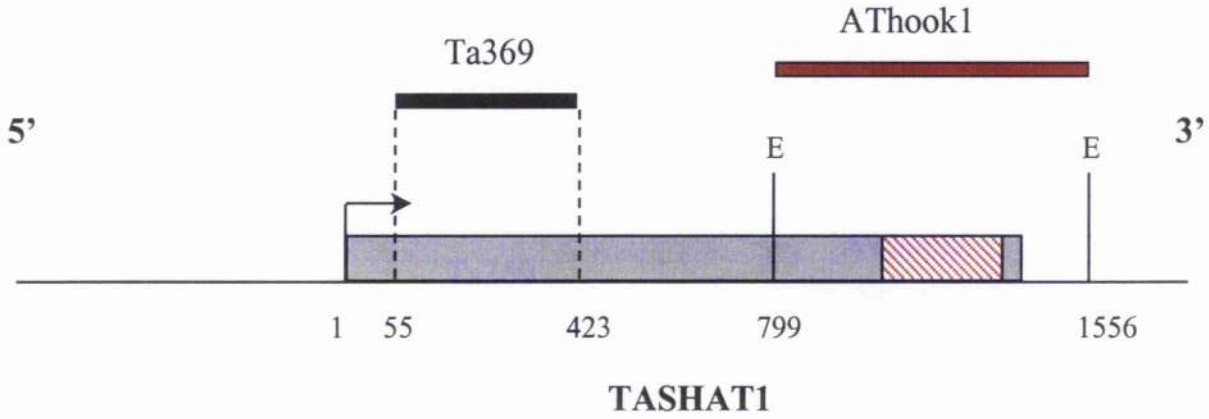


Fig. 3.21: Position of probes Ta369 and AThook1 in relation to TashAT1. Arrow at position 1 indicates the putative translation start site. Grey box represents the TashAT1 gene; E: EcoRI sites. Red hatched box indicates AT hook encoding region.

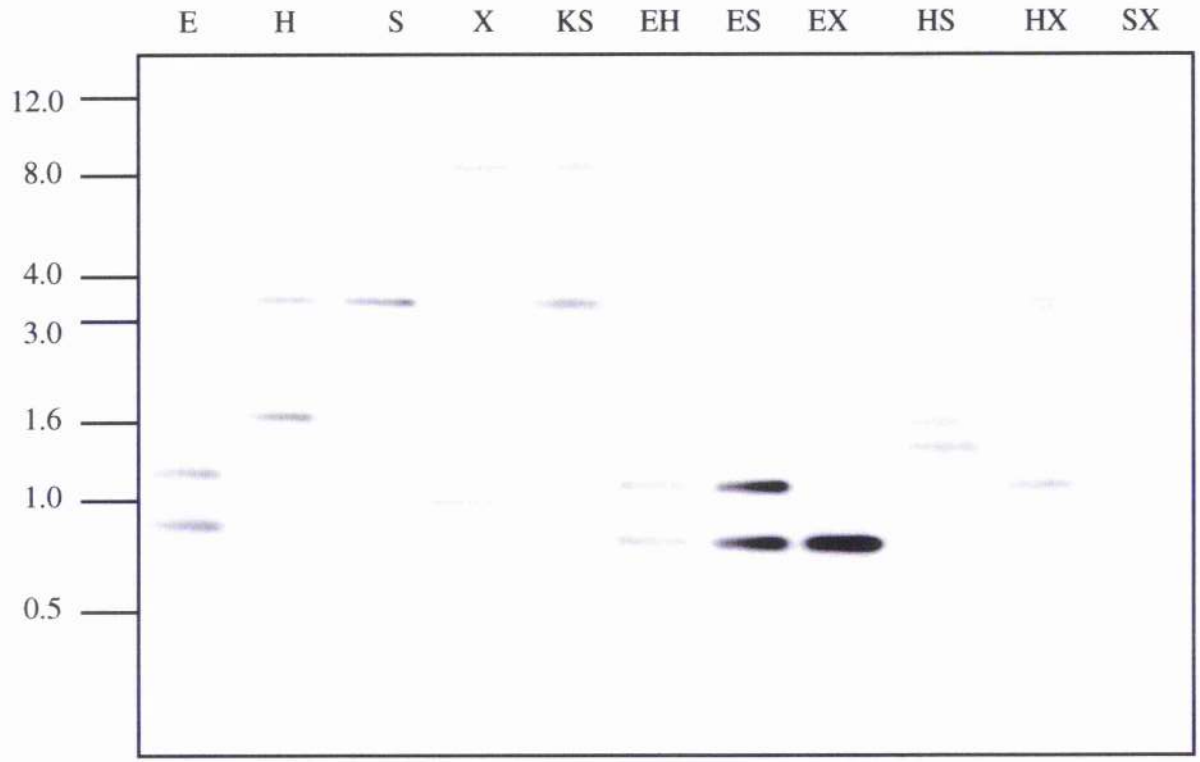


Fig. 3.22: **Southern blot analysis of λdash 13 DNA, digested with various restriction enzymes and hybridised with probe AThook1**. E: EcoRI; H: HindIII; K:KpnI; S: SpeI; X:XbaI. Molecular weight markers are indicated (in kb) to the left of the figure.

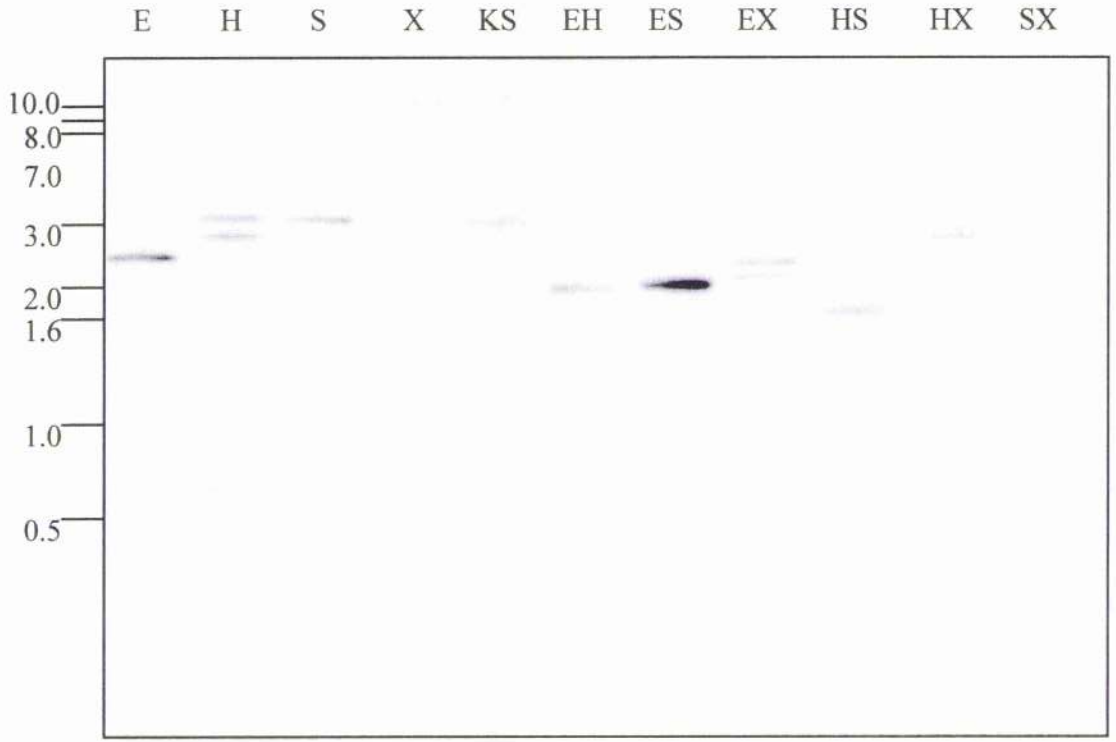


Fig. 3.23: Southern blot analysis of λ dash 13 DNA, digested with various restriction enzymes and hybridised with probe Ta369. E: EcoRI; H: HindIII; S: SpeI; X:XbaI. Molecular weight markers are indicated (in kb) to the left of the figure.

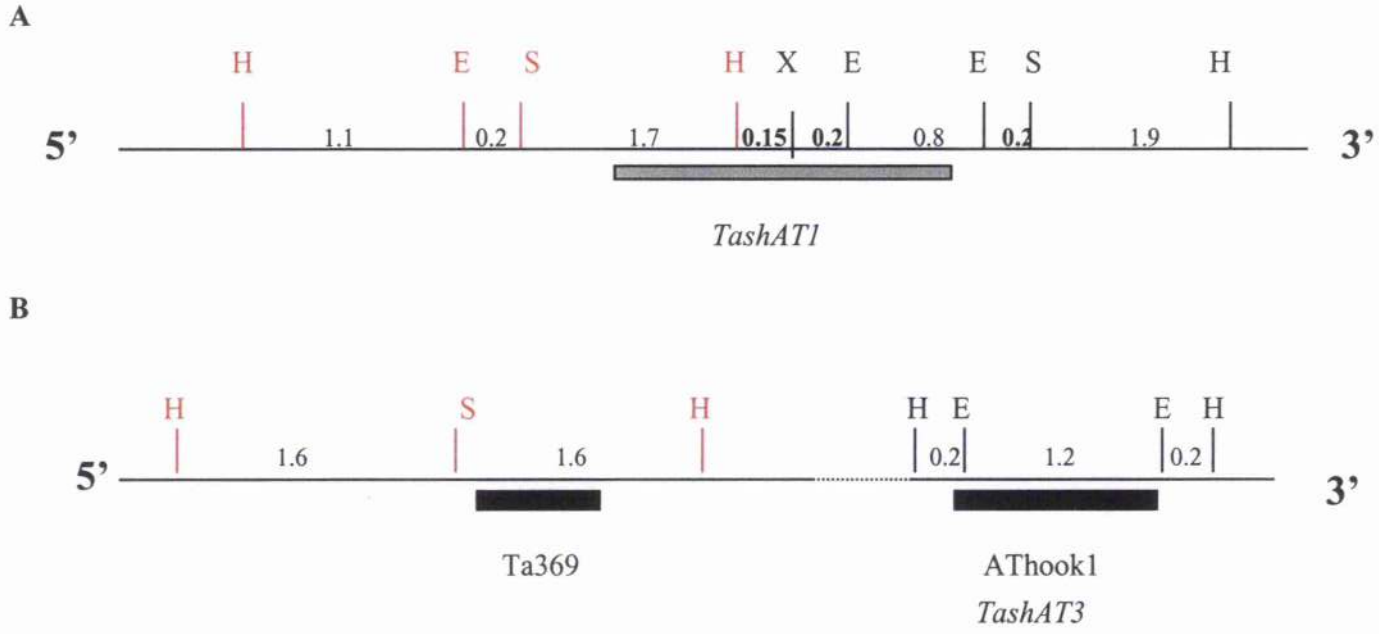


Fig. 3.24: Deduced restriction enzyme map derived from Southern blot analysis of λ dash 13 hybridised to probes AThook1 and Ta369. A and B: deduced restriction maps for *TashAT1* and *TashAT3*, respectively. Grey box indicates the deduced position of *TashAT1*; black box indicates the location of both probes. Restriction enzyme sites mapped from the Ta369 Southern analysis are indicated in red type. Restriction fragment positions, marked in bold, were obtained from the λTa1 sequence (see appendix K for restriction fragment sizes). Note, in A, the 0.2kb EcoRI/SpeI fragment three prime of TashAT1 is not shown in λTa1 sequence.

3.2.3.2 T7 or T3 mapping

T7 or T3 mapping is a technique that reveals the sites of any restriction fragment in a unidirectional, sequential order from a λ clone insert (shown in Fig. 3.25, for T7 mapping only). The insert is excised from the λ vector by NotI - located in the multiple cloning sites (MCS) of the λ arms that flank either side of the insert. NotI sites are unlikely to be found within the insert because it recognises an 8bp sequence. The T7 and T3 primer sequences are also present within the MCS within the left and right λ arms respectively, and remain with the insert after excision with NotI. The insert is partially digested by the restriction enzyme of interest and as larger fragments are digested to completion, only small regions of the insert adjoin the T7 or T3 sequence, in the MCS. The order of restriction sites from the MCS can be read from the smallest to the largest, like a sequential ladder, upon hybridisation with a T3 or T7 oligonucleotide.

λ dash 13 was mapped with respect to the T7 primer, initially. In order to compare the restriction fragments directly, the same restriction enzymes were chosen to map λ dash 13 by T7 mapping as those used in the Southern blot analysis of this clone (see Fig. 3.26). From this southern blot, a restriction map was then generated (see Fig. 3.27). The 3.2kb SpeI (λ Ta1) fragment was identified from the T7 restriction map, and was consistent with that of the sequenced λ Ta1 fragment (see Fig. 3.11). Hence the location of *TashAT1* was mapped in relation to λ dash 13.

The position of *TashAT3* was found by identifying the 1.2kb EcoRI and 1.6kb HindIII fragments, both previously found to be derived from the AT hook-encoding region of *TashAT3*. The 1.6kb HindIII fragment, which contained the 1.2kb EcoRI fragment was located 2.2kb downstream of the 3' end of *TashAT1*. This confirmed the location of the AT hook-encoding region of *TashAT3*. Inspection of the restriction map, upstream of the 1.6kb HindIII fragment to the 3' end of *TashAT1*, revealed that it was identical to that of the 5' region of *TashAT1*.

Previous Southern blot analysis with probe Ta369 identified a 3.2kb HindIII fragment (see Fig. 3.23). The only fragment of this size was found adjacent to the 1.6kb HindIII fragment at the 5' end, and encompassed the AT hook encoding region of *TashAT1* (see Fig. 3.27). In addition, a 3.2kb HindIII fragment was mapped to the 3' end of *TashAT1*, with the probe, AThook1 (see Fig. 3.22). Thus, it appears that the 3.2kb HindIII fragment detected by probes Ta369 and AThook1 comprise the second Ta369 sequence and part of *TashAT3*.

Given the close proximity between the second Ta369 sequence and the AThook3 sequence, the Ta369 sequence may be part of *TashAT3*. Thus, *TashAT1* and *TashAT3*, separated by a 1.7kb region, appeared to have almost identical sequences, although the existing data at the time was insufficient to comprehensively map the 5' or 3' end of *TashAT3*. However, once the completed *TashAT3* gene was sequenced, it was confirmed to be identical to *TashAT1* over the first 1.4kb (Swan *et al.*, 2001a). Inspection of the EcoRI and HindIII restriction map at the 5' terminus of λ dash 13 was shown to be identical to that of the 3' end of *TashAT2* (Swan *et al.*, 2001a). This indicated that *TashAT2* was approximately 2kb from the 5' end of *TashAT1*.

The restriction map shown in Fig 3.27 revealed that all three *TashAT* genes were clustered over a 13.5kb region (inclusive of the entire TashAT2 gene), and that *TashAT1* and the five prime region of the putative *TashAT3* gene were almost identical in sequence. The restriction map of the *TashAT* genes, generated from the T7 mapping technique confirmed data produced by sequence analysis (Swan *et al.*, 2001a).

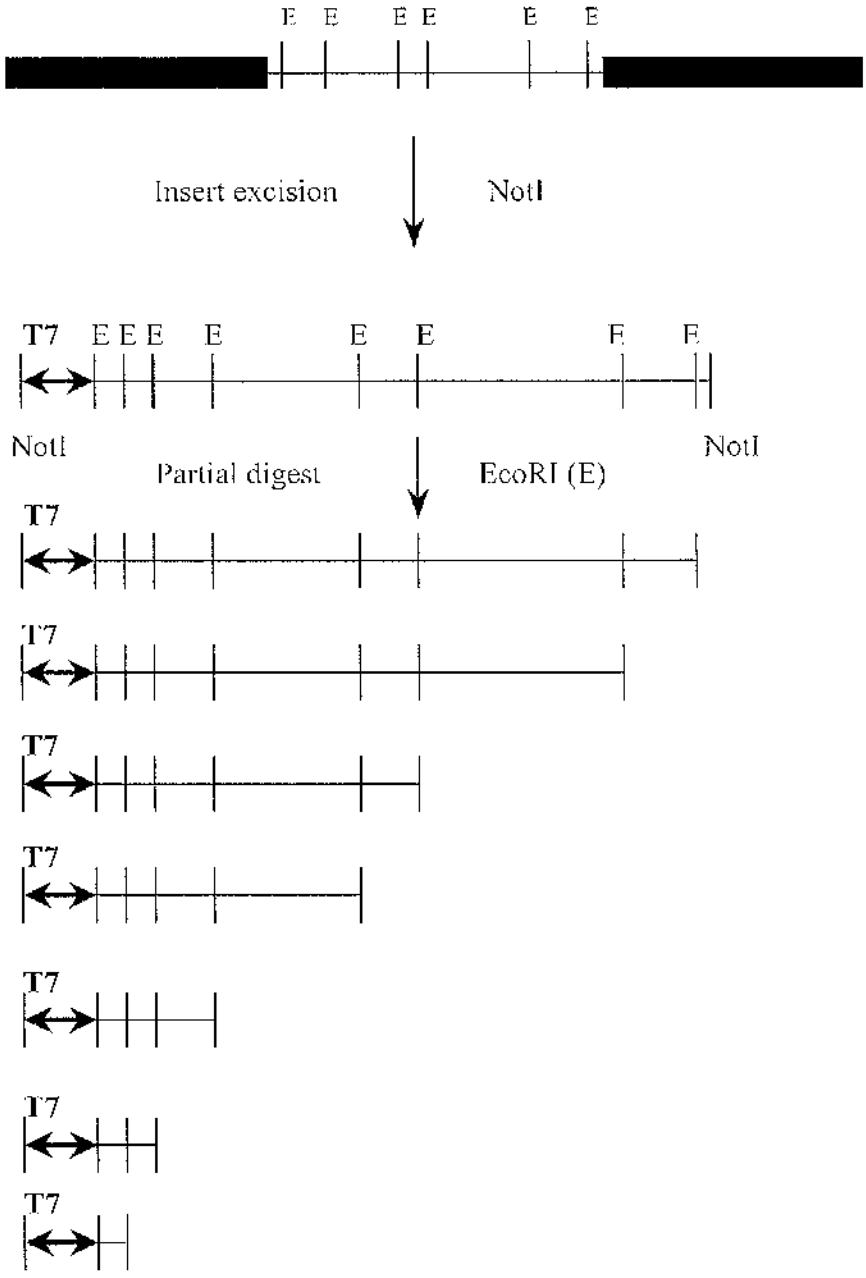


Fig. 3.25: T7 restriction mapping technique (Smith and Birnstiel, 1976), modified from Birren and Lai (1996). Black boxes represent the left and right arms of λ dash II. Double arrow represents the T7 hybridisation site. Only the intermediary fragments that adjoin theT7 site will be detected upon Southern blotting and hybridisation with the T7 oligonucleotides.

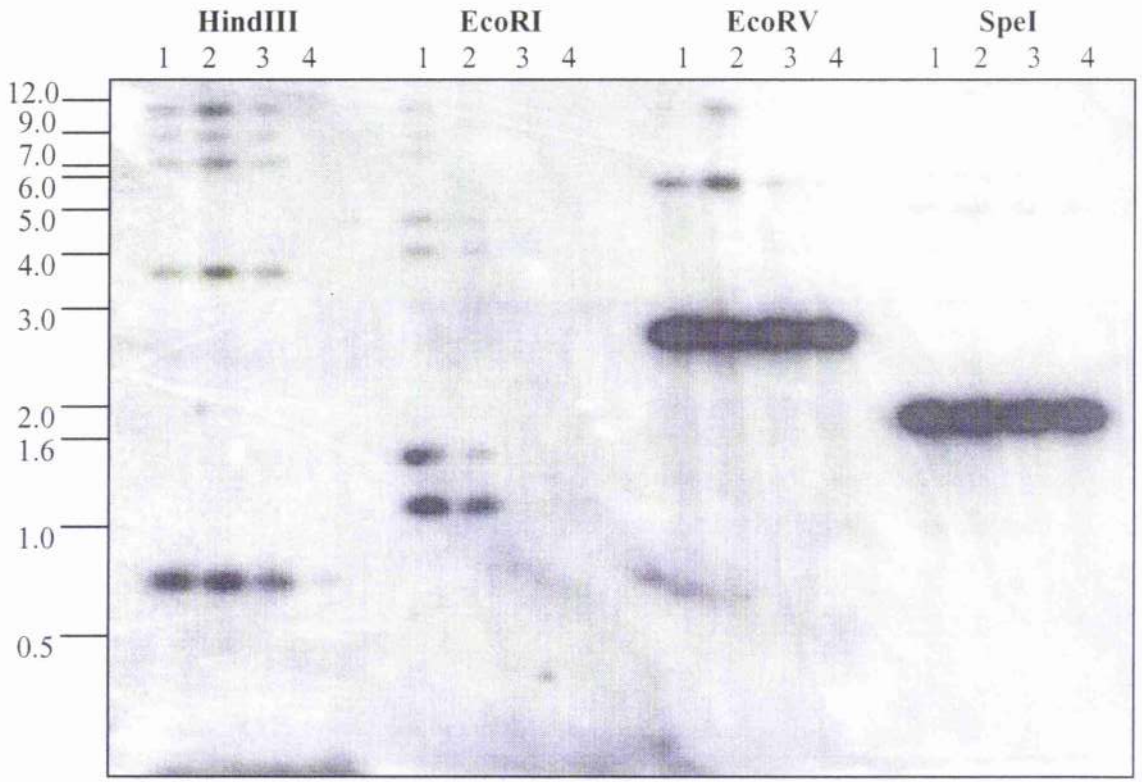


Fig. 3.26: T7 mapping analysis of λ dash 13 DNA. DNA was digested with four restriction enzymes at 0.5 minutes (1); 2 minutes (2); 10 minutes (3) and 60 minutes (4). Molecular weight markers are indicated (in kb) to the left of the figure.

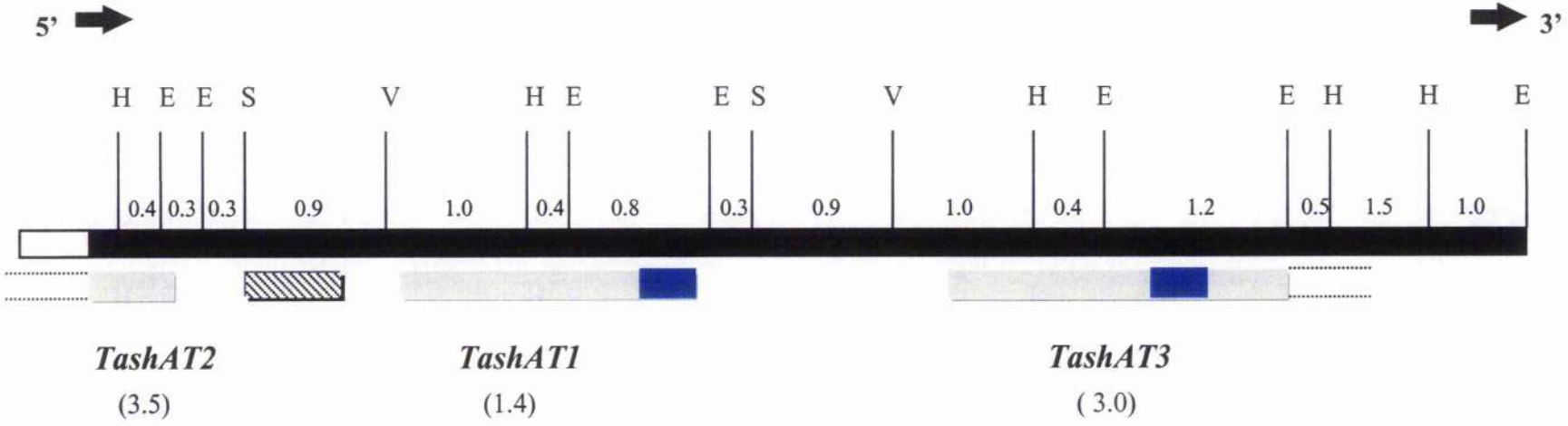


Fig. 3.27: Restriction map of the TashAT gene cluster from λ dash 13 DNA, based on T7 mapping.
Numbers represent the size of the DNA fragments or genes (in kb); E: EcoRI; H: HindIII; S: SpeI; V: EcoRV sites. Grey boxes represent the genes or gene fragments identified. White box represents the λ dash II vector sequence; black region represents the cloned insert; striped box represents the 900bp partial ORF. Blue bloxes represent AT hook encoding regions. Dashed lines indicates that the gene is expected to continue.

3.2.4 Identification of TashAT1 and TashAT3 within the T. annulata genome

To verify that the *TashAT* locus identified within λ dash 13 was a true representation of the *TashAT* locus in the *T. annulata* genome, Southern blot analysis was performed on D7 cell genomic DNA (see Fig. 3.28), using the same restriction enzymes and probes as those used to map λ dash 13. The resultant restriction fragments were compared to that of λ dash 13 to identify any possible rearrangements that could have occurred when λ dash 13 was cloned.

Inspection of the lane containing genomic DNA from non-infected bovine lymphosarcoma (BL20) cells digested with EcoRI did not reveal any bands. This confirmed that the AT hook encoding fragments were derived from the parasite only. By contrast, D7 genomic DNA digested with EcoRI produced a 0.8kb, a 1.2kb and a 3.3kb fragment. The 0.8kb and 1.2kb EcoRI fragments were previously found to derive from *TashAT1* and *TashAT3*, respectively, and the 3.0kb EcoRI was derived from *TashAT2* (Swan *et al.*, 2001a), so it appeared that the AT hook encoding regions of all three *TashAT* genes were present in the D7 genome.

Inspection of Fig. 3.28 showed that the SpeI restriction enzyme produced three fragments at 3.2kb, 3.4kb and one at approximately 10kb. Southern blot analysis of λ dash 13 with SpeI produced two fragments at 3.2kb and over 12kb. The 3.2kb SpeI fragment was the same size as the λTa1 fragment, and likely to contain *TashAT1*. The 10kb genomic SpeI fragment is likely to be the SpeI fragment (over 12kb) from λ dash 13, which contains *TashAT3*; the 2kb or more size difference might be due to the polylinker sequence derived from the λ dash II vector. The 3.4kb fragment is likely to be derived from *TashAT2*, from the restriction map generated by Swan *et al.* (2001a).

Two restriction fragments at 1.6kb and 3.2kb were generated from the HindIII digestion of genomic DNA, confirming that the HindIII fragments within λ dash 13, identified by probe AThook1 were not rearrangements. The KpnI single digest was not observed and XbaI single digest failed, so restriction digests of genomic DNA with these enzymes were not included in this analysis. Overall, the results of these single digests suggest that there are only three copies of the AT hook encoding sequence within the parasite genome, likely to be from *TashAT1*, *TashAT2* and *TashAT3*.

The double restriction digests of EcoRI, HindIII and SpeI (see Fig. 3.28), were analysed to confirm the location of these fragments in the D7 cell genome. The EcoRI and SpeI double digest, removed all the SpeI fragments and the 3.3kb EcoRI fragment, leaving the 1.2kb, 0.8kb and a 2.4kb fragment. The 1.2kb and 0.8kb fragments also remained after EcoRI/SpeI digestion of λ dash 13 (see Fig 3.22) and it is likely the corresponding genomic fragments are derived from the 3.2kb (containing *TashAT1*) and 10kb SpeI (containing *TashAT3*) fragments respectively. The 3.3kb EcoRI fragment is likely to be derived from the 3.4kb SpeI fragment as both fragments are derived from *TashAT2* and would be predicted to produce a 2.4kb fragment if digested from the restriction map of *TashAT2* (Swan *et al.*, 2001a). The double digest between EcoRI and HindIII (Fig. 3.28) failed as only the HindIII fragments were detected so these fragments could not be mapped. The HindIII/SpeI double digest removed the 3.2kb HindIII and all the SpeI fragments, leaving a 1.6kb fragment (from HindIII digest only) and a 1.4kb fragment. This suggests that the 3.2kb HindIII fragment was derived from the 3.2kb SpeI fragment containing *TashAT1*, producing the 1.4kb SpeI/HindIII fragment, consistent with the restriction map generated for λ dash 13 DNA (see Fig. 3.27). The data from Fig. 3.28 suggests that the1.6kb HindIII fragment was either derived from the 3.4kb SpeI fragment or the 10kb SpeI fragment.

To verify the restriction map corresponding to the 5' region of *TashAT1*, Southern blot analysis was performed on restriction digested D7 genomic DNA, hybridised to a 2kb EcoRI/SpeI probe that spanned the 5' coding region of *TashAT1* (see Fig. 3.29). The results revealed that the EcoRI, HindIII and SpeI digests produced fragments identical in size to those produced upon restriction digestion of λ dash 13 DNA (see Fig. 3.23), indicating that the restriction map of the five prime region of *TashAT1* and *TashAT3* is likely to be accurate. There was an additional 10kb SpeI fragment absent in the corresponding λ dash 13 DNA digest (see Fig. 3.23), possibly due to the weak hybridisation of the smaller Ta369 probe. Since the two SpeI fragments at 10kb and 3.2kb were identical in size to those detected by probe AThook1 in D7 genomic DNA, these sequences are likely to be derived *TashAT1* and *TashAT3*, contained within the 3.2kb SpeI and the 10kb fragments respectively. The 3.4kb SpeI fragment was absent, indicating that this fragment does not contain the 5' *TashAT1* sequence, consistent with the sequence of *TashAT2*. Three fragments at 3.2kb, 4.2kb and 7kb were identified in the EcoRV genomic digest. It is likely that the 3.2kb fragment is derived from *TashAT1*, as a fragment of similar size extends over this gene from the restriction map of λ dash 13 DNA. Since the HindIII and SpeI digestion only produced two copies of the 5' *TashAT1* sequence, the 4.2kb or the 7kb

fragment could be derived from *TashAT3*. One of these EcoRV fragment could represent another copy of the 900bp ORF previously identified in Fig. 3.11 and within the 2kb EcoR1/SpeI probe (see Fig. 3.27).

In summary, the fragments produced from restriction digestion with EcoRI, SpeI and HindIII in genomic DNA are maintained in the λ dash 13 clone, indicating that the restriction map of *TashAT1* and *TashAT3* within λ dash 13 is likely to be accurate. The number of AT hook encoding fragments in the genome, suggests that there are only three *TashAT* genes within the parasite genome.

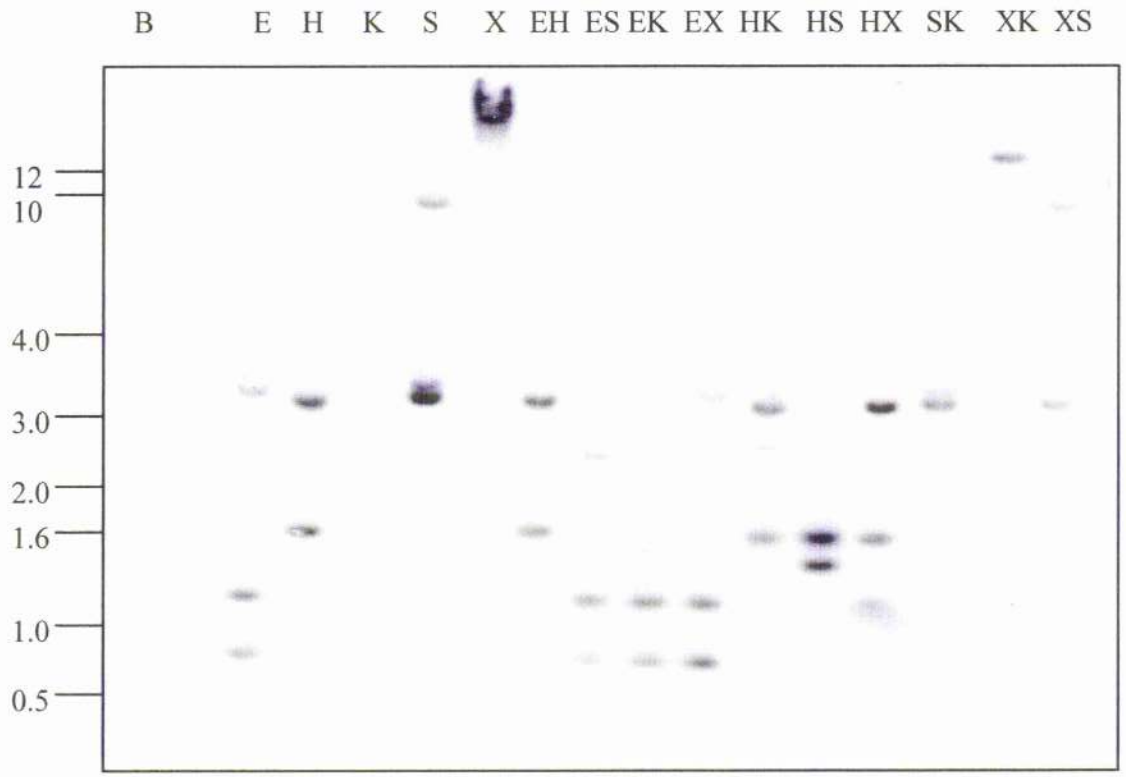


Fig. 3.28: Southern blot analysis of *T.annulata* genomic DNA from D7 cells, hybridised with probe AThook1.
B: BL20 control DNA digested with EcoRI. D7 genomic DNA was digested with EcoRI (E); HindIII (H); KpnI (K); SpeI (S); XbaI (X) alone or in combination. The molecular weight markers are indicated (in kb) to the left of the figure.

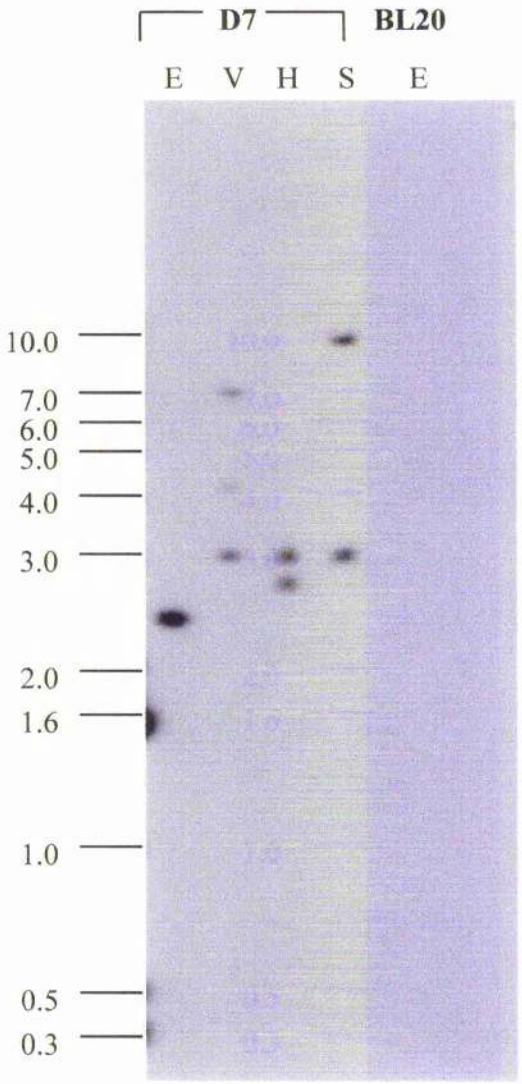


Fig 3.29: Southern blot analysis of D7 genomic DNA from *T. annulata* hybridised to a 2kb EcoRI/SpeI fragment corresponding to the 5' coding region of *TashAT1*. E: EcoRI, V: EcoRV, H: HindIII, S: SpeI. DNA molecular weight markers are indicated (in kb) on the right.

3.3 Discussion

The full length *TashAT1* gene was isolated from a 3.2kb SpeI fragment derived from λ dash 13 DNA by Southern blot analysis with an AT hook encoding probe corresponding to the *TashAT1* ORF. Sequence analysis revealed that *TashAT1* encoded a 1401bp predicted ORF, and was identical to DNA fragments derived from the *TashAT1* ORF originally identified within the λgt11 clone, cl-12. When the predicted peptide sequence of *TashAT1* was compared with that of *TashAT2* (Swan *et al.*, 1999), the two genes shared 46.1% overall identity with each other over the AT hook encoding region, confirming that they were separate, but related genes.

Southern blot analysis of λ dash 13 restriction digested DNA with the p600 probe, comprising the AT hook domains from *TashAT1*, also identified other DNA fragments that pointed to the existence of a third *TashAT* gene. One of these fragments was a 1.2kb EcoRI fragment, which was not found within *TashAT1* or *TashAT2*, but was almost identical to the AT hook domain of *TashAT1* and a region of *TashAT2*. The 1.2kb fragment was subsequently mapped to *TashAT3* (Swan *et al.*, 2001a). In fact, the first 1.4kb of *TashAT3* was found to be 99% identical to the entire *TashAT1* gene, including the AT hook encoding domain and shared 99% identity to *TashAT2* over an adjacent 1.4kb adjacent region (Swan *et al.*, 2001a).

The predicted peptide sequence of TashAT1 or TashAT3 (TashAT1/3) has a number of features that suggests the gene product of TashAT1 may be a transcription factor. These features include four AT hook motifs, several nuclear localisation signals (NLS), a signal peptide sequence and a transactivation domain.

The presence of both classical ("pat 4" and "pat 7") and bipartite NLS suggests that *TashAT1* is transported to the host nucleus. Studies have shown that the basic amino acids of the NLS interact with the nuclear binding protein (NBP) to facilitate transport through the nuclear pore complex, in an ATP dependent manner (Whiteside and Goodbourn, 1993; Hicks and Raikhel, 1995). This process is conserved amongst lower and higher eukaryotes (Hicks and Raikhel, 1995). Whilst there is a danger of false positive results, inspection of the peptide sequence of the related nuclear located transcription factor, *HMGI(Y)* (Eckner and Birnstiel, 1989; Johnson *et al.*, 1989; Karlson *et al.*, 1989), revealed 4 potential "pat 4" NLS located within the AT hook region. Moreover, several nuclear localisation signals that were found in the predicted TashAT1 polypeptide (see Appendix D) were also identified

within the putative TashAT2 polypeptide (Swan *et al.*, 1999). TashAT2 was later detected in the host nucleus of infected cells by IFAT (Swan *et al.*, 1999), and also in the nucleus of uninfected COS7 cells, previously transfected with a TashAT2 construct. This evidence might support the theory that TashAT1 and TashAT3 polypeptides may also be transported to the host nucleus. However, to determine if these signals confer host nuclear localisation, BL20 cells could be transfected with a TashAT1 fusion protein construct, with and without the NLS domains, followed by IFAT analysis.

Further evidence that TashAT1/3 might be transported out of the macroschizont is indicated by the presence of a predicted 24-residue short stretch of hydrophobic amino acids located at the extreme N-terminus, recognised as a signal sequence. The signal sequence was predicted for TashAT1 in all but one of three signal sequence prediction programs, including TargetP, the best predictor of N-terminal sorting signals (Bannai *et al.*, 2002). The signal sequence belongs to the classical eukaryotic secretory pathway and is essential for secretion of the polypeptide to the endoplasmic reticulum, for further targeting to subcellular compartments of the cell (Nakai, 1996). A similar signal sequence also exists in TashAT2, which has been shown to be located within the host nucleus of *Theileria* infected cells. Moreover, the TargetP results showed that there was a good likelihood that the signal sequence of TashAT1/3 was secreted. In this case, the signal sequence might enable the TashAT1/3 polypeptide to be translocated out of the parasite into the host endoplasmic reticulum, where, potentially, NLSs could direct the polypeptide to the host nucleus.

The AT hook motifs found in the potential peptide sequences of TashAT1, TashAT3 and TashAT2 all contain class II AT hook domains, which have a lower DNA binding affinity compared to the class I and III AT hooks (Aravind and Landsman, 1998). Thus, the TashAT polypeptides would be expected to have a relatively low DNA binding affinity. Since *TashAT1* and *TashAT3* possesses an additional AT hook domain as well as two RKRP elements, potentially capable of binding DNA (Aravind and Landsman, 1998) compared to TashAT2, it would be expected that TashAT1 and TashAT3 polypeptides would have a greater DNA binding capacity than TashAT2.

When TashAT1 was compared to protein domain families within the ProDom database, the only obvious homology found was to the AT hook motifs of the HMG protein family. This suggested that the main function of TashAT1 was DNA binding. The similarity of the AT hook encoding motifs of the *TashAT* genes and HMGI(Y) proteins (which also encode

class II AT hooks) might suggest that TashAT polypeptides perform a similar function to the HMGI(Y) proteins. HMGI(Y) proteins have been implicated in the transcriptional activation of genes associated with the immune system and cell growth, including the α subunit of the IL-2 receptor (IL-2Rα) (Reeves, 2001) and NF-κB (Thanos and Maniatis, 1992, 1995; John *et al.*, 1995). Studies have shown that HMGI(Y) proteins are involved in normal cell growth regulation, such as adipocytes (Melillo *et al.*, 2001) but are low or undetectable in fully differentiated cells (Lundberg *et al.*, 1989; Reeves and Bustin, 1996). High levels of HMGI(Y) have consistently been found to be associated with tumour formation (Tallini and Dal Cin, 1999; Reeves, 2001). Since TashAT2 polypeptide expression was highest during the macroschizont stage, when the host cell becomes immortalised (Swan *et al.*, 1999), it is possible that the TashAT2 or TashAT1/3 polypeptides might also be involved in host cell proliferation of *T. annulata* infected cell. Alternatively they might be involved in negative regulation of differentiation from the macroschizont to the merozoite.

TashAT1 was found to contain a potential transactivation domain, as these domains are thought to elicit transcriptional activation through protein-protein interactions (Triezenberg *et al.*, 1995), this might suggests that TashAT1 might also interact with other proteins. Studies have shown that transactivation domains have a significant role in transcriptional activation: isoforms of NF-AT transcription factors that do not possess transactivation domains show a third of transcriptional activity compared to NF-AT factors that do contain transactivation domains (Imamura *et al.*, 1998). The potential transactivation domains in TashAT1/3 may suggest that the TashAT polypeptides interact with one or more polypeptide, when binding to the host DNA target sequence.

The numerous phosphorylation sites predicted in the TashAT1, TashAT3 and TashAT2 peptide sequences are of particular interest, as many transcription factors are known to be regulated by phosphorylation/de-phosphorylation. For example, HMG I(Y) proteins are phosphorylated by cdc2 kinase, at the G2/M stage, which resulted in a decrease in DNA-binding activity (Nissen *et al.*, 1991; Reeves *et al.*, 1991). Specific environmental stimuli also trigger phosphorylation of HMGI(Y) proteins via the signal transduction pathway, particularly by Casein Kinase 2 and protein kinase C within B-lymphocytes (Wang *et al.*, 1995,1997; Banks *et al.*, 2000; Xiao *et al.*, 2000). Interestingly, Casein Kinase II (CKII) was also found to be associated with the transformation of host cells by *T .parva* (ole-MoiYoi, *et al.*, 1995). It is possible that the function of the TashAT polypeptides may be regulated by phosphorylation, possibly by CKII. However, these results would need to be

confirmed experimentally because these sites are defined by a small number of residues.

The predicted N-glycosylation sites suggest that TashAT1 might be secreted to a number of destinations within the cell (Helenius and Aebi, 2001; Grogan *et al.*, 2002) and these modifications might aid TashAT1 secretion out of the parasite. However, these modifications would have to be verified experimentally, particularly as N-glycosylation sites are only defined by four conserved amino acids (see appendix H). This is particularly important for the myristolation site prediction as the PSORTII and PROSITE programs appear to give conflicting results. In addition, there are complex considerations regarding the sequence context when predicting a true myristolation site (Grogan *et al.*, 2002; Towler *et al.*, 1988).

The detection of a possible PEST sequence within the predicted amino acid sequence of TashAT1, indicated that TashAT1 might be targeted for proteolytic degradation. Studies on PEST sequences (reviewed by Rechsteiner and Rogers, 1996) revealed that these sequences are present in metabolic enzymes, protein kinases and phosphatases. Transcription factors such as Fos, Jun, p53, IκB and cyclins are also targeted for degradation, as a form of biochemical regulation mostly via the 26S-ubiquitin protease pathway. PEST sequences can be either constitutive proteolytic sequences or conditional signals; where, in the latter case they are targeted after modification, such as phosphorylation. One theory put forward is that phosphorylation of serine and threonine residues activate latent PEST signals for proteolysis (Rechsteiner and Rogers, 1996). Phosphorylation sites overlap with the potential pest sequences of TashAT1/3. Thus, if these phosphorylation sites are proven to exist experimentally, then it seems likely that the predicted PEST sequence of TashAT1 would be a conditional signal. Further studies involving mutational analysis of the PEST sequence are required to verify if there is an active PEST sequence within TashAT1/3. One interesting observation is the high abundance of PEST sequence encoding polypeptides that give rise to immunogenic peptides presented on MHC I molecules (Rechsteiner and Rogers, 1996). If TashAT1/3 is shown to be expressed in this way, then these polypeptides could potential be vaccine candidates.

The 5' upstream regions of *TashAT1* and *TashAT3*, were identical, but had poor identity with the 5' upstream region of *TashAT2*, apart from a 31bp motif, common to all three *TashAT* genes and another macroschizont encoded gene, *Tash1* (Swan *et al.*, 2001b). Thus, this motif might bind a common regulatory factor that controls the transcription of all three *TashAT* genes, and maybe other macroschizont genes. The TashUM sequence was not

similar to any other eukaryotic transcription factor binding site, in common with CAT1, a possible regulatory motif of the gene encoding the *T. annulata* merozoite surface protein, *Tams1* (Shiels *et al.*, 2000) and is therefore more likely to be unique to *T. annulata.* This is supported by evidence of other potential regulatory regions in the related apicomplexan, *Plasmodium*, which were dissimilar to other eukaryotic transcription factor binding sites (Horrocks *et al.*, 1998).

It has been shown that host factors are translocated into the parasite (Carrington *et al.*, 1995), so the lymphocyte specific factors identified by the MatInspector search are of particular interest, especially as one factor, Pu.1, associated with cell growth and IKRS with lymphoid differentiation. However, this analysis only gives a theoretical possibility There is high probability of mistaken matches as some of the transcription factor binding sites are degenerate or A.T rich (e.g. SATB). This latter case might produce biased results as *Theileria* contains more A and T nucleotides in its genome compared to higher eukaryotes. Distance and orientation between promoter elements are also considerations as many promoter elements are regulated synergistically or antagonistically. Further investigations into the lymphocyte associated transcription factors would be required to determine if they are expressed in *Theileria* infected cells and are translocated to the parasite.

Secondary structural predictions have predicted that the potential TashAT1 polypeptide is composed of mainly looped structures, particularly over the AT hook encoding domains, which is consistent with structural studies of AT hook motifs (Reeves, 2001). The looped structures, together with the high solvent accessibility values over this region and between residues at the C-terminal 139 amino acids suggest that these regions may be exposed on the surface of the TashAT1 protein. It is possible that this exposed looped structure might enable the AT hook motifs to interact with the target DNA sequence. The extended β sheet structure over the N- terminus of TashAT1 suggests that the N-terminus is internal.

To identify the genomic organisation of *TashAT1* in relation to *TashAT2* and *TashAT3*, Southern blot analysis was performed on λ dash 13 DNA, using two probes derived from *TashAT1*. One of the most striking findings was that all the *TashAT* genes are located in close proximity to each other. Also, the N-terminal region of *TashAT3* is 99% identical to the entire *TashAT1* gene, whilst *TashAT3* shares 83% overall identity with *TashAT2*, which increases to almost 100% in certain regions. These observations suggest that there has been a relatively recent, tandem duplication event involving *TashAT3* and an adjacent gene

TashHN (Stadler, unpublished, 2000) that was partially encoded by the 900bp ORF (see fig. 3.30). Studies by Stadler (unpublished, 2000) have shown that *TashAT1* is not present in all *Theileria* infected cell lines, supporting the evidence of a recent duplication event. It is likely that *TashAT1* arose from an intrachromosomal duplication, as these duplicated sequences tend to be physically close together and more similar and shorter than other types of sequence repeats (Achaz *et al.*, 2001). One possible reason for this duplication event could be a response to selective pressure. This reason has been put forward to explain changes in gene expression detected within infected cells after prolonged culture by many groups (Adamson *et al.*, 2000ab; Sutherland *et al.*, 1996; Preston *et al.*, 1998; Oura *et al.*, 2001). The possibility that *TashAT1* was duplicated under selective pressure of *in vitro* cultured conditions, is unlikely given that TashAT1 mRNA was found in sporozoites (see Swan *et al.*, 2001a). The advantages of this duplication event may be to increase the gene dosage of *TashAT3* and/or *TashHN*. Alternatively, TashAT1 might be able to perform a slightly different function to TashAT3 that is advantageous to the parasite. Gene duplication has been reported in other apicomplexan parasites such as TgPCNA1 and 2 in *Toxoplasma gondii* (Guerini *et al.*, 2000) and the Merozoite Surface Protein (MSP) gene family in *Plasmodium chabaudi* (Black *et al.*, 1999) and this event may be more common in apicomplexans. One study found more gene repeats in *Plasmodium* compared to other eukaryotes (Achaz *et al.*, 2001), which was hypothesised to be due to the high selective pressures for sequence variation. A high number of repeats might allow for more recombination events, leading to greater genetic variation and evolutionary rate (Achaz *et al.*, 2001).

A comparison of the predicted amino acid sequence of TashAT1 with that derived from the *T. parva genome* revealed no equivalent *TashAT1* or *TashAT3* genes in *T. parva*, despite the fact that *T. annulata* and *T. parva* diverged from a common ancestor (Chansiri *et al.*, 1999). The lack of *TashAT* genes and AT hook motifs found in the *T. parva* genome compared to *T. annulata*, might suggest these gene products are not essential for *T. parva* and may relate to the different cell types these species infect. However, the sequencing of *T. parva* genome is not yet complete and may yet reveal a *T. parva* homologue of a *TashAT* gene.

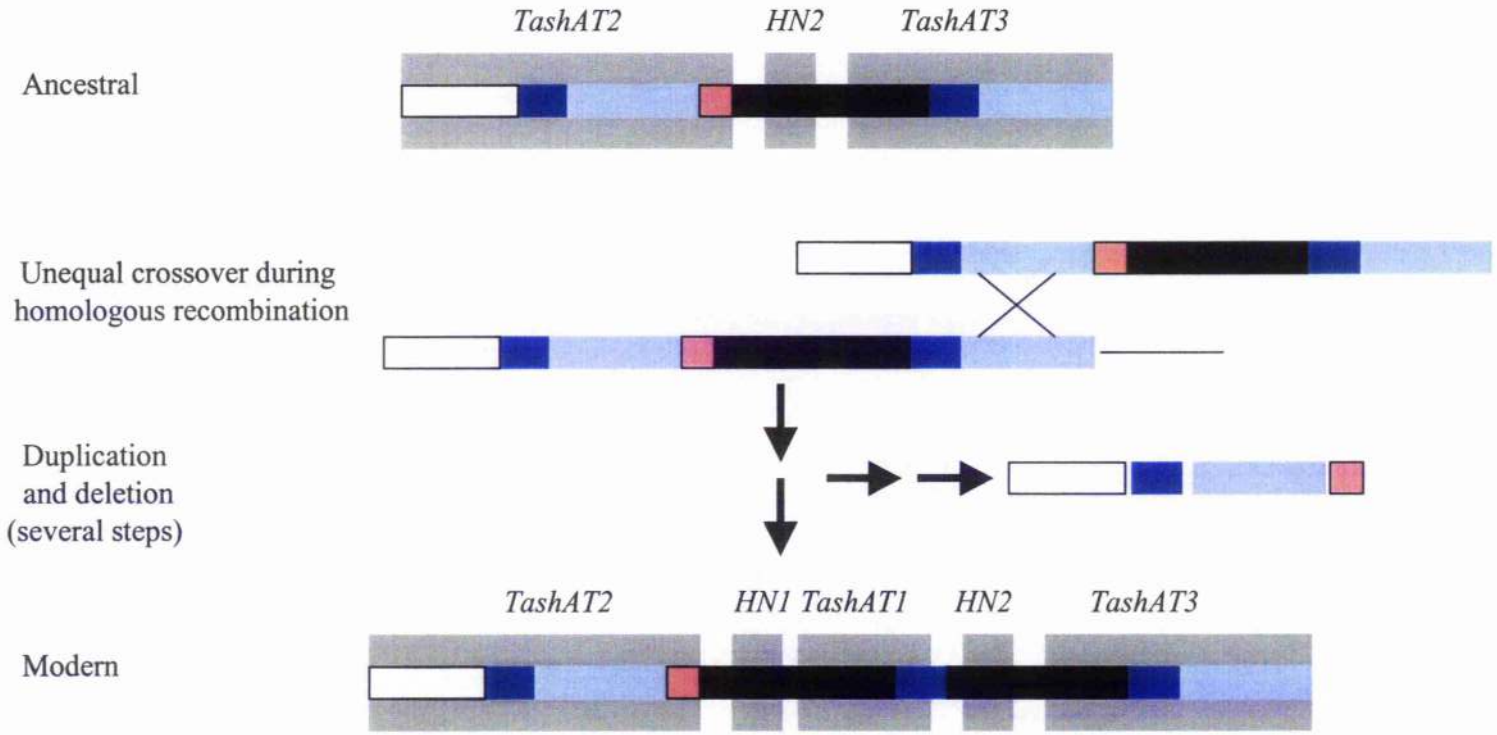


Fig. 3.30: A possible evolutionary mechanism for the existence of *TashAT1* by one or a series of non-equal crossover event(s) between the ancestral *TashAT2*, *TashAT3* or other genes , losing the C-terminal part of *TashAT2*. Coloured boxes represent areas of homology, with purple box indicating the AT hook encoding region. Grey areas represent open reading frames. *HN1* and *2* represent *TashHN1* and *2* respectively.

4. Identification and Expression of the TashAT mRNA species and Analysis of a TashAT Upstream Motif.

4.1 Introduction

In chapter three, two further *TashAT* genes, *TashAT1* and *TashAT3* were identified. RFLP mapping located these genes to a 13.5kb gene cluster together with *TashAT2*. The sequences of *TashAT1/3* were homologous with *TashAT2*, within the AT hook domain region. Later sequence comparisons showed that the entire *TashAT1* ORF was 99.9% identical to the five prime end of *TashAT3*.

The striking sequence similarities between the *TashAT* genes and polypeptides coupled to the likely identification of TashAT2 in the host cell nucleus imply that TashAT1 and TashAT3 may also be involved in host/parasite interactions. As the predicted TashAT1 polypeptide may be involved in regulation of the host cell environment, it was important to define the mRNA species encoded by *TashAT1*, and to determine the expression profile of TashAT1 and TashAT3 during differentiation to the merozoite. This could be achieved by Northern blot analysis on RNA from infected cells undergoing a differentiation time course, using gene specific and common probes. Any variability in the function of the *TashAT* genes could be investigated by comparing the relative levels of TashAT mRNA transcripts between members of this gene family: a similar profile would provide further evidence that these genes may be under a common regulatory mechanism.

A further aim of this chapter was to confirm regulation of parasite differentiation by Northern blot analysis on parasite infected cell lines attenuated for merogony, such as the D7B12 cell line. If the *TashAT* genes were associated with differentiation and/or host cell proliferation, such experiments might be predicted to show a lack of down regulation of the TashAT mRNA in the attenuated cell line.

In addition to determining if the expression of TashAT mRNA was associated with differentiation, it was of relevance to investigate how *TashAT1/3* gene expression is regulated. Such evidence could provide an insight into the manner by which the parasite down regulates macroschizont gene expression during differentiation to the merozoite.

This may due to the simple removal of macroschizont-specific regulators, or it could be due to the inhibition of those factors by regulators specific to the merozoite stage.

Analysis of genes from related apicomplexan parasites shows that most genes have a classical eukaryotic bipartite structure, consisting of a basal promoter and upstream enhancer elements. Examples of such genes are the *pfs25* and *GBP130* genes in *P. falciparum* (Horrocks *et al.*, 1998; Horrocks and Lanzer, 1999) and the *NPT3* gene in *Toxoplasma gondii* (Nakaar *et al.*, 1998). A recent review by Horrocks *et al.* (1998) and van Lin *et al.* (2000) indicated that *Plasmodium* gene transcription occurs in a monocistronic manner. However, whilst there is similarity in the structure of apicomplexan and higher eukaryotic genes, there is growing evidence that there are differences between higher eukaryotic and apicomplexan transcription mechanisms. For example, the large number of putative transcription factor binding motifs upstream of *P. falciparum* genes have not yielded functional, homologous binding sites. Moreover, only one putative TATA box binding protein (PfTBP) has been cloned in *P. falciparum* to date (McAndrew *et al.*, 1993). Although similar in structure to the TBP of higher eukaryotes, the *P. falciparum* TBP had a low homology with the TBP of higher eukaryotes, despite the normally high evolutionary conservation of this protein amongst other eukaryotes (Nikolov *et al.*, 1994). Furthermore, a comparison of putative enhancer regions of *P. falciparum* genes show no homology to any other known eukaryote (reviewed by Horrocks *et al.*, 1998; van Lin *et al.*, 2000), and implied that *P. falciparum* contains a distinct, unique set of transcription factors. Evidence that Plasmodial transcription differs from that of higher eukaryotes was demonstrated when the common eukaryotic SV40 promoter, was transfected into *P. falciparum* but failed to result in any reporter gene expression. Conversely promoters from *P. falciparum* also fail to drive reporter gene expression in COS7 cells (Horrocks *et al.*, 1998 and references therein).

Recent analysis of the *T. annulata* gene, *Tams1*, identified a unique, 9bp motif (called CAT1) upstream of the transcription start site (TSS) of *Tams1*, which was postulated to function in control of *Tams1*. The CAT1 motif generated a specific band shift in an Electromobility band shift assay (EMSA) associated with parasite enriched nuclear extracts from infected cells compared to control, uninfected BL20 extracts (Shiels *et al.*, 2000). In addition, a nucleotide search failed to produce any strong identity to any known promoters/enhancers, indicating that CAT1 may be a motif unique to *Theileria* species.

While the CAT-1 motif of *Tams1* was initially defined by EMSA, it was also shown to be conserved in the *Tams1* orthologues of three *Theileria* species. In chapter three, a highly conserved region located upstream of the TSS of all the *TashAT* genes and also in an unrelated macroschizont gene, *Tash1* (Swan *et al.*, 2001b) was identified. Overall, the upstream region of *TashAT1/3* shared 78.8% identity with the upstream motif of *TashAT2*, and 87.1% identity with the upstream motif of *Tash1* (see Fig. 3.18). This motif is known as TashUM (*TashAT1/3* upstream motif). TashUM and its homologues upstream of *TashAT2* and *Tash1* appear to be separated into a proximal region with a relatively abundant GC rich content (70%) and a distal region with an average relatively rich AT content of 86.2%. One striking feature of the GC rich region is that the upstream regions of all four genes are almost identical bar one nucleotide. Thus the TashUM motif was chosen for further analysis because the similarity of the TashUM motif of *TashAT1/3* with the upstream motif of *TashAT2* and *Tash1* clearly implied functional conservation, possibly of a regulatory factor-binding site. In the absence of a transfection and *in vitro* transcription system for *Theileria*, it was decided that a preliminary study of the capability of the TashUM motif to bind nuclear proteins would be performed, using EMSA. If this motif is involved in the regulation of *TashAT1*, such an experiment could be predicted to show a band shift with parasite extracts only. However, the possibility that specific shifts could be generated with host enriched extracts cannot be ruled out, particularly as there is evidence of translocation of parasite proteins to the host (Swan *et al.*, 1999), and host proteins to the parasite (Carrington *et al.*, 1995).

Structural studies of genes from *Plasmodium* indicated that transcription is not initiated by a TATA box (Horrocks *et al.*, 1998) but by an initiator element (Inr) ; a 7-9bp alternative or complementary TATA box found in some eukaryotes (Smale and Baltimore, 1989). Potential Inr elements have also been identified upstream of *T. annulata* genes, such as *Tams1* (Shiels *et al.*, 2000), *TashAT2* and *Tash1* (Swan, unpublished, 2001). To characterise the upstream region of *TashAT1*, the method of 5' Rapid Amplification of cDNA Ends (RACE) would be used to locate the transcription start site (TSS) of *TashAT1*. This would enable the identification of any potential Inr element and locate the position of the TashUM motif relative to the TSS. 5' RACE was chosen because it is a rapid, sensitive technique for cloning the five prime end of any mRNA species, by PCR amplification. This technique has the advantages of amplifying low concentrations of mRNA and reduces non-specific products by the use of nested primers and results in a product that can be cloned and sequenced. Other methods that determine the transcription start site are also available, such as primer extension (Goodbourn, 1996) and S1 nuclease mapping (Berk and Sharp,

1977). However, the use of S1 nuclease (which is specific for single stranded nucleic acids) to map an AT rich DNA sequence often leads to inaccurate results due to partial removal of nucleotides within AT rich sequences that have transiently become single stranded. Furthermore, the primer extension method is insensitive and prone to background problems (Goodbourn, 1996).

In summary the work in this chapter was aimed at determining how TashAT1 expression is regulated during differentiation to the merozoite and whether this was similar to other members of the *TashAT* cluster. In addition the work set out to define the upstream region of *TashAT1* by identification of the TSS. This would allow further comparison with the upstream regions of *TashAT2* and *Tash1* by locating the position of the 31bp TashUM motif. Analysis of this motif for binding parasite nuclear factors would then be carried out by EMSA.

4.2 Results

4.2.1 Identification of the mRNA species encoded by TashAT1 and TashAT3

In order to identify the mRNA species encoded by *TashAT1*, Northern blot analysis was performed with RNA from non-differentiating *T. annulata* infected D7 cells. BL20 RNA was also included as a negative control to demonstrate that the mRNA detected was specific to the parasite. Hybridisation was carried out using two gene specific probes to confirm the identity of *TashAT1*. Probe AThook1 was located 799-1556bp relative to the translation start site of the *TashAT1*. Probe Ta369 was situated 55-423bp relative to the translation start site of the *TashAT1*. Initial hybridisation of D7 RNA with probe AThook1 identified three transcripts (shown in Fig. 4.1, lane 2) which were estimated to be 2.1kb, 3.6kb and 4.0kb relative to size standards. No bands were detected with BL20 RNA (Fig. 4.1, lane 1), confirming that these mRNA species were not derived from the host cell. The sizes of two of these transcripts corresponded approximately to the open reading frames of *TashAT1* (1.4 kb) and *TashAT2* (3.6 kb), and after RNA processing, suggested that the 2.1kb and 4.0kb transcripts represent TashAT1 and TashAT2 mRNA respectively. The subsequent discovery of the AT hook-encoding gene, *TashAT3*, with an ORF of 3.0 kb made TashAT3 the most likely candidate for the 3.6kb transcript. However, it is possible that the higher bands could represent an unprocessed TashAT1 transcript or another highly related gene. TashAT1 appeared to be the most abundant messages followed by TashAT2 and lastly TashAT3.

To confirm the identity of the TashAT1 transcript the second gene specific probe, Ta369, was used to probe a Northern blot of RNA from D7 cells (see Fig. 4.1, lane 3). Two transcripts were detected with this probe at 2.1kb and a feinter band at 3.6kb; the 4.0kb signal was absent. No signal was detected in the BL20 RNA tract with this probe (see Fig. 4.1, lane 4). Both these transcripts were identical in size to the previous transcripts in Fig. 4.1 (lane 1), deduced to be TashAT1 and TashAT3. In addition, the higher abundance of TashAT1 relative to TashAT3 detected with the AThook1 probe was also observed with probe Ta369. The mRNA species detected by probe Ta369 were therefore deduced to be TashAT1 (2.1kb) and TashAT3 (3.6kb).

In summary, two common transcripts at 2.1kb and 3.6kb were detected using two probes derived from *TashAT1* by Northern blot analysis. These messages were deduced to be TashAT1 and TashAT3 mRNA respectively on the basis of size.

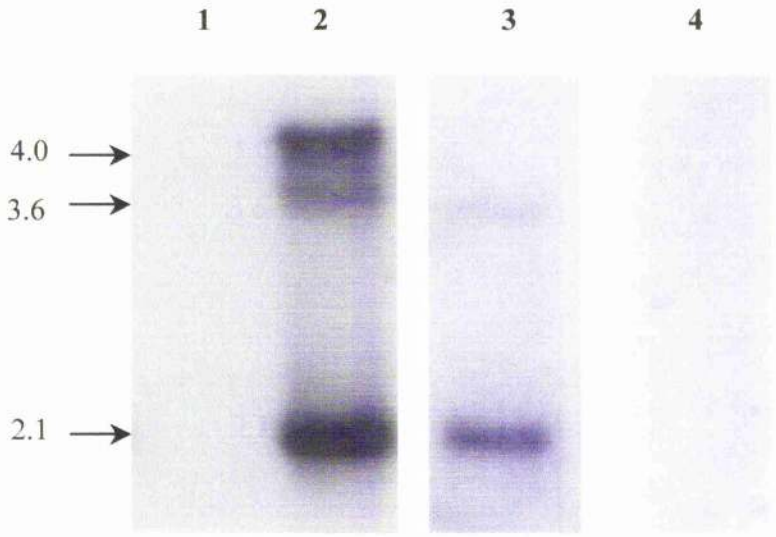


Fig. 4.1: Northern blot analysis of total RNA derived from D7 cell line, hybridised to probe AThook1 and Ta369. 10μg of RNA was used in each lane. Lanes 1 and 2: D7 and BL20 RNA hybridised to probe AThook1, respectively. Lanes 3 and 4: D7 and BL20 RNA hybridised to probe Ta369, respectively. Arrows indicate the size of the mRNA transcripts (in kb).

4.2.2 Determination of the expression profile of TashAT1 mRNA during differentiation to the merozoite.

To establish if TashAT1/3 mRNA expression was down regulated during differentiation to the merozoite, a time course experiment was performed with D7 cells, which have an enhanced ability to differentiate (Fig. 4.2). Total RNA from D7 cells was purified from a differentiation time course, incorporating cells cultured at 37°C (day 0), and from cells grown for 2, 4 and 6 days at 41°C, transferred onto a northern blot and hybridised with probe Ta369. To control for the amount of parasite RNA present, the same blot was hybridised to a probe representing the *T. annulata* 18S sn rRNA gene, called 2p3 (Swan *et al.*, 1996) The blot was also hybridised to a gene probe encoding a rhoptry protein, which is known to be up regulated at the mRNA level during merogony (Swan, unpublished, 2001) and is therefore a control for differentiation of the parasite (Fig. 4.2C).

Hybridisation with the probe Ta369 (Fig. 4.2A) only detected the TashAT1 transcript (at 2.1kb), and not the TashAT3 gene transcript, which may have been present at levels too low to detect in this experiment. Hybridisation of these RNA samples with the 2p3 and Rhoptry gene probes revealed a 3.5kb and 3.3kb transcript in both Fig. 4.2B and 4.2C, which corresponded correctly to the sizes of the 18S snRNA of *T. annulata* and the Rhoptry gene mRNA species respectively. The signal intensities of hybridised TashAT1, 2p3 and Rhoptry gene RNA transcripts were measured by densitometric analysis (shown in Fig. 4.3) to determine the percentage change in the expression levels of each RNA species (see Fig. 4.4).

The data presented in both Fig. 4.3C and 4.4C, showed that Rhoptry gene RNA had increased markedly from day 0 to day 6, confirming that the D7 cells were differentiating. By comparison, TashAT1 mRNA levels appeared to increase from day 0 to day 2 (by 42.3%), then decreased from day 2 to day 4 (by 59.6%), and remained relatively constant from day 4 to day 6 (see Fig. 4.4A). From day 0 to day 2, the levels of 2p3 RNA also increased from day 0 to day 2 (by 10.3%), and then increased by 15.3% from day 2 to day 4 but only increased marginally (by 1.7%) from day 4 to day 6 (Fig. 4.3B). Thus, from day 2 to day 4, TashAT1 levels had actually decreased during merogony, in comparison to 2p3 RNA levels, which actually increased during this time period (see Fig. 4.4A and B), with both RNA species remaining approximately constant from day 4 to day 6. However, from day 0 to day 2, it was unclear whether TashAT1 mRNA levels had truly increased relative to the levels of 2p3 RNA as both TashAT1 and 2p3 levels increased. Moreover, in the

absence of an RNA standard of known concentration for each autoradiograph, it was not possible to compare the relative increase in 2p3 RNA levels with TashAT1 mRNA levels.

To compare the expression levels of TashAT1 mRNA with other TashAT transcripts, Northern blot analysis was performed on differentiating D7 time course cultures using the *TashAT1* probe, AThook1 and with probe 2p3 (see Fig. 4.5). The smear shown with the 2p3 hybridisation (Fig. 4.5B) was likely to be caused by excess unincorporated radioactive label, or a saturated 2p3 signal because no degradation was detected upon hybridisation with AThook1 (Fig. 4.5A). As such, densitometry measurements for the 2p3 hybridised blot were carried out on the 3.5kb band, excluding the smeared area. The results of the densitometric analysis (see Fig. 4.6), showed that all three TashAT mRNA expression levels increased from day 0 to day 2 but all decreased between day 2 to day 6. Meanwhile, 2p3 RNA levels had increased from day 0 to day 6, but particularly from day 0 to day 2. However, the level of increase in 2p3 RNA could not be determined as the densitometry readings in Fig 4.6B appear to show that the 2p3 signal has reached saturation point at day 2. Thus, from day 2 to day 6, the levels of TashAT1, along with TashAT2 and TashAT3 had decreased in comparison to the total parasite RNA load.

Comparing the relative changes in abundance between the TashAT transcripts from Fig. 4.7 revealed that from day 0 to day 2, the increase in TashAT1 (56.4%) was less rapid than TashAT2 (73%) and TashAT3 (77%). However, from day 0 to day 6, the decrease in TashAT1 mRNA levels were similar to those of TashAT2 and TashAT3. Fig. 4.7B shows that 2p3 RNA levels had increased by 46.5% from day 0 to 2 and continued to increase by 4.8% and 9.8% from day 2 to 4 and from day 4 to 6, respectively. This confirmed that the TashAT transcripts had decreased relative to the levels of parasite RNA during this period. It may be noted that the signal intensity of TashAT1 mRNA was greater using probe AThook1 compared to probe Ta369. This may be due to the presence of nicked RNA in the Northern blot hybridised with probe Ta369, which does not contribute to the signal, but does contribute to the background (Goodbourn, 1996). Furthermore, probe AThook1 is more than twice the size of probe Ta369, and is therefore likely to produce a stronger signal than probe Ta369.

In summary, TashAT1 mRNA appeared to increase from day 0 to day 2, but decreased from day 2 to day 6 with respect to merogony. Northern blotting with probe AThook1 showed that TashAT2 and TashAT3 mRNA increased at a greater rate than TashAT1 mRNA from day 0 to day 2, although the overall abundance of TashAT1 was larger than

TashAT2 and TashAT3 at day 0 and day 2. The rate of decrease of all three TashAT transcripts were approximately similar from day 2 to day 6.

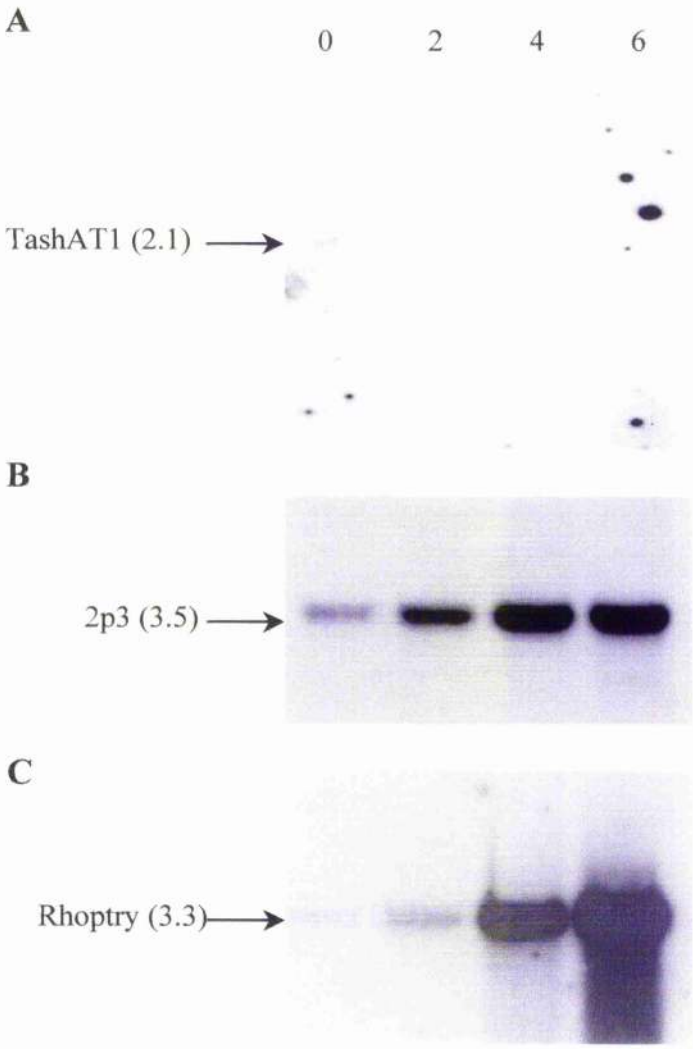


Fig. 4.2: Northern blot analysis of total RNA derived from D7 cells taken during a differentiation time course experiment and hybridised to Ta369. 10µg of RNA was used in each lane. Panels A: Ta369; B: 2p3 and C: Rhoptry gene. The time points for each RNA sample are indicated above each lane (in days). The detected RNA transcripts (in kb) are indicated by arrows.

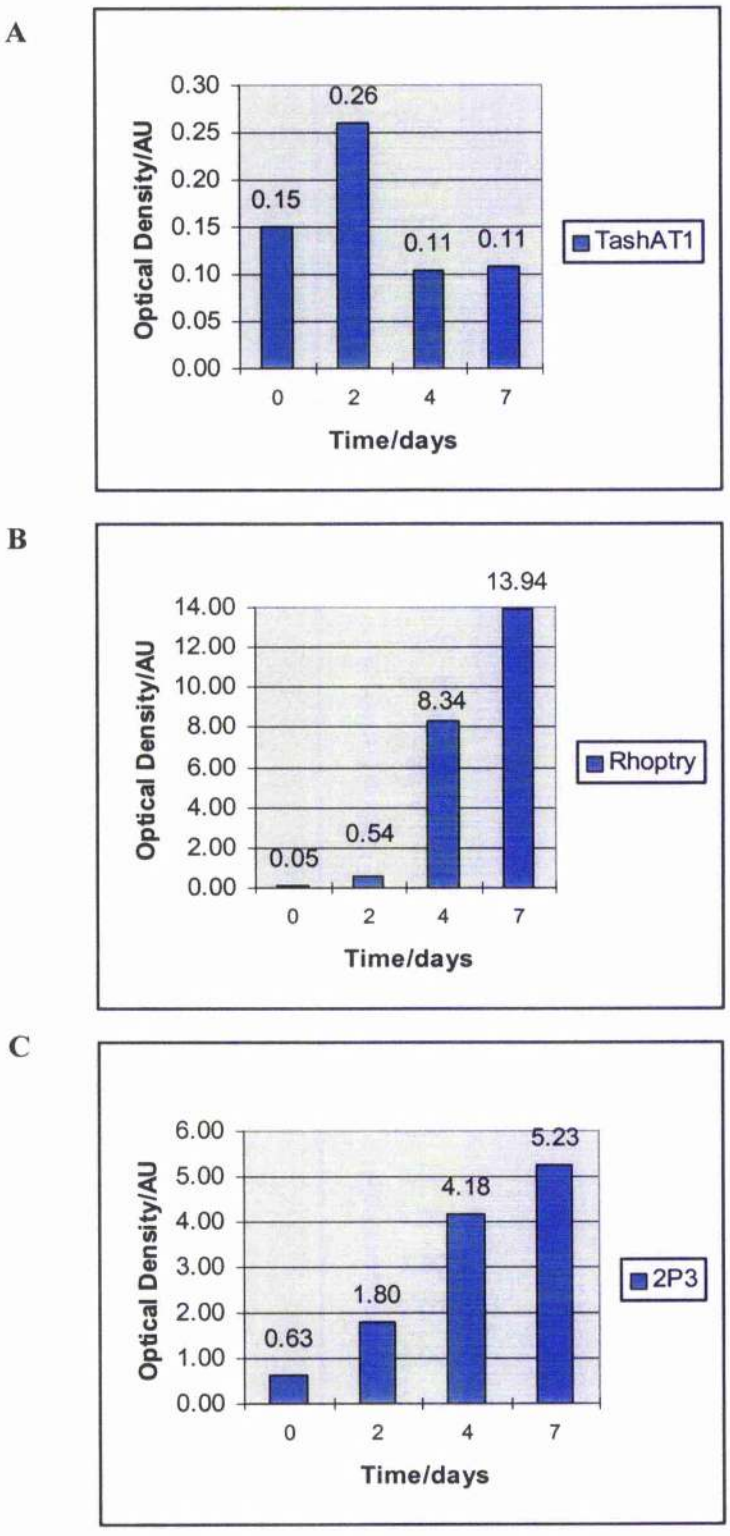
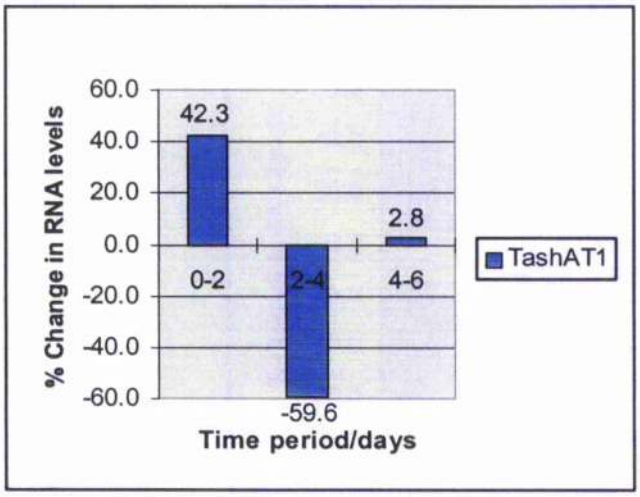
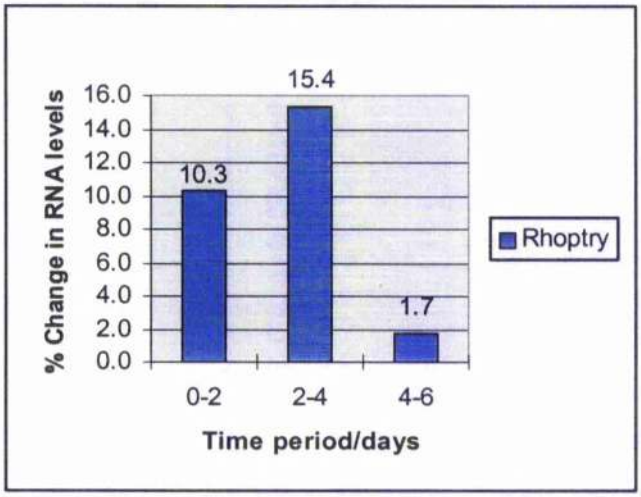


Fig. 4.3: Densitometric analysis of TashAT1 compared to 2p3 and Rhoptry gene RNA levels during a differentiation time course of D7 cells. A: TashAT1 mRNA; B: Rhoptry gene RNA; C: 2p3 RNA; AU: Arbitrary Units.

A



B



C

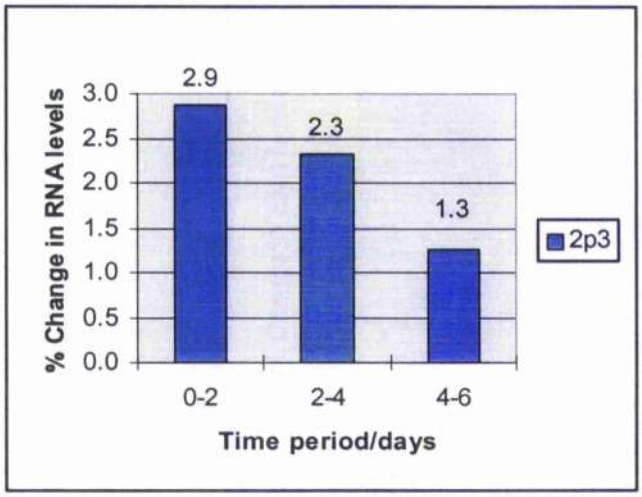


Fig. 4.4: Percentage change in TashAT1, Rhoptry gene and 2p3 RNA levels during a D7 cell differentiation time course.
Panels A: TashAT1 mRNA; B: Rhoptry gene RNA and C: 2p3 RNA.

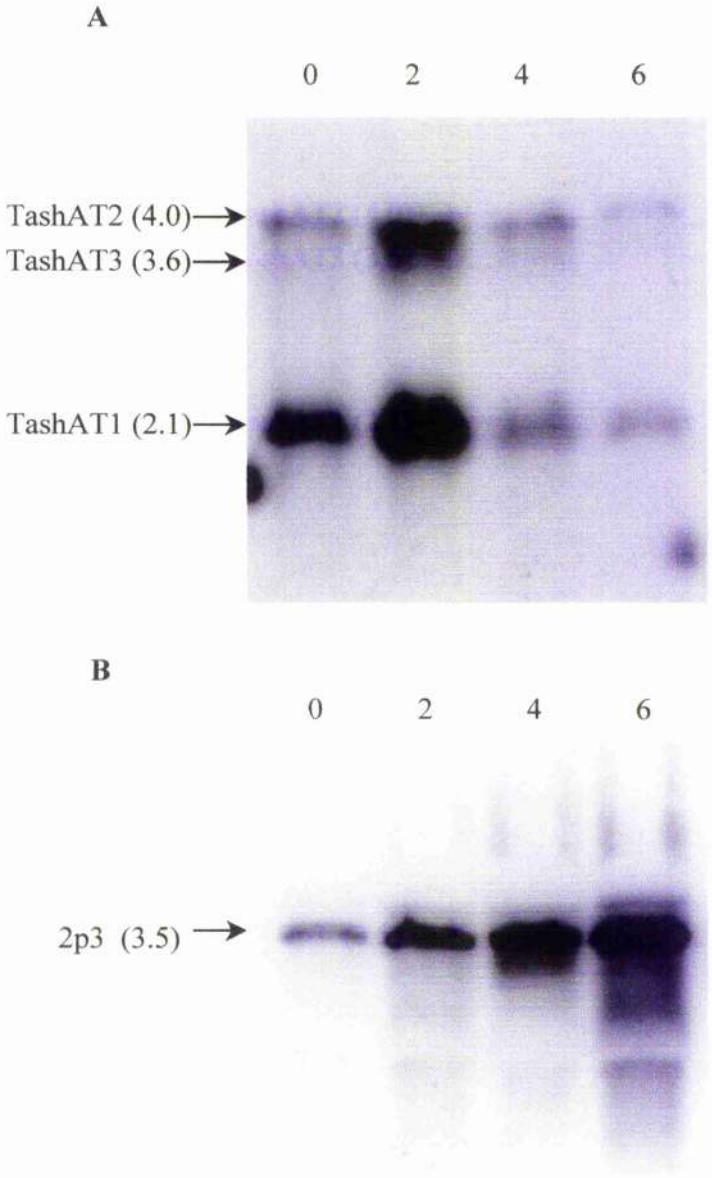


Fig. 4.5: Northern blot analysis of RNA from D7 cells, during a differentiation time course, hybridised to probe AThook1. 10μg of RNA was used in each lane. Panel A: probe AThook1 and panel B: probe 2p3. The time points for each RNA sample are indicated above each lane (in days). Arrows indicate the mRNA species detected with its corresponding size (in kb).

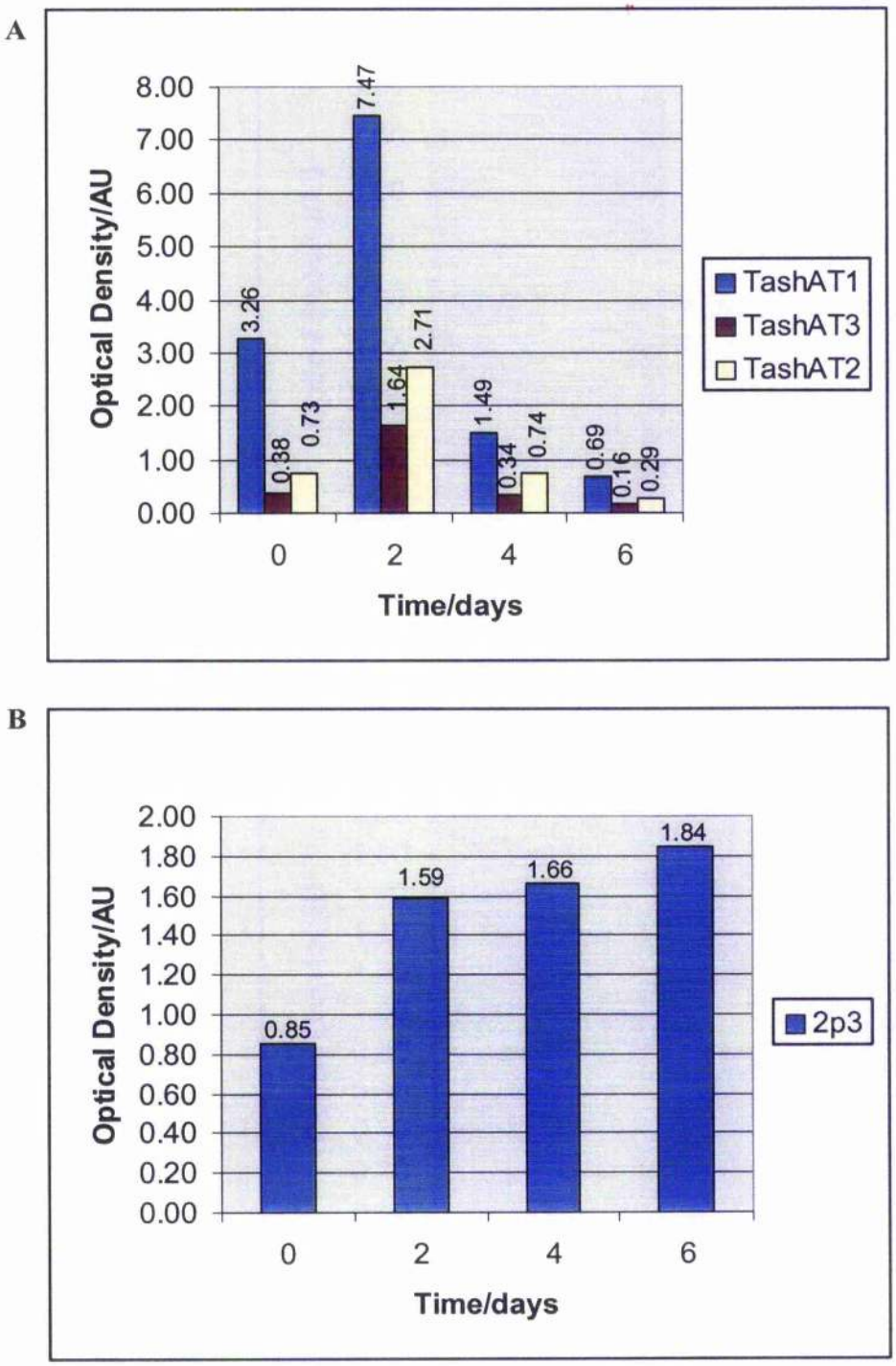


Fig. 4.6: Densitometric analysis of TashAT mRNA compared to 2p3 RNA levels from D7 cells, during a differentiation time course.
Panel A: TashAT mRNA and panel B: 2p3 RNA; AU: Arbitrary Units.

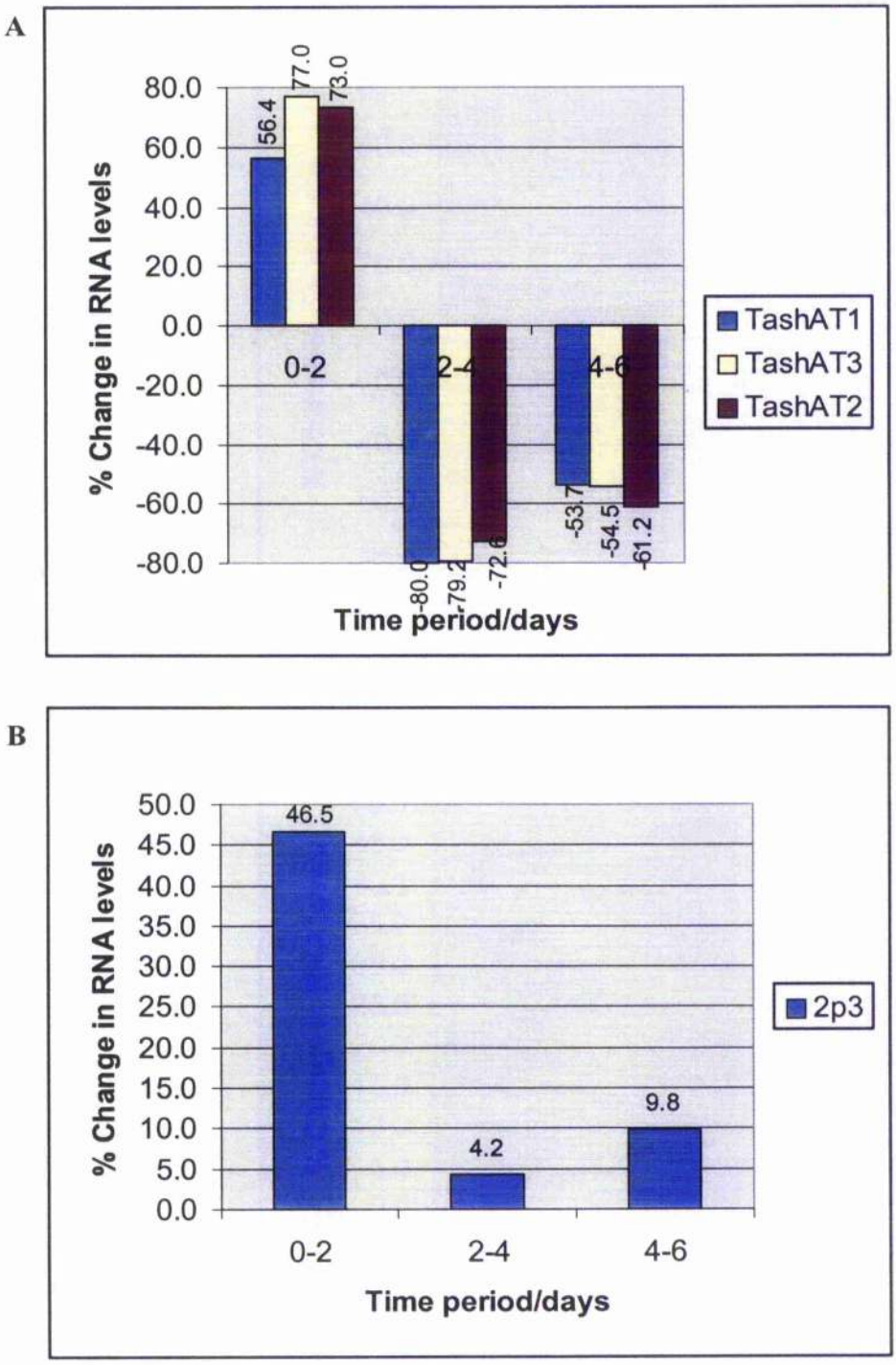


Fig. 4.7: Percentage change in TashAT mRNA and 2p3 RNA levels from D7 cells during a differentiation time course.
Panels A: TashAT mRNA and B: 2p3 RNA.

4.2.3 Comparison of the expression profile of TashAT transcripts between cell lines that are competent (D7) or attenuated (D7B12) for differentiation.

To determine if the TashAT mRNA species were differentially expressed in infected cell lines with altered capacities to differentiate, TashAT mRNA profiles were compared using Northern blot analysis. This was achieved using cell lines that are enhanced (D7) and attenuated (D7B12) for differentiation which were then hybridised with probe AThook1, and 2p3 as a control for parasite RNA load. The results (Fig. 4.8) showed that D7B12 and D7 cells expressed the same overall TashAT mRNA profile, with the exception of TashAT2 and TashAT3 mRNA levels, which differed between the two cell lines.

Densitometric analysis of this autoradiograph (see Fig. 4.9) confirmed that TashAT1 was the most abundant of the TashAT transcripts in both D7 and D7B12 cell lines. In D7B12 cells, the abundance of the transcripts relative to the TashAT1 message was 21.6% for TashAT3 mRNA, and 15.1% for TashAT2 mRNA. However, the situation was reversed with respect toTashAT2 and TashAT3 within the D7 cell line. Here, TashAT2 was the second most abundant message after TashAT1, with 27.8% of the signal intensity of the TashAT1 transcript. TashAT3 was the least abundant message in D7 cells, being only 14.03% as intense as the TashAT1 mRNA spp.. These figures were derived by calculating the proportion of TashAT2 or TashAT3 as a percentage of the total amount of TashAT1 within each cell line.

A comparison of the level of increase of each TashAT mRNA spp. in D7 cells with D7B12 cells (see Fig. 4.9), revealed that TashAT1 mRNA levels may be equivalent in D7 cells relative to D7B12 cells. This was because the intensity of the TashAT1 message was 2.8 times greater in D7B12 cells compared to D7 cells. However, this figure was also approximately equivalent to the overall increase of *T. annulata* 2p3 RNA in D7B12 cells compared to D7 cells. Although the densitometric readings showed elevated levels of TashAT2 (by a factor of 1.6) in D7B12 cells compared to D7 cells, there was 2.5 times more 2p3 RNA in D7B12 cells compared to D7 cells. Therefore, there may not be any real increase in TashAT2 mRNA levels in D7 cells compared to D7B12 cells. By contrast TashAT3 mRNA levels were 4.3 times more abundant in D7B12 cells compared to D7 cells. This figure was greater than the overall increase in RNA load in D7B12 cells compared to D7 cells, suggesting there may be a real increase in TashAT3 mRNA levels in

D7B12 cells compared to D7 cells.

In summary, TashAT2 and TashAT3 mRNA are differentially expressed in D7 and D7B12 cells, whilst TashAT1 mRNA levels show little change in both cell lines. TashAT1 mRNA species appears to be present at higher levels relative to TashAT2 and TashAT3 mRNA.

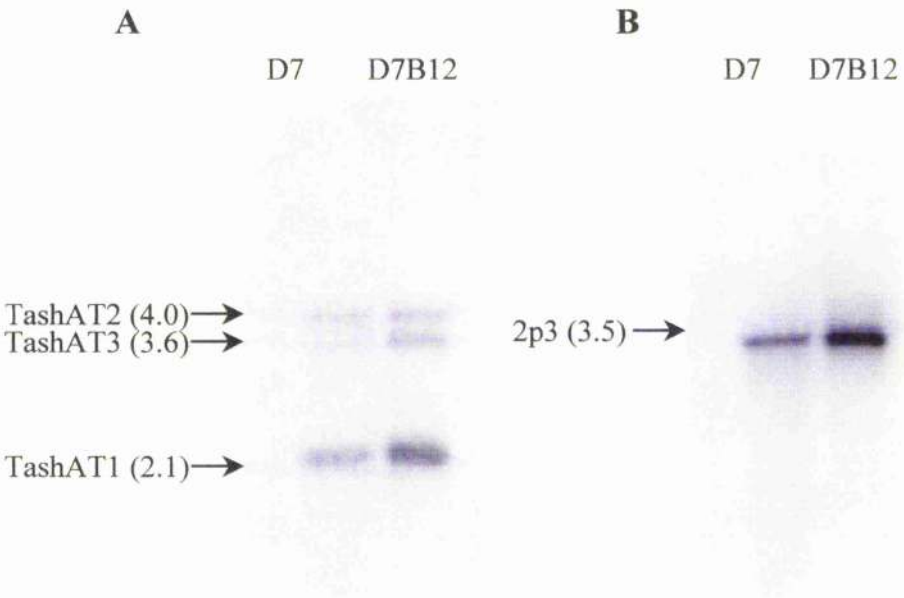


Fig. 4.8: Northern blot analysis of RNA from D7 and D7B12 cell lines at 37°C hybridised to probe AThook1. 10μg of RNA was used in each lane. Panels A: probe AThook1 and B: probe 2p3. The sizes (in kb) of the detected RNA species are indicated by arrows.

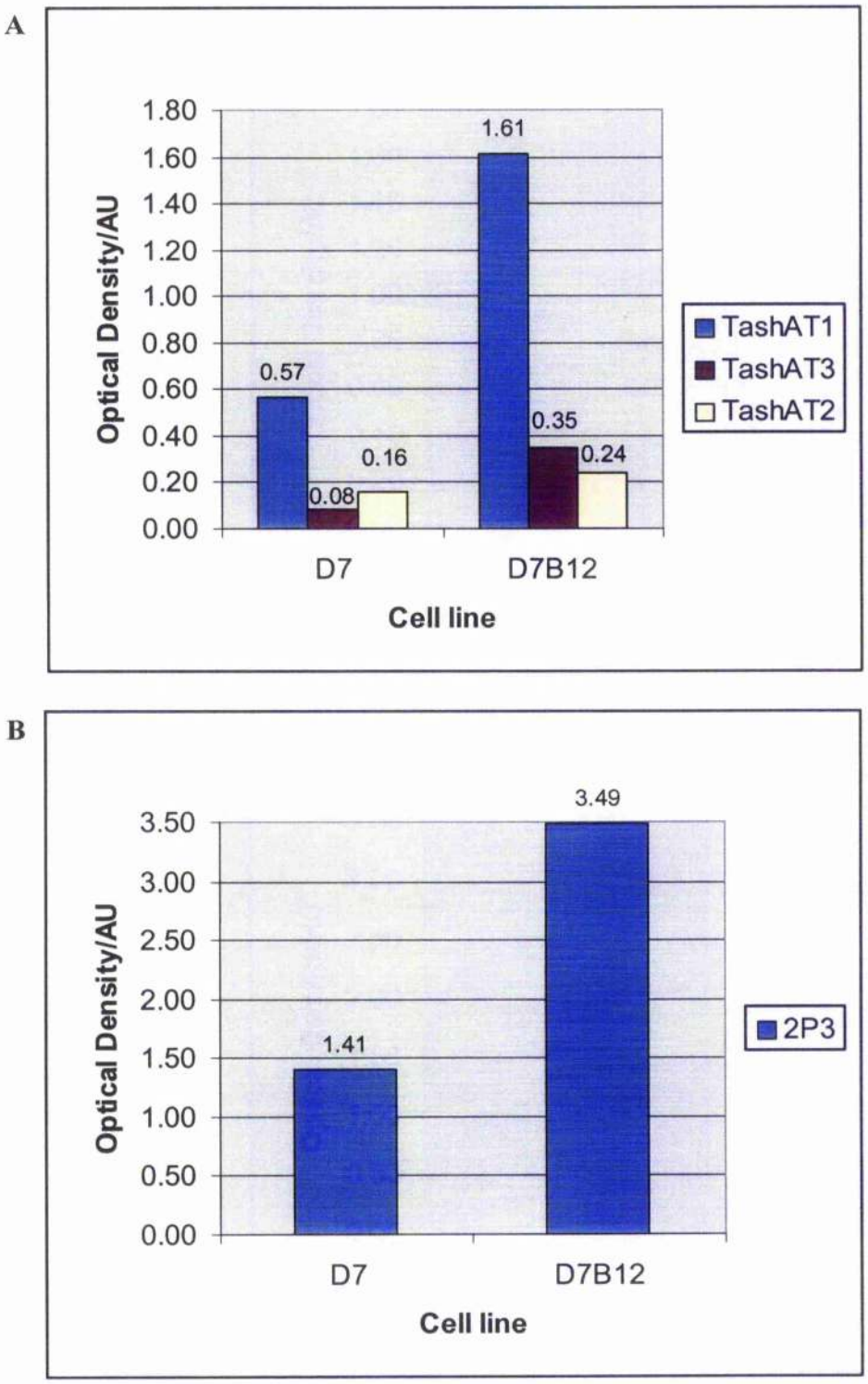


Fig. 4. 9: A comparison of TashAT mRNA and 2p3 RNA levels between D7 and D7B12 cell lines by densitometric analysis.
Panels A: TashAT mRNA and B: 2p3 RNA; AU: Arbitrary Units.

4.2.4 The expression profile of the TashAT mRNA transcripts in D7B12 cells.

To determine if TashAT mRNA expression is associated with parasite differentiation RNA samples were extracted from D7B12 cells following culture at 41°C to induce differentiation. Northern blot analysis was performed with the AThook1 probe (Fig. 4.10A) and compared with the equivalent levels of the control RNA species, 2p3 (Fig. 4.10B). These results showed that the abundance of all three TashAT RNA spp. were elevated at day 2 and day 6 compared to day 0 and day 4. Similarly the levels of 2p3 were also elevated at day 2 and day 6 compared to day 0 and day 4.

TashAT mRNA and 2p3 RNA expression profiles were quantified by densitometric analysis of the autoradiograph (see Fig. 4.11) and the corresponding percentage changes in their expression over time was calculated (see Fig. 4.12). These results confirmed the data presented in Fig. 4.10. From day 0 to day 2 all the three mRNA transcripts increased by an average of 78%. From day 2 to day 4 all TashAT mRNA levels decreased by an average of 61% but then increased again by an average of 39% from day 4 to day 6. This profile was reflected approximately by the rate of change of 2p3 RNA levels, which increased from day 0 to day 2 by 72%, decreased by 45% from day 2 to day 4 and finally increased from day 4 to day 6 by 32%. The only exception to this was the rate of decrease in TashAT mRNA abundance, which exceeded that of 2p3 by 16% from day 2 to day 4. Thus, the variations in TashAT mRNA profile do not appear to be decreasing during differentiation but follow the profile of parasite RNA, overall. Densitometric readings also show that the levels of all three transcripts remain constant relative to each other from day 0 to day 6. However, TashAT3 only increased by 25% compared to TashAT1 and TashAT2 mRNA from day 4 to day 6.

In summary, there did not appear to be any evidence of a decrease in TashAT mRNA levels when Northern blot analysis was performed on D7B12 RNA from a differentiation time course, in contrast to TashAT mRNA levels during a D7 differentiation time course. Little change was detected in the relative abundance of any TashAT mRNA relative to each other from D7B12 cells during a differentiation time course.

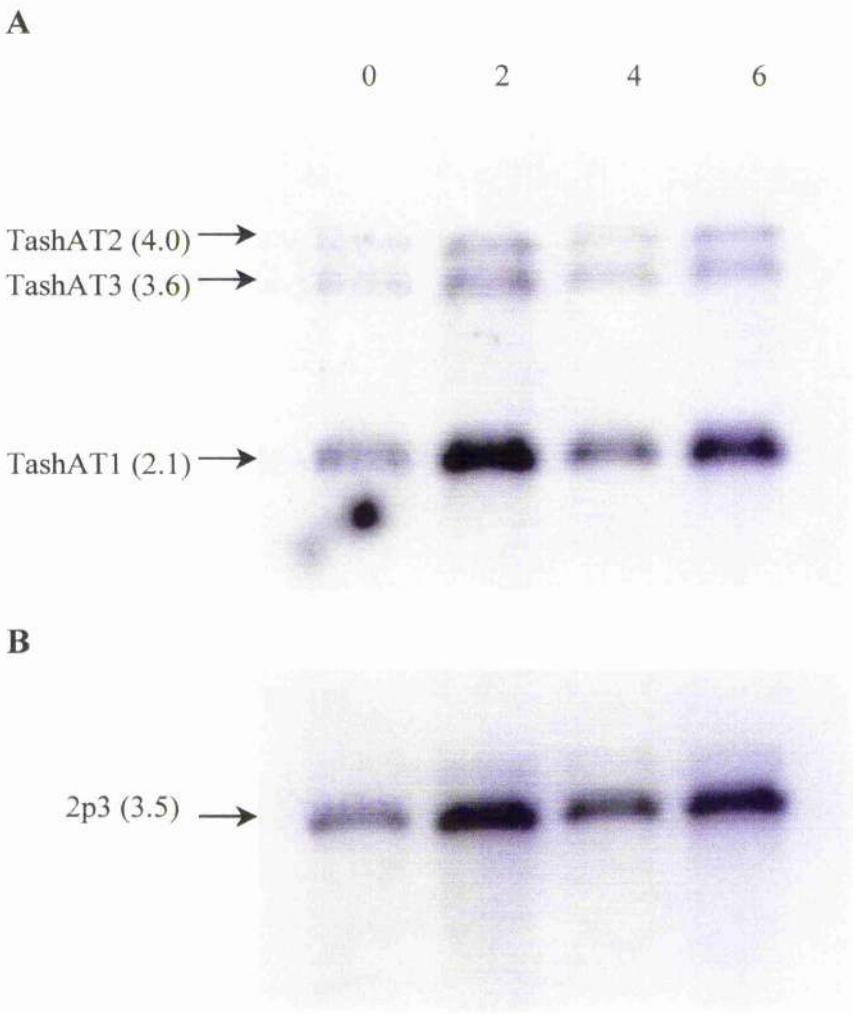


Fig. 4.10: Northern blot analysis of D7B12 RNA from a differentiation time course, hybridised to probe AThook1. 10μg of RNA was used in each lane. Panels A: probe AThook1 and B: probe 2p3. Each lane represents a time point (days). Arrows indicate the RNA species and it's corresponding size (in kb).

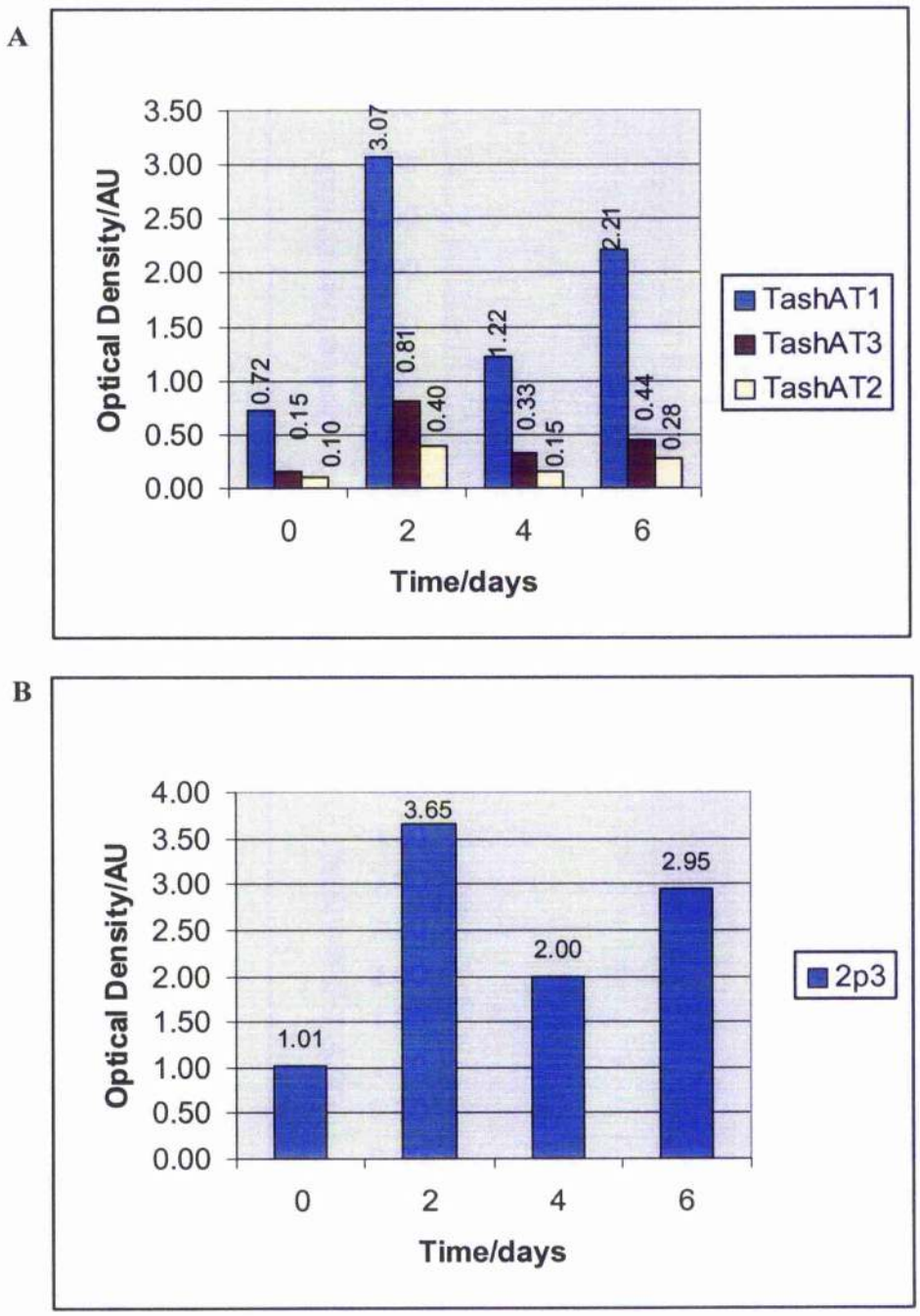


Fig. 4.11: Densitometric analysis of TashAT mRNA and 2p3 RNA levels from D7B12 cells over a differentiation time course. Panels A: TashAT mRNA and B: 2p3 RNA; AU: Arbitrary Units.

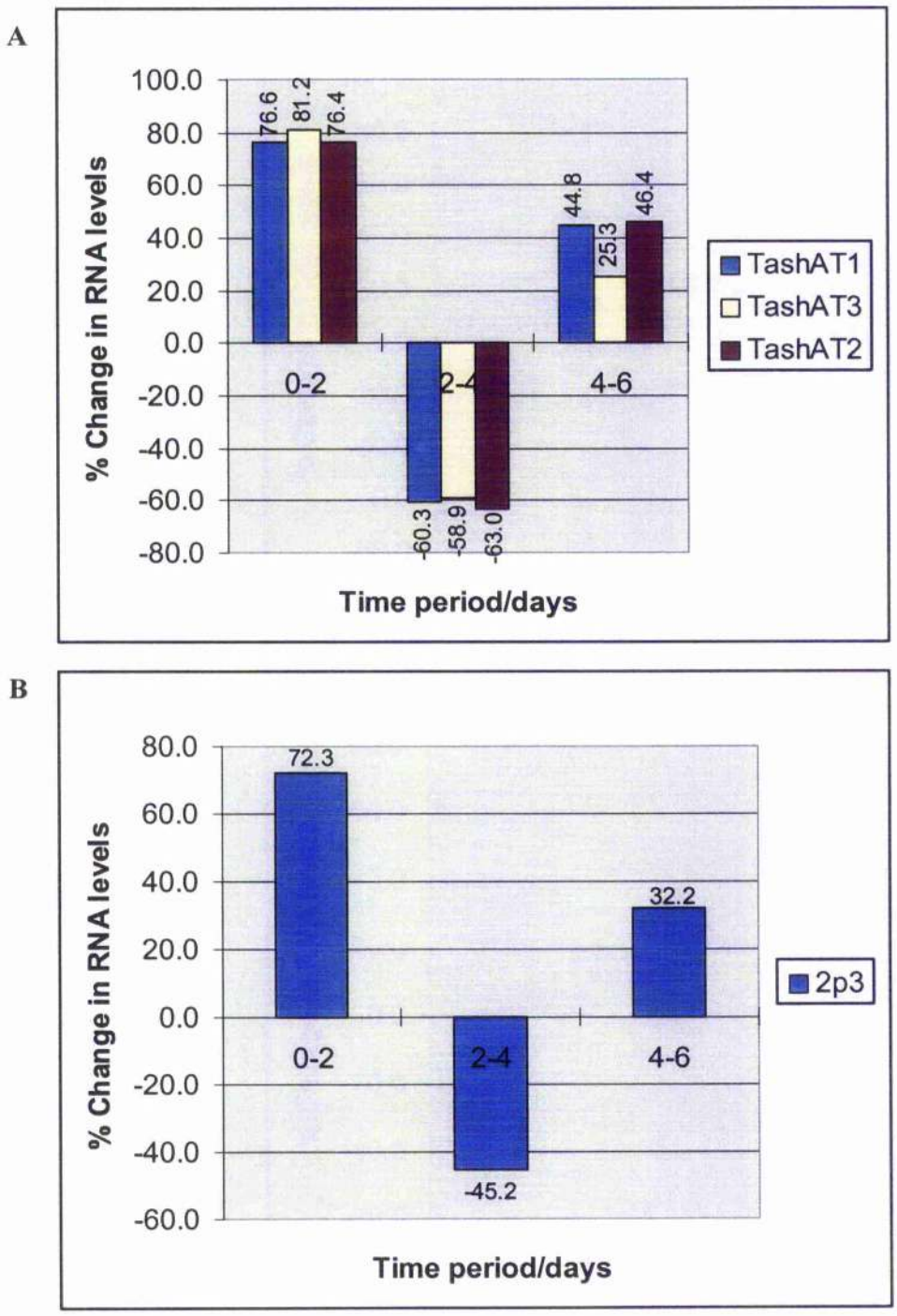


Fig. 4.12: Percentage change in TashAT mRNA and 2p3 RNA levels from D7B12 cells over a differentiation time course. Panels A: TashAT mRNA and B: 2p3 RNA.

4.2.5 Identification of the Transcription Start site of TashAT1

To map the Transcription start site (TSS) of *TashAT1*, which could be predicted to be within an Inr element, the technique of five prime RACE was employed as outlined in Fig. 4.13. Three nested primers rsp1, 2 and 3 (see Appendix A) positioned at 341bp, 270bp and 219bp downstream of the translation start site of *TashAT1*, respectively were designed to produce a final product of approximately 250bp after PCR amplification of the upstream region of TashAT1 cDNA.

The results of the first 5' RACE amplification revealed a band approximately 250bp after two rounds of PCR amplification (Fig. 4.14) upon agarose gel electrophoresis. Chloramphenicol acetyl transferase (CAT) control RNA provided by the 5'RACE kit was used to check the reverse transcription reaction (see lane 4) and subsequent PCR amplification steps (see lane 3) were successful, using control primers GSP1,GSP2 and GSP3 provided by the kit. The control RT-PCR reaction in lane 4 produced an RT-PCR product at approx. 500bp (adjusting for uneven DNA electrophoresis), which was the correct size according to the 5'RACE kit manual. The control reaction in lane 3 generated four bands at approximately 500bp, 700bp, 800bp and 900bp. The 500bp band correctly related to the 500bp control RT-PCR product described in the manual. The 700bp band may relate to the 711bp PCR intermediate PCR product with GSP2. The 900bp product could have been a result of mis-priming from the original 891bp control RNA template or poor quality control RNA. To control for genomic DNA contamination, water was used instead of RNA (see lane 5), and the absence of any products indicated that there was no DNA contamination present.

The 250bp PCR generated band was excised, purified and cloned into the vector pGem T-easy (Promega). Twenty-four of the cloned PCR products were sequenced to find the TSS of *TashAT1*. The results (shown in experiment 1 of Fig. 4.15) revealed that 8 clones (33%) contained sequences locating the TSS to 30bp (-30) upstream of the A nucleotide of the translation start site codon for *TashAT1*: this position appeared to be the most frequent of all the TSSs mapped. The TSS loci of 7 sequences were mapped near to the putative –30 TSS locus: at –26, –27, -31, -33, –35 relative to the translation start site. This experiment appeared to have a large spread of TSSs, with eight sequences showing a TSS position of under 26bp or over 36bp relative to the translation start site, although each of these mapped TSS positions only occurred once.

To confirm the results of the first 5' RACE experiment, a second 5'RACE experiments was performed. The results of the second PCR amplification of this experiments also revealed a diffuse band approximately 200 or 300bp (data not shown), which was cloned into the vector pGem T-easy. After restriction digestion, twenty clones containing inserts of 200 to 300bp were sequenced and the putative TSS determined. The results, shown in Fig. 4.15 (experiment 2) revealed that almost half of the sequenced clones (9 clones) showed that the TSS of *TashAT1* was located 30bp upstream of the A nucleotide of the translation start site codon. Three sequences showed putative TSS loci at -28bp and -29bp relative to the translation start site of *TashAT1*. Another three sequenced clones revealed TSS loci of 35 and 36bp relative to the translation start site. A further five more sequences terminated at 17, 19, 21, 26 and 76bp upstream of the translation start site of *TashAT1*. In total fourteen of the twenty viable clones, had TSS loci between 28 and 36bp relative to the translation start site. However, it appeared that the most frequently found start site locus for the TSS was, again, 30bp upstream of the translation start site of *TashAT1*.

In summary, there were a total of 15 out of 46 TSS loci mapped between 26 and 35bp (excluding the –30 position), upstream of the translation start site in both experiments, although the frequency of each TSS loci was low (3 or less). Whilst it appears that the major TSS for *TashAT1* could be predicted to be 30bp upstream of the translation start site, the variability in TSS loci mapped in both 5' RACE experiments, cannot exclude other TSS positions near to the –30bp locus.

To determine if the sequence surrounding the TSS of *TashAT1* conformed to a typical Inr motif, the seven nucleotides surrounding the TSS of *TashAT1* were compared to the consensus Inr sequence. Javahery *et al.* (1994) determined a loose consensus for the Inr element to be a pyrimidine rich sequence of $YYA_{+1}NT/AYY$, where the +1 position denotes the TSS. Using this system the putative Inr sequence of *TashAT1* was deduced to be 5'-GTG_{+1}TTAT-3'.

To identify if *TashAT1* contained an additional TATA box, the upstream region of *TashAT1* was analysed using the software program MatInspector (see Fig. 3.19). This program identified a number of TATA-like boxes, but none conformed to the strict consensus TATA motif. The nearest TATA-like box was also located outside of the normal range of a TATA box (in most cases 30bp or less from the translation start site).

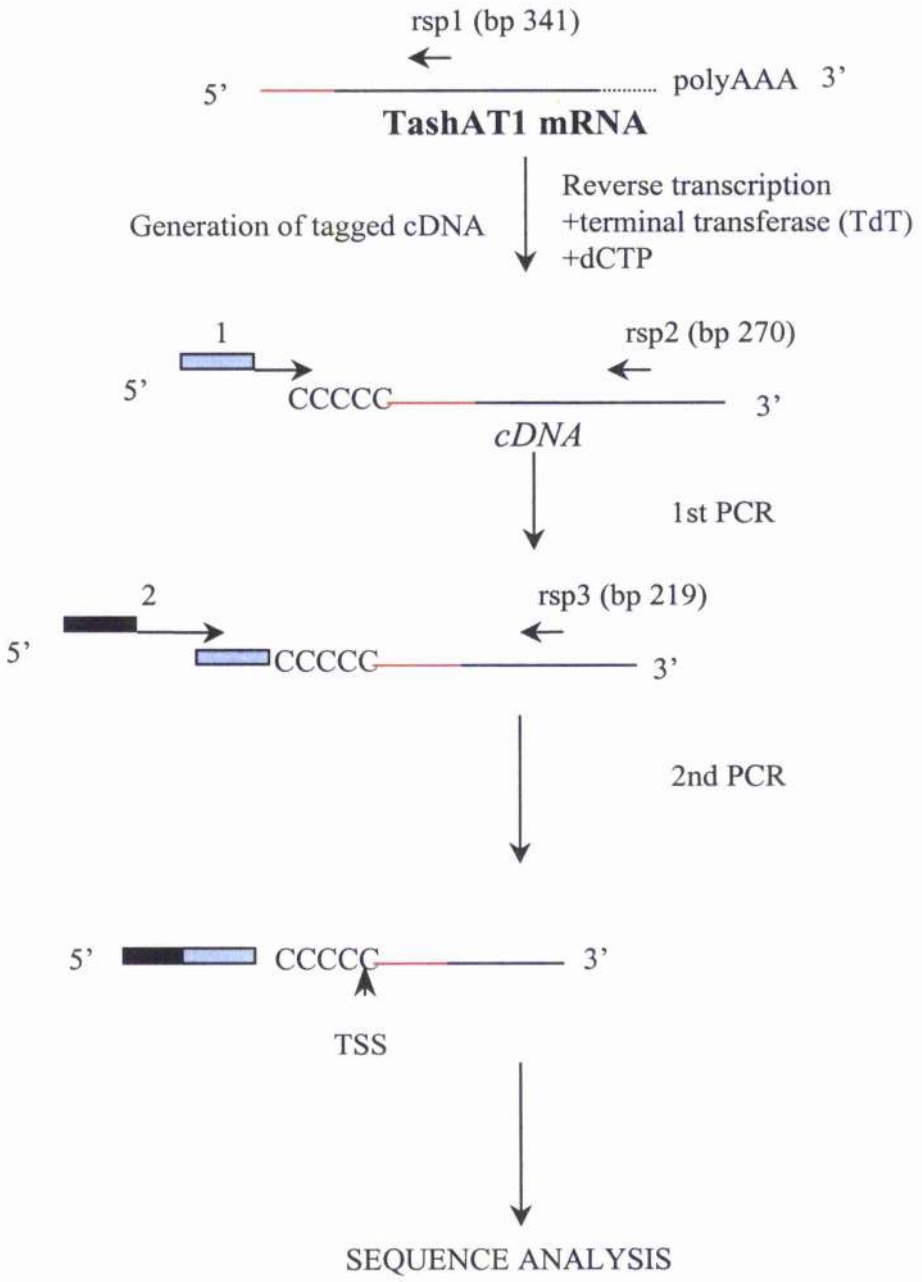


Fig. 4.13: **Diagram of five prime RACE analysis performed for TashAT1 RNA**. rsp1,2,3 were primers designed against the 5' region of TashAT1. Primers 1 (purple box) and 2 (black box) represent general primers (supplied with 5' RACE kit). The red line indicates the untranslated portion of the transcript. The position of transcription start site (TSS) is indicated by an arrowhead. The dotted line represents the Remaining TashAT1 message three prime to the rsp1 primer.

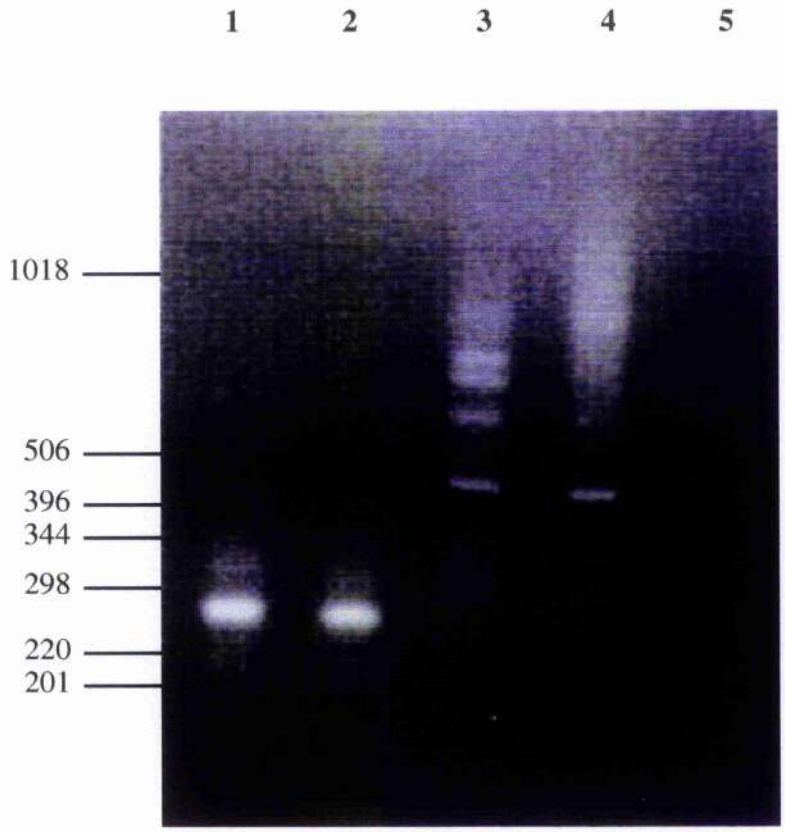


Fig. 4.14: PCR products from Five Prime RACE analysis of TashAT1. Lanes 1 and 2 show the resultant PCR products after the second round of amplification from D7 mRNA. Lane 4 (control) contained CAT cDNA product after reverse transcriptase amplification reaction using GSP1 primer, lane 3 contained CAT cDNA products after the second PCR amplification reaction using primers GSP2 followed by GSP3 , lane 5 contained no DNA. All control reagents were provided by the kit. The DNA markers sizes (bp) are indicated on the right.

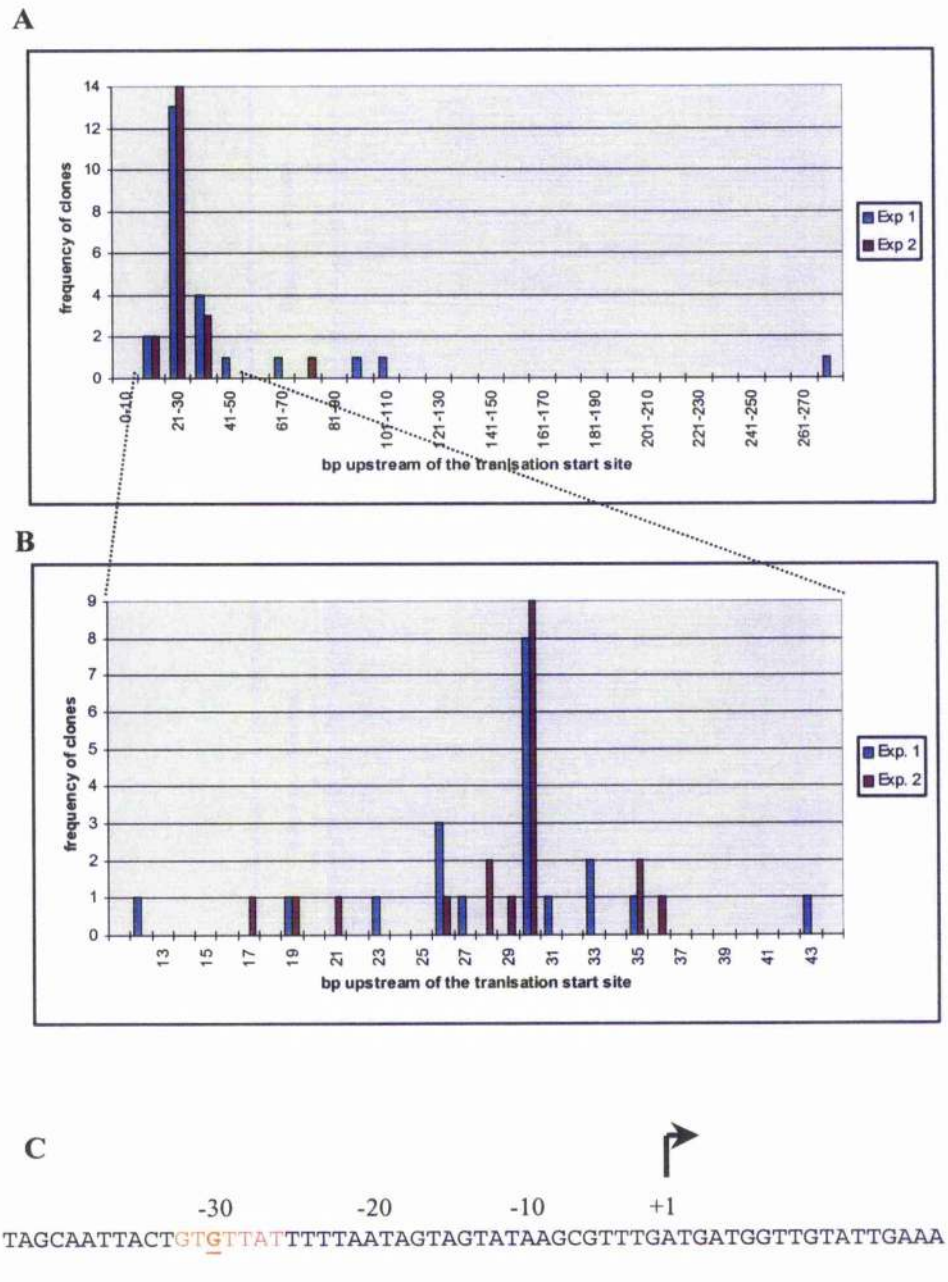


Fig. 4.15: Frequency of transcription start sites (TSS) mapped to the upstream region of TashAT1. Panel A shows the entire range of TSS found; panel B shows the most frequently occurring TSS mapped. The site of the most frequently found TSS is shown (underlined) upstream of the TashAT1 gene (panel C). The putative translation start site (at +1) of TashAT1 is indicated by an arrow. The sequence highlighted in red shows the consensus sequence of a possible Inr element.

4.2.6 Investigation of possible nuclear factors binding to a DNA motif upstream of TashAT1 and TashAT3 by Electrophoretic Mobility Shift Assay

To determine if the TashUM motif, upstream of *TashAT1* and *TashAT3* was capable of binding to parasite nuclear factors from infected cells, TashUM was end-labelled, incubated with D7 and D7B12 cell nuclear extracts and the resulting interactions with nuclear factors examined by EMSA (Fig. 4.16). Control reactions were also carried out with non-infected BL20 nuclear extracts to investigate whether host factors were responsible for any band shifts with macroschizont infected cell nuclear extracts.

The results of this analysis, in Fig. 4.16A, revealed a large band shift (band shift 1) that showed a small retardation with the parasite enriched nuclear fractions derived from both D7 and D7B12 cells. This band was clearly retarded relative to the migration of the TashUM probe with no nuclear extract, and indicated the presence of a factor within the parasite-enriched fraction, which bound to TashUM. No conclusive results could be determined for the EMSA containing BL20 and host nuclear extracts (Fig. 4.16A) because the concentration of these extracts were too high in both D7 and D7B12 nuclear extracts. However, the results from Fig. 4.16B, which had reduced levels of D7 host-enriched nuclear and BL20 extracts, showed that a band shift was also present in host enriched nuclear extracts. This was retarded to the same extent as the band shift observed in the parasite-enriched nuclear extracts (Fig. 4.16A, band shift 1). In addition, a band of lower intensity was obtained in reactions containing nuclear extracts from uninfected cells (see Fig. 4.16B, band shift 2), although it should be noted that a band in a similar position was present in the probe alone track. Furthermore, both BL20 and probe alone bands migrated with a slower mobility than the major shift obtained with the extracts derived from infected cells. Therefore it was unclear whether the band shift observed in the BL20 track was the same as the band shift detected with host-enriched nuclear extract.

To assess if this band shift was due to a factor specifically binding to the TashUM oligonucleotide, a "cold competition" band shift assay was included, where a 100 fold excess of non-radioactively labelled TashUM probe was added to the reaction before the radioactive TashUM probe was added. The excess cold probe should bind to all remaining factors, leaving the radioactively labelled probe unbound, resulting in an absence of the band shift upon autoradiography. When the cold TashUM oligonucleotide was added to the

parasite extract (see Fig. 4.16: H1, P1 lanes) the major band shift was not observed. This preliminary result might indicate that the band shift was caused by a specific interaction between nuclear factors and the TashUM probe.

Two more major band shifts were also detected in both host (band shift 4) and parasite (band shift 3) fractions in Fig. 4.16. The parasite band shift disappeared with the addition of excess cold probe, indicating that there might be specific binding of the factor to the probe, but it was unclear whether this shift was also detected in BL20 nuclear extract, because of the excess BL20 nuclear extract present in the BL20 track. In Fig. 4.16B, the host band shift also appeared to be present in BL20 nuclear extract, but at a greater intensity.

A separate experiment carried out in collaboration with S. McKellar, using two non-specific DNA competitors, poly dI.dC and poly dG.dC, showed that the major band shift (band shift 1) only formed in the presence of poly dG.dC (Fig. 4.17). This indicated that the factor preferentially bound AT rich DNA. In addition, when D7 host and parasite-enriched nuclear extracts from a differentiating time course were analysed by EMSA (Fig. 4.18, band shift 1), a marked decrease was observed in the signal intensity obtained with the day 6 extracts relative to the day 0 extracts.

In summary, a major band shift, has been identified with both host and parasite-enriched nuclear extracts that is out-competed with the addition of excess cold probe. From the preliminary analysis that was performed, the origin of this factor, host or parasite, could not be identified. However the data demonstrates that the factor(s) responsible for the band shift (1) has an affinity for AT rich DNA. In addition the level of factor(s) responsible for the band shift had decreased during the differentiation process. Other band shifts have also been identified in both host and parasite nuclear fractions, which require further analysis.

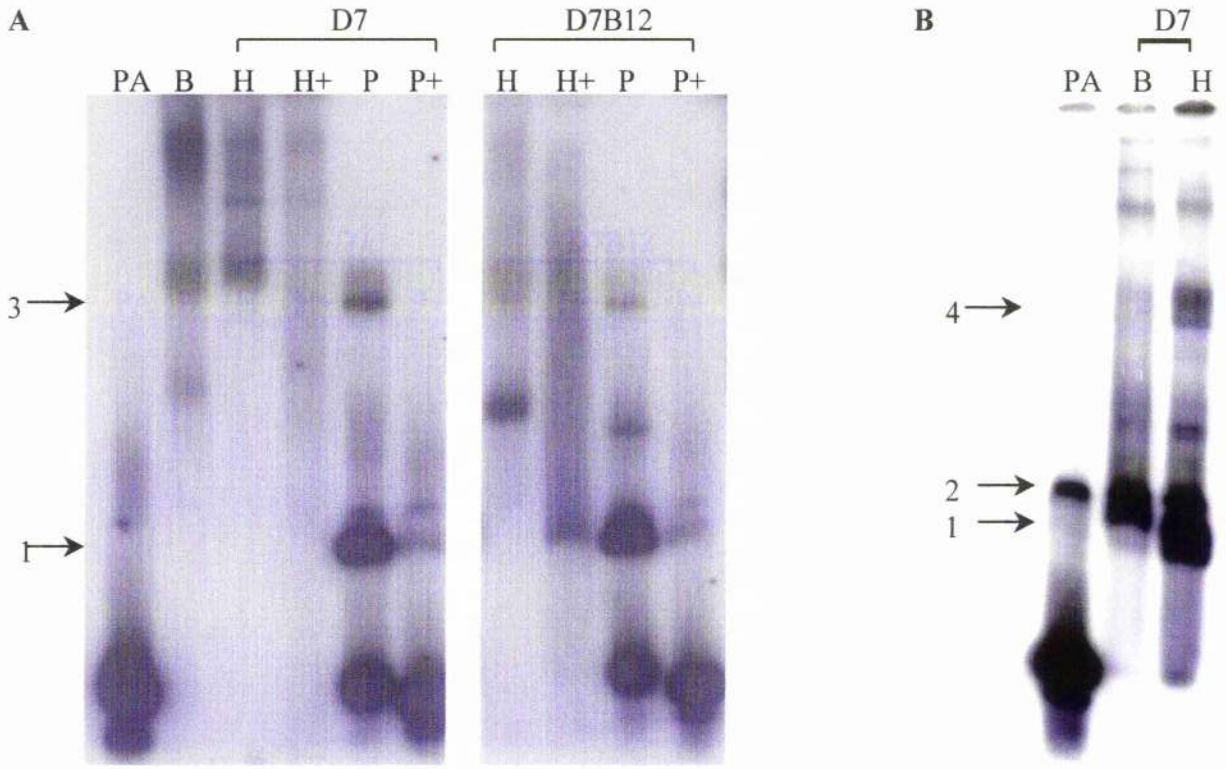


Fig. 4.16: **EMSA analysis of D7, D7B12 and uninfected nuclear extracts at 37°C**. Panels A and B represent two separate EMSA experiments. PA represents probe alone, B represents BL20 control nuclear extract. H and P represent host and parasite nuclear extract respectively. The + sign indicates that excess non-radioactively labelled probe was added to the reaction. Numbered arrows represent the specific band shifts detected.

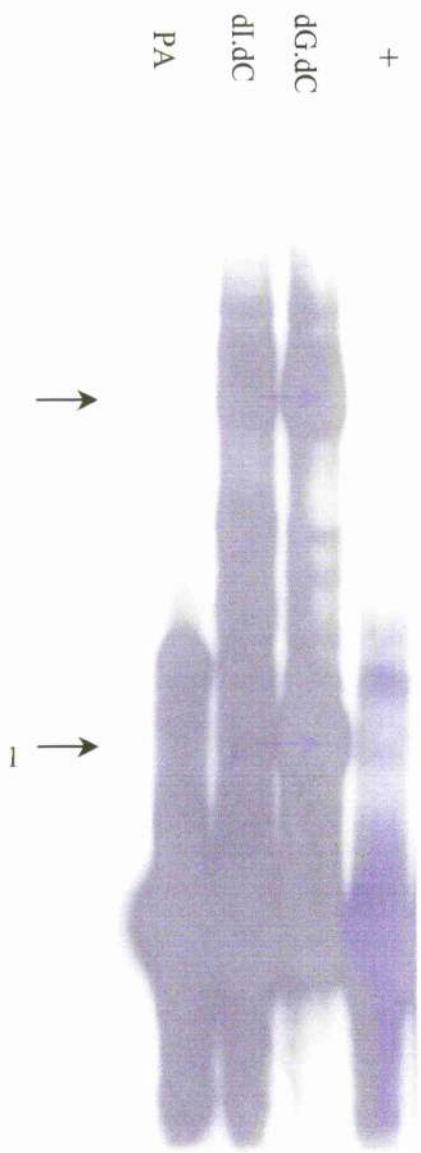


Fig. 4.17: EMSA analysis of D7 parasite nuclear extract at 37°C with non-specific competitors poly (dG.dC) and (dI.dC).
Excess cold probe: +; PA: probe alone (with no nuclear extract).
An arrow denotes band shift 1 (Courtesy of S. McKellor).

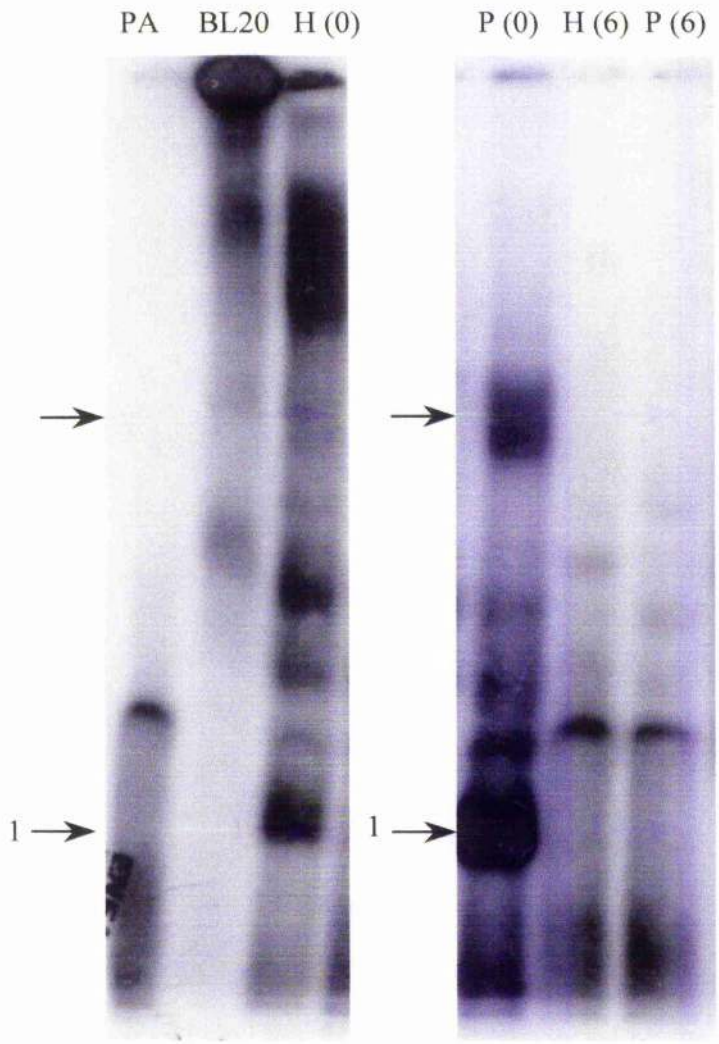


Fig. 4.18: EMSA of D7 host and parasite enriched nuclear extracts during a differentiation time course. Host and parasite enriched nuclear extracts are denoted by H and P, respectively. Time points (days) are indicated in brackets above each lane. PA: probe alone (with no nuclear extracts). The arrow indicates the major band shifts detected.

4.3 Discussion

The first aim of this chapter was to identify the mRNA species encoded by *TashAT1*. This was demonstrated by Northern blot analysis using two probes derived from *TashAT1* and *TashAT3*, which showed that the 2.1kb RNA message was likely to be TashAT1, and the 3.6 kb TashAT3, as they were approximately close to the predicted open reading frames of the corresponding genes. The results of Northern blot analysis with the AThook1 probe in Fig. 4.1A confirmed the sizes of the TashAT1/3 messages. The 4.0 kb mRNA detected with probe AThook1 was previously concluded to encode TashAT2 mRNA (Swan *et al.*, 1999). The apparent size difference between the TashAT messages and their corresponding genes might be explained by the presence of a 5' or 3' untranslated region (UTR) in these transcripts. These are common feature of apicomplexan mRNA species (Fan and Davidson, 1996; Odberg-Ferragut *et al.*, 1996), where they have been shown to play a role in protein translation and/or message stability (Furger *et al.*, 1997; Hotz *et al.*, 1997).

Studies by Swan *et al.* (1999) revealed that *TashAT2* gene expression was down regulated early on during differentiation towards the merozoite, and as such could be a possible candidate for a gene involved in determining specific gene expression of a macroschizont infected cell. To assess if TashAT1 mRNA was expressed in a similar manner, Northern blot analysis was performed from a differentiating time course. The results revealed that TashAT1, as well as TashAT2 and TashAT3 mRNA were down regulated from day 2 to day 6, in a similar manner to each other as host cell division subsides (Shiels *et al.*, 1992). Work by Shiels *et al.* (1994) demonstrated that the down regulation of macroschizont polypeptides is temporally linked to an increase in merozoite gene expression. Thus if the down regulation of TashAT mRNA is reflected at the polypeptide level, it is possible that the *TashAT* genes may be involved in the negative regulation of parasite differentiation or be associated with the control of host cell division. This is supported by Northern blot analysis of TashAT transcripts in the D7B12 cell line, severely attenuated for differentiation, which did not detect any notable changes in the overall abundance of TashAT mRNA from day 0 to day 6, although it was noted that TashAT3 mRNA levels had not decreased as much as TashAT1 or TashAT2 mRNA from day 4 to 6. Further work is required to determine if this difference is statistically significant. Therefore, the results so far might suggest that the co-ordinated down regulation of the TashAT transcripts is regulated by common regulatory factor(s).

Whilst TashAT1 mRNA is expressed at high levels in both D7 and D7B12 cells, there is some evidence to suggest that *TashAT1* may not be essential to the control of macroschizont gene expression or host cell proliferation. Swan *et al.* (2001a) found reduced or a complete absence of TashAT1 mRNA expression in different *T. annulata* cell lines, with some cell lines lacking the *TashAT1* gene in their genome (Stadler, unpublished 2000). One possible explanation for the variability in TashAT1 mRNA expression may be an alteration of the control of parasite gene expression as the infected cells adapt to culture. Many workers (Adamson *et al.*, 2000ab; Sutherland *et al.*, 1996; Oura *et al.*, 2001) have described such variability in host cell gene expression. This alteration could affect *TashAT* gene expression which could then, theoretically, go on to alter host gene expression. Further experiments would be required on the ability of these cell lines to proliferate or differentiate to elucidate which *TashAT* genes are essential in maintaining the proliferative state of the infected host cell or involved in blocking differentiation of D7B12 cells.

The increase in TashAT mRNA levels between day 0 and day 2 could not be determined as a genuine up-regulation since 2p3 RNA levels had also increased during this time period. If this is the case, one explanation for this phenomenon may be a general increase in macroschizont gene expression upon differentiation. Certainly, there is a general increase in parasite size and infected cell growth in some *Theileria* infected cell lines during this period (Shiels *et al.*, 1992). Thus, 2p3 levels might also be expected to increase, although this would need to be verified since there is one example of a ribosomal protein in *Theileria* (Oura *et al.*, 2002) and other examples of *Plasmodium* rRNA genes which are differentially expressed (reviewed by Waters *et al.*, 1989). If the *TashAT* genes are up-regulated, this response is unlikely to be directly linked to a classical heat shock response as HSP70 mRNA levels were found to be similar in both D7 and D7B12 cells, the latter of which are attenuated for differentiation (Swan *et al.*, 2001a and references therein). Therefore, to ascertain whether up-regulation has truly occurred, the increase in TashAT mRNA levels relative to their DNA levels could be compared over this time period.

Densitometric analysis of the northern blots of D7 and D7B12 revealed that the expression profiles of TashAT2 mRNA and TashAT3 message were also altered in D7B12 cells compared to D7 cells. TashAT3 was more abundant than TashAT2 mRNA in D7B12 cells, but less abundant in D7 cells, whilst TashAT1 mRNA levels apparently remained constant between the two cell lines. This observation might suggest that differential expression of TashAT mRNA is caused, by differential binding of regulatory factors to the upstream regions of these genes; either to the TashUM motif and/or to other important control

regions of each *TashAT* gene. It is possible that the apparent switch between TashAT2 and TashAT3 mRNA could possibly mean that these genes could act as functional substitutes for each other in the two cell lines, and therefore may be controlled in part by shared regulatory factors. Further dominant negative analysis with a *TashAT* gene mutant in which the AT hook and transactivation domain were removed would help define the role of the *TashAT* genes *in vivo*. It would also be interesting to determine if the quantitative changes in TashAT2/3 gene expression of D7 and D7B12 cells were related to the altered pattern of gene expression recently defined for the cell lines described by Oura *et al.* (2001).

So far, all studies on *Theileria*, and other apicomplexan genes have shown that all gene transcription occurs monocistronically (Horrocks *et al.*, 1998). This data and the presence of the TashUM motif upstream of *TashAT1* suggest that TashAT1 may also be transcribed in a monocistronic fashion. However, nuclear run on analysis as performed for Tams1 (Shiels *et al.*, 2000) is necessary to confirm this for TashAT1.

In order to define the structure of the upstream region of *TashAT1*, five prime RACE was performed on D7 mRNA. Two separate experiments showed that the most frequent transcription start site (approximately 33% and 45%) was 30bp upstream of the deduced translation start site. The TSS and adjacent sequence was compared to the consensus Inr motif as described by Javahery *et al.* (1994). The predicted Inr of *TashAT1* conforms in general to the eukaryotic Inr motif, with the exception of a G instead of an A nucleotide at position +1 (the position of the TSS itself). Deletion studies by Javahery *et al.* (1994) have shown that at A at +1 is an important nucleotide for Inr activity, if this is mutated the promoter strength is reduced to a greater extent. Although low Inr activity can be imparted even in the absence of an A at +1, this evidence suggests that the putative TSS found for *TashAT1* would be weak. This result was in contrast to the predicted Inr elements of *Tams1* (Shiels *et al*, 2000) which conformed totally to the predicted Inr sequence. However, inspection of the upstream sequence of *TashAT1* revealed a potential consensus Inr elements 29-23bp upstream of the translation start site (TTA(+1)TTTT), where A is 27bp upstream from the translation start site. This fact, and the number of TSSs mapped with 10bp of the TSS found suggests it is possible that TSS might be located 27bp upstream of the translation start site of *TashAT1*. If this is the case, then the frequency of the site at -30bp could be due to a strong block of the reverse transcriptase enzyme due to secondary structure formation at the 5' end of TashAT1 mRNA (Goodbourn, 1996). This could mask the true TSS and Inr, located further upstream, which could be a few nucleotides. This may even be likely given the longer products detected in the 5'RACE experiments, which must

either be from TashAT3 mRNA or another very closely related gene transcript, not identified clearly by southern or northern blotting carried out to date. Further experiments would be required to confirm the exact position of the potential Inr element of *TashAT1*. However, the transcription start site data generated for *TashAT2* and *Tash1*, located at -37 and -33bp from their translation start sites respectively (Swan, unpublished, 2001), suggest that the TSS of *TashAT1* is likely to be within the 30bp range from the translation start site. In common with other apicomplexan parasite genes (see Horrocks *et al.*, 1998), no TATA box was found within a 30bp range of the translation start site of *TashAT1*. Therefore, it appears unlikely that transcription initiation occurs solely or partly from a TATA box, but further functional analysis using *in vitro* transcription or transfection techniques would be required.

To obtain further data to show that the TashUM motif may function (possibly in association with an Inr element) to control gene expression, EMSA analysis was performed with host and parasite enriched nuclear extracts and revealed a major band shift in both extracts. Unfortunately, the presence of this band shift in reactions using both parasite and host enriched nuclear extracts prevented the designation of the binding factors to either the host or parasite. Moreover, given the close proximity of the lower intensity BL20 band shift, it is possible that the shift is generated by a host molecule that is up-regulated in parasite infected cells.

Despite these warning signals and the requirement of further experiments with appropriate controls, for example, including an unrelated probe, it is possible that the TashUM band shift detected with host-enriched nuclear extracts is due to a parasite factor. This factor may have a similar cellular localisation profile as the TashAT polypeptides and be transported into the host cell nucleus. Swan *et al.* (1999) presented evidence that an antisera raised against the AT hook region of TashAT2 was detected within host and parasite compartments and it is therefore conceivable that DNA binding factors could also be present in both parasite and host enriched nuclear extracts. A more remote possibility is that the pattern of band shifts is related to a host cell factor that is transported to the parasite compartment and such events have been reported by Carrington *et al.* (1995). In future, EMSAs could include reactions with host or parasite-enriched nuclear extract and oligonucleotides containing binding sites that correspond exclusively to host transcription factors such as NF-κB. These reactions would determine the relative level of cross-contamination in host or parasite-enriched nuclear extracts. Ideally, the development of pure host and parasite nuclear extracts would allow a clearer identification of host or

parasite associated band shifts, including the three band shifts identified in section 4.2.6. The microtubule inhibitor nocodazole, which has been used to produce high quality purified parasite extracts from infected cells (Baumgartner *et al.*, 1999) could be used to this end, although the effect of this drug on TashUM factor binding would have to be tested.

A further result of EMSA analysis was that the major TashUM shift had decreased when incubated with nuclear extracts derived from cultures undergoing merogony, and therefore may be down regulated. The result is interesting regardless of whether the origin of the factor responsible is derived from the host or parasite as it could relate to a factor that determines parasite gene expression at the macroschizont stage. Alternatively this factor could be a parasite controlled host factor involved in the regulation of leukocyte gene expression in actively proliferating macroschizont-infected cells.

Experiments using different non-specific DNA competitors, revealed that the low molecular weight band shift only appeared in the presence of poly dG.dC, which competes out factors which preferentially bind GC rich DNA, and was blocked by the addition of poly dI.dC. These results showed that the factor(s) binding to the TashUM sequence preferentially bind AT rich sequence. Recent work by Swan *et al.* (2001a) purified DNA that showed strong binding to GST-TashAT2 and found that this DNA was AT rich, as expected for an AT hook binding protein. Interestingly, the TAAAT motif, defined by Swan *et al.* (2001a) is also present in the TashUM oligonucleotide. Swan *et al.* (2001a) also showed that a TashAT1 fusion protein, containing the AT hook encoding motifs preferentially binds to AT rich DNA. Thus, the major band shift could represent a TashAT or closely related polypeptide binding to the TashUM probe. Indeed a preliminary experiment using a concatenated TashUM probe showed that this probe bound to a λgt11 expression clone, expressing the AT hook region of TashAT1 (McKellar and Shiels, unpublished, 2001). In contrast to these results, EMSA experiments with the TashUM homologue of *TashAT2* does not generate a major band shift with infected cell nuclear extracts (Swan and Phillips, unpublished, 2001). There is only a 6bp difference between the TashUM of *TashAT1/3* compared to the TashUM of *TashAT2*; which all occur in the AT rich region of the motif, and this region may be the site of specific binding to transcription factors. Thus, the 6bp difference is enough to change the binding properties of the TashUM region in *TashAT2* and might possibly account for the altered expression levels of TashAT2 and TashAT3 in D7 cells compared with D7B12 cells.

Further experiments for this work could involve EMSA analysis with the TashUM probe with deletions or mutations introduced into the probe. Verification of whether TashATs can bind to the TashUM probe could be established by supershift analysis using nuclear extracts and antisera generated against TashAT1/3 or TashAT2 or South-western analysis. UV fixation of the TashUM band shift to gain an estimate of the molecular mass of the associated polypeptides proved unsuccessful (McKellar and Shiels, unpublished, 2001). Ultimately, the identity of this factor could be obtained by purification of the factor(s) using the probe bound to a sepharose column followed by peptide sequencing. Alternatively, other proteomic methods could be used to fix the TashUM oligonucleotide onto a membrane, followed by mass spectrometry/protein sequencing. It is conceivable that the transcriptional activators of HMGI(Y) genes might also modulate TashAT gene expression. Some of these factors, including a number of cytokines and the transcription factor AP-1, have been implicated in attenuation of *Theileria* infected cell lines. Therefore, it would be useful to perform supershift assays with antisera generated against these transcription factors, to identify any host factors that bind to the TashUM motif upstream of *TashAT1/3*.

The work in this chapter has shown that all TashAT mRNA species are down regulated during differentiation to the merozoite, in a similar fashion. TashAT2 and TashAT3 mRNA appear to be differentially expression in D7 and D7B12 cell lines. A common motif upstream of the *TashAT1/3* genes and the *Tash1* gene (Swan *et al.*, 2001b) appeared to bind to a down regulated nuclear factor(s), which preferentially bound AT rich DNA.

5. Identification and Expression of TashAT1 and TashAT3 Polypeptide

5.1 Introduction

Data presented in previous chapters described the characterisation of a gene, *TashAT1* with considerable sequence identity to *TashAT2* including conserved AT hook binding domains found in a wide range of eukaryotic polypeptides. *TashAT1*, *TashAT2* and the newly discovered *TashAT3* were found to be clustered together in the *T. annulata* genome. Northern blot analysis presented in chapter 4 showed that all three genes are down regulated at the mRNA level during differentiation and that this occurs at a time point coincident with the initial reduction of host cell proliferation, between days 2-4 (Shiels *et al.*, 1992).

Sequence analysis of *TashAT1* revealed four AT hook DNA binding domains, nuclear localisation signals and an N-terminal signal peptide. This combination of motifs indicate that the polypeptide encoded by *TashAT1* could be transported from the parasite to the host nucleus. Evidence that one of the TashAT polypeptides, TashAT2, is transported to the host nucleus was reported by Swan *et al.* (1999). Antisera raised against two regions of TashAT2 polypeptide, the AT hook domain and the N-terminal section, that showed no identity with mammalian polypeptides in the database, did react against the host nucleus in addition to the macroschizont. Further experiments showed that TashAT2 was actively translocated to the nucleus of COS7 cells following transfection with a GFP-TashAT2 fusion construct (Swan *et al.*, 1999).

Given the level of sequence identity between *TashAT2* and *TashAT1*, it is likely that the TashAT1 polypeptide performs a similar function to the TashAT2 polypeptide. In particular, the presence of the putative nuclear localisation signals in both TashAT1/3 and TashAT2 putative polypeptides might suggest that TashAT1 is also transported to the host cell nucleus to modulate bovine gene expression and alter the proliferative status of the leukocyte. Indeed, there is evidence that AT hook encoding genes are involved in cell transformation. Many forms of leukaemia are associated with chromosomal translocations which generate chimaeric polypeptides containing AT hooks that bind to DNA and a transactivation domain which binds to other transcription factors, for example the *HRX* gene (Tkachuk *et al.*, 1992; Waring and Cleary, 1997). Thus, the first important aim was to

verify if the TashAT1 polypeptide was actually expressed in *Theileria* infected cells, and secondly, to determine the location of TashAT1 within the macroschizont infected cell. In order to achieve this, it was necessary to generate antisera specific to TashAT1 and then use this reagent to perform Western blot analysis and indirect fluorescent antibody tests (IFAT), respectively.

In chapter 4, TashAT1 mRNA was found to be expressed at the highest level at day 2 during a differentiating time course, but decreased significantly with respect to the 18S sn rRNA after this time point. Whilst it is likely that this data points to a down regulation of the polypeptide encoded by the mRNA, it is possible that a stable protein could remain functional after this time point. If the latter situation was true, it would indicate that TashAT1 may not be directly involved in the establishment of the cellular status of the macroschizont or the control of host proliferation. A third aim of this chapter was, therefore, to determine the expression profile of the TashAT1 polypeptide during merogony by Western blot and IFAT analysis.

Differences in the levels of TashAT1 mRNA were also discovered between D7 cells and D7B12 cells (a cell line attenuated for the differentiation process) under differentiation conditions. The attenuated D7B12 cells are able to proliferate at 41°C and show minimal levels of individual differentiating cells, following prolonged culture at an elevated temperature (Shiels *et al.*, 1994). If the TashAT1 polypeptide was down regulated in differentiating cells (D7), but remained effectively constant in non differentiating cells (D7B12), this would provide an interesting link between the presence of the TashAT1 polypeptide and the inability of the parasites to differentiate. Thus, the fourth aim of this chapter would be to determine the expression levels of TashAT1 polypeptide in D7B12 cells under differentiation conditions.

The fifth and final aim of this chapter was to investigate if TashAT1 was expressed in other *T. annulata* stages. The sporozoite stage was of particular interest as TashAT1 mRNA were identified in sporozoites by Swan *et al.* (2001a) where it was suggested that the gene could be involved in sporoblastogenesis or in the establishment of the parasite after sporozoite entry into the leukocyte. If TashAT1 was found to be expressed at the polypeptide level in sporozoites, this could indicate that TashAT1 has more than one function during the parasite life cycle or is involved in the initial phase of establishing infection within the white blood cell.

5.2 Results

5.2.1 Production of a GST-TashAT1 fusion protein and anti-sera generation

To analyse the polypeptide encoded by the *TashAT1* gene, it was necessary to generate antisera against a portion of this gene. The deduced peptide sequences of *TashAT1* and *TashAT2* were similar (46.1%) over the AT hook encoding regions in the C-terminal half of *TashAT1* (see chapter 3). Hence, the amino acid sequences were screened to find a unique region of *TashAT1*, which would be specific for *TashAT1* alone. To eliminate regions of strong homology, a DNA sequence comparison was performed between *TashAT1* and *TashAT2* and also with the 1.2kb EcoR1 fragment from λ dash 13, which were found to encode parts of another TashAT gene, subsequently identified as *TashAT3*. From these comparisons, a 368bp region located at the five prime end of the *TashAT1* ORF was chosen because it was least similar to any of the related sequences available at that time (see Fig. 5.1). This fragment was positioned 55bp downstream of the first putative translational start site, excluding most of the potential signal sequence, reducing the possibility of cross reactivity with other *Theileria* polypeptides carrying a similar signal sequence. Furthermore, as the signal sequence may be cleaved off during membrane translocation, in a similar manner described for the surface polypeptide Tams1 (Gubbels *et al.*, 2000), it would not be a good target for the generation of antisera for immunolocalisation studies.

Two primers (1 and 2, shown in Appendix A) were designed to amplify the 368bp region using the λ dash 13 clone as template DNA. These primers respectively incorporated EcoRI and BamH1 cloning sites into the final PCR fragment, so that the PCR fragment could be directionally cloned into the expression vector pGEX2TK, a well established system for expression of fusion proteins from recombinant DNA. The resultant PCR product amplified from λ dash 13 (see Fig. 5.2) and λ dash 1 DNA (which both contain *TashAT1)* gave a single strong band upon agarose gel electrophoresis confirming that this region of DNA was unique in these clones. Despite attempts to optimise PCR conditions, an additional band at 600bp was observed in lane 1, containing D7 genomic DNA. This band may represent a gene sharing identity with *TashAT1* or could have been an artefact due to non-specific annealing of the primers to genomic DNA. The 368bp PCR product was called TashAT1-N, and cloned into pGEX2TK vector. A 0.9kb HindIII/EcoRV probe

(which overlapped with the five prime region of *TashAT1*) detected a DNA fragment of this size by Southern blot analysis in a restriction digested TashAT1-N clone (see Fig. 5.3), indicating that this product was derived from *TashAT1* (lane 2). Although the pGEX2TK vector was also detected by this probe (lane 1), this was likely to be due to hybridisation of remaining vector sequence in the probe. Sequence analysis confirmed this insert was 100% identical to the corresponding region of *TashAT1* (Fig. 5.4), and was therefore suitable for expression of recombinant TashAT1-N.

Initial experiments to induce expression of TashAT1-N in *E. coli* XL-1 blue cells produced minimal levels of fusion protein, which may have been caused by high rates of degradation of the foreign fusion protein in these bacterial cells. Therefore the cells were induced at 30° C and 37°C over a 4 hour period, in an attempt to reduce the rate of degradation of TashAT1-N. The data presented in Fig. 5.5 showed the presence of a product at 45kDa (lanes 10 and 11) after 60 minutes, with a marked increase in levels at 120 minutes (lanes 13 and 14). However a further 120 minutes of incubation did not increase the yield of the 45kDa polypeptide (lanes 17 and 18). The 45kDa product was considered to be the TashAT1-N fusion protein because the size of this product approximated the predicted molecular mass of the TashAT1-N fusion protein (41kDa) and it was only detected in the induced cell extracts. The yields of TashAT1-N fusion protein appeared to be equal at both 30°C and 37°C at all time points (see Fig. 5.5), and indicated that temperature was not the reason for the low yields. The yields of TashAT1-N alone (lane 14) were markedly lower compared to those of the GST polypeptide (lane 15) or the fusion protein, cl10, at approximately 70kDa (lane 16). However, the levels of TashAT1-N polypeptide were estimated to be sufficient for protein purification and antisera production.

To purify TashAT1-N product, XL1-blue cells expressing TashAT1-N fusion protein were sonicated to disrupt the bacterial cell walls (described in section 2.2.8) and the resulting supernatant and pellet fractions analysed by SDS-PAGE as shown in Fig. 5.6. This result revealed that the fusion protein was present in both the supernatant and in the pellet fractions, most likely due to TashAT1-N being partially insoluble, as other polypeptides in the sample were fully separated into their respective fractions. Purification of the fusion protein was performed over a glutathione column (see section 2.2.8). Bound fusion protein was then washed and eluted from the column by reduced glutathione solution. SDS PAGE analysis in Fig. 5.6 demonstrated that TashAT1-N was successfully eluted at a greater concentration (lane 4) compared to the induced bacterial supernatant (lane 2). Some minor contamination by residual bacterial proteins that were eluted with TashAT1-N were also

detected (lane 4, Fig. 5.6). To remove the excess salts and to concentrate the fusion protein, the sample was dialysed against PBS as described in section 2.2.8. The concentration of TashAT1-N was increased by approximately 40% (lane 5, Fig. 5.6).

The total concentration of TashAT1-N was 1.5 mg/ml. Antisera (called 107 and 104) was generated against TashAT1-N from 2 New Zealand White rabbits after three boosts and antiserum 107 tested on fusion protein extract, shown in Fig. 5.7. The strongest reactivity observed was against purified TashAT1-N fusion protein, in the induced XL-1 blue cell extract (lanes 4 and 5), and to a lesser extent, GST alone, at 27kDa, (lane 2). Reactivity was also detected against the uninduced XL-1 blue cell extract (lane 1). Thus, reactivity of this antisera appeared to be directed against the 41kDa TashAT1-N polypeptide in cells expressing TashAT1-N fusion protein. Reactivity was also observed with another parasite fusion protein, cl10, at approximately 70kDa (lanes 6 and 7), which was probably due to antibodies generated against GST. Other minor bacterial proteins were also detected by antiserum 107. These results showed that antiserum 107 reacted strongly to TashAT1-N, and could be of use in the identification of TashAT1 by Western blot analysis of infected cell extracts.

In summary, a GST-fusion protein was produced from a region of *TashAT1*, TashAT1-N, which showed no homology to *TashAT2* or any other sequence at the time. Antisera were then generated against the TashAT1-N fusion protein, purified from bacterial cell extract. Western blot analysis of this antisera showed an increased reactivity to TashAT1-N fusion protein, making it suitable for analysis on infected cell cultures.

A

```
  1 MMVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKI
 51 MILSTPEDKITEIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIG
101 NAVDEAMKFIYVSGNFYKYINKSEFEDYYKSFCSVFIKIPPGKLPIPRIK
151 KNVKTEKVDKRKLKRDRQRKDKPQSEQHDKNVDIVSQSLAEEGIDLEKKI
201 VGREEPTQQTEKQQEPTELEPETIPVELESDDEEIDESNVSKPKESDCIL
251 TQNRYTQTDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVKD
301 IDTQTDIQELENIGIQTIGNFSDITEVTKKHEQPEVPKRRPGRPRKQKPE
351 PEQPKRKRGRPRKQKYETKKTWLLRPRNMKTETKKTWLLRPRKQKPEPEQ
401 PKRKRGRPRKQKYETKKTWLLRPRNMKTETKKTWLLRPRKHKPEPEQPKR
451 KRCRPRKQKPEPSSDT
```

B

```
                170        180        190        200        210        220
TashAT1-N TATCTACACCAGAAGATAAGATAACTGAAATACGTTCTAAAACGAAACTAATTTGGGGAA
                                                          ||  ||||||  ||||||
TashAT2   CCATCAAACAATATACCTTTAGGTGGAGTGTAAGTTCTATAAGGAACACCTTTATGTATT
                290        300        310        320        330        340

                230        240        250        260        270        280
TashAT1-N GCGATCGAGGTGAATATGTTAAATGTTTTACTAGATTTTCATTTGAATCGTCCGATAAGA

TashAT2   CCTTCTATACAATTCTCATTATTTATTACTTGCAAAGGTACTAAAATAAAATCATCTCGT
                350        360        370        380        390        400
```

Fig. 5.1: Putative peptide sequence of TashAT1 (panel A) and sequence comparison between the TashAT1-N fusion protein and *TashAT2* (panel B). The region designed for the TashAT1-N fusion protein is underlined; the putative signal sequence is indicated in bold. The identity score for the sequence comparison in panel B was 87.5% identity in a 16bp overlap.

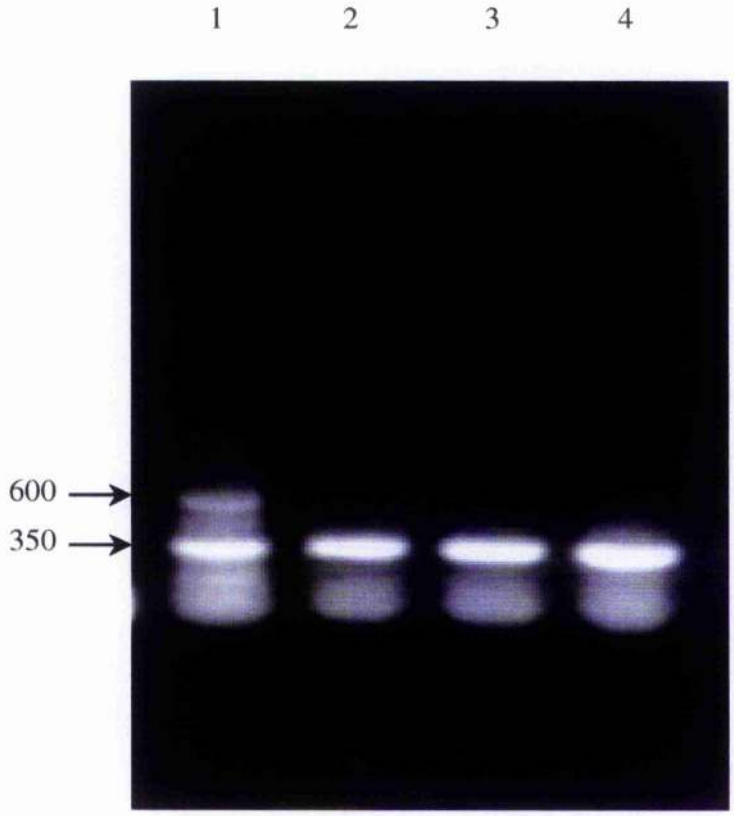


Fig. 5.2: PCR amplification of Ta369. D7 genomic DNA (lane 1); λ dash 13 (lane 2); λ dash 1(lane 3) and λ Ta1 (lane 4). Sizes of DNA fragments are indicated in bp.

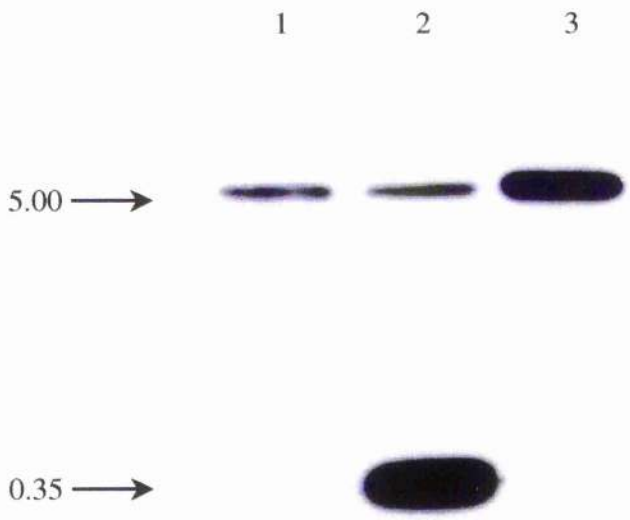


Fig. 5.3: Southern blot of Ta369 clone, hybridised with a 0.9kb HindIII/EcoRV probe overlapping the five prime region of TashAT1. Lane 1: pGEX2TK; lane 2: Ta369 cloned into pGEX2TK; lane 3: λTa1, all digested with EcoRI and BamHI. Arrows indicate the DNA fragments detected (kb).

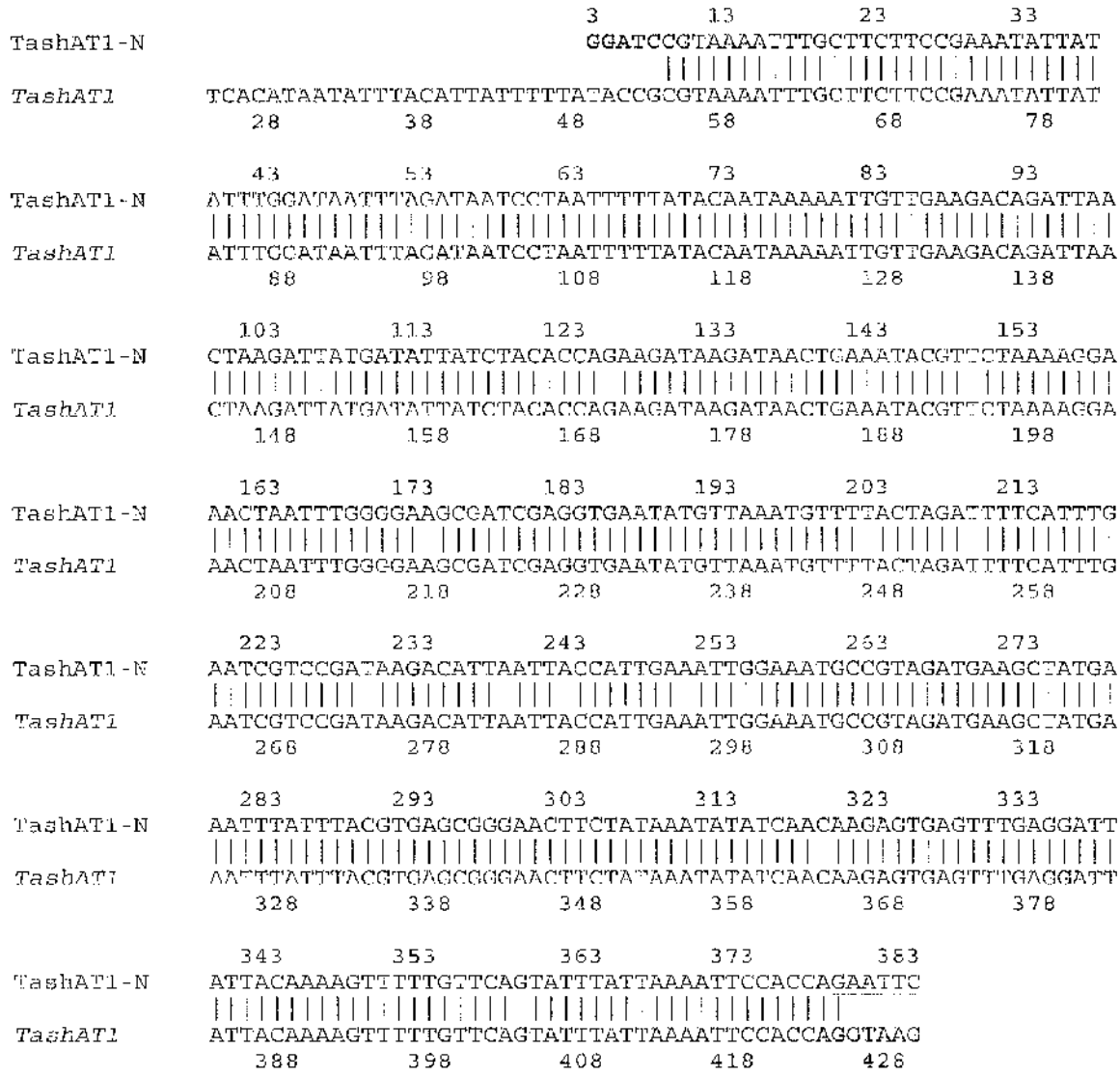


Fig. 5.4: DNA sequence comparison between *TashAT1* and the PCR product of the potential TashAT1 fusion protein, Ta369. Sequence comparison shows 100% identity over a 374 nucleotide overlap. Sequence in bold type and underlined indicate the BamHI and EcoRI restriction enzyme sites respectively.

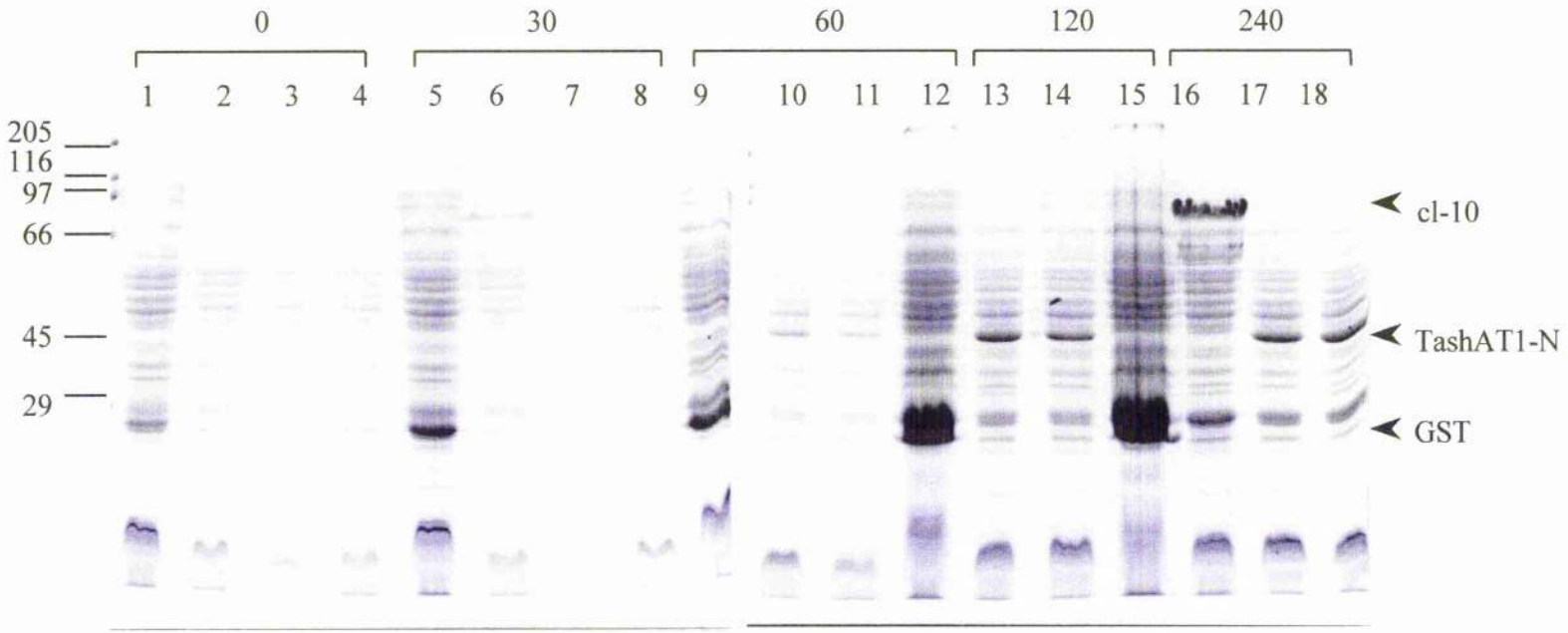


Fig. 5.5: Expression of TashAT1-N fusion protein in XL-1 blue *E.coli* cell extracts at 5 time points .
Time points indicated in minutes above the lanes. Lanes 1-4 contains extracts derived from un-induced cells: lane 1 and 5, 9, 12, 15 contain extracts derived from un-induced and induced pGex vector control cells, respectively. Lanes 2, 6 and 16 contain extracts of cells expressing the Tash1 fusion protein.
Lanes 3, 7, 10, 13 and 17 contain extracts of cells expressing TashAT1-N at 37°C, whereas lanes 4, 8, 11, 14 and 18 contain extracts of cells expressing TashAT1-N at 30 °C. Arrows on the left indicate the molecular weight markers (in kDa); the arrow head on the right indicates the fusion proteins detected.

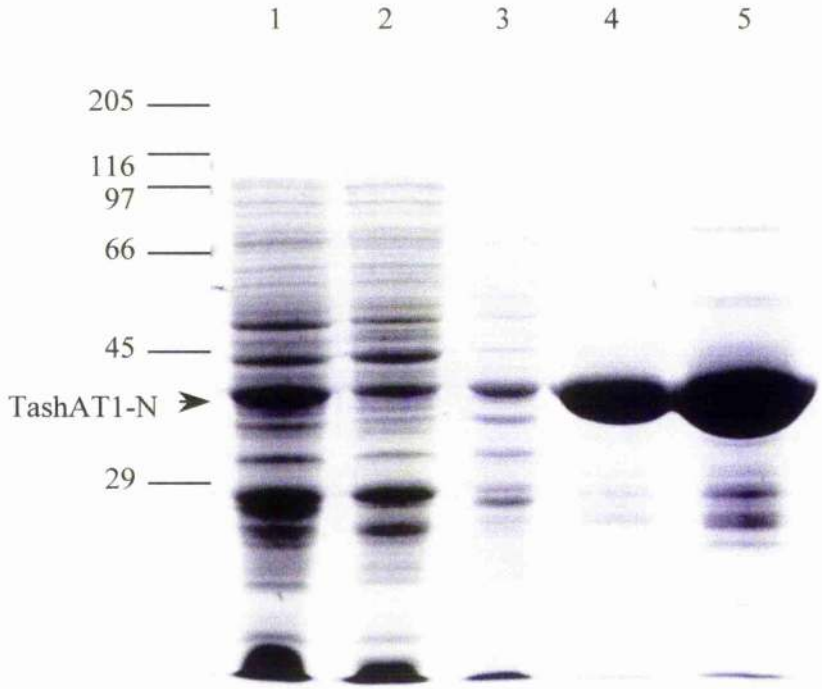


Fig. 5.6: Purification of the fusion protein TashAT1-N shown by SDS-PAGE. Lane 1 shows the induced whole bacterial cell extract, lanes 2 and 3 contain the supernatent and resuspended pellet, respectively after sonication. Lane 4 contains the eluate after column purification of the bacterial lysate. Lane 5 contains the eluate after dialysis. Numbers on the left indicate the molecular weight markers (in kDa). The arrow head indicates theTashAT1-N-GST fusion protein.

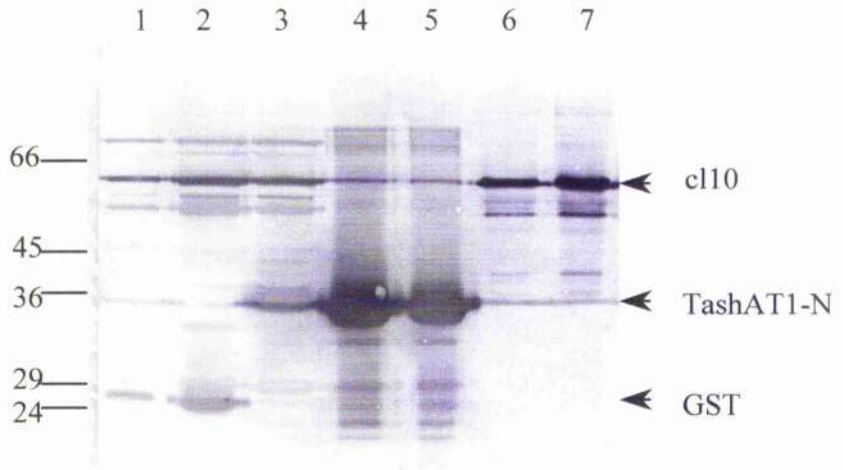


Fig. 5.7: **Western blot analysis of *E.coli* extracts expressing TashAT1-N, with antiserum 107**. Lane 1: uninduced XL-1 blue cell extract containing pGex alone; lane 2: induced XL-1 blue pGex cell extract expressing GST. Lane 3: induced XL-1 blue TashAT1-N-GST cell extract. Lanes 4 and 5: purified TashAT1-N-GST. Lanes 6 and 7: XL-1 blue cell extract, induced to express the cl-10 GST fusion protein. Numbers on the left indicate the molecular weight markers (in kDa). The arrow heads on the right, indicates the fusion proteins detected.

5.2.2 Detection and localisation of the TashAT1 polypeptide

Western blot analysis was performed with D7 and BL20 protein extracts using antisera 104 and 107 to detect the TashAT1 polypeptides and/or the presence of any related polypeptides. Each antiserum was used at a different dilution to determine the optimal concentration. The results, in Fig. 5.8A and B, showed that both antisera reacted to a polypeptide at approximately 66kDa. From sequence data, the predicted molecular mass of TashAT1 is 55.3kDa, and although the polypeptide detected did not match the size of the predicted mass of TashAT1, it was within range taking into account aberrant mobility and any possible post-translational modifications of this polypeptide. A comparison of sera 107 and 104 revealed that serum 107 was the most sensitive, showing reactivity at a dilution of 1:200, whereas serum 104 only reacted with the polypeptide at 1:50. Further Western blot and IFA studies with serum 104 (data not shown) showed that serum 104 recognised the same polypeptides but more faintly. Since serum 104 was found to be less sensitive than 107, the latter was chosen for future work. In lane 1 (Fig. 5.8), a polypeptide of approximately 66kDa and a larger polypeptide at 125kDa were also present in the BL20 extract reacted with antisera 107. This result indicated that cross-reactivity had occurred with host derived polypeptides.

When further Western blot analysis was performed with D7 and D7B12 using serum 107 (shown in Fig. 5.9), two polypeptides were detected at approximately 66kDa and 71kDa and a larger polypeptide at 180kDa was also detected in D7B12 extract (Fig. 5.9A). The only polypeptide detected in D7 extracts was at 66kDa in Fig. 5.9A. However, a separate Western blot (Fig. 5.9B) revealed the same three polypeptides, estimated to be approximately 180kDa and 71kDa and 66kDa in D7 cell extract, within the limits of the percentage gel used. This demonstrated that detection of polypeptides in D7 cell extracts was variable using antisera 107.

Based on size considerations, the 180kDa band is unlikely to represent the TashAT1 polypeptide. Given that this band appeared to be specific to *Theileria* infected cells, it is possible that antisera 107 detected shared amino acid epitopes with a related polypeptide, such as TashAT3. The 180kDa polypeptide was unlikely to be TashAT2 firstly because of the sequence divergence of TashAT1 and TashAT2 over the region used to generate the TashAT1 fusion protein, and secondly because the TashAT2 polypeptide was found to be 150kDa (Swan *et al.*, 1999). Indeed, following the production of the antisera it became apparent that TashAT3 was the most likely candidate for the 180kDa band because the

TashAT1-N polypeptide was found to be 100% identical to the N –terminal section of TashAT3. Also this band was within range of the predicted molecular weight of TashAT3 polypeptide (113kDa). The difference in the predicted and actual sizes of TashAT3 could be due to aberrant molecular mobility of the TashAT polypeptides combined with less accurate size estimation with the percentage of acrylamide used for the SDS-PAGE gel or post-translational modifications.

Due to the recognition of polypeptides in the BL20 cell extract, a peptide blocking experiment was performed to determine if serum 107 reacted specifically with the polypeptides detected in D7 cell extracts, by Western blot analysis. In this assay, serum 107 was pre-incubated with its cognate fusion protein, TashAT1-N, before being used to probe a blot of D7 protein extract. If the reactivity was specific and related to epitopes in the fusion protein then it would be expected that all the antisera would bind to TashAT1-N, resulting in no, or significantly reduced reactivity compared to the reactivity of antisera alone with D7 extracts. Fig. 5.10 demonstrates that all three polypeptides detected by non-blocked serum 107 in the D7 extract (Panel C) were blocked upon pre-incubation with TashAT1-N (panel E). When serum 107 was pre-incubated with induced XL-1 blue cell extract expressing GST detection of the 180, 71 and 66kDa polypeptides was not blocked (Panel D). Thus this experiment showed that specific detection of the three polypeptides by the 107 antiserum was due to specific epitopes binding in the cognate fusion protein and not due to cross reactivity with epitopes present in GST or a contaminating bacterial polypeptide.

To investigate the localisation of polypeptides detected by serum 107 within a macroschizont-infected cell, IFAT analysis was performed with cells derived from D7 and D7B12, cultured at 37°C. The results (Fig. 5.11) showed strong recognition of both host cell nuclei and the macroschizont in D7B12 cells. Antiserum fluorescence was also detected in host nucleus and macroschizont of D7 cells (panel D) as confirmed by DAPI staining of the host and parasite nuclei of D7 cells (panel E), but to a lesser extent. This was similar to the pattern obtained with the EL24 or DE39 antisera raised against TashAT2 (Swan *et al.*, 1999). No reactivity was observed using pre-immune serum with D7 cells (panel B), confirming that antibodies were not generated against polypeptides in the infected cells, other than TashAT1-N. In panel A, no reactivity was detected in the uninfected bovine BL20 with serum 107, indicating that antiserum 107 did not react to host derived polypeptides by IFAT. In contrast repeated Western blot experiments sometimes revealed the presence of a 66kDa band in BL20 extracts (as shown in Fig. 5.8) and by

Swan *et al.* (2001a). Therefore based on this data antiserum 107 could not be used to distinguish TashAT1 and TashAT3 by IFAT and was re-named anti-TashAT1/3 serum.

To eliminate the possibility that the IFAT fluorescence was due to recognition of TashAT2, the anti- TashAT1/3 serum was incubated with a Bovine Macrophage cell line (BoMacs) that had been transfected with a construct expressing the TashAT2 polypeptide (apart from 23 N-terminal amino acids). Fig. 5.12 showed that anti-TashAT1/3 sera did not react with BoMac cells expressing TashAT2 (Fig. 5.12A), compared to DE39, an antiserum raised against TashAT2, which showed strong nuclear reactivity (Fig. 5.12C). As anti-TashAT1/3 serum was shown previously to react against D7 cells, the results indicated that TashAT2 is not recognised by the anti-TashAT1/3 serum by IFAT.

Western blot analysis was also performed on D7 extracts using antisera DE39 raised against the AT hook encoding region of TashAT2, that is common to all *TashAT* genes, to determine if the polypeptides detected by anti-TashAT1/3 serum are also recognised by this antisera. The results in Fig. 5.13 showed that antisera DE39 predominantly detected a number polypeptides, including four major bands estimated to be 58kDa, 90kDa, 160kDa and over 200kDa in D7B12 cell extracts, only the 90kDa polypeptide was detected in D7 cell extracts. However, none of the polypeptides, detected by antiserum DE39 corresponded to the polypeptides detected by anti-TashAT1/3 serum, apart from the 58kDa polypeptide. This latter polypeptide might be a TashAT1 candidate, although antiserum DE39 detected a number of polypeptides at this size. It was also possible that the 160kDa polypeptide in lane 4 corresponded to the 180kDa polypeptide, identified in Fig. 5.9, but the 180kDa band was not detected in either D7 or D7B12 cell extracts, (possible due to poor polypeptide transfer onto the membrane) so this could not be verified. The position of the 205kDa molecular weight marker was deduced from a double log graph of the molecular weight markers (data not shown) as this band was not detected.

Although anti-TashAT1/3 serum detected three polypeptides, it was not known if the host and/or parasite IFAT reactivity corresponded with these polypeptides. Parallel studies performed at this time (Stadler, unpublished, 2000) identified two *T. annulata* cell lines (Ta46A and TBL20) which did not contain *TashAT1* specific restriction fragments by Southern blotting. However, the *TashAT3* gene was thought to be present in Ta46A cells by Southern blot analysis (Stadler, unpublished, 2000) and small amounts of TashAT3 mRNA were detected in TBL20 cells by Northern blot analysis (Swan *et al.*, 2001a).

In order to determine if the 66kDa or 71kDa polypeptides and the 180kDa polypeptides were in fact TashAT1 and TashAT3 respectively, Western blot analysis was performed with Ta46A and TBL20 cell extracts and compared with those of D7 and D7B12 extracts (Fig. 5.14). The results revealed two bands at 71kDa and 66kDa (seen as a doublet) and a band at 55kDa in both the Ta46A and TBL20 cell lines. These bands were identical to the polypeptide profile of D7 and D7B12 respectively (see lanes 1 and 2). In Fig. 5.14 (lane 5) the same polypeptides and an additional band at 70kDa were also present in BL20 cell extract (which shares the same lineage as TBL20 cells). This clearly demonstrated that the polypeptides detected by anti-TashAT1/3 serum at 66kDa or 71kDa were not TashAT1, as they were detected in two cell lines that lack the *TashAT1* gene and in uninfected cells. The origin of the 66kDa and 71kDa polypeptides in D7 cells may be host cell derived, or could be homologues of the *TashAT* genes.

To investigate further whether either the 71kDa or the 180kDa were responsible for the reactivity observed in D7 and D7B12 cells by IFAT, Ta46A and TBL20 cells were subjected to IFAT analysis. The results (see Fig. 5.15) showed that anti-TashAT1/3 serum produced no reactivity against the host nucleus or the macroschizont of Ta46A cells (Fig. 5.15, panel C), in contrast to D7 and D7B12 cells (Fig. 5.11, panels C and D). However, TBL20 cells, which contained *TashAT3* but not *TashAT1*, showed weak reactivity in the host nucleus but not in the macroschizont (Fig. 5.15, panel B). These results revealed that neither the 66kDa band nor the 71kDa polypeptides could be responsible for the macroschizont and host reactivity detected in D7 cells as no reactivity was detected in Ta46A cells by IFAT, which lacked the *TashAT1* gene. The identity of the polypeptide responsible for host nuclear reactivity with anti-TashAT1/3 serum in TBL20 cells could be the 180kDa polypeptide but this could not be confirmed by Western blot analysis because this molecule was not detected in Fig. 5.14. As no reactivity was detected in BL20 cells (Fig. 5.11, panel A), the host reactivity in TBL20 cells is unlikely to be due to a host polypeptide.

In summary, three polypeptides were initially detected in D7 and D7B12 cell extracts by anti-TashAT1/3 serum. Two of these molecules at 66kDa and 71kDa were similar in size to the predicted molecular weight of TashAT1 (55.3kDa). A comparison of these cell lines with other cell lines, lacking the *TashAT1* gene, showed that the 66kDa and 71kDa polypeptides did not represent the TashAT1 polypeptide. The third polypeptide detected at 180kDa, was predicted to be TashAT3, based on its size and sequence identity with TashAT1 predicted polypeptide. IFAT analysis with anti-TashAT1/3 serum showed

reactivity in both the host nucleus and the macroschizont in D7 cells, and to a greater degree in D7B12 cells.

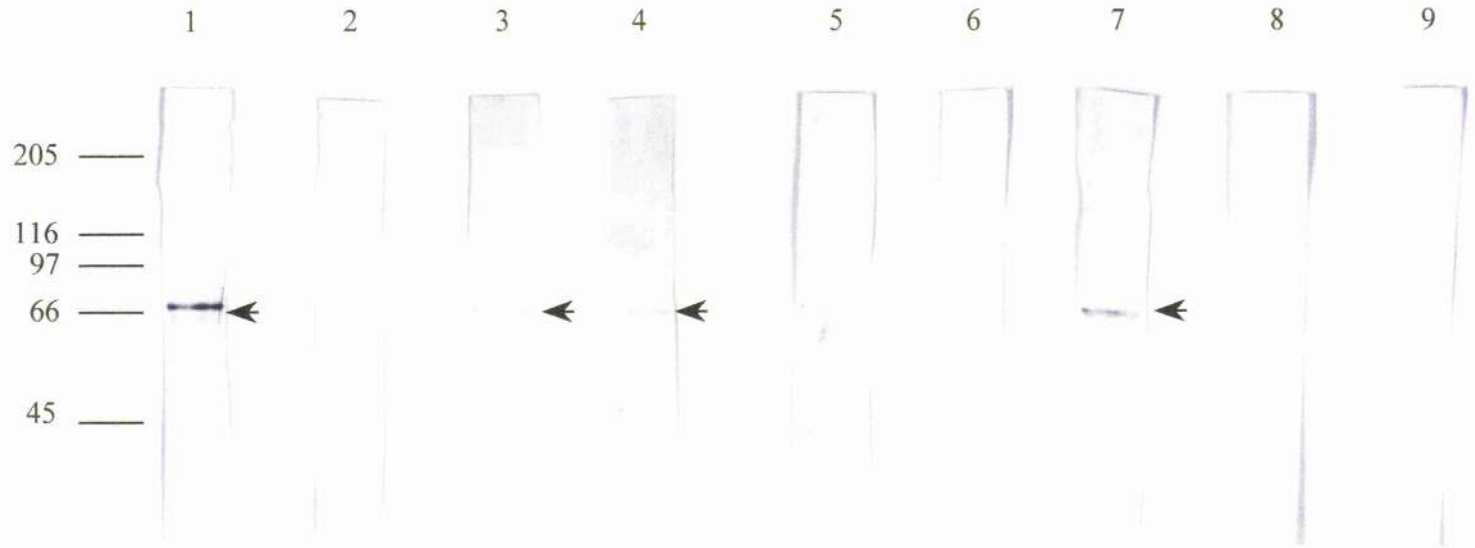


Fig. 5.8 : Western blot analysis of D7 cell protein extract at 37°C with antisera 107 and 104. Lanes 1 and 9: BL20 control extract incubated with 107 (1/200) and 104 (1/200) respectively. Lanes 2-4: D7 cell extract incubated with 107 at 1/1000 (lane 2), 1/200 (lane 3) and 1/50 (lane 4). Lane 5 : D7 extract incubated with pre-immune 107 serum at 1/200. Lane 6 and 7: D7 extract incubated with 104 at 1/200 and 1/50 respectively. Lane 8: D7 extract incubated with pre-immune 104 serum (1/50). Numbers on the left represent molecular weight markers (in kDa). The arrow head indicates the polypeptide detected.

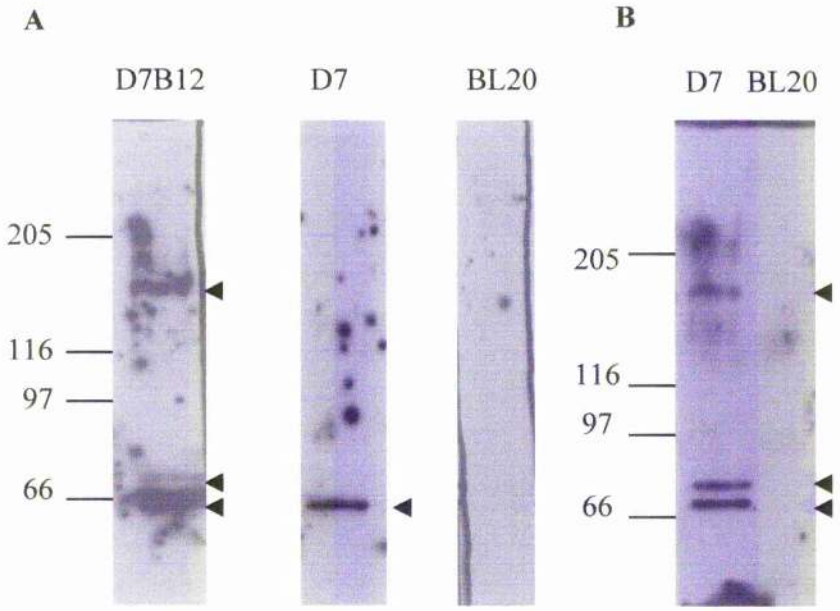


Fig. 5.9: Western blot analysis of *T. annulata* infected leukocyte and BL20 control extract with antiserum 107. Panels A and B refer to two separate western analyses. Numbers on the left indicate the molecular weights markers (in kDa). Arrow heads on the right indicate the polypeptides detected.

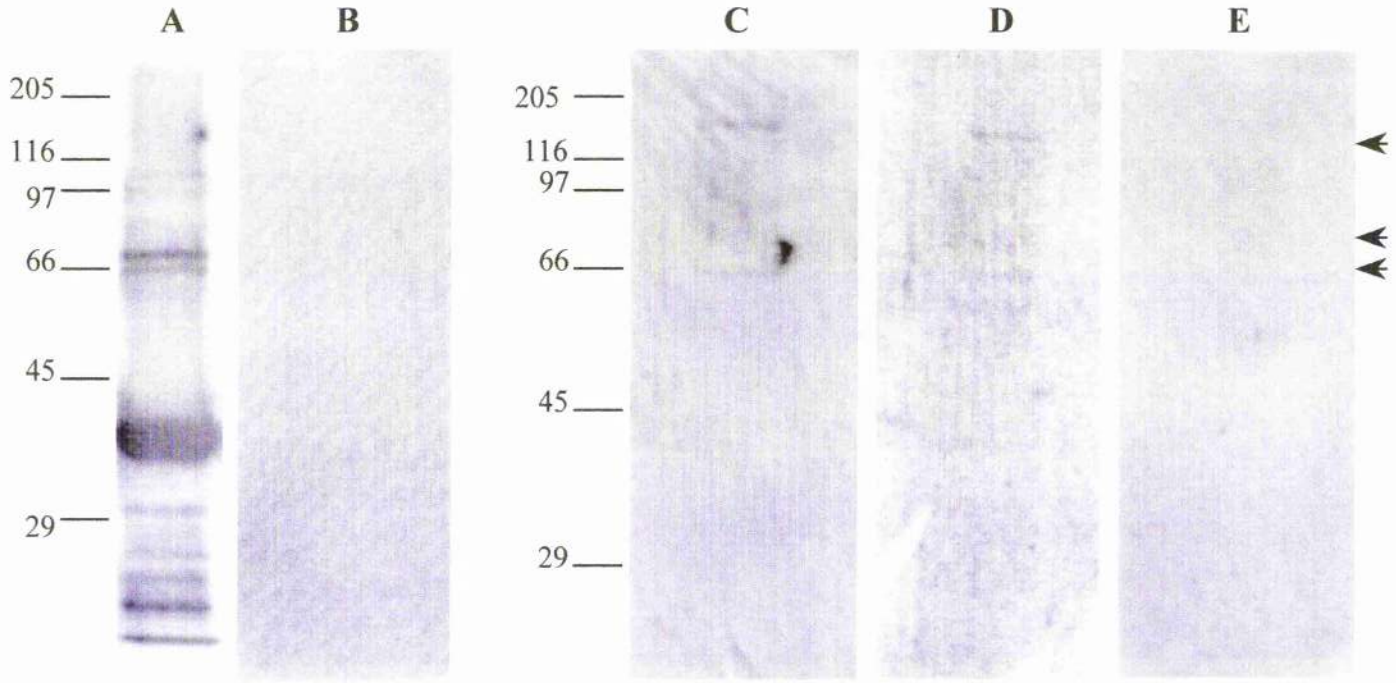


Fig. 5.10: Western blot analysis of a D7 cell extract blocked with TashAT1-N using antiserum 107. Panel A: protein extract from XL-1 blue cells expressing TashAT1-N-GST. Panel B: pre-immune serum; panel C: antiserum 107; Panel D: antiserum 107 pre-blocked with XL-1 blue cell extract expressing the GST polypeptide; Panel E: antiserum 107 pre-blocked with TashAT1-N-GST. Numbers indicate the molecular weight markers (in kDa). Arrow heads indicate the polypeptides detected.

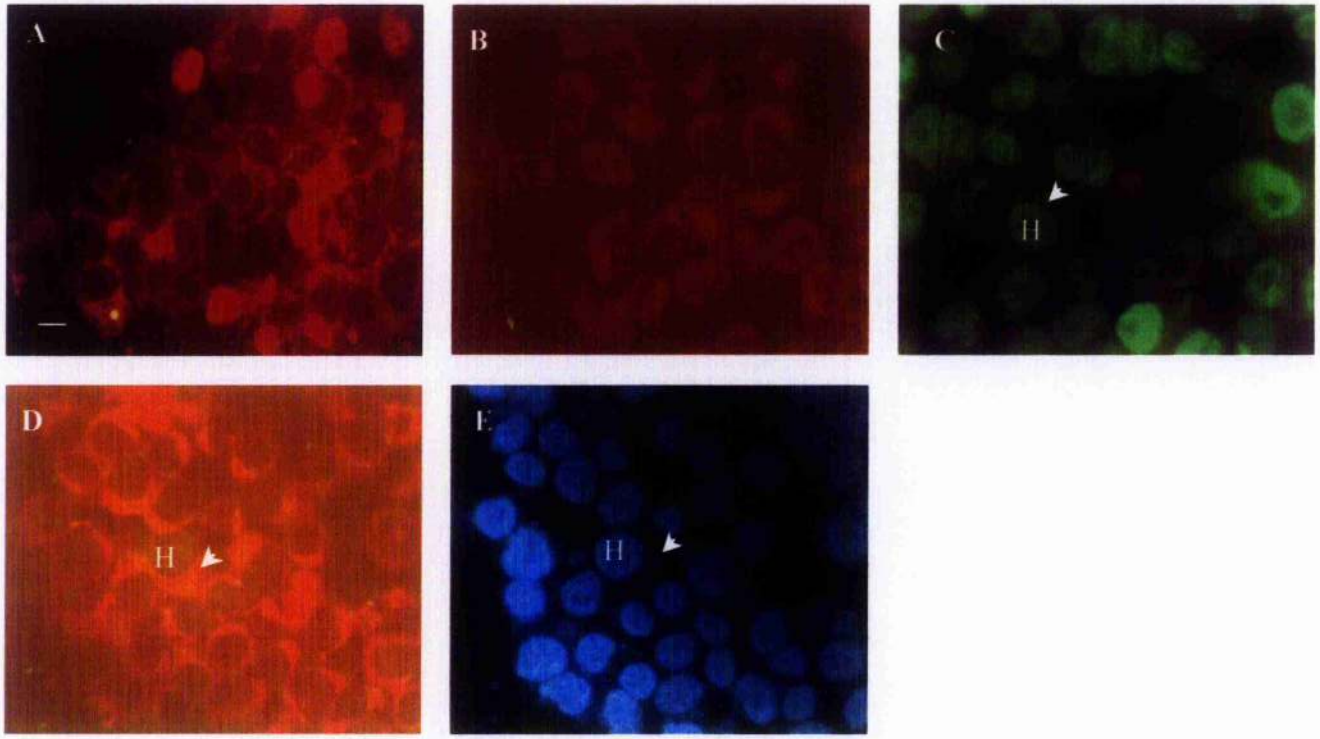


Fig. 5.11: IFAT analysis of *T. annulata* macroschizont infected leukocytes and BL20 cells with various antisera. Cells were counter-stained in red, antisera reactivity is shown in green. Panel A: BL20 cells reacted with anti-TashAT1/3 serum; B: D7 cells reacted with pre-immune serum; C: D7B12 cells reacted with anti-TashAT1/3 serum; D: D7 cells reacted with anti-TashAT1/3 serum; E: same as panel D treated with DAPI nuclear stain (blue). H indicates the host nucleus; the arrow indicates the macroschizont. Bar= 10μm

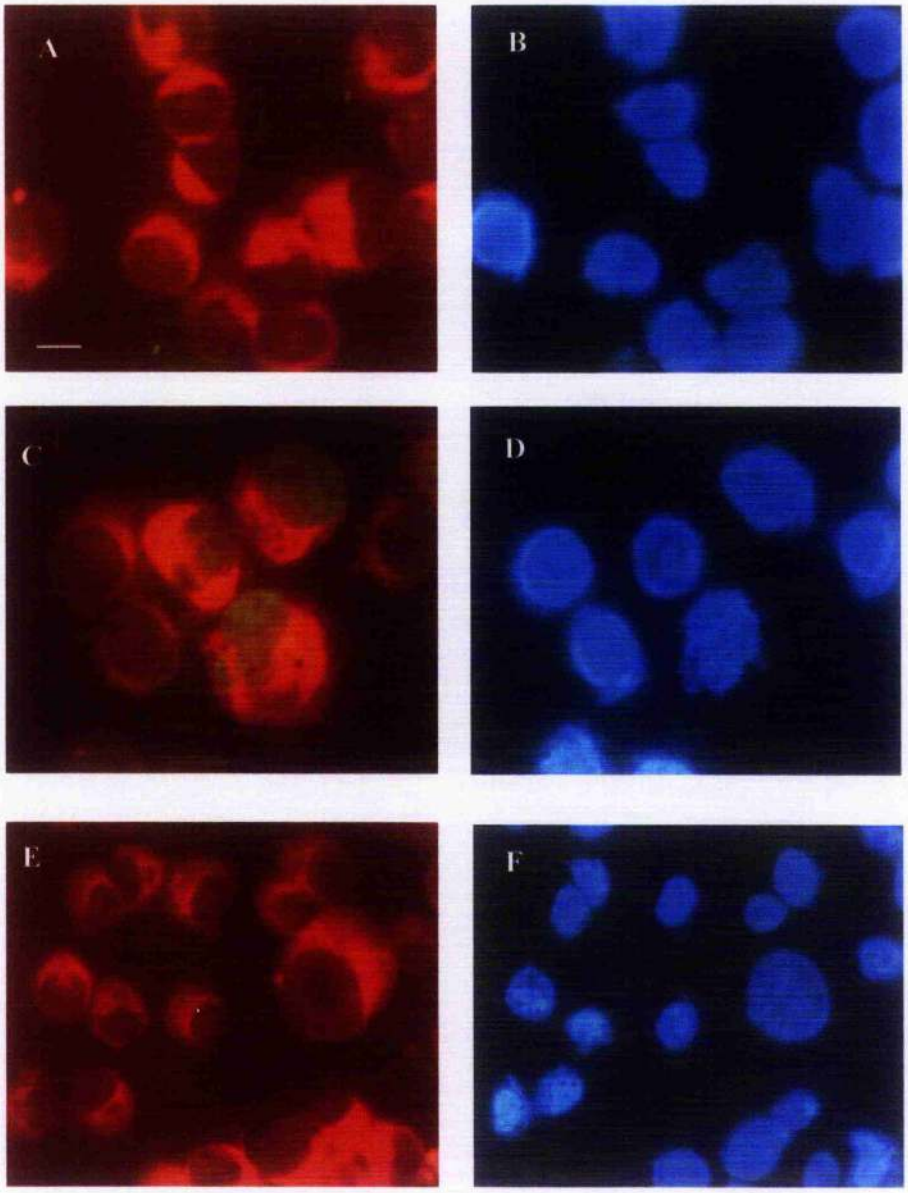


Fig. 5.12: **IFAT analysis of Bovine macrophage cells (BoMacs) with sera 107 and DE39**. Cells were counter-stained in red; antisera reactivity is shown in green. Panels A and C: BoMacs transfected with TashAT2 with sera 107 and DE39 respectively. Panel E:Non-transfected BoMac cells with DE39. Panels B, D and F are the corresponding DAPI stained cells to panels A, C and E respectively. Bar= 10μm.

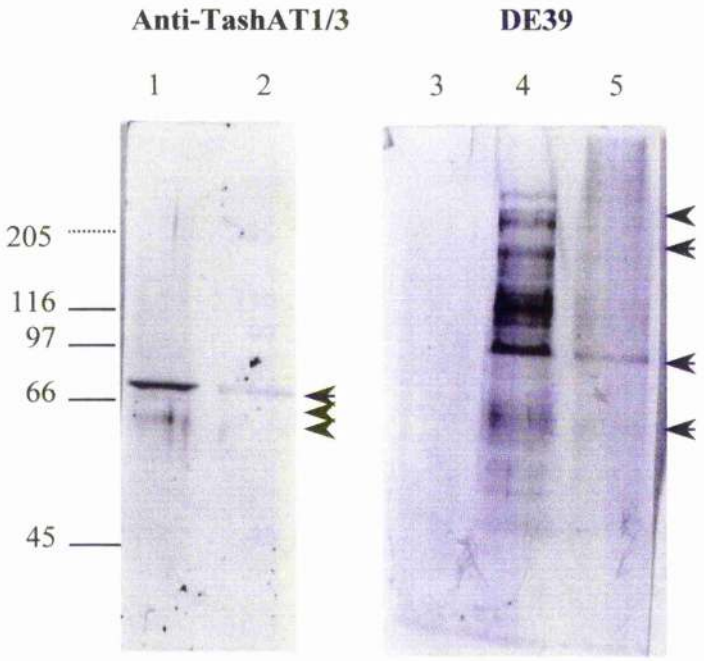


Fig. 5.13: Western blot analysis of macroschizont infected leukocytes with anti-TashAT1/3 and DE39 antisera. Cell extracts: D7B12 (lane 1); D7 (lane 2), BL20 (lane 3); D7B12 (lane 4) and D7 (lane 5). Numbers on the left indicate molecular weight markers (in kDa); dotted line indicates deduced marker position. Arrow heads indicate the major polypeptides detected.

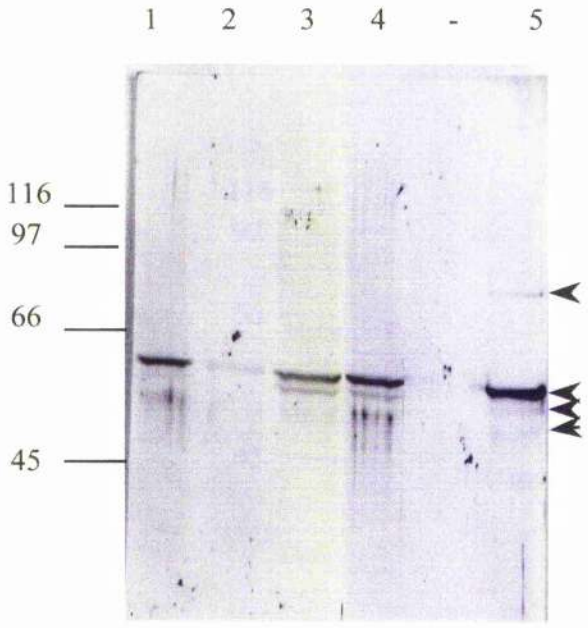


Fig. 5.14: Western blot analysis of cell extracts derived from four *T.annulata* cell lines and BL20 control cells reacted with TashAT1/3 antiserum. Lane 1:D7B12 cells; lane 2: D7 cells; lane 3: Ta46A cells; lane 4: TBL20 cells and lane 5: BL20 cells. Numbers on the left represent molecular weight markers (in kDa); arrows on the right indicate the polypeptides detected.

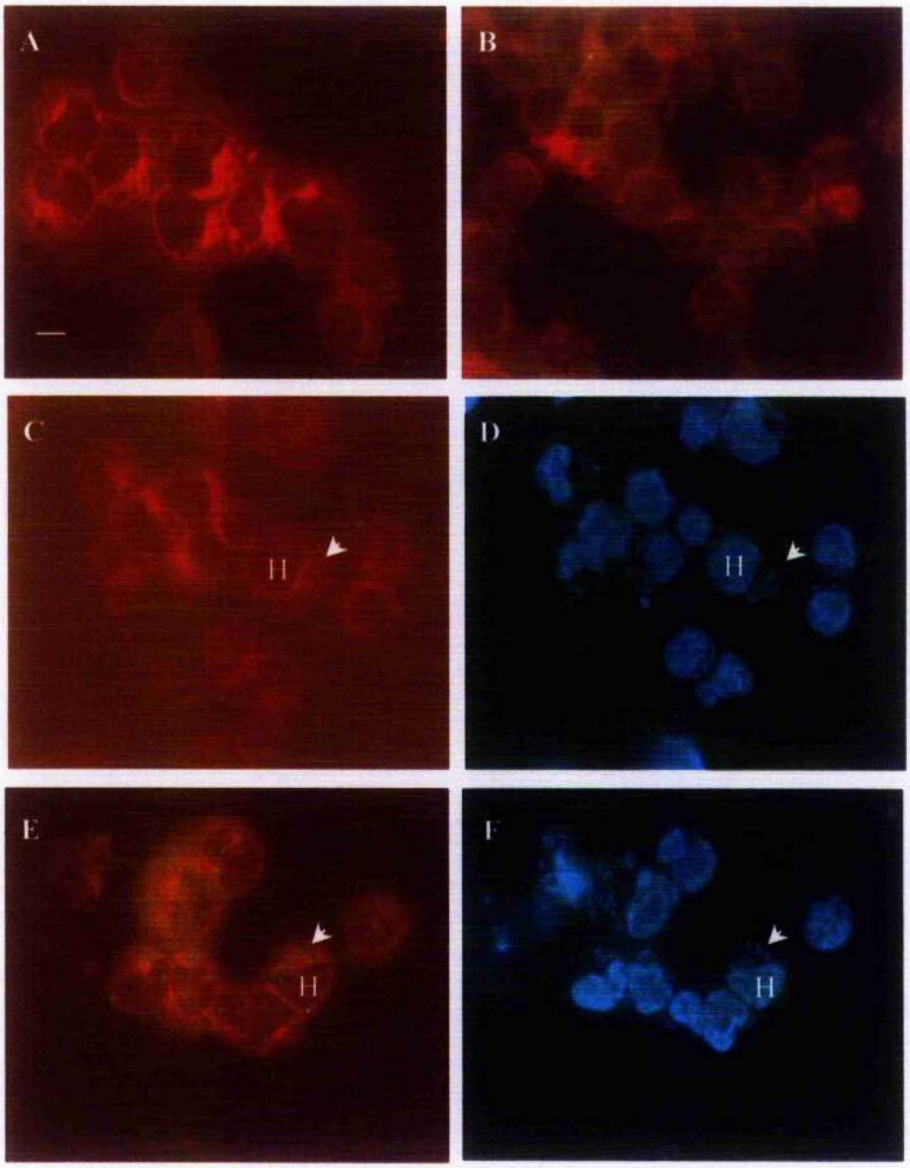


Fig. 5.15: IFAT analysis of *T.annulata* infected cell lines TBL20, Ta46A with various antisera. Cells were counter-stained in red; anti-sera reactivity is shown in green. Panel A: TBL20 cells with pre-immune serum and anti-TashAT1/3 serum in panel B. Panel C: Ta46A cells reacted with anti-TashAT1/3 serum; panel D: same as panel D treated with DAPI nuclear stain (blue). Panel E:Ta46A cells reacted with positive control serum R881; panel F: same as panel E treated with DAPI nuclear stain. H: host nucleus; arrow head indicates a macroschizont. Bar=10μm.

5.2.3 TashAT3 polypeptide expression during differentiation to the merozoite

To compare the expression profile of the polypeptides detected by anti-TashAT1/3 serum with that of the TashAT1/3 mRNA transcripts during merogony, Western blot analysis was performed on cell extracts from a D7 cell differentiation time course (Fig 5.16). Two bands were detected in the D7 cell extract, one at 180kDa and a second at approximately 71kDa that was previously eliminated as a TashAT polypeptide candidate. The levels of the 180kDa polypeptide increased from 6 hours to 24 hours and stayed at this elevated level until 48 hours, except at the 31 hour time point, which showed diminished levels of this polypeptide. At day 4 (96 hours), only traces of the 180kDa polypeptide were visible; by day 6 (144 hours), no traces of this polypeptide were observed.

To show that the reduction in the levels of the 180kDa polypeptide was associated with differentiation, Western blot analysis was performed on the D7B12 cell line (attenuated for differentiation), incubated under differentiation time course conditions (see Fig. 5.17). These results showed two bands at approximately 180kDa and the non-TashAT polypeptides at 66kDa, and at 71kDa (from days 4 and 6 only). The levels of the 66kDa polypeptide remained constant over the time course apart from day 4, where it was slightly elevated. The 180kDa polypeptide showed little variation from day 0 to day 6, although there was a noticeable increase in levels at day 4. The levels of total protein extract, as measured by Ponceau staining (data not shown), did not increase at day 4. Thus, no progressive reduction in the levels of the 180kDa polypeptide in D7B12 cells was detected from day 0 to day 6, unlike its profile in D7 cells.

To confirm the down regulation of host nuclear and/or parasite polypeptides detected by anti-TashAT1/3 during merogony, IFAT analysis was performed on a D7 cell differentiation time course. In general, these results (Fig. 5.18) revealed that both host nucleus and macroschizont reactivity decreased as the macroschizont differentiated towards the merozoite. However, small differences were observed between the reactivity of the host nucleus and the parasite in this experiment. Reactivity in the host nucleus was relatively weaker than in the macroschizont at day 0. At day 2 the reactivity in the macroschizont had disappeared in the majority of cells, whereas host reactivity became stronger at this time point. No reactivity was detected in the macroschizont at day 4 and day 6. Host nuclear reactivity became fainter in each cell at day 4, and the number of cells showing host nuclear reactivity also became less frequent. This trend continued to day 6,

although some host nuclear reactivity was still visible at the last time point, consistent with an asynchronous culture.

To confirm that the decrease in reactivity in the host nucleus and parasite detected by IFAT was due to differentiation, IFAT analysis was performed on D7B12 cells under differentiation time course conditions (see Fig. 5.19). The levels of host nuclear and parasite fluorescence remained constant from day 0 to day 6, although a slight decrease in parasite reactivity at day 6 was observed in D7B12 cells, but not to the same degree as the reactivity in D7, day 6 cells. Parasite and host nuclear reactivity remained unchanged from day 0 to day 2, unlike their counterparts in D7 cells at this time point.

In summary, the results of Western blot analysis showed a reduction in the levels of the 180kDa polypeptide in D7 cells undergoing merogony. This reduction was not so apparent in D7B12 cells that are unable to undergo differentiation to the merozoite. The results of IFAT analysis of D7 and D7B12 cells under differentiation conditions showed that reactivity of anti-TashAT1/3 serum had decreased with respect to both the host nucleus and the parasite in D7 cells overall, however, this reduction was not observed for D7B12 cells.

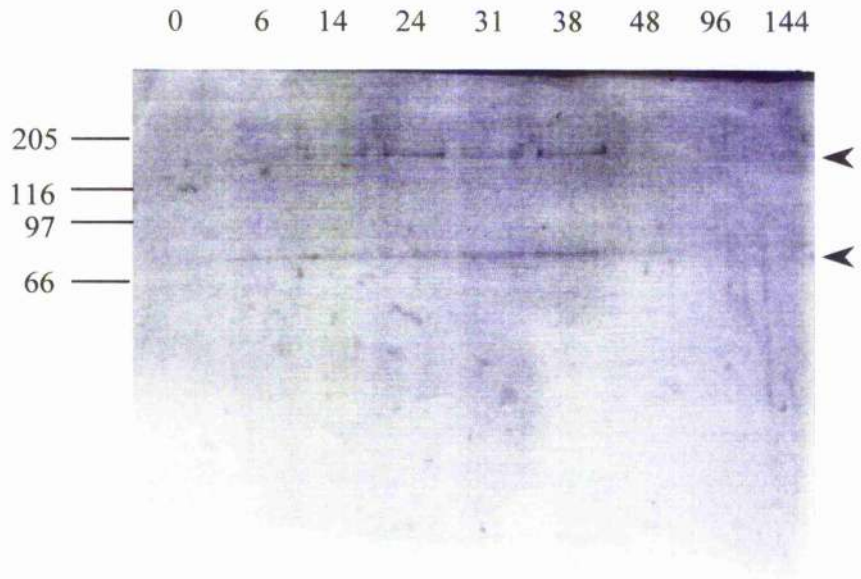


Fig. 5.16: Western blot analysis of D7 cell extracts taken from a differentiating time course incubated with anti-TashAT1/3 serum. Time points are indicated (in hours) above each lane; numbers on the left indicate the molecular weight markers (in kDa); the polypeptides detected are indicated by arrows.

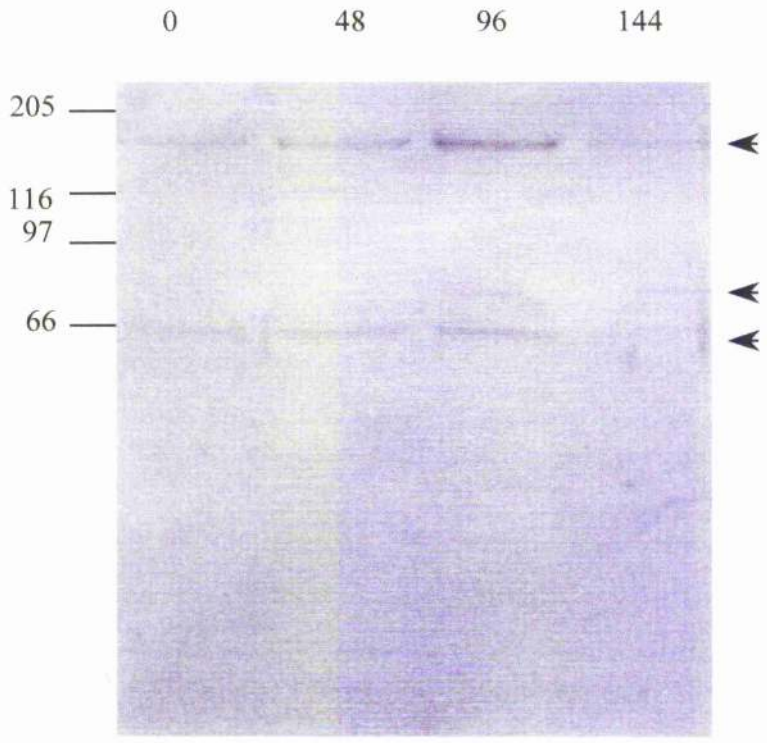


Fig. 5.17: Western blot analysis of a D7B12 differentiation time course, incubated with anti-TashAT1/3 serum. Time points (hours) are indicated above each lane; day 0= D7 cells at 37°C. Numbers indicate molecular weight markers (in kDa). The polypeptides detected are indicated by arrow heads.

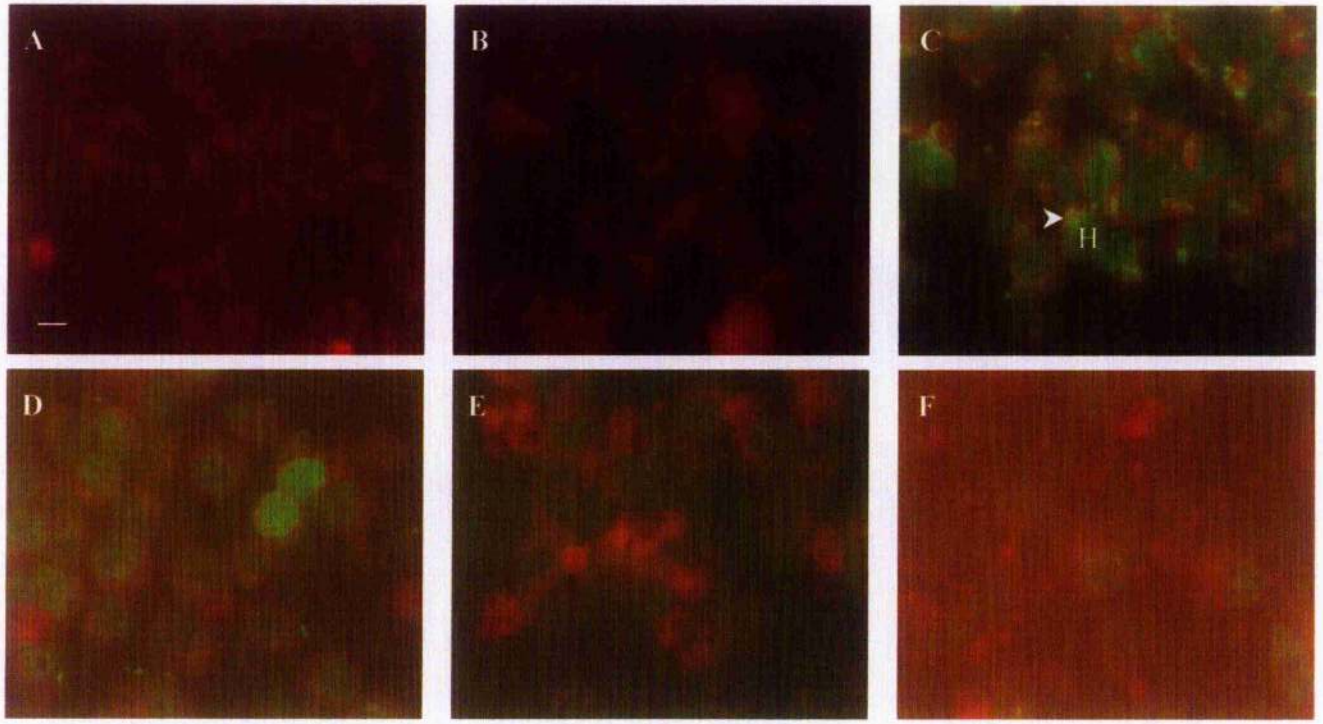


Fig. 5.18: IFAT analysis of a D7 cell culture time course at 41°C. Cells were counter-stained in red; antisera reactivity is shown in green. BL20 cells (panel A); D7 cells reacted with pre-immune serum (panel B) and anti-TashAT1/3 serum at day 0 (panel C); day 2 (panel D); day 4 (panel E) and day 6 (panel F). H represents the host nucleus and the arrow indicates the macroschizont. Bar=10μm.

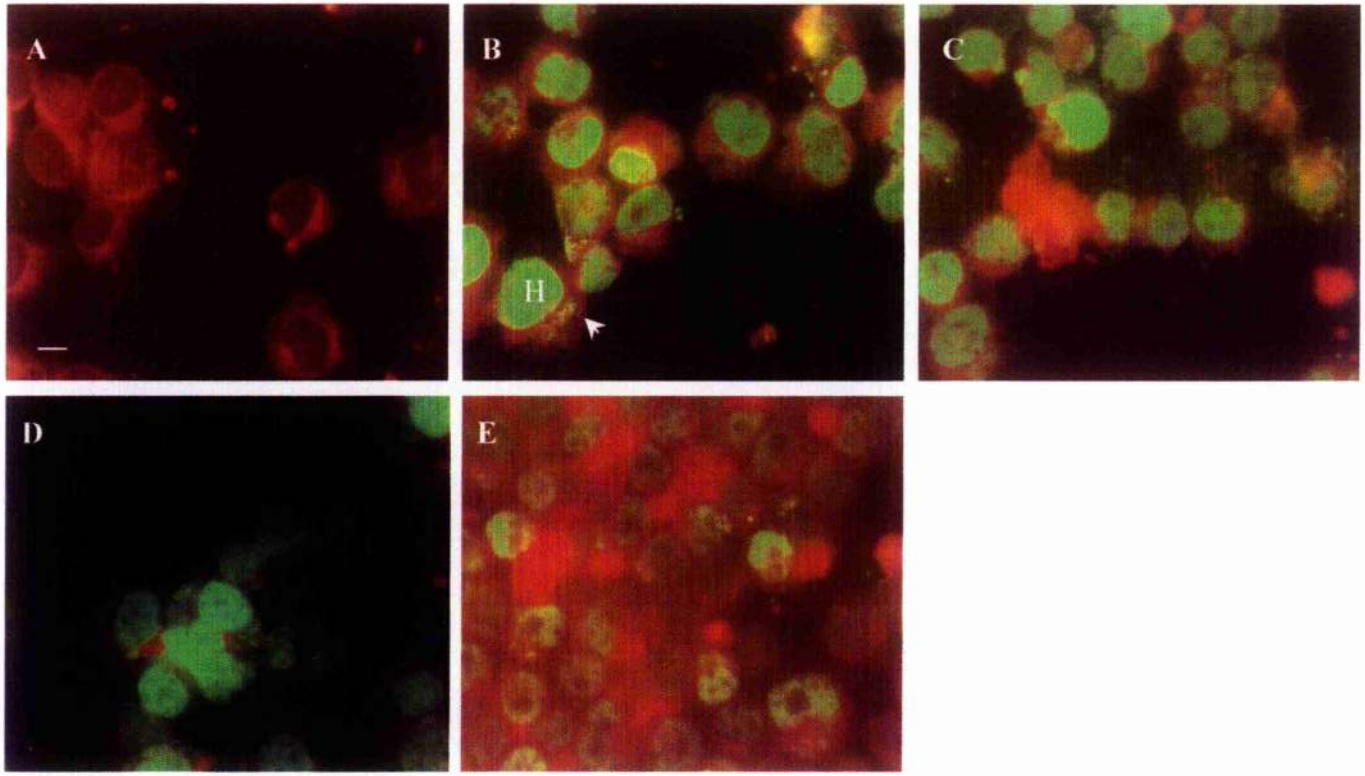


Fig. 5.19: IFAT analysis of D7B12 cells under differentiation time course conditions. Cells were counter-stained in red; antisera reactivity is shown in green. Panel A: reacted with pre-immune serum; panels B, C,D and E: reacted with anti-TashAT1/3 serum at day 0, 2, 4 and 6 respectively.
H: indicates host nucleus; arrow indicates a macroschizont. Bar= 10μm.

5.2.4 Localisation of the TashAT3 polypeptide during differentiation to the merozoite

Previous Western blot analysis identified a 180kDa polypeptide with anti-TashAT1/3 serum, but IFAT analysis revealed reactivity in the host nucleus and in the parasite. In order to determine where the 180kDa polypeptides was located during differentiation, Western blot analysis was performed with host and parasite enriched nuclei extracts over differentiation time course conditions in D7 and D7B12 cells. The results (Fig. 5.20A) identified five main bands at approximately 55kDa, 66kDa, 71kDa, 80kDa and finally 180kDa polypeptide. The size of the latter polypeptide was calculated from the deduced position of the 205kDa molecular weight marker, derived from a double log graph of the standard molecular weight markers because it was not detected by ponceau staining. The 180kDa polypeptide was enriched in the D7B12 host nuclear extracts at day 0 and day 9. By contrast the 180kDa polypeptide was absent in whole and parasite extracts at both time points, although no reactivity was detected in the parasite extracts at day 0, so this result would have to be repeated with a control for parasite reactivity. The levels of the 180kDa polypeptide in the host nuclear fraction had decreased from day 0 to day 9: this was consistent with the ponceau stained filter (data not shown), which showed reduced protein concentration in this lane.

Analysis of the 180kDa polypeptide in D7 cells (Fig. 5.20B) also showed an enrichment in host nuclear extracts at day 0 and day 2, but this band disappeared after day 2. The decrease in 180kDa polypeptide levels occurred at an early stage in D7 cells compared to D7B12 cell, where this polypeptide was still detected at day 9. The profile of the 180kDa polypeptide was also similar to the equivalent band in whole cell extracts by Western blot analysis (Fig. 5.16) and to the host nuclear reactivity by IFAT (Fig. 5.18), although residual reactivity was also seen in the host nucleus at days 4 and 6 by IFAT. The lack of reactivity with anti-TashTA1/3 serum at day 4 and day 6 in Fig. 5.20B may be due to the relative insensitivity of Western blot analysis compared to IFAT.

In order to confirm nuclear reactivity and verify the purity of the host and parasite fractions in Fig. 5.20, IFAT analysis was performed with enriched nuclear extracts. The results, seen in Fig. 5.21, revealed that the host nuclei are relatively pure, particularly in D7B12 extracts but do contain some parasite material. There appeared to be a lot of host cell debris in the parasite fractions of both D7 and D7B12 extracts, but very few fluorescent host nuclei were observed in the parasite extract. Nonetheless it is probable that more host material was

present in the parasite fraction. The parasite fractions of D7 and especially D7B12 extracts showed fluorescent particles that could be macroschizont nuclei. Although this would need further verification by counter staining with DAPI, this might suggest that anti-TashAT1/3 serum reacted against the macroschizont nuclei. The apparently low level of contamination in the host cell fraction indicates that the Western blot results are likely to be reliable with respect to host nuclear enrichment.

In summary, Western blot analysis indicated that the 180kDa polypeptide was located in the host nuclei and its levels decreased during merogony, consistent with previous data. IFAT analysis also showed reactivity from extracts enriched for host nuclei: there was very little contamination of host and parasite nuclei fractions from each other, indicating the results of Western blot analysis were reliable.

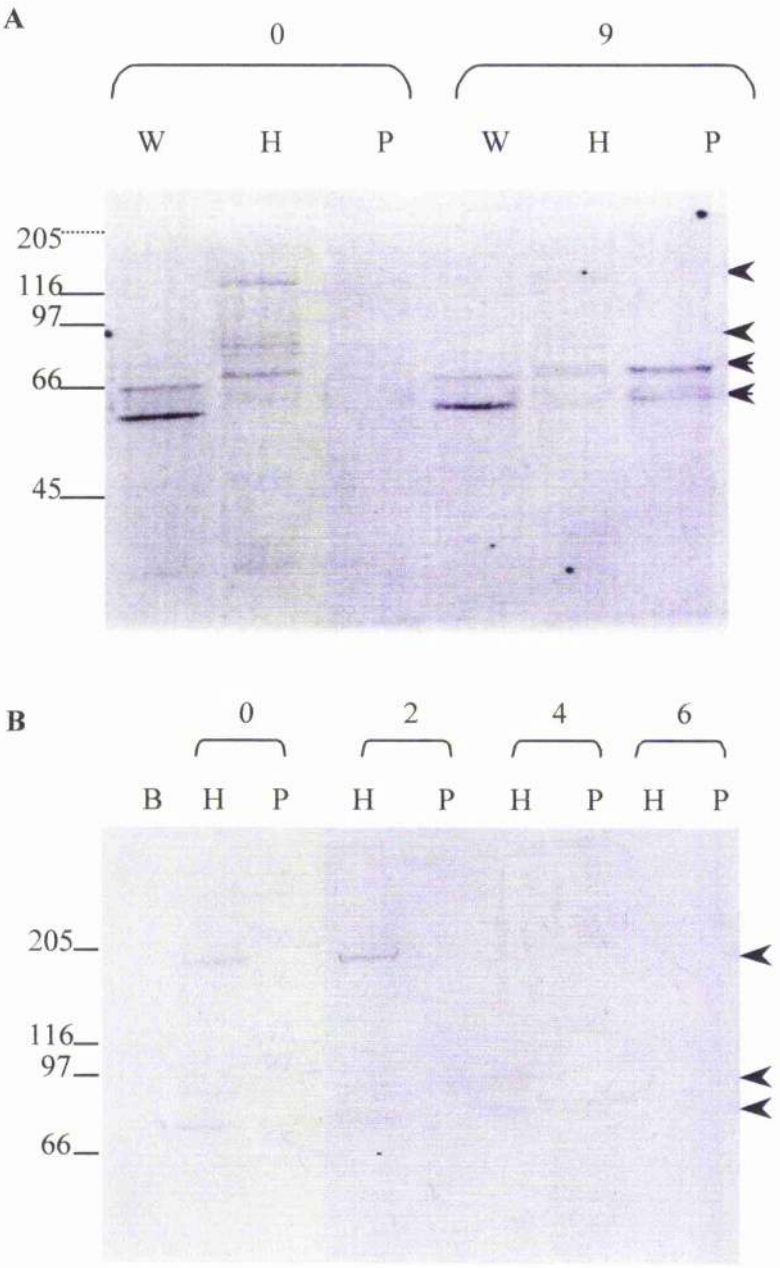


Fig. 5.20 : Western blot analysis of host and parasite enriched nuclear extracts from *T. annulata* infected leukocytes reacted with anti-TashAT1/3 serum during a differentiation time course. Panel A: D7B12 cell extracts at day 0 (0) and day 9 (9). Panel B: D7 cell extracts from day 0 to day 6. B: BL20 cell extract; W, H and P: whole cell, host and parasite enriched nuclear extracts, respectively. Numbers indicate molecular weight markers (in kDa); dotted line indicates deduced marker position. Arrow heads indicate the polypeptides detected.

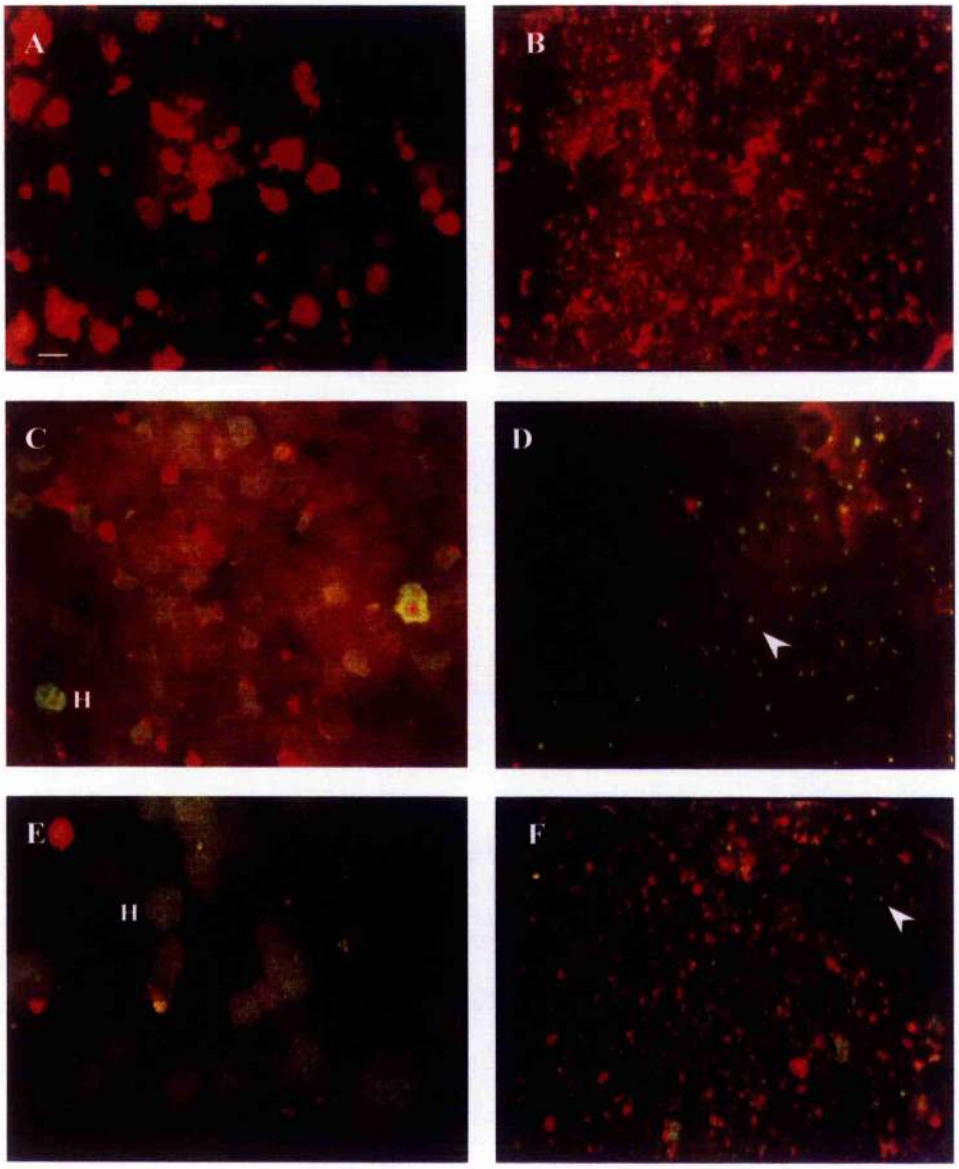


Fig. 5.21: IFAT analysis of host and parasite enriched nuclear fractions at 37°C from macroschizont infected leukocytes. Cellular material was counter-stained in red; antisera reactivity is shown in green. Panels A and B: D7 host and parasite nuclei respectively with control pre-immune serum. Panels C and D: D7 host and parasite nuclei respectively using anti-TashAT1/3 serum. Panels E and F: D7B12 host and parasite nuclei respectively with anti-TashAT1/3 serum. Arrow indicates parasite nuclei. H indicates a host nucleus. Bar=10μm.

5.2.5 TashAT1 and TashAT3 expression in sporozoites

Previous investigations by Swan *et al.* (2001a) revealed that TashAT1 and TashAT3 were expressed at the mRNA level in sporozoites. As IFAT analysis appeared to be the most reliable method to detect TashAT3 and possibly TashAT1, *T. annulata* sporozoites were subjected to IFAT analysis to determine if TashAT3 or TashAT1 were expressed at this stage (see Fig. 5.22).

The results showed no reactivity in the *T. annulata* sporozoite infected tick salivary gland preparation with the anti-TashAT1/3 serum (Fig. 5.22, panel A). To identify the sporozoite nuclei, the infected tick preparation was also counter-stained with DAPI (Fig. 5.22, panel C). A composite image was prepared from the images in panel A and C (Fig. 5.22, Panel E), which confirmed that the non-reactive material corresponded to the sporozoite nuclei. In contrast, the positive control, monoclonal antibody 1A7 (Boulter *et al.*, 1994), which detects the sporozoite surface antigen, SPAG1, showed reactivity in panel B (Fig. 5.22) that corresponded to the DAPI-stained sporozoite nuclei (see Fig. 5.22, panels D and F). Assuming the IFAT reactivity detected in macroschizont infected cells is caused by recognition of TashAT1 or TashAT3, this indicated that the TashAT3 or TashAT1 polypeptide might not be expressed at this stage.

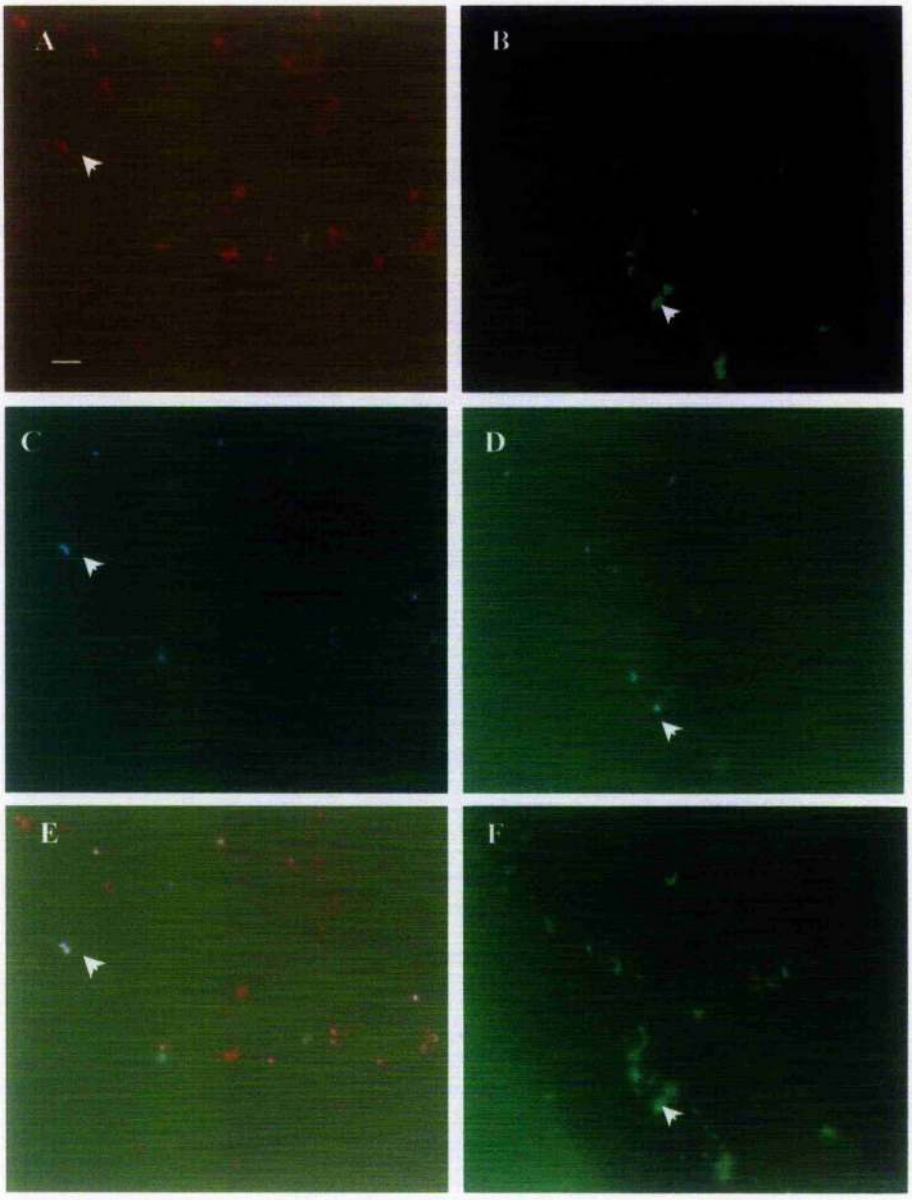


Fig. 5.22: IFAT analysis of *T. annulata* sporozoites using anti-TashAT1/3 serum. Sporozoites were counter-stained in red; antisera reactivity is shown in green. Panel A shows reactivity against anti-TashAT1/3 serum; panel C shows the corresponding DAPI stained nuclei (blue); panel E is a composite image of panels A and C. Panel B shows reactivity against mAb, 1A7, with the corresponding DAPI stained nuclei in panel D; F represents the composite image of panels B and D.Arrow indicates the sporozoites and their corresponding nuclei. Bar=10μm.

5.3 Discussion

In order to study the expression of the TashAT1 polypeptide, antisera was generated to a unique region of the gene. One of these antisera detected at least three polypeptides at approximately 66kDa, 71kDa and 180kDa in both D7 and D7B12 cells. However, the 66kDa and 71kDa polypeptides were eliminated as candidates for TashAT1 when they were detected in two cell lines, (Ta46A and TBL20) which lacked the *TashAT1* gene. It seems likely that these polypeptides are derived from the host as they were also detected in BL20 cell extract. It is possible that the 58kDa polypeptide detected by antiserum DE39 may relate to TashAT1, but this would require further verification as DE39 detected a large number of polypeptides, some of which were of the same size. In addition, a 55kDa polypeptide was also detected in Ta46A and TBL20 cells.

Based on size considerations, detection of the polypeptide at 180kDa was likely to be due to cross recognition of a distinct polypeptide with related epitopes rather than specific recognition of the TashAT1 protein. Given the sequence divergence between *TashAT1* and *TashAT2* over the region used to generate the fusion protein it was concluded that recognition of TashAT2 was not represented by the 180kDa polypeptide. This was confirmed by the failure of serum 107 to detect BoMAC cells transfected with the *TashAT2* gene. However, following the production of the antisera it became apparent that TashAT3 was a likely candidate for the 180kDa band as the TashAT1-N polypeptide was later found to be 100% identical to the N –terminal section of TashAT3. Moreover, the profile of this molecule closely matches the profile of TashAT3 mRNA with respect to differentiation. However, it is noticeable that there is a size difference in the predicted size of TashAT3 and the actual size. This difference may be only be accounted for by aberrant migration of this polypeptide through the SDS-PAGE gel or that TashAT3 may be post-translationally modified. A more remote possibility is that the 180kDa molecule was in fact a related TashAT cluster polypeptide. Confirmation that this molecule is TashAT3 could be achieved by immunoprecipitation followed by peptide sequence analysis (Zalut *et al.*, 1980) or mass spectrometry (Ross *et al.*, 2002).

Western blot analysis with host nuclear and parasite enriched extracts together with IFAT on TBL20 cells provided some evidence that reactivity in the host nucleus was due to the 180kDa polypeptide, the most likely candidate for TashAT3. When host and parasite enriched fractions of macroschizont infected cell extracts were analysed by Western blot analysis, the 180kDa polypeptide was found only in the host nuclear enriched fraction.

IFAT analysis also showed reactivity in the host nucleus as well as the macroschizont in D7 and D7B12 cells. However, the strongest evidence that detection of TashAT3 and not TashAT1, was probably responsible for the host nuclear reactivity was the fact that fluorescence was observed in TBL20 host nuclei, a cell line which lacked the *TashAT1* gene, but expressed TashAT3 mRNA. No reactivity was observed in BL20 control cells from the same host lineage, reducing the possibility that cross reactivity of a host polypeptide was responsible. Moreover, the weak host reactivity conforms to the weak detection of TashAT3 mRNA in the TBL20 cells (Swan *et al.*, 2001a). Thus, the reactivity in the host nucleus suggests that TashAT3 is translocated to the host nucleus. If TashAT3, like TashAT2 is translocated to the host cell nucleus, then it seems likely that it's function is to bind host DNA and modulate host cell gene expression. However, a more remote possibility is that host reactivity could be due to a further gene related to the *TashAT* cluster, also translocated to the host nucleus. To verify that TashAT3 is responsible for host cell reactivity, a *TashAT3* construct could be transfected into BoMAC cells to prove the polypeptide has the ability to enter the host nucleus. However, the development of a transfection system in *Theileria* infected cells would provide conclusive evidence that TashAT3 is transported into the host cell nucleus if performed in a cell line that lacks the *TashAT3* gene.

The identity of the polypeptide(s) responsible for the IFAT reactivity against the macroschizont in D7 was not clear since the reactivity detected in the macroschizont of infected cells with anti-TashAT1/3 serum could not be attributed to any of the polypeptides detected by Western blot analysis. The 66kDa and 71kDa polypeptides were eliminated as possible candidates as they were present in cell lines that do not contain the *TashAT1* gene and were also present in BL20 cells. The 180kDa TashAT3 polypeptide was unlikely to be responsible as it was found to be preferentially localised to the host cell nucleus. Time course studies of D7 cells by IFAT, showed that macroschizont reactivity was switched off just prior to merogony and was barely detectable at the day 2. It seemed possible that the polypeptide responsible for the macroschizont reactivity in D7 and D7B12 cells might be TashAT1 but is not recognised by anti-TashAT1/3 serum by Western blot analysis, possibly masked by the presence of other polypeptides that are detected by anti-TashAT1/3 serum, of the same size. This theory is supported by the fact that macroschizont reactivity is not detected in cell lines that lack the *TashAT1* gene, in contrast to that observed in D7 and D7B12 cells with anti-TashAT1/3. However it is possible that macroschizont reactivity in D7 and D7B12 cells by IFAT is caused by a distinct but related polypeptide to TashAT1, which is present at reduced levels in Ta46A and TBL20 cells. Indeed, there is evidence of

other genes with similarity to the N-terminal region of *TashAT1*, which are now being identified by Shiels and McKellar (2000). A more remote option is that macroschizont reactivity in D7 and D7B12 cells was due to a host factor translocated to the parasite. This possibility has been reported by Carrington *et al.* (1995), but is highly unlikely in this case as no reactivity was detected against anti-TashAT1/3 in BL20 cells.

If the IFAT reactivity against the macroschizont is not due to TashAT1, the likely absence of a TashAT1 polypeptide in Western blots could indicate that *TashAT1* is a pseudogene, which may be transcribed but not translated. Some evidence to support this theory comes from previous work by which revealed that *TashAT1* is unlikely to be an essential gene due to its absence in certain cell lines (Swan *et al.*, 2001a; Stadler, unpublished. 2000). The fact that *TashAT1* is almost identical to the N- terminus of *TashAT3* might suggest that *TashAT3* is a functional substitute for TashAT1. To identify a TashAT1 polypeptide further studies could involve 2 dimensional gel electrophoresis followed by Western blot analysis to determine the approximated size and isoelectric point of the polypeptide. Future studies could also involve transfection of Ta46A cells (lacking the *TashAT1* gene) with an epitope tagged-*TashAT1* construct, followed by IFAT and Western blot analysis. This would confirm the existence of the TashAT1 polypeptide using anti-TashAT1/3 serum.

Western blot analysis of the TashAT3 polypeptide revealed that this polypeptide was also down regulated from day 2 to day 6, during differentiation to the merozoite, in a similar fashion to its cognate mRNA. Interestingly, the increased levels of TashAT3 polypeptide is also reflected by increased levels of TashAT3 mRNA in D7B12 cells, between day 0 and day 2. By contrast, no decrease was observed for this polypeptide by Western blot analysis or IFAT in the attenuated D7B12 cells, although the levels of this polypeptide did fluctuate. This could be explained by previous, unpublished observations by Shiels (2001) that show the levels of polypeptides in D7B12 can vary when these cells are cultured at 41°C. Therefore, down regulation of TashAT3 polypeptide may occur by reduction in mRNA production or stability. The early down regulation of TashAT3 during differentiation suggests that the TashAT3 polypeptide may be associated with differentiation and maintenance of the macroschizont infected cell, either by blocking differentiation in D7B12 cells or by stimulating cytokinesis.

The elevated levels of TashAT3 in D7B12 cells might indicate that its expression is linked with attenuation. This does not seem likely, as low levels of TashAT3 mRNA were detected in other attenuated cell lines (Swan *et al.*, 2001a). A more likely explanation for

the high and stable levels of TashAT3 in D7B12 cells is that an unknown and variable alteration occurred when the D7B12 cell line was cloned that caused changes to TashAT expression. One possible outcome of such variable expression of *TashAT* genes in different *Theileria* infected cloned cell line, is that it could lead to changes in bovine gene expression and molecular phenotypes that have previously been described for these cells (Preston *et al.*, 2001, 1998; Sutherland *et al.*, 1996; Somerville *et al.*, 1997, 1998b; Adamson *et al.*, 2000ab; Oura *et al.*, 2001). To obtain more quantitative data on how TashAT polypeptide levels vary over a differentiation time course, their levels could be compared by densitometric analysis, with those of a constitutively expressed polypeptide using the same samples. In addition, the microtubule inhibitor nocodazole, which induces (reversible) cell cycle arrest in the infected cell (Baumgartner *et al.*, 1999), could measure the accumulation of TashAT polypeptides compared to a control over a time point and therefore determine the changes in TashAT expression with respect to parasite differentiation.

The detection of TashAT1 and TashAT3 mRNA in the sporozoite material raises the possibility that these genes are involved in sporoblastogenesis in the tick or in the initial establishment of the parasite once the sporozoite has invaded the host cell. However, the preliminary IFAT results showed no reactivity in sporozoites with anti-TashAT1/3 serum, although these results would have to be confirmed with pre-immune controls for 1A7 mAb or by Western blot analysis. It could be postulated, on the basis of these preliminary findings, that sporozoite specific post-translational modification of these polypeptides could mask detection by anti-TashAT1/3 serum or there may be problems with the sporozoite material tested. However, another possibility could be that these mRNA transcripts are transcribed by the sporozoite but stored in a stable form in preparation for their rapid translation following invasion of the host cell. Alternatively, their expression may be confined to sporogony within the salivary gland of the tick. It would be of interest to test for the presence of TashAT1/3 polypeptides, by IFAT, immediately following invasion and differentiation to the trophozoite stage. In addition, other experiments could analyse infected salivary glands to determine, in particular, if the *TashAT1/3* gene products are transported to the nucleus of the tick host cell.

The finding that TashAT2 and possibly TashAT3 are transported out of the parasite, and may be processed and presented on the surface of the infected cell make them a potential target(s) for a schizont vaccine. Moreover, TashAT1/3 contain PEST sequences that are capable of targeting a polypeptide for proteolysis. There is evidence that PEST sequence-

encoding proteins give rise to immunogenic peptides presented on MHC I molecules (Rechsteiner and Rogers, 1996 and references therein). This possibility could be verified by evaluating the T-cell responses of infected cattle to recombinant TashAT2 or TashAT3 antigens, as described for *Toxoplasma* antigens (Fatoohi *et al.*, 2002). Swan *et al.* (2001a) demonstrated that TashAT2 is common to all macroschizont infected cell lines tested, and would therefore be the best candidate for such a vaccine. So far these genes appear to be highly conserved, eliminating the problem of immune evasion, but further work would be required to show that any protective epitopes are not polymorphic in other *T. annulata* stocks. Any potential cross-reactivity with bovine polypeptides, as detected by anti-TashAT1/3 against BL20 extract (Fig. 5.8) would have to be checked in order to prevent the generation of antibodies against bovine polypeptides. In theory, given these tests, inoculation of the host with a live Attenuated Macroschizont Vaccine (AMV) should elicit an immune response against the TashAT polypeptide fragments on the surface of the infected cell, to protect the host against any future infection by *Theileria* sporozoites. Indeed, previous AMV trials have been successful (reviewed by Shkap and Pipano, 2000) and in one such trial, all AMV immunised cattle were resistant to field challenge, whilst 50% of the non-immunised cattle developed tropical theileriosis (Viseras *et al.*, 1997).

.

6. General Discussion

Infection of bovine leukocytes by the protozoan parasite, *T. annulata* reversibly induces the host cell to transform into a neoplastic-like phenotype (Brown, 1990). The transformation is associated with, and almost certainly involves, changes to host cell gene expression by the parasite (Adamson *et al.*, 2000a; Hall *et al.*, 1999; Sutherland *et al.*, 1996; Oura *et al.*, 2001). In contrast, the parasite factors that induce host cell division have not been identified, although it has been proposed that they are likely to be down regulated during differentiation to the merozoite (Carrington *et al.*, 1995). Thus examination of down regulated macroschizont genes, could provide an insight into the parasite-encoded mechanism that modulates lymphoproliferation (Carrington *et al.*, 1995). Previously, Swan *et al.* (1999) identified a small family of genes, whose expression is down regulated during merogony. One member of this family, *TashAT2*, encodes a protein that was found to be localised to the host's nucleus and encodes DNA binding AT hook motifs. It was therefore postulated to be involved in controlling bovine gene expression and possibly the induction of host cell proliferation.

6.1 Analysis of the TashAT1/3 Predicted Polypeptide

One of the primary aims of this project was to clone and characterise the second member of the *TashAT* gene family, *TashAT1*. The results revealed a predicted TashAT1 polypeptide encoded by a single 1.4kb open reading frame. The *TashAT1* gene was found to be located within a small, gene rich cluster containing *TashAT2* and a third *TashAT* gene, *TashAT3*. Two identical, macroschizont specific genes were also identified adjacent to the 5' and 3' ends of *TashAT1* (Swan *et al.*, 2001a; Stadler, 2000, unpublished). Sequence analysis showed that, like *TashAT2*, *TashAT1* was predicted to encode AT hook domains, a signal peptide sequence and nuclear localisation signals, which indicated that the TashAT1 polypeptide may be a DNA binding factor that is transported from the parasite to the infected host cell nucleus. A further feature common to the predicted TashAT1 and TashAT2 polypeptides was the large number of predicted phosphorylation sites that might suggest some form of post-translational control of the TashAT1 polypeptide. Such modifications are known to be involved in regulation of transcriptional activation, such as cyclin dependent kinases, which were found to regulate the AT hook encoding HMGI(Y) molecules via positive and negative phosphorylation. One such example is cdc2 kinase, which modulates the Human HMGI(Y) protein DNA binding in the G2/M-phase of the cell

cycle (Nissen *et al.*, 1991; Reeves *et al.*, 1991). The presence of a putative trancriptional transactivation domain in TashAT1 and TashAT3 provides further evidence for possible protein-protein interactions with other transcription factors. Certainly, these domains are present in a number of transcription factors including the HMG domain-encoding Sox proteins such as LEF-1, and the POU domain proteins, e.g. Oct-1, which interact in a number of combinations to regulate metazoan embryogenesis (Dailey and Basilico, 2001).

6.2 *TashAT* Gene Organisation

Striking similarities between the *TashAT* genes were identified at the predicted polypeptide level, particularly between *TashAT1* and *TashAT3*, in which the five prime region of *TashAT3* was almost identical to *TashAT1*. The predicted polypeptide sequence of TashAT3 was also virtually identical with that of TashAT2 over base pairs 1320-2715 of *TashAT3* and 1386-2781 of *TashAT2* (Swan *et al.*, 2001a). The close proximity of these genes and their sequence homology, taken together, suggested a recent gene duplication event in this cluster (see Fig. 3.30). This is supported by the identification of the two identical ORFs (TashHN) that flank *TashAT1* and the demonstration of parasite genomes that lack *TashAT1*. It is reasonable to suggest that of the *TashAT* genes, *TashAT1* is the most recent copy and possibly originated from a gene duplication event involving *TashAT3* and *TashHN*. It is unclear what advantage this gene duplication event might confer, but given the lack of identification of a TashAT1 polypeptide it could be that this advantage is conferred by an extra copy of *TashHN*. The sequence data, location of the *TashAT2* polypeptide in the hosts nucleus and duplication of these genes suggests that the *TashAT* cluster has an important function for the parasite. However, a comparison of the predicted TashAT1 polypeptide with the preliminary sequence of the most related species, *T. parva,* did not retrieve a *TashAT* homologue with a conserved AT hook domain. The lack of AT hook motifs in the *T. parva* genome compared to *T. annulata* may, in part, be due to differences in the cell type that each parasite infects which, in turn, may influence the mechanism each parasite species deploys to achieve a transformed phenotype. It is of interest, therefore, that while *T. annulata* can infect and transform bovine dendritic cells (Stephens and Howard, 2002), *T. parva* is unable to induce host cell immortalisation, even though it has been shown to invade the cell and form macroschizonts (Wells and McKeever, 1998; Shaw *et al.*, 1993). Thus it might be that the different parasite species are limited by interactions with different subsets of cell specific gene targets or cell specific accessory factors.

6.3 TashAT mRNA Identification

While the *TashAT1* gene may not be present in all *T. annulata* genomes, it was shown to be expressed at the mRNA level in certain infected cell lines. Two *TashAT1* derived probes detected a 2.1kb transcript, which was closest to the predicted size of the TashAT1 mRNA transcript. The probe containing AT hook encoding motifs, also detected two larger transcripts at 3.6kb and 4.0kb which corresponded approximately to the size predicted for *TashAT3* and identified for *TashAT2*, respectively. Given this, it seems likely that the 2.1kb transcript represents *TashAT1*; however, isolation of the corresponding TashAT1 cDNA would be required to demonstrate that *TashAT1* is represented by the 2.1kb transcript. In addition, nuclear run on experiments with a judicious choice of probes is necessary to determine if the *TashAT1* gene is transcribed in a monocistronic or polycistronic fashion. However, it would seem likely that the former case is most likely based on studies of apicomplexan parasite genes, including *T. annulata*, which have all demonstrated that transcription is regulated in a monocistronic fashion (Lanzer *et al.*, 1993; Horrocks *et al.*, 1998; Jean *et al.*, 2001; Shiels *et al.*, 2000).

6.4 TashAT mRNA Expression

In this study, the profile of all three TashAT mRNA species all showed early negative regulation (between day 2 and day 4) of parasite differentiation towards the merozoite. Research has shown that this period corresponds with a decrease in host cell division, and also coincides with marked changes to the expression of stage specific genes (Shiels *et al.*, 1994). Swan *et al.* (1999) found that the reduction in TashAT mRNA levels was earlier in comparison to other macroschizont genes and suggested that early down regulated macroschizont genes could encode regulatory factors involved in the maintenance of the macroschizont status, perhaps by regulating host cell division and/or parasite differentiation. In the D7B12 cell line, which is derived from the same lineage as D7 cells but is severely attenuated for differentiation, there appears to be greater amount of TashAT3 mRNA compared to TashAT2. By contrast, the levels of TashAT2 mRNA exceed TashAT3 mRNA levels in D7 cells. This indicates that *TashAT2* and *TashAT3* may be differentially expressed at the mRNA level at least, and could be functional substitutes. The causes of these alterations might have occurred after the cell line was cloned and such *in vitro* modifications have been widely reported (Hall *et al.*, 1999; Adamson *et al.*, 2000ab; Sutherland *et al.*, 1996; Somerville *et al.*, 1998b). This might account for the

altered host cell gene expression, such as MMP9, observed in cell lines of the same lineage (Adamson *et al.*, 2000ab; Sutherland *et al.*, 1996; Somerville *et al.*, 1998b). Another possibility is that the increased levels of TashAT3 mRNA detected in the severely attenuated D7B12 cell line are associated with negative control of parasite differentiation in *Theileria* infected cell lines. However, at present, there is no evidence of a direct correlation between TashAT3 mRNA levels and the ability of infected cells to differentiate and further experiments would be required to establish such an association.

6.5 *TashAT* Gene Regulation

Sequence analysis of the upstream regions of *TashAT1* with *TashAT3*, revealed an identical upstream sequence motif (TashUM) at position -45 from the putative transcription start site. Similar TashUM-like motifs were also found upstream of *TashAT2* and an unrelated macroschizont gene, *Tash1*. Such conservation suggests an important functional role for TashUM. EMSA analysis of nuclear extracts and a double stranded oligonucleotide containing the TashUM sequence, revealed binding with a factor that was enriched in parasite-enriched extracts but was also present in host-enriched nuclear extracts. Analysis of parasite extracts derived from cells that were undergoing parasite differentiation suggested that the nuclear factor was down regulated with respect to merogony. It is possible that this factor represents a parasite-encoded transcription factor that binds to TashUM to modulate *TashAT1* expression at the macroschizont stage, although the results obtained cannot exclude that it is a host factor transported to the parasite (Carrington *et al.*,1995). Further experiments with good controls for binding specificity and host or parasite enrichment are required to complete this work.

EMSA analysis of TashUM probe using polydG.dC and dI.dC revealed that the TashUM binding factor had an affinity for AT rich DNA. This fact and the finding that TashAT2 binds to AT rich DNA sequences (Swan *et al.*, 2001a) raises the possibility that that one or more of the TashAT polypeptides bind to TashUM in some form of self regulation. Experiments could attempt to isolate and characterise this factor by purification through a column containing the TashUM oligonucleotide and determine it's origin, (host or parasite) and identity by proteomic techniques. To investigate whether the binding factor is in fact a TashAT polypeptide(s), binding assays using TashAT1 and TashAT2 fusion proteins and/or supershift assays using antisera generated against the TashAT polypeptides could be performed.

It is possible that the factor(s) binding to the TashUM motif might be at least in part, responsible for the differential expression of *TashAT2* and *TashAT3*. This is because work by Swan and Phillips (unpublished, 2001) failed to obtain a band shift in D7 nuclear extracts using a probe derived from the TashUM-like sequence, upstream of *TashAT2*. These results indicate that each *TashAT* gene may be under separate control, but that this control must be under some form of co-ordination during differentiation to the merozoite. It would be interesting to perform further EMSA analysis of the five prime region of *TashAT1* and *TashAT2* to discover possible common and gene specific control elements that could be involved in transcriptional regulation of these genes. In the case of *TashAT1* and *TashAT3*, which have identical upstream sequences, differential expression is likely to be achieved via other mechanisms, possibly mediated by the 3' untranslated region (Hotz *et al.*, 1997; Furger *et al.*, 1997) or some other form of post-translational control. One such method could be by targeted destruction, possibly via the potential PEST sequence within TashAT1/3, which is capable of targeting proteins for proteolysis (Rechsteiner and Rogers, 1996). Serine or Threonine phosphorylation is known to activate conditional PEST sequences for destruction (Rechsteiner and Rogers, 1996). As TashAT1/3 have many potential phosphorylation sites, it could be envisaged that TashAT1/3 levels are modulated by differential phosphorylation. One such candidate could be Casein Kinase II (CKII), which can potentially phosphorylate TashAT1/3 and was found to be elevated in *T. parva* infected cells (Ole-MoiYoi, 1995). Moreover, CKII has been shown to phosphorylate HMGI(Y) proteins *in vivo* (reviewed by Reeves *et al.*, 2001).

6.6 Identification and Expression of the TashAT1 and TashAT3 Polypeptides

Attempts to identify the TashAT1 polypeptide proved inconclusive: several polypeptides close to 55.3kDa, (the predicted size of the TashAT1 polypeptide) were detected, however these polypeptides were discounted as they were detected in a cell line that lacked the *TashAT1* gene and in the non-infected BL20 cell line. However, the IFAT data suggested that TashAT1 could be responsible for macroschizont reactivity in D7 and D7B12 cells as no macroschizont reactivity was detected in cell lines that lacked *TashAT1* gene. It is possible that the *bona fide* TashAT1 polypeptide is the same size as these cross recognised polypeptides and so cannot be discriminated by Western blot analysis. This possibility could be verified by performing immunoblot analysis following by 2 dimensional (2-D) electrophoresis, or 2-D analysis and protein sequencing of polypeptides obtained by

immunoprecipitation with the anti-TashAT1 serum. Alternatively, anti-TashAT1/3 serum could be used to determine reactivity against an *in vitro* translation TashAT1 product derived from TashAT1 mRNA (Pelham and Jackson, 1976).

The antisera generated against TashAT1 also detected a polypeptide at 180kDa, which corresponded approximately to the predicted size of TashAT3 (113kDa). This polypeptide was found to be enriched in host cell nuclei by IFAT and Western blot analysis and was down regulated during time course experiments when the parasite was undergoing differentiation to the merozoite. Anti-TashAT1/3 serum did not detect TashAT2 by any method so it was postulated that this polypeptide most likely represented TashAT3. During a time course experiment, reactivity against both subcellular compartments was reduced after day 2 at 41°C. The reactivity in the host nucleus was consistent with the presence and expression profile of TashAT3, but it was not clear whether the reactivity against the parasite was due to TashAT1 or a further macroschizont polypeptide related to TashAT1. Preliminary IFAT results showed no TashAT1/3 polypeptide reactivity against anti-TashAT1/3 serum in sporozoites, despite the presence of TashAT1/3 mRNA at that stage. The reasons for these are unclear, further experiments are required to confirm these results and establish their presence within the tick salivary gland or immediately following invasion of the sporozoite into the host leukocyte.

6.7 Possible Functions of the TashAT Polypeptides

Sequence analysis of all *TashAT* genes revealed that they had closest homology to HMGI(Y) proteins over the AT-hook encoding region. Studies of other AT hook encoding genes showed that these proteins bind to a wide range of genes. The HMGI(Y) proteins bind upstream of the cytokines IL-2Rα receptor and IL-4 (Chuvpilo *et al.*, 1993) whilst Sox-4 and LEF-1, containing HMG-binding domains, are involved in T and B cell differentiation by eliciting enhancer activity of CD2 and TCR-α gene (reviewed by Fitzsimmons and Hagman,1996). It has been widely reported that HMG proteins are known to bind to genes involved in cell growth e.g. the leptin promoter, the human insulin receptor (Brunetti *et al.*, 2001) and the viral VHP 18 protein (Bouallaga *et al.*, 2000). Thus, in *Theileria* infected cells, the function of the *TashAT* genes may be to target cytokine genes and/or their downstream targets to modulate host cell division, such as IL-2Rα, which is known to stimulate proliferation of host infected cells via the IL-2/IL-2R autocrine loop (Dobbelaere and Heussler, 1999). Inappropriate IL-2R expression and down

regulation of IL-4 mRNA have been shown to occur in the host's T-cell response to *T. annulata* infected cells, which promotes a Th1 skewed inappropriate host immune response (Dobbelaere and Heussler, 1999; Campbell and Spooner, 1999). It is possible that this process may be mediated by the TashAT polypeptides since the HMGI(Y) proteins are known to negatively regulate IL-4 expression (Chuvpilo *et al.*, 1993) and stimulate IL-2Rα gene expression.

The early negative regulation of the *TashAT* genes at the mRNA level in D7 cells during differentiation, but not in D7B12 cells, which, effectively cannot differentiate, raises the possibility that these genes might have a role in the negative regulation of macroschizont differentiation to the merozoite. If the reactivity detected in D7 macroschizont is in fact TashAT1, then this gene would be the likely candidate. Although TashAT1 mRNA levels do not decrease from day 0 to day 2 during differentiation, it is possible that down regulation of the TashAT1 polypeptide could be achieved at the protein level. Thus TashAT1 or a related TashAT polypeptide could regulate parasite gene expression in a stage specific manner.

A third possible role for the *TashAT* genes might be to modulate host cell proliferation or infected cell survival. This could be achieved by regulation of genes encoding polypeptides that function in these processes. There are a number of studies that show that aberrant expression of AT hook proteins can generate neoplasia. Abnormally high levels of HMGI(Y) have consistently been found to be associated with neoplastic cellular transformation in many forms of cancer and are now viewed as a diagnostic marker for neoplastic transformation (Giancotti *et al.*, 1989, 1991). The similarity between HMGI(Y) and the predicted TashAT proteins, coupled to the likely nuclear localisation of TashAT2 and TashAT3 suggests that the *TashAT* genes may be involved in host cell proliferation. Indeed preliminary evidence has shown that TashAT2 might be involved in cell division as transfection of the TashAT2-TashHN construct into BoMAC cells, stimulates cell division (Oura, unpublished, 2001). It is also conceivable that the *TashAT* genes could be involved in the initial establishment of the parasite within the host infected cells. For example, they may be involved in the expression of IFN-γ within the early stages of the parasite infected cell, and thereby stimulate the production of cytokines, such as IL-1 and TNF-α. These cytokines were shown to be directly correlated to infected cell proliferation, and could maintain the parasite within the host cell and/or and prevent apoptosis (Campbell and Spooner, 1999). This may account for the expression of TashAT1/3 mRNA at the sporozoite stage (Swan *et al.*, 2001a), and, based on preliminary results which show no

TashAT1/3 polypeptide in the sporozoite, these mRNA species might be rapidly translated shortly after invasion. Alternatively, it is possible that the TashAT proteins bind directly to NF-κB, which is instrumental in blocking host cell apoptosis in *T. parva* infected cells (Dobbelaere, *et al.*, 2000); as there are NF-κB binding sites within the HMGI(Y) molecule (Reeves, 2001).

6.8 Summary

In summary, the studies carried out in this thesis provide further evidence that the genome of *T. annulata* encodes polypeptides which contain motifs that allow their transport from the parasite to the host nucleus where they could bind to AT rich stretches of DNA. Experimental evidence that the TashAT3 polypeptide is located in the host nucleus was obtained. The relative levels of TashAT mRNA were shown to be altered in cells that have lost the ability to differentiate. It can be concluded that it is likely that the TashAT family of proteins are involved in controlling bovine gene expression of a macroschizont infected cell.

6.9 Future Work

Clearly there are a number of future directions of research on the *TashAT* genes. This would include transfection of a *TashAT* gene(s) construct into uninfected host cells and analysis on the effect it may have on host cell division and/or host cell gene expression, followed by identification of the target genes and the TashAT polypeptide binding domains. It will also be important to identify any bovine polypeptides that could bind to the TashAT polypeptides as this may assist in the identification of their function and mechanism of action. Ultimately, studies will need to address functionality of the TashAT proteins *in situ*. This would involved disrupting expression of the *TashAT* genes in the parasite either by gene knockout technology (Pellicer *et al.*, 1980) or by RNA gene silencing methodology (reviewed by Fire *et al.*, 1998). Such technology will enable further insight into how apicomplexan parasites manipulate their intracellular environments and allow identification of the molecular mechanisms involved. It can be predicted that the TashAT polypeptides are important molecules that perform this function in *T. annulata* infected cells.

List of References

Achaz, G., Netter, P., & Coissac, E. 2001, "Study of Intrachromosomal Duplications Among the Eukaryote Genomes", *Mol. Biol. Evol.* , vol. 18, no. 12, pp. 2280-2288.

Adalar, S., Gunay, M., & Alp, H. 1992, "Protection afforded by inoculation of 10^6 attenuated cells.", *Int. Symp. Parasite Immunology*, 7-9th October, 1992. Pendik, Istanbul, pp. 25-28.

Adamson, R. E. & Hall, F. R. 1996, "Matrix metalloproteinases mediate the metastatic phenotype of *Theileria annulata*-transformed cells", *Parasitol.*, vol. 113 (Pt 5), pp. 449-455.

Adamson, R. E., Ilhan, T., Wilkie, G. F., Preston, P. M., Brown, C. G., & Hall, F. R. 2000a, "*Theileria annulata*: reduced host MMP activity profile of long-term culture is transmitted as a parasite-encoded trait", *Exp. Parasitol.*, vol. 94, no. 1, pp. 45-47.

Adamson R, Logan M, Kinnaird J, Langsley G, Hall R. 2000b, "Loss of matrix metalloproteinase 9 activity in *Theileria annulata*-attenuated cells is at the transcriptional level and is associated with differentially expressed AP-1 species." *Mol. Biochem. Parasitol.*, vol. 106 (Pt 1), pp. 51-61.

Aggarwal, A., Kumar, S., Jaffe, R., Hone, D., Gross, M., & Sadoff, J. 1990, "Oral Salmonella: malaria circumsporozoite recombinants induce specific CD8+ cytotoxic T cells", *J. Exp. Med.*, vol. 172, no. 4, pp. 1083-1090.

Ahmed, J. S., Diesing, L., Oechtering, H., Ouhelli, H., & Schein, E. 1988, "The role of antibodies in immunity against *Theileria annulata* infection in cattle", *Zentralbl. Bakteriol. Mikrobiol. Hyg.*, vol. 267, no. 3, pp. 425-431.

Ahmed, J. S., Shayan, P., Hugel, F. U., Biermann, R., Ewald, C., Schein, E., & Gerdes, J. 1997, "Macroschizonts of *Theileria annulata* as vaccine and diagnostic tools", *Trop. Anim. Health Prod.*, vol. 29, no. 4 Suppl., pp. 128S-132S.

Adler, S. & Ellenbogen, V. 1935, "Observations on theileriosis in Palestine", *Arch. Inst. Pasteur*, vol. 13, pp. 451-471.

Aravind, L. & Landsman, D. 1998, "AT-hook motifs identified in a wide variety of DNA-binding proteins", *Nucleic Acids Res.*, vol. 26, no. 19, pp. 4413-4421.

Ayoubi, T. A., Jansen, E., Meulemans, S. M., & Van de Ven, W. J. 1999, "Regulation of HMGIC expression: an architectural transcription factor involved in growth control and development", *Oncogene*, vol. 18, no. 36, pp. 5076-5087.

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., Morzaria, S. P., & . 1988, "Bovine T cells, B cells, and null cells are transformed by the protozoan parasite *Theileria parva*", *Infect. Immun.*, vol. 56, no. 2, pp. 462-467.

Banks, G. C., Li, Y., & Reeves, R. 2000, "Differential in vivo modifications of the HMGI(Y) nonhistone chromatin proteins modulate nucleosome and DNA interactions", *Biochem.*, vol. 39, no. 28, pp. 8333-8346.

Bannai, H., Tamada, Y., Maruyama, O., Nakai, K., & Miyano, S. 2002, "Extensive feature detection of N-terminal protein sorting signals", *Bioinformatics.*, vol. 18, no. 2, pp. 298-305.

Barnett, S. F. 1963, "The biological races of the bovine (*Theileria*) and their host-parasite relationship.," in P. C. C. Garnham, A. E. Pierce, & I. Roitt, eds. *Immunity to Protozoa*, Blackwell, Oxford, pp. 180-195.

Barnett, S. F. 1977, "*Theileria*," in J. P. Kreier, ed. *Parasitic Protozoa*, 4 edn. Academic Press, New York, San Francisco, London, pp. 77-113.

Baumgartner, M., Tardieux, I., Ohayon, H., Gounon, P., & Langsley, G. 1999, "The use of nocodazole in cell cycle analysis and parasite purification from *Theileria parva*-infected B cells", *Microbes Infect.*, vol. 1, no. 14, pp. 1181-1188.

Baumgartner, M., Chaussepied, M., Moreau, M. F., Werling, D., Davis, W. C., Garcia, A., & Langsley, G. 2000, "Constitutive PI3-K activity is essential for proliferation, but not survival, of *Theileria parva*-transformed B cells", *Cell Microbiol.*, vol. 2, no. 4, pp. 329-339.

Baylis, H. A., Megson, A., Brown, C. G., Wilkie, G. F., & Hall, R. 1992, "*Theileria annulata*-infected cells produce abundant proteases whose activity is reduced by long-term cell culture", *Parasitol.*, vol. 105 (Pt 3), pp. 417-423.

Baylis, H. A., Megson, A., & Hall, R. 1995, "Infection with *Theileria annulata* induces expression of matrix metalloproteinase 9 and transcription factor AP-1 in bovine leucocytes", *Mol. Biochem. Parasitol.*, vol. 69, no. 2, pp. 211-222.

Beckerbauer, L., Tepe, J. J., Cullison, J., Reeves, R., & Williams, R. M. 2000, "FR900482 class of anti-tumor drugs cross-links oncoprotein HMG I/Y to DNA in vivo", *Chem. Biol.*, vol. 7, no. 10, pp. 805-812.

Ben Miled, L., Dellagi, K., Bernardi, G., Melrose, T. R., Darghouth, M., Bouattour, A., Kinnaird, J., Shiels, B., Tait, A., & Brown, C. G. 1994, "Genomic and phenotypic diversity of Tunisian *Theileria annulata* isolates", *Parasitol.*, vol. 108 (Pt 1), pp. 51-60.

Beniwal, R. K., Sharma, R. D., & Nichani, A. K. 2000, "Determination of duration of immunity of calves vaccinated with the *Theileria annulata* schizont cell culture vaccine", *Vet. Parasitol.*, vol. 90, no. 1-2, pp. 25-35.

Berk, A. J. & Sharp, P. A. 1977, "Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease-digested hybrids", *Cell*, vol. 12, no. 3, pp. 721-732.

Berlingieri, M. T., Manfioletti, G., Santoro, M., Bandiera, A., Visconti, R., Giancotti, V., & Fusco, A. 1995, "Inhibition of HMGI-C protein synthesis suppresses retrovirally induced neoplastic transformation of rat thyroid cells", *Mol. Cell Biol.*, vol. 15, no. 3, pp. 1545-1553.

Beutler, B. & Cerami, A. 1986, "Cachectin and tumour necrosis factor as two sides of the same biological coin", *Nature*, vol. 320, no. 6063, pp. 584-588.

Bhattacharyulu, Y., Chaudhri, R. P., Gill, B. S. 1975, "Studies on the development of *Theileria annulata* Dschunskowsky and Luhs, 1904 in the tick-Hyalomma anatolicum anatolicum Koch, 1844", *Ann. Parasitol. Hum. Comp.*, vol. 50, pp. 397-408.

Bishop, R., Allsopp, B., Spooner, P., Sohanpal, B., Morzaria, S., & Gobright, E. 1995, "*Theileria*: improved species discrimination using oligonucleotides derived from large subunit ribosomal RNA sequences", *Exp. Parasitol.*, vol. 80, no. 1, pp. 107-115.

Black, C. G., Wang, L., Hibbs, A. R., Werner, E., & Coppel, R. L. 1999, "Identification of the *Plasmodium chabaudi* homologue of merozoite surface proteins 4 and 5 of *Plasmodium falciparum*", *Infect. Immun.*, vol. 67, no. 5, pp. 2075-2081.

Bohne, W., Heesemann, J., & Gross, U. 1994, "Reduced replication of *Toxoplasma gondii* is necessary for induction of bradyzoite-specific antigens: a possible role for nitric oxide in triggering stage conversion", *Infect. Immun.*, vol. 62, pp. 1761-1767.

Botteron, C. & Dobbelaere, D. 1998, "AP-1 and ATF-2 are constitutively activated via the JNK pathway in *Theileria parva*-transformed T-cells", *Biochem. Biophys. Res. Commun.*, vol. 246, no. 2, pp. 418-421.

Bouallaga, I., Massicard, S., Yaniv, M., & Thierry, F. 2000, "An enhanceosome containing the JunB/Fra-2 heterodimer and the HMG-I(Y) architectural protein controls HPV-18 transcription.", *EMBO Rep.*, vol. 1, pp. 422-427.

Boulter, N., Knight, P. A., Hunt, P. D., Hennessey, E. S., Katzer, F., Tait, A., Williamson, S., Brown, D., Baylis, H. A., & Hall, R. 1994, "*Theileria annulata* sporozoite surface antigen (SPAG-1) contains neutralizing determinants in the C terminus", *Parasite Immunol.*, vol. 16, no. 2, pp. 97-104.

Boulter, N. R., Glass, E. J., Knight, P. A., Bell-Sakyi, L., Brown, C. G., & Hall, R. 1995, "*Theileria annulata* sporozoite antigen fused to hepatitis B core antigen used in a vaccination trial", *Vaccine*, vol. 13, no. 13, pp. 1152-1160.

Boulter, N. R. 1996, "Assessment of SPAG-1, a Sporozoite Surface Antigen of *Theileria annulata* as a Vaccine Candidate.", PhD thesis. University of York.

Boulter, N. R., Brown, C. G., Kirvar, E., Glass, E., Campbell, J., Morzaria, S., Nene, V., Musoke, A., d'Oliveira, C., Gubbels, M. J., Jongejan, F., & Hall, F. R. 1998, "Different vaccine strategies used to protect against *Theileria annulata*", *Ann. N.Y. Acad. Sci.*, vol. 849, pp. 234-246.

Boulter, N. & Hall, R. 1999, "Immunity and vaccine development in the bovine theilerioses", *Adv. Parasitol.*, vol. 44, pp. 41-97.

Bradford, M. M. 1976, "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding", *Anal. Biochem.*, vol. 72, pp. 248-254.

Brown, C. G., Stagg, D. A., Purnell, R. E., Kanhai, G. K., & Payne, R. C. 1973, "Infection and transformation of bovine lymphoid cells in vitro by infective particles of *Theileria parva*", *Nature*, vol. 245, no. 5420, pp. 101-103.

Brown, C. G. 1990, "Control of tropical theileriosis (*Theileria annulata* infection) of cattle", *Parassitologia*, vol. 32, no. 1, pp. 23-31.

Brun, R. & Schonenberger, M. 1981, "Stimulating effect of citrate and cis-Aconitate on the transformation of *Trypanosoma brucei* bloodstream forms to procyclic forms in vitro", *Z. Parasitenkd.*, vol. 66, no. 1, pp. 17-24.

Brunetti, A., Manfioletti, G., Chiefari, E., Goldfine, I. D., & Foti, D. 2001, "Transcriptional regulation of human insulin receptor gene by the high-mobility group protein HMGI(Y)", *FASEB J.*, vol. 15, no. 2, pp. 492-500.

Bullock, W. O., Fernandez, J. M., & Short, J. M. 1987, "XL1-blue: A High Efficiency Plasmid Transforming Reca Es", *Biotechniques*, vol. 5, no. 4, pp. 376-379.

Burridge, M. J. 1971, "Application of the indirect fluorescent antibody test in experimental East Coast fever (*Theileria parva* infection of cattle)", *Res. Vet. Sci.*, vol. 12, no. 4, pp. 338-341.

Burridge, M. J. & Kimber, C. D. 1972, "The indirect fluorescent antibody test for experimental East Coast Fever (*Theileria parva* infection of cattle). Detection of antibodies to schizonts in dried blood samples", *Z. Tropenmed. Parasitol.*, vol. 23, no. 3, pp. 328-331.

Bustin, M., Lehn, D. A., & Landsman, D. 1990, "Structural features of the HMG chromosomal proteins and their genes", *Biochim. Biophys. Acta*, vol. 1049, no. 3, pp. 231-243.

Bustin, M. & Reeves, R. 1996, "High-mobility-group chromosomal proteins: architectural components that facilitate chromatin function", *Prog. Nucleic Acid Res. Mol. Biol.*, vol. 54, pp. 35-100.

Büttner, D. W. 1967, "Fine structure of merozoites of *Theileria parva*", *Z. Tropenmed. Parasitol.*, vol. 18, no. 2, pp. 224-244.

Campbell, J. D., Brown, D. J., Glass, E. J., Hall, F. R., & Spooner, R. L. 1994, "*Theileria annulata* sporozoite targets", *Parasite Immunol.*, vol. 16, no. 9, pp. 501-505.

Campbell, J. D., Howie, S. E., Odling, K. A., & Glass, E. J. 1995, "*Theileria annulata* induces aberrant T cell activation in vitro and in vivo", *Clin. Exp. Immunol.*, vol. 99, no. 2, pp. 203-210.

Campbell, J. D., Nichani, A. K., Brown, D. J., Howie, S. E., Spooner, R. L., & Glass, E. J. 1997, "Parasite-mediated steps in immune response failure during primary *Theileria annulata* infection", *Trop. Anim. Health Prod.*, vol. 29 (Suppl 4), pp. 133S-135S.

Campbell, J. D., Nichani, A. K., Brown, D. J., Glass, E. J., & Spooner, R. L. 1998, "A stage-specific, parasite-induced, "window" of in vivo interferon-gamma production is associated with pathogenesis in *Theileria annulata* infection", *Ann. N.Y. Acad. Sci.*, vol. 849, pp. 152-154.

Campbell, J. D. & Spooner, R. L. 1999, "Macrophages behaving badly: infected cells and subversion of immune responses to *Theileria annulata*", *Parasitol. Today*, vol. 15, no. 1, pp. 10-16.

Carrington, M., Allsopp, B., Baylis, H. A., Malu, N.-M., Shochat, Y., & Sohal, S. 1995, "Lymphoproliferation caused by *Theileria parva* and *Theileria annulata*", in *Molecular Approaches to Parasitology*, vol. 12, J. C. Boothroyd & R. Komuniecki, eds., Wiley-Liss, New York, pp. 43-56.

Carson, C. A., Timms, P., Cowman, A. F., & Stewart, N. P. 1990, "*Babesia bovis:* evidence for selection of subpopulations during attenuation", *Exp. Parasitol.*, vol. 70, no. 4, pp. 404-410.

Carter, R. & Miller, L. H. 1979, "Evidence for environmental modulation of gametocytogenesis in *Plasmodium falciparum* in continuous culture.", *Bull WHO*, vol. 57(suppl. 1), pp. 37-52.

Chansiri, K., Kawazu, S., Kamio, T., Terada, Y., Fujisaki, K., Philippe, H., & Sarataphan, N. 1999, "Molecular phylogenetic studies on *Theileria* parasites based on small subunit ribosomal RNA gene sequences", *Vet. Parasitol.*, vol. 83, no. 2, pp. 99-105.

Chaudhri, S. S. & Subramanian, G. 1992, "Cell-mediated immune responses to sporozoites and macroschizonts of *Theileria annulata*", *Vet. Parasitol.*, vol. 41, no. 1-2, pp. 23-34.

Chaussepied, M. & Langsley, G. 1996, "*Theileria* transformation of bovine leukocytes: a parasite model for the study of lymphoproliferation", *Res. Immunol.*, vol. 147, no. 3, pp. 127-138.

Chaussepied, M., Lallemand, D., Moreau, M. F., Adamson, R., Hall, R., & Langsley, G. 1998, "Upregulation of Jun and Fos family members and permanent JNK activity lead to constitutive AP-1 activation in *Theileria*-transformed leukocytes", *Mol. Biochem. Parasitol.*, vol. 94, no. 2, pp. 215-226.

Chin, M. T., Pellacani, A., Wang, H., Lin, S. S., Jain, M. K., Perrella, M. A., & Lee, M. E. 1998, "Enhancement of serum-response factor-dependent transcription and DNA binding by the architectural transcription factor HMG-I(Y)", *J. Biol. Chem.*, vol. 273, no. 16, pp. 9755-9760.

Chizyuka, H. G. & Mulilo, J. B. 1990, "Methods currently used for the control of multi-host ticks: their validity and proposals for future control strategies", *Parassitologia*, vol. 32, no. 1, pp. 127-132.

Church, G. M. & Gilbert, W. 1984, "Genomic sequencing", *Proc. Natl. Acad. Sci. U.S.A*, vol. 81, no. 7, pp. 1991-1995.

Chuvpilo, S., Schomberg, C., Gerwig, R., Heinfling, A., Reeves, R., Grummt, F., & Serfling, E. 1993, "Multiple closely-linked NFAT/octamer and HMG I(Y) binding sites are part of the interleukin-4 promoter", *Nucleic Acids Res.*, vol. 21, no. 24, pp. 5694-5704.

Cmarik, J. L., Li, Y., Ogram, S. A., Min, H., Reeves, R., & Colburn, N. H. 1998, "Tumor promoter induces high mobility group HMG-Y protein expression in transformation-sensitive but not -resistant cells", *Oncogene*, vol. 16, no. 26, pp. 3387-3396.

Conrad, P. A., Kelly, B. G., & Brown, C. G. 1985, "Intraerythrocytic schizogony of *Theileria annulata*", *Parasitol.*, vol. 91 (Pt 1), pp. 67-82.

Conrad, P. A., Denham, D., & Brown, C. G. 1986, "Intraerythrocytic multiplication of *Theileria parva* in vitro: an ultrastructural study", *Int. J. Parasitol.*, vol. 16, no. 3, pp. 223-229.

Corpet, F., Gouzy, J., & Kahn, D. 1998, "The ProDom database of protein domain families", *Nucleic Acids Res.*, vol. 26, no. 1, pp. 323-326.

Cowman, A. F., Timms, P., & Kemp, D. J. 1984, "DNA polymorphisms and subpopulations in *Babesia bovis*", *Mol. Biochem. Parasitol.*, vol. 11, pp. 91-103.

Creemers, P. 1982, "Lack of reactivity of sera from *Theileria parva*-infected and recovered cattle against cell membrane antigens of *Theileria parva* transformed cell lines", *Vet. Immunol. Immunopathol.*, vol. 3, no. 4, pp. 427-438.

d'Oliveira, C., van der, W. M., Habela, M. A., Jacquiet, P., & Jongejan, F. 1995, "Detection of *Theileria annulata* in blood samples of carrier cattle by PCR", *J. Clin. Microbiol.*, vol. 33, no. 10, pp. 2665-2669.

d'Oliveira, C., Feenstra, A., Vos, H., Osterhaus, A. D., Shiels, B. R., Cornelissen, A. W., & Jongejan, F. 1997, "Induction of protective immunity to *Theileria annulata* using two major merozoite surface antigens presented by different delivery systems", *Vaccine*, vol. 15, no. 16, pp. 1796-1804.

Dailey, L. & Basilico, C. 2001, "Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes", *J. Cell Physiol.*, vol. 186, no. 3, pp. 315-328.

DeBenedette, M. & Snow, E. C. 1991, "Induction and regulation of casein kinase II during B lymphocyte activation", *J. Immunol.*, vol. 147, no. 9, pp. 2839-2845.

Devandra, C. 1995, "Animal production systems in south-east and east Asia: potential challenges for research", in P. Gardner & C. Devandra, eds. *Global Agenda for Livestock Research, Proceedings of a Consultation, ILRI, Nairobi, Kenya,* 18–20 January 1995. ILRI (International Livestock Research Institute), Nairobi, Kenya, pp. 41-48.

Dhar, S. & Gautam, O. P. 1978, "A note on use of hyperimmune serum in bovine tropical theileriosis.", *Indian. Vet. J.*, vol. 56, pp. 738-740.

Dickson, J. & Shiels, B. R. 1993, "Antigenic diversity of a major merozoite surface molecule in *Theileria annulata*", *Mol. Biochem. Parasitol.*, vol. 57, no. 1, pp. 55-64.

Dobbelaere, D. A., Coquerelle, T. M., Roditi, I. J., Eichhorn, M., & Williams, R. O. 1988. "*Theileria parva* infection induces autocrine growth of bovine lymphocytes", *Proc. Natl. Acad. Sci. U.S.A*, vol. 85, no. 13, pp. 4730-4734.

Dobbelaere, D. & Heussler, V. 1999, "Transformation of leukocytes by *Theileria parva* and *T. annulata*", *Annu. Rev. Microbiol.*, vol. 53, pp. 1-42.

Dobbelaere, D., Fernandez, P., Machado, J., Botteron, C., & Heussler, V. 1999, "Interference by the intracellular parasite *Theileria parva* with T-cell signal transduction pathways induces transformation and protection against apoptosis", *Vet. Immunol. Immunopathol.*, vol. 72, no. 1-2, pp. 95-100.

Dobbelaere, D. A., Fernandez, P. C., & Heussler, V. T. 2000, "*Theileria parva*: taking control of host cell proliferation and survival mechanisms", *Cell Microbiol.*, vol. 2, no. 2, pp. 91-99.

Dolan, T. T., Teale, A. J., Stagg, D. A., Kemp, S. J., Cowan, K. M., Young, A. S., Groocock, C. M., Leitch, B. L., Spooner, R. L., & Brown, C. G. 1984, "A histocompatibility barrier to immunization against East Coast fever using *Theileria parva*-infected lymphoblastoid cell lines.", *Parasite Immunol.*, vol. 6, no. 3, pp. 243-250.

Drummond, R. O., Ernst, S. E., Trevino, J. L., Gladney, W. J., & Graham, O. H. 1976, "Tests of acaricides for control of *Boophilus annulatus* and *B. microplus*", *J. Econ. Entomol.*, vol. 69, no. 1, pp. 37-40.

Dschunkowski, E. & Luhs, J. 1904, "Die Piroplasmosen der rinder", *Zentralbl. Bakteriol. Mikrobiol. Hyg.*, vol. 35, pp. 486-492.

Du, W. & Maniatis, T. 1994, "The high mobility group protein HMG I(Y) can stimulate or inhibit DNA binding of distinct transcription factor ATF-2 isoforms", *Proc. Natl. Acad. Sci. U.S.A*, vol. 91, no. 24, pp. 11318-11322.

Dyer, M. & Tait, A. 1987, "Control of lymphoproliferation by *Theileria annulata.*", *Parasitol. Today*, vol. 3, no. 10, pp.309-311.

Dyer, M., Hall, R., Shiels, B., & Tait, A. 1992, "*Theileria annulata*: alterations in phosphoprotein and protein kinase activity profiles of infected leukocytes of the bovine host, *Bos taurus*", *Exp. Parasitol.*, vol. 74, no. 2, pp. 216-227.

Eckner, R. & Birnstiel, M. L. 1989, "Cloning of cDNAs coding for human HMG I and HMG Y proteins: both are capable of binding to the octamer sequence motif", *Nucleic Acids Res.*, vol. 17, no. 15, pp. 5947-5959.

Emanuelsson, O., Nielsen, H., Brunak, S., & von Heijne, G. 2000, "Predicting subcellular localization of proteins based on their N-terminal amino acid sequence", *J. Mol. Biol.*, vol. 300, no. 4, pp. 1005-1016.

Emery, D. L. 1981, "Adoptive transfer of immunity to infection with *Theileria parva* (East Coast fever) between cattle twins", *Res. Vet. Sci.*, vol. 30, no. 3, pp. 364-367.

Emery, D. L., Eugui, E. M., Nelson, R. T., & Tenywa, T. 1981, "Cell-mediated immune responses to *Theileria parva* (East Coast fever) during immunization and lethal infections in cattle", *Immun.*, vol. 43, no. 2, pp. 323-336.

Emery, D. L., Morrison, W. I., Buscher, G., & Nelson, R. T. 1982, "Generation of cell-mediated cytotoxicity to *Theileria parva* (East Coast fever) after inoculation of cattle with parasitized lymphoblasts", *J. Immunol.*, vol. 128, no. 1, pp. 195-200.

Ernst, P., Wang, J., Huang, M., Goodman, R. H., & Korsmeyer, S. J. 2001, "MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein", *Mol. Cell Biol.*, vol. 21, no. 7, pp. 2249-2258.

Fan, J. Y. & Davidson, E. A. 1996, "Molecular cloning and antigenic mapping of heat-shock protein 70 from the malaria species *Plasmodium berghei*", *Am. J. Trop. Med. Hyg.*, vol. 55 (pt 5), pp. 570-576.

Fashena, S. J., Reeves, R., & Ruddle, N. H. 1992, "A poly(dA-dT) upstream activating sequence binds high-mobility group I protein and contributes to lymphotoxin (tumor necrosis factor-beta) gene regulation", *Mol. Cell Biol.*, vol. 12, no. 2, pp. 894-903.

Fatoohi, A. F., Cozon, G. J., Greenland, T., Ferrandiz, J., Bienvenu, J., Picot, S., & Peyron, F. 2002, "Cellular immune responses to recombinant antigens in pregnant women chronically infected with *Toxoplasma gondii*", *Clin. Diagn. Lab Immunol.*, vol. 9, no. 3, pp. 704-707.

Fawcett, D. W., Doxsey, S., Stagg, D. A., & Young, A. S. 1982, "The entry of sporozoites of *Theileria parva* into bovine lymphocytes in vitro. Electron microscopic observations", *Eur. J. Cell Biol.*, vol. 27, no. 1, pp. 10-21.

Fawcett, D., Musoke, A., & Voigt, W. 1984, "Interaction of sporozoites of *Theileria parva* with bovine lymphocytes in vitro. I. Early events after invasion", *Tissue Cell*, vol. 16, no. 6, pp. 873-884.

Fawcett, D. W., Young, A. S., & Leitch, B. L. 1985, "Sporogony in *Theileria* (Apicomplexa: Piroplasmida)", *J. Submicrosc. Cytol.*, vol. 17, pp. 299-314.

Fawcett, D. W. & Stagg, D. A. 1986, "Passive endocytosis of sporozoites of *Theileria parva* in macrophages at 1-2 degrees C", *J. Submicrosc. Cytol.*, vol. 18, no. 1, pp. 11-19.

Fawcett, D. W., Conrad, P. A., Grootenhuis, J. G., & Morzaria, S. P. 1987, "Ultrastructure of the intra-erythrocytic stage of *Theileria* species from cattle and waterbuck", *Tissue Cell*, vol. 19, no. 5, pp. 643-655.

Fell, A. H., Preston, P. M., & Ansell, J. D. 1990, "Establishment of *Theileria*-infected bovine cell lines in scid mice", *Parasite Immunol.*, vol. 12, no. 3, pp. 335-339.

Fich, C., Klauenberg, U., Fleischer, B., & Broker, B. M. 1998, "Modulation of enzymatic activity of Src-family kinases in bovine T cells transformed by *Theileria parva*", *Parasitol.*, vol. 117 (Pt 2), pp. 107-115.

Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. 1998, "Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*", *Nature*, vol. 391, no. 6669, pp. 806-811.

Fitzsimmons, D. & Hagman, J. 1996, "Regulation of gene expression at early stages of B-cell and T-cell differentiation", *Curr. Opin. Immunol.*, vol. 8, no. 2, pp. 166-174.

Fivaz, B. H., Norval, R. A., & Lawrence, J. A. 1989, "Transmission of *Theileria parva* bovis (Boleni strain) to cattle resistant to the brown ear tick *Rhipicephalus appendiculatus* (Neumann)", *Trop. Anim. Health Prod.*, vol. 21, no. 2, pp. 129-134.

Flach, E. J. & Ouhelli, H. 1992, "The epidemiology of tropical theileriosis (*Theileria annulata* infection in cattle) in an endemic area of Morocco", *Vet. Parasitol.*, vol. 44, no. 1-2, pp. 51-65.

Forsyth, L. M., Jackson, L. A., Wilkie, G., Sanderson, A., Brown, C. G., & Preston, P. M. 1997, "Bovine cells infected in vivo with *Theileria annulata* express CD11b, the C3bi complement receptor", *Vet. Res. Commun.*, vol. 21, no. 4, pp. 249-263.

Forsyth, L. M., Minns, F. C., Kirvar, E., Adamson, R. E., Hall, F. R., McOrist, S., Brown, C. G., & Preston, P. M. 1999, "Tissue damage in cattle infected with *Theileria annulata* accompanied by metastasis of cytokine-producing, schizont-infected mononuclear phagocytes", *J. Comp. Pathol.*, vol. 120, no. 1, pp. 39-57.

Fox, M. 1997, "Regulation of TamS1 gene expression during differentiation to the merozoite in *Theileria annulata*.", PhD thesis, University of Glasgow.

Friedhoff, K., Scholtyseck, E. 1968, "Fine structure of *Babesia ovis* (Piroplasmidea) in *Rhipicephalus bursa* (Ixodoidea): transformation from spheroid to vermicule form", *Z Parasitenkd*; vol. 30, no. 4, pp. 347-359.

Friedhoff, K. T. 1988, "Transmission of *Babesia*", in M. Ristic, ed. *Babesiosis of Domestic Animals and Man*, CRC Press, Boca Raton, FL, pp. 23-52.

Furger, A., Schurch, N., Kurath, U., Roditi, I. 1997, "Elements in the 3' untranslated region of procyclin mRNA regulate expression in insect forms of *Trypanosoma brucei* by modulating RNA stability and translation", *Mol. Cell. Biol.*, vol. 17, no. 8, pp. 4372-4380.

Giancotti, V., Buratti, E., Perissin, L., Zorzet, S., Balmain, A., Portella, G., Fusco, A., & Goodwin, G. H. 1989, "Analysis of the HMGI nuclear proteins in mouse neoplastic cells induced by different procedures", *Exp. Cell Res.*, vol. 184, no. 2, pp. 538-545.

Giancotti, V., Bandiera, A., Buratti, E., Fusco, A., Marzari, R., Coles, B., & Goodwin, G. H. 1991, "Comparison of multiple forms of the high mobility group 1 proteins in rodent and human cells. Identification of the human high mobility group I-C protein", *Eur. J. Biochem.*, vol. 198, no. 1, pp. 211-216.

Gill, B. S., Bhattacharyulu, Y., & Kaur, D. 1976, "Immunisation against bovine tropical theileriosis (*Theileria annulata* infection)", *Res. Vet. Sci.*, vol. 21, no. 2, pp. 146-149.

Glascodine, J., Tetley, L., Tait, A., Brown, D., & Shiels, B. 1990, "Developmental expression of a *Theileria annulata* merozoite surface antigen", *Mol. Biochem. Parasitol.*, vol. 40, no. 1, pp. 105-112.

Glass, E. J., Innes, E. A., Spooner, R. L., & Brown, C. G. 1989, "Infection of bovine monocyte/macrophage populations with *Theileria annulata* and *Theileria parva*", *Vet. Immunol. Immunopathol.*, vol. 22, no. 4, pp. 355-368.

Goodbourn, S. 1996, "Transcript mapping," in K. Docherty, ed. *Gene Transcription, RNA Analysis*, John Wiley and Sons, Chichester, New York, Brisbane, Toronto, Singapore, pp. 38-44.

Gray, M. A. & Brown, C. G. 1981, "*In vitro* neutralisation of theilerial sporozoite infectivity with immune serum.," in A. D. Irvin, M. P. Cunningham, & A. S. Young, eds. *Advances in the Control of Theileriosis*, Martinus Nijhoff Publishers, The Hague, Boston, London, pp. 127-131.

Grogan, M. J., Pratt, M. R., Marcaurelle, L. A., Bertozzi, C. R. 2002, "Homogeneous glycopeptides and glycoproteins for biological investigation" *Annu. Rev. Biochem.*, vol. 71, pp. 593-634.

Grootenhuis, J. G., Morrison, W. I., Karstad, L., Sayer, P. D., Young, A. S., Murray, M., & Haller, R. D. 1980, "Fatal theileriosis in eland (Taurotragus oryx): pathology of natural and experimental cases", *Res. Vet. Sci.*, vol. 29, no. 2, pp. 219-229.

Gubbels, J. M., de Vos, A. P., van der Weide, M., Viseras, J., Schouls, L. M., de Vries, E., & Jongejan, F. 1999, "Simultaneous detection of bovine *Theileria* and *Babesia* species by reverse line blot hybridization", *J. Clin. Microbiol.*, vol. 37, no. 6, pp. 1782-1789.

Gubbels, M. J., d'Oliveira, C., & Jongejan, F. 2000, "Development of an indirect Tams1 enzyme-linked immunosorbent assay for diagnosis of *Theileria annulata* infection in cattle", *Clin. Diagn. Lab Immunol.*, vol. 7, no. 3, pp. 404-411.

Gubbels, M. J., Katzer, F., Shiels, B. R., & Jongejan, F. 2001, "Study of *Theileria annulata* population structure during bovine infection and following transmission to ticks", *Parasitol.*, vol. 123 (Pt 6), pp. 553-561.

Guerini, M. N., Que, X., Reed, S. L., & White, M. W. 2000, "Two genes encoding unique proliferating-cell-nuclear-antigens are expressed in *Toxoplasma gondii*", *Mol. Biochem. Parasitol.*, vol. 109, no. 2, pp. 121-131.

Hagiwara, K., Tsuji, M., Ishihara, C., Tajima, M., Kurosawa, T., Iwai, H., & Takahashi, K. 1993, "The Bo-RBC-SCID mouse model for evaluating the efficacy of anti-theilerial drugs", *Int. J. Parasitol.*, vol. 23, no. 1, pp. 13-16.

Hall, F. R. 1988, "Antigens and immunity in *Theileria annulata* infections", *Parasitol. Today*, vol. 4, pp. 257-261.

Hall, R. & Baylis, H. A. 1993, "Tropical Theileriosis", *Parasitol. Today*, vol. 9, no. 9, pp. 310-312.

Hall, R., Ilhan, T., Kirvar, E., Wilkie, G., Preston, P. M., Darghouth, M., Somerville, R., & Adamson, R. 1999, "Mechanism(s) of attenuation of *Theileria annulata* vaccine cell lines", *Trop. Med. Int. Health*, vol. 4, no. 9, p. A78-A84.

Hall, R., Boulter, N. R., Brown, C. G., Wilkie, G., Kirvar, E., Nene, V., Musoke, A. J., Glass, E. J., & Morzaria, S. P. 2000, "Reciprocal cross-protection induced by sporozoite antigens SPAG-1 from *Theileria annulata* and p67 from *Theileria parva*", *Parasite Immunol.*, vol. 22, no. 5, pp. 223-230.

Hanafusa, H., Garber, E. A., & Hanafusa, T. 1986, "Transformation by p60src with altered N-terminal sequences", *Princess Takamatsu Symp.*, vol. 17, pp. 233-240.

Harlow, E. & Lane, D. 1988, "*Antibodies: A Laboratory Manual*", Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Hashemi-Fesharki, R. 1991, "Chemotherapeutic value of parvaquone and buparvaquone against *Theileria annulata* infection of cattle", *Res. Vet. Sci.*, vol. 50, no. 2, pp. 204-207.

Heath, S., Hieny, S., & Sher, A. 1990, "A cyclic AMP inducible gene expressed during the development of infective stages of *Trypanosoma cruzi*", *Mol. Biochem. Parasitol.*, vol. 43, no. 1, pp. 133-141.

Helenius, A. & Aebi, M. 2001, "Intracellular functions of N-linked glycans" *Science*, vol. 291, no. 5512, pp. 2364-2369.

Heussler, V. T., Machado, J., Jr., Fernandez, P. C., Botteron, C., Chen, C. G., Pearse, M. J., & Dobbelaere, D. A. 1999, "The intracellular parasite *Theileria parva* protects infected T cells from apoptosis", *Proc. Natl. Acad. Sci. U.S.A*, vol. 96, no. 13, pp. 7312-7317.

Heussler, V. T., Kuenzi, P., Fraga, F., Schwab, R. A., Hemmings, B. A., & Dobbelaere, D. A. 2001, "The Akt/PKB pathway is constitutively activated in *Theileria*-transformed leucocytes, but does not directly control constitutive NF-kappaB activation", *Cell Microbiol.*, vol. 3, no. 8, pp. 537-550.

Hicks, G. R. & Raikhel, N. V. 1995, "Protein import into the nucleus: an integrated view", *Annu. Rev. Cell Dev. Biol.*, vol. 11, pp. 155-188.

Hinz, M., Krappmann, D., Eichten, A., Heder, A., Scheidereit, C., & Strauss, M. 1999, "NF-κB function in growth control: regulation of cyclin D1 expression and G0/G1-to-S-phase transition.", *Mol. Cell. Biol.*, vol. 19, pp. 2690-2698.

Hoffman, K., Bucher, P., Falquet, L., & Bairoch, A. 1999, "The PROSITE database, its status in 1999.", *Nucleic Acids Res.*, vol. 27, pp. 215-219.

Homer, M.J., Aguilar-Delfin, I., Telford, S.R., Krause, P.J., Persing, D.H. 2000, "Babesiosis", *Clin Microbiol Rev*, vol.13, no. 3, pp. 451-69.

Hooshmand-Rad, P. 1976, "The pathogenesis of anaemia in *Theileria annulata* infection", *Res. Vet. Sci.*, vol. 20, no. 3, pp. 324-329.

Horrocks, P., Dechering, K., & Lanzer, M. 1998, "Control of gene expression in *Plasmodium falciparum*", *Mol. Biochem. Parasitol.*, vol. 95, no. 2, pp. 171-181.

Horrocks, P. & Lanzer, M. 1999, "Mutational analysis identifies a five base pair cis-acting sequence essential for GBP130 promoter activity in *Plasmodium falciparum*", *Mol. Biochem. Parasitol.*, vol. 99, no. 1, pp. 77-87.

Höllsberg, P. 1999, "Mechanisms of T-cell activation by human T-cell lymphotrophic virus type I.", *Microbiol. Mol. Biol. Rev.*, vol. 63, pp. 308-333.

Hotz, H. R., Hartmann, C., Huober, K., Hug, M., Clayton, C. 1997, "Mechanisms of developmental regulation in *Trypanosoma brucei*: a polypyrimidine tract in the 3' untranslated region of a surface protein mRNA affects RNA abundance and translation", *Nucleic. Acids Res.*, vol. 25, no.15, pp. 3017-3026.

Hulliger, L., Wilde, J. K., Brown, C. G., Turner, L. 1964, "Mode of multiplication of *Theileria* in cultures of bovine lymphocytic cells" *Nature,* vol. 203, pp. 728–730.

Hulliger, L. 1965, "Cultivation of three species of *Theileria* in lymphoid cells in vitro", *J. Protozool.*, vol. 12, no. 4, pp. 649-655.

Hulliger, L., Brown, C. G., & Wilde, J. K. 1966, "Transition of developmental stages of *Theileria parva* in vitro at high temperature", *Nature*, vol. 211, no. 46, pp. 328-329.

Hunter, T. & Sefton, B. M. 1980, "Transforming gene product of Rous sarcoma virus phosphorylates tyrosine", *Proc. Natl. Acad. Sci. U.S.A*, vol. 77, no. 3, pp. 1311-1315.

Hunter, T., Alexander, C. B., & Cooper, J. A. 1985, "Protein phosphorylation and growth control", *Ciba Found. Symp.*, vol. 116, Wiley Press, New York, pp. 188-204.

Imamura, R., Masuda, E. S., Naito, Y., Imai, S., Fujino, T., Takano, T., Arai, K., & Arai, N. 1998, "Carboxyl-terminal 15-amino acid sequence of NFAT-1 is possibly created by tissue-specific splicing and is essential for transactivation activity in T cells", *J. Immunol.*, vol. 161, no. 7, pp. 3455-3463.

Innes, E. A., Millar, P., Brown, C. G., & Spooner, R. L. 1989, "The development and specificity of cytotoxic cells in cattle immunized with autologous or allogeneic *Theileria annulata*-infected lymphoblastoid cell lines", *Parasite Immunol.*, vol. 11, no. 1, pp. 57-68.

Innis, M. A. & Gelfand, D. H. 1990, "Optimisation of PCRs," in M. A. Innis, J. J. Sninsky, & T. J. White, eds. *PCR Protocols*, Academic Press, New York, pp. 3-12.

Irvin, A. D., Brown, C. G., Kanhai, G. K., & Stagg, D. A. 1975, "Comparative growth of bovine lymphosarcoma cells and lymphoid cells infected with *Theileria parva* in athymic mice.", *Nature*, vol. 255, pp. 713-714.

Irvin, A. D., Ocama, J. G., & Spooner, P. R. 1982, "Cycle of bovine lymphoblastoid cells parasitised by *Theileria parva*", *Res. Vet. Sci.*, vol. 33, no. 3, pp. 298-304.

Irvin, A. D. & Morrison, W. I. 1987, "Immunopathology, immunology and immunoprophylaxis of *Theileria* infections," in E. J. L. Soulsby, ed. *Immune Responses in Parasitic Infections: Immunology, Immunopathology and Immunoprophylaxis*, vol. 3, CRC Press, Boca Raton, Florida, pp. 223-273.

Irvin, A. D. 1987, "Characterization of species and strains of *Theileria*", *Adv. Parasitol.*, vol. 26, pp. 145-197.

Ivanov, V., Stein, B., Baumann, I., Dobbelaere, D. A., Herrlich, P., & Williams, R. O. 1989, "Infection with the Intracellular Protozoan Parasite *Theileria parva* Induces Constitutively High Levels of NF-κB in Bovine T Lymphocytes", *Mol. Cell. Biol.*, vol. 9, no. 11, pp. 4677-4686.

Jahnke, H. E. 1982, "Livestock production systems and livestock development in tropical Africa." Kieler Wissenschaftsverlag Vauk, Kiel.

Janse, C. J., Ramesar, J., van den Berg, F. M., & Mons, B. 1992, "*Plasmodium berghei*: in vivo generation and selection of karyotype mutants and non-gametocyte producer mutants", *Exp. Parasitol.*, vol. 74, no. 1, pp. 1-10.

Jarrett, W. F. & Brocklesby, D. W. 1966, "A preliminary electron microscopic study of East Coast Fever (*Theileria parva* infection)", *J. Protozool.*, vol. 13, no. 2, pp. 301-310.

Jarrett, W. F., Crighton, G. W., & Pirie, H. M. 1969, "*Theileria parva*: kinetics of replication", *Exp. Parasitol.*, vol. 24, no. 1, pp. 9-25.

Javahery, R., Khachi, A., Lo, K., Zenzie-Gregory, B., & Smale, S. T. 1994, "DNA sequence requirements for transcriptional initiator activity in mammalian cells", *Mol. Cell Biol.*, vol. 14, no. 1, pp. 116-127.

Jean, L., Péry, P., Dunn, P., Bumstead, J., Billington, K., Ryan, R., & Tomley, F. 2001, "Genomic organisation and developmentally regulated expression of an apicomplexan aspartyl proteinase", *Gene*, vol. 262, no. 1-2, pp. 129-136.

John, S., Reeves, R. B., Lin, J. X., Child, R., Leiden, J. M., Thompson, C. B., & Leonard, W. J. 1995, "Regulation of cell-type-specific interleukin-2 receptor alpha-chain gene expression: potential role of physical interactions between Elf-1, HMG-I(Y), and NF-kappa B family proteins", *Mol. Cell Biol.*, vol. 15, no. 3, pp. 1786-1796.

Johnson, K. R., Lehn, D. A., & Reeves, R. 1989, "Alternative Processing of mRNAs Encoding Mammalian Chromosomal High-Mobility-Group Proteins HMG-I and HMG-Y", *Mol. Cell Biol.*, vol. 9, pp. 2114-2123.

Joyce, D., Albanese, C., Steer, J., Fu, M., Bouzahzah, B., & Pestell, R. G. 2001, "NF-kappa B and cell-cycle regulation: the cyclin connection", *Cytokine Growth Factor Rev.*, vol. 12, no. 1, pp. 73-90.

Jura, W. G., Brown, C. G., & Kelly, B. 1983, "Fine structure and invasive behaviour of the early developmental stages of *Theileria annulata* in vitro", *Vet. Parasitol.*, vol. 12, no. 1, pp. 31-44.

Kachani, M., & Spooner, R. L. 1992, "Anti-lymphocyte antibodies generated in animals immunised with *Theileria annulata*-infected cells", *Vet. Immunol. Immunopathol.*, vol. 33, no's. 1-2, pp. 163-169.

Karlson, J. R., Mork, E., Holtlund, J., Laland, S. G., & Lund, T. 1989, "The amino acid sequence of the chromosomal protein HMG-Y, its relation to HMG-I and possible domains for the preferential binding of the proteins to stretches of A-T base pairs", *Biochem. Biophys. Res. Commun.*, vol. 158, no. 3, pp. 646-651.

Katende, J., Morzaria, S., Toye, P., Skilton, R., Nene, V., Nkonge, C., Musoke, A. 1998, "An enzyme-linked immunosorbent assay for detection of *Theileria parva* antibodies in cattle using a recombinant polymorphic immunodominant molecule", *Parasitol. Res.*, vol. 84, no. 5, pp. 408-416.

Katzer, F., Carrington, M., Knight, P., Williamson, S., Tait, A., Morrison, I. W., & Hall, R. 1994, "Polymorphism of SPAG-1, a candidate antigen for inclusion in a sub-unit vaccine against *Theileria annulata*", *Mol. Biochem. Parasitol.*, vol. 67, no. 1, pp. 1-10.

Katzer, F., McKellar, S., Ben Miled, L., d'Oliveira, C., & Shiels, B. 1998, "Selection for antigenic diversity of Tams1, the major merozoite antigen of *Theileria annulata*", *Ann. N.Y. Acad. Sci.*, vol. 849, pp. 96-108.

Kawamoto, S., Takahashi, K., Kurosawa, T., Sonoda, M., & Onuma, M. 1990, "Intraerythrocytic schizogony of *Theileria sergenti* in cattle", *Nippon Juigaku. Zasshi*, vol. 52, no. 6, pp. 1251-1259.

Kinnaird, J. H., Logan, M., Kirvar, E., Tait, A., & Carrington, M. 1996, "The isolation and characterization of genomic and cDNA clones coding for a cdc2-related kinase (ThCRK2) from the bovine protozoan parasite *Theileria*", *Mol. Microbiol.*, vol. 22, no. 2, pp. 293-302.

Kinnaird, J., Logan, M., Tait, A., & Langsley, G. 2001, "TaCRK3 encodes a novel *Theileria annulata* protein kinase with motifs characteristic of the family of eukaryotic cyclin dependent kinases: a comparative analysis of its expression with TaCRK2 during the parasite life cycle", *Gene*, vol. 279, no. 2, pp. 127-135.

Kirvar, E., Ilhan, T., Katzer, F., Hooshmand-Rad, P., Zweygarth, E., Gerstenberg, C., Phipps, P., & Brown, C. G. 2000, "Detection of *Theileria annulata* in cattle and vector ticks by PCR using the Tams1 gene sequences", *Parasitol.*, vol.120 (Pt. 3), pp. 245-254.

Knight, P., Musoke, A. J., Gachanja, J. N., Nene, V., Katzer, F., Boulter, N., Hall, R., Brown, C. G., Williamson, S., Kirvar, E., Bell-Sakyi, L., Hussain, K., & Tait, A. 1996, "Conservation of neutralizing determinants between the sporozoite surface antigens of *Theileria annulata* and *Theileria parva*", *Exp. Parasitol.*, vol. 82, no. 3, pp. 229-241.

Laemmli, U. K. 1970, "Cleavage of structural proteins during the assembly of the head of bacteriophage T4", *Nature*, vol. 227, pp. 680-685.

Lanzer, M., de Bruin, D., Ravetch, J.V. 1992, "A sequence element associated with the Plasmodium falciparum KAHRP gene is the site of developmentally regulated protein-DNA interactions.", *Nucleic Acids Res.*, vol. 20 (pt 12), pp. 3051-3056.

Lanzer, M., Wertheimer, S. P., de Bruin, D., & Ravetch, J. V. 1993, "*Plasmodium*: control of gene expression in malaria parasites", *Exp. Parasitol.*, vol. 77, no. 1, pp. 121-128.

Lawrence, J. A., Norval, R. A., & Uilenberg, G. 1983, "*Rhipicephalus zambeziensis* as a vector of bovine Theileriae", *Trop. Anim. Health Prod.*, vol. 15, no. 1, pp. 39-42.

Lawrence, J. A. 1990, "Summing-up strategies for tick-borne diseases control as applies to Africa, especially East Africa.", *Parassitologia*, vol. 32, pp. 113-115.

Leger, H., Sock, E., Renner, K., Grummt, F., & Wegner, M. 1995, "Functional interaction between the POU domain protein Tst-1/Oct-6 and the high-mobility-group protein HMG-I/Y", *Mol. Cell Biol.*, vol. 15, no. 7, pp. 3738-3747.

Levine, N. D., 1973, "*Protozoan parasites of domestic animals and of man.*", 2 edn, Burgess Pub. Co., Minneapolis.

Levine, N. D. 1988, "*Protozoa Phylum Apicomplexa.*", CRC, Boca Raton, FL.

Lewis, E. A. & Fotheringham, W. 1941, "The transmission of *Theileria parva* by ticks", *Parasitol.*, vol. 33, pp. 251-277.

Lewis, E. A. 1950, "Conditions affecting the East Coast Fever parasites in the tick and in cattle", *East Afr. Agric. J.*, vol. 16, pp. 65-77.

Lundberg, K., Karlson, J. R., Ingebrigtsen, K., Holtlund, J., Lund, T., & Laland, S. G. 1989, "On the presence of the chromosomal proteins HMG I and HMG Y in rat organs", *Biochim. Biophys. Acta*, vol. 1009, no. 3, pp. 277-279.

Mallick, K. P., Dhar, S., Malhotra, D. V., Bhushan, C., & Gautam, O. P. 1987, "Immunization of neonatal bovines against *Theileria annulata* by an infection and treatment method", *Vet. Parasitol.*, vol. 24, no. 3-4, pp. 169-173.

Maloo, S. H., Rowlands, G. J., Thorpe, W., Gettinby, G., & Perry, B. D. 2001, "A longitudinal study of disease incidence and case-fatality risks on small-holder dairy farms in coastal Kenya", *Prev. Vet. Med.*, vol. 52, no. 1, pp. 17-29.

Manuja, A., Nichani, A. K., Kumar, R., Rakha, N. K., Kumar, B., & Sharma, R. D. 2000, "Comparison of cellular schizont, soluble schizont and soluble piroplasm antigens in ELISA for detecting antibodies against *Theileria annulata*", *Vet. Parasitol.*, vol. 87, no. 2-3, pp. 93-101.

Mason, P. J., Shiels, B. R., Tait, A., Beck, P., & Hall, R. 1989, "Sequence and expression of a gene from *Theileria annulata* coding for a 70-kilodalton heat-shock protein", *Mol. Biochem. Parasitol.*, vol. 37, no. 1, pp. 27-35.

McAndrew, M. B., Read, M., Sims, P. F., & Hyde, J. E. 1993, "Characterisation of the gene encoding an unusually divergent TATA-binding protein (TBP) from the extremely A+T-rich human malaria parasite *Plasmodium falciparum*", *Gene*, vol. 124, no. 2, pp. 165-171.

McHardy N. & Rae D. G. 1982, "Antitheilerial activity of the coccidiostat monensin.", *Trop. Anim. Health Prod.*, vol. 14, no. 1, pp. 13-14.

McHardy, N. & Morgan, D. W. 1985, "Treatment of *Theileria annulata* infection in calves with parvaquone", *Res. Vet. Sci.*, vol. 39, no. 1, pp. 1-4.

McKeever, D. J., Nyanjui, J. K., & Ballingall, K. T. 1997, "In vitro infection with *Theileria parva* is associated with IL10 expression in all bovine lymphocyte lineages", *Parasite Immunol.*, vol. 19, no. 7, pp. 319-324.

McKellar, S. & Shiels, B. R. 2001, Unpublished Work.

Mehlhorn, H. & Schein, E. 1976, "Electron microscope studies on developmental stages of *Theileria parva* (Theiler, 1904) in the intestine of the tick *Hyalomma anatolicum excavatum* (Koch, 1844)", *Tropenmed. Parasitol.*, vol. 27, no. 2, pp. 182-191.

Mehlhorn, H. & Schein, E. 1984, "The piroplasms: life cycle and sexual stages", *Adv. Parasitol.*, vol. 23, pp. 37-103.

Meisner, H. & Czech, M. P. 1991, "Phosphorylation of transcriptional factors and cell-cycle-dependent proteins by casein kinase II", *Curr. Opin. Cell Biol.*, vol. 3, no. 3, pp. 474-483.

Melillo, R. M., Pierantoni, G. M., Scala, S., Battista, S., Fedele, M., Stella, A., De Biasio, M. C., Chiappetta, G., Fidanza, V., Condorelli, G., Santoro, M., Croce, C. M., Viglietto, G., & Fusco, A. 2001, "Critical role of the HMGI(Y) proteins in adipocytic cell growth and differentiation", *Mol. Cell Biol.*, vol. 21, no. 7, pp. 2485-2495.

Melrose, T. R., Brown, C. G., Morzaria, S. P., Ocama, J. G., & Irvin. A. D. 1984, "Glucose phosphate isomerase polymorphism in *Theileria annulata* and *T. parva*", *Trop. Anim. Health Prod.*, vol. 16, no.4, pp. 239-245.

Mercurio, F. & Manning, A. M. 1999, "Multiple signals converging on NF-κB", *Curr. Opin. Cell. Biol.*, vol. 11, pp. 226-232.

Middendorf, L. R., Bruce, J. C., Bruce, R. C., Eckles, R. D., Grone, D. L., Roemer, S. C., Sloniker, G. D., Steffens, D. L., Sutter, S. L., Brumbaugh, J. A., & . 1992, "Continuous, on-line DNA sequencing using a versatile infrared laser scanner/electrophoresis apparatus", *Electrophoresis*, vol. 13, no. 8, pp. 487-494.

Moltmann, U. G., Mehlhorn, H., Schein, E., Rehbein, G., Voigt, W. P., & Zweygarth, E. 1983, "Fine structure of *Babesia equi* Laveran, 1901 within lymphocytes and erythrocytes of horses: an in vivo and in vitro study", *J. Parasitol.*, vol. 69, no. 1, pp. 111-120.

Morrison, W. I., Buscher, G., Murray, M., Emery, D. L., Masake, R. A., Cook, R. H., & Wells, P. W. 1981, "*Theileria parva*: kinetics of infection in the lymphoid system of cattle", *Exp. Parasitol.*, vol. 52, no. 2, pp. 248-260.

Morrison, W. I., MacHugh, N. D., & Lalor, P. A. 1996, "Pathogenicity of *Theileria parva* is influenced by the host cell type infected by the parasite", *Infect. Immun.*, vol. 64, no. 2, pp. 557-562.

Morzaria, S. P., Young, A. S., & Kimber, C. D. 1977, "The serological relationship of a British *Theileria* with other *Theileria* species using the indirect fluorescent antibody test", *Res. Vet. Sci.*,vol. 22, no. 3, pp. 330-333.

Morzaria, S., Roeder, P. L., Roberts, D. H., Chasey, D., & Drew, T. W. 1982, in O. C. Straub, ed. "Characteristics of a continuous suspension cell line derived from a calf with sporadic bovine leukosis", Commission of the European Communities, Luxembourg, pp. 519-528.

Morzaria, S. & Nene, V. 1990, "Bovine theileriosis: progress in immunization methods", *Int. J. Anim. Sci.*, vol. 5, pp. 1-14.

Mottram, J. C., Kinnaird, J. H., Shiels, B. R., Tait, A., & Barry, J. D. 1993, "A novel CDC2-related protein kinase from *Leishmania mexicana*, LmmCRK1, is post-translationally regulated during the life cycle", *J. Biol. Chem.*, vol. 268, no. 28, pp. 21044-21052.

Muhammed, S. I., Lauerman, L. H., Jr., & Johnson, L. W. 1975, "Effect of humoral antibodies on the course of *Theileria parva* infection (East Coast fever) of cattle", *Am. J. Vet. Res.*, vol. 36, no. 4 Pt.1, pp. 399-402.

Musisi, F. L. 1990, "Methods currently used for the control of East Coast fever: their validity and proposals for future control strategies", *Parassitologia*, vol. 32, no.1, pp. 15-22.

Musoke, A., Morzaria, S., Nkonge, C., Jones, E., & Nene, V. 1992, "A recombinant sporozoite surface antigen of *Theileria parva* induces protection in cattle", *Proc. Natl. Acad. Sci. U.S.A*, vol. 89, no. 2, pp. 514-518.

Musoke, A. J., Palmer, G. H., McElwain, T. F., Nene, V., & McKeever, D. 1996, "Prospects for subunit vaccines against tick-borne diseases", *Br. Vet. J.*, vol. 152, no. 6, pp. 621-639.

Naessens, J., Newson, J., Bensaid, A., Teale, A. J., Magondu, J. G., & Black, S. J. 1985, "De novo expression of T cell markers on *Theileria parva*-transformed lymphoblasts in cattle", *J. Immunol.*, vol. 135, no. 6, pp. 4183-4188.

Nakaar, V., Bermudes, D., Peck, K. R., & Joiner, K. A. 1998, "Upstream elements required for expression of nucleoside triphosphate hydrolase genes of *Toxoplasma gondii*", *Mol. Biochem. Parasitol.*, vol. 92, no. 2, pp. 229-239.

Nakai, K. 1996, "Refinement of the Prediction Methods of Signal Peptides for the Genome Analysis of *Saccharomyces cerevisiae* and *Bacillus subtilis*", in T. Akutsu, K. Asai, M. Hagiya, S. Kuhara, S. Miyano & K. Nakai. eds. *Genome Informatics Workshop, 1996*. Universal Academic Press, Tokyo, pp. 72-81.

Nakai, K. & Horton, P. 1999, "PSORT: a program for detecting sorting signals in proteins and predicting their subcellular localization", *Trends Biochem. Sci.*, vol. 24, pp. 34-35.

Neitz, W. O. 1957, "Theileriosis, gonderiosis and cytauxzoonoses: a review.", *Onderstepoort J. Vet. Res.*, vol. 27, pp. 319-338.

Nielsen, H., Engelbrecht, J., Brunak, S., & von Heijne, G. 1997, "Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites.", *Protein Eng.*, vol. 10, pp. 1-6.

Nikolov, D. B. & Burley, S. K. 1994, "2.1 A resolution refined structure of a TATA box-binding protein (TBP)", *Nat. Struct. Biol.*, vol. 1, no. 9, pp. 621-637.

Nissen, M. S., Langan, T. A., & Reeves, R. 1991, "Phosphorylation by cdc2 kinase modulates DNA binding activity of high mobility group I nonhistone chromatin protein", *J. Biol. Chem.*, vol. 266, no. 30, pp. 19945-19952.

Norval, R.A.I., Perry, B.D. & Young, A.S. 1992, "The epidemiology of theileriosis in Africa.", Academic Press, London, pp. 63-97, 131-154, 181-202.

Odberg-Ferragut, C., Soete, M., Engels, A., Samyn, B., Loyens, A., Van Beeumen, J., Camus, D., Dubremetz, J. F. 1996, "Molecular cloning of the *Toxoplasma gondii* sag4 gene encoding an 18 kDa bradyzoite specific surface protein", *Mol. Biochem. Parasitol.*, vol. 82, issue 2, pp. 237-244.

ole-MoiYoi, O.K., Sugimoto, C., Conrad, P.A., Macklin, M.D. 1992 "Cloning and characterization of the casein kinase II alpha subunit gene from the lymphocyte-transforming intracellular protozoan parasite *Theileria parva*.", *Biochem.* vol. 31, no. 27, pp. 6193-202.

ole-MoiYoi, O. K. 1995, "Casein kinase II in theileriosis", *Science*, vol. 267, no. 5199, pp. 834-836.

Ouhelli, H., Innes, E. A., Brown, C. G., Walker, A. R., Simpson, S. P., & Spooner, R. L. 1989, "The effect of dose and line on immunisation of cattle with lymphoblastoid cells infected with *Theileria annulata*", *Vet. Parasitol.*, vol. 31, no. 3-4, pp. 217-228.

Oura, C. A. 2001, Unpublished Work.

Oura, C. A., Tait, A., & Shiels, B. R. 2001, "*Theileria annulata*: identification, by differential mRNA display, of modulated host and parasite gene expression in cell lines that are competent or attenuated for differentiation to the merozoite", *Exp. Parasitol.*, vol. 98, no. 1, pp. 10-19.

Oura, C. A., Kinnaird, J., Tait, A., & Shiels, B. R. 2002, "Identification of a 40S Ribosomal protein (S17) that is differentially expressed between the macroschizont and piroplasm stages of *Theileria annulata*", *Int. J. Parasitol.*, vol. 32, issue 1, pp. 73-80.

Ozkok, U. & Pipano, E. 1981, "Trials with cell-culture vaccine against theileriosis in Turkey.", in J. Irvin, M. Cunningham, & A. Young, eds. *Advances in the Control of Theileriosis*. Martinus Nijhoff Publishers, The Hague, Boston, London, pp. 256-258.

Pelham, H. R. & Jackson, R.J. 1976, "An efficient mRNA-dependent translation system from reticulocyte lysates", *Eur. J. Biochem*, vol. 67, pp. 247.

Pellicer, A., Robins, D., Wold, B., Sweet, R., Jackson, J., Lowy, I., Roberts, J. M., Sim, G. K., Silverstein, S., & Axel, R. 1980, "Altering genotype and phenotype by DNA-mediated gene transfer", *Science*, vol. 209, no. 4463, pp. 1414-1422.

Perkins, N. D., Felzien, L. K., Betts, J. C., Leung, K., Beach, D. H., Nabel, G. J. 1997, "Regulation of NF-kappaB by cyclin-dependent kinases associated with the p300 coactivator" *Science*, vol. 275, no.5299, pp. 523-527.

Persing, D. H., & Conrad, P. A. 1995, "Babesiosis: new insights from phylogenetic analysis", *Infect. Agents Dis.*, vol. 4, no. 4, pp. 182-195.

Pinna, L. A. 1990, "Casein kinase 2: an 'eminence grise' in cellular regulation?", *Biochim. Biophys. Acta*, vol. 1054, no. 3, pp. 267-284.

Pipano, E. & Tsur, I. 1966, "Experimental Immunisation against *Theileria annulata* with a tissue culture vaccine.", *Lab. Trials Refuah Vet.*, vol. 23, pp. 186-194.

Pipano, E. 1976, "Control of tropical theileriosis and anaplasmosis in Israel", *Bull. Off. Int. Epizoot.*, vol. 86, pp. 55-61.

Pipano, E., Samish, M., Kriegel, Y., & Yeruham, I. 1981, "Immunization of Friesian cattle against *Theileria annulata* by the infection-treatment method", *Br. Vet. J.*, vol. 137, no. 4, pp. 416-420.

Pipano, E. 1989a, "Vaccination against *Theileria annulata* theileriosis," in I. Wright, ed. *Veterinary Protozoan and Hemoparasite Vaccines*. CRC Press, Boca Raton, FL, pp. 203-234.

Pipano, E. 1989b, "Bovine theileriosis in Israel", *Rev. Sci. Tech. Off. Int. Epiz.*, vol. 8, pp. 79-87.

Pipano, E. 1995, "Live vaccines against hemoparasitic diseases in livestock", *Vet. Parasitol.*, vol. 57, no. 1-3, pp. 213-231.

Polla, B. S. 1991, "Heat shock proteins in host-parasite interactions", *Immunol. Today*, vol. 12(3), pp. A38-41.

Preston, P. M., Brown, C. G., & Spooner, R. L. 1983, "Cell-mediated cytotoxicity in *Theileria annulata* infection of cattle with evidence for BoLA restriction", *Clin. Exp. Immunol.*, vol. 53, no. 1, pp. 88-100.

Preston, P. M. & Brown, C. G. 1985, "Inhibition of lymphocyte invasion by sporozoites and the transformation of trophozoite infected lymphocytes in vitro by serum from *Theileria annulata* immune cattle", *Parasite Immunol.*, vol. 7, no. 3, pp. 301-314.

Preston, P. M., Brown, C. G., Bell-Sakyi, L., Richardson, W., & Sanderson, A. 1992a, "Tropical theileriosis in Bos taurus and Bos taurus cross Bos indicus calves: response to infection with graded doses of sporozoites of *Theileria annulata*", *Res. Vet. Sci.*, vol. 53, no. 2, pp. 230-243.

Preston, P. M., Brown, C. G., & Richardson, W. 1992b, "Cytokines inhibit the development of trophozoite-infected cells of *Theileria annulata* and *Theileria parva* but enhance the proliferation of macroschizont infected cell lines.", *Parasite Immunol.*, vol. 14, pp. 125-141.

Preston, P. M., Brown, C. G., Entrican, G., Richardson, W., & Boid, R. 1993, "Synthesis of tumour necrosis factor-alpha and interferons by mononuclear cells from *Theileria annulata*-infected cattle", *Parasite Immunol.*, vol. 15, no. 9, pp. 525-534.

Preston, P. M., Jackson, L. A., Sutherland, I. A., Bell-Sakyi, L., Wilkie, G., Brown, D. J., Schofield, J., Melrose, T. R., Sanderson, A., & Brown, C. G. 1998, "*Theileria annulata*: the expression of two novel macroschizont antigens on the surface of infected mononuclear cells differs during in vitro attenuation of a virulent cell line", *Exp. Parasitol.*, vol. 89, no. 2, pp. 228-240.

Preston, P. M., Hall, F. R., Glass, E. J., Campbell, J. D., Darghouth, M. A., Ahmed, J. S., Shiels, B. R., Spooner, R. L., Jongejan, F., & Brown, C. G. 1999, "Innate and adaptive immune responses co-operate to protect cattle against *Theileria annulata*", *Parasitol. Today*, vol. 15, no. 7, pp. 268-274.

Preston, P. M., Jackson, L. A., Sutherland, I. A., Brown, D. J., Schofield, J., Bird, T., Sanderson, A., & Brown, C. G. 2001, "*Theileria annulata*: attenuation of a schizont-infected cell line by prolonged in vitro culture is not caused by the preferential growth of particular host cell types", *Exp. Parasitol.*, vol. 98, no. 4, pp. 188-205.

Quandt, K., Frech, K., Karas, H., Wingender, E., & Werner, T. 1995, "MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data", *Nucleic Acids Res.*, vol. 23, no. 23, pp. 4878-4884.

Radley, D. E., Brown, C. G., Burridge, M. J., Cunningham, M. P., Peirce, M. A., & Purnell, R. E. 1974, "East Coast fever: quantitative studies of *Theileria parva* in cattle", *Exp. Parasitol.*, vol. 36, no. 2, pp. 278-287.

Rafyi, A., Maghami, G., & Houshmand, P. 1965, "On the virulence of *Theileria annulata* (Dschunkowsky and Luhs, 1904) and the premunition against bovine theileriosis in Iran", *Bull. Off Int. Epizoot.*, vol. 64, pp. 431-446.

Rao, A. 1994, "NF-ATp: A transcription factor required for the co-ordinate induction of several cytokine genes.", *Immunol. Today*, vol. 15, no. 6, pp. 274-281.

Rechsteiner, M. & Rogers, S. W. 1996, "PEST sequences and regulation by proteolysis", *Trends Biochem. Sci.*, vol. 21, no. 7, pp. 267-271.

Reeves, R. & Nissen, M. S. 1990, "The A.T-DNA-binding domain of mammalian high mobility group I chromosomal proteins. A novel peptide motif for recognizing DNA structure", *J. Biol. Chem.*, vol. 265, no. 15, pp. 8573-8582.

Reeves, R., Langan, T. A., & Nissen, M. S. 1991, "Phosphorylation of the DNA-binding domain of nonhistone high-mobility group I protein by cdc2 kinase: reduction of binding affinity", *Proc. Natl. Acad. Sci. U.S.A*, vol. 88, no. 5, pp. 1671-1675.

Reeves, R., Leonard, W. J., & Nissen, M. S. 2000, "Binding of HMG-I(Y) imparts architectural specificity to a positioned nucleosome on the promoter of the human interleukin-2 receptor alpha gene", *Mol. Cell Biol.*, vol. 20, no. 13, pp. 4666-4679.

Reeves, R. 2001, "Molecular biology of HMGA proteins: hubs of nuclear function", *Gene*, vol. 277, no. 1-2, pp. 63-81.

Richardson, J. O., Forsyth, L. M., Brown, C. G., & Preston, P. M. 1998, "Nitric oxide causes the macroschizonts of *Theileria annulata* to disappear and host cells to become apoptotic", *Vet. Res. Commun.*, vol. 22, no. 1, pp. 31-45.

Robinson, P. M., 1982, "Theileriosis annulata and its transmission-a review", *Trop. Anim. Health Prod.*, vol. 14, no. 1, pp. 3-12.

Rogers, S., Wells, R., & Rechsteiner, M. 1986, "Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis", *Science*, vol. 234, no. 4774, pp. 364-368.

Ross, A. R., Lee P. J., Smith, D. L., Langridge, J. I., Whetton, A. D., & Gaskell, S. J. 2002, "Identification of proteins from two-dimensional polyacrylamide gels using a novel acid-labile surfactant", *Proteomics*, vol. 2, issue 7, pp.928-936.

Rost, B. 1996, "PHD: predicting one-dimensional protein structure by profile-based neural networks", *Methods Enzymol.*, vol. 266, pp. 525-539.

Sambrook, J., Fritsch, E. F., & Maniatis, T. 1989, "Molecular Cloning: A Laboratory Manual", 2 edn, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Samish, M. 1977, "Infective *Theileria annulata* in the tick without a blood meal stimulus", *Nature*, vol. 270, no. 5632, pp. 51-52.

Samish, M., & Pipano, E. 1978, "Development of infectivity in Hyalomma detritum (Schulze, 1919) ticks infected with *Theileria annulata* (Dchunkowsky and Luhs, 1904)", *Parasitol.*, vol. 77, (Pt. 3), pp. 375-379.

Schein, E. & Voigt, W. P. 1979, "Chemotherapy of bovine theileriosis with Halofuginone. Short communication", *Acta Trop.*, vol. 36, no. 4, pp. 391-394.

Schein, E., Rehbein, G., Voigt, W. P., Zweygarth, E. 1981, "*Babesia equi* (Laveran 1901) 1. Development in horses and in lymphocyte culture", *Tropenmed. Parasitol.*, vol. 32, no. 4 pp. 223-227.

Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P. 1998, "SMART, a simple modular architecture research tool: identification of signaling domains", *Proc. Natl. Acad. Sci. U.S.A*, vol. 95, no. 11, pp. 5857-5864.

Schwanbeck, R., Gymnopoulos, M., Petry, I., Pickielko, A., Szewczuk, Z., Heyduk, T., Zechel, K., & Wisniewski, J. R. 2001, "Consecutive steps of phosphorylation affect conformation and DNA binding of the chironomus high mobility group A protein", *J. Biol. Chem.*, vol. 276, no. 28, pp. 26012-26021.

Seidman, L.A. & Moore, C.J. 2000, "Basic Laboratory Methods for Biotechnology: textbook and laboratory reference", Prentice Hall, Upper Saddle River, N.J., pp. 401-440.

Seldin, D. C., & Leder, P. 1995, "Casein kinase II alpha transgene-induced murine lymphoma: relation to theileriosis in cattle", *Science*, vol. 267, no. 5199, pp. 894-897.

Sergent, E., Donatien, A., Parrot, L., & Lestoquard, F. 1945, "Études sur les Piroplasmoses Bovines", Institut Pasteur d'Algérie, Algiers.

Shapira, M., McEwen, J. G., & Jaffe, C. L. 1988, "Temperature effects on molecular processes which lead to stage differentiation in Leishmania", *EMBO J.*, vol. 7, no. 9, pp. 2895-2901.

Shaw, M. K., Tilney, L. G., & Musoke, A. J. 1991, "The entry of *Theileria parva* sporozoites into bovine lymphocytes: evidence for MHC class I involvement", *J. Cell Biol.*, vol. 113, no. 1, pp. 87-101.

Shaw, M. K. & Tilney, L. G. 1992, "How individual cells develop from a syncytium: merogony in *Theileria parva* (Apicomplexa)", *J. Cell Sci.*, vol. 101 (Pt 1), pp. 109-123.

Shaw, M. K., Tilney, L. G., & McKeever, D. J. 1993, "Tick salivary gland extract and interleukin-2 stimulation enhance susceptibility of lymphocytes to infection by *Theileria parva* sporozoites", *Infect. Immun.*, vol. 61, no. 4, pp. 1486-1495.

Shaw, M. K., Tilney, L. G., Musoke, A. J., & Teale, A. J. 1995, "MHC class I molecules are an essential cell surface component involved in *Theileria parva* sporozoite binding to bovine lymphocytes", *J. Cell Sci.*, vol. 108 (Pt 4), pp. 1587-1596.

Shayan, P., & Ahmed, J. S. 1997, "*Theileria*-mediated constitutive expression of the casein kinase II-alpha subunit in bovine lymphoblastoid cells", *Parasitol. Res.*, vol. 83, issue 6, pp. 526-532.

Shiels, B., McDougall, C., Tait, A., & Brown, C. G. 1986, "Antigenic diversity of *Theileria annulata* macroschizonts", *Vet. Parasitol.*, vol. 21, no. 1, pp. 1-10.

Shiels, B., Hall, R., Glascodine, J., McDougall, C., Harrison, C., Taracha, E., Brown, D., & Tait, A. 1989, "Characterization of surface polypeptides on different life-cycle stages of *Theileria annulata*", *Mol. Biochem. Parasitol.*, vol. 34, no. 3, pp. 209-220.

Shiels, B., Kinnaird, J., McKellar, S., Dickson, J., Miled, L. B., Melrose, R., Brown, D., & Tait, A. 1992, "Disruption of synchrony between parasite growth and host cell division is a determinant of differentiation to the merozoite in *Theileria annulata*", *J. Cell Sci.* , vol. 101 (Pt 1), pp. 99-107.

Shiels, B. R., Smyth, A., Dickson, J., McKellar, S., Tetley, L., Fujisaki, K., Hutchinson, B., Kinnaird, J. 1994, " A Stoichiometric Model of Stage Differentiation in the Protozoan Parasite *Theileria annulata*", *Mol. Cell. Different.*, vol. 2, no. 2, pp. 101-125.

Shiels, B. R., d'Oliveira, C., McKellar, S., Ben Miled, L., Kawazu, S., & Hide, G. 1995, "Selection of diversity at putative glycosylation sites in the immunodominant merozoite/piroplasm surface antigen of *Theileria* parasites", *Mol. Biochem. Parasitol.*, vol. 72, no. 1-2, pp. 149-162.

Shiels, B., Aslam, N., McKellar, S., Smyth, A., & Kinnaird, J. 1997, "Modulation of protein synthesis relative to DNA synthesis alters the timing of differentiation in the protozoan parasite *Theileria annulata*", *J. Cell Sci.*, vol. 110 (Pt 13), pp. 1441-1451.

Shiels, B., Swan, D., McKellar, S., Aslam, N., Dando, C., Fox, M., Ben Miled, L., & Kinnaird, J. 1998, "Directing differentiation in *Theileria annulata*: old methods and new possibilities for control of apicomplexan parasites", *Int. J. Parasitol.*, vol. 28, no. 11, pp. 1659-1670.

Shiels, B., Fox, M., McKellar, S., Kinnaird, J., & Swan, D. 2000, "An upstream element of the TamS1 gene is a site of DNA-protein interactions during differentiation to the merozoite in *Theileria annulata*", *J. Cell Sci.*, vol. 113 (Pt 12), pp. 2243-2252.

Shiels, B.R. & McKellar, S. 2000, "Reciprocal regulation of DNA-binding proteins during stage differentiation in *Theileria annulata*", *XI Molecular Parasitology Meeting* 17-21st September, 2000. Woods Hole, MA, U.S.A., p. 114.

Shiels, B.R. 2001, Unpublished Work.

Shkap, V., & Pipano, E. 2000, "Vaccination against tropical theileriosis", *Ann. N.Y. Acad. Sci.*, vol. 916, pp. 484-500.

Shukla, P. E. & Sharma, R. D. 1991, "Immunization of bovine calves with cell culture vaccine against *Theileria annulata*", *Acta. Vet. BRNO*, vol. 60, pp. 79-86.

Singh, D. K., Jagdish, S., Gautam, O. P., & Dhar, S. 1979, "Infectivity of ground-up tick supernates prepared from Theilerai annulata infected Hyalomma anatolicum anatolicum", *Trop. Anim. Health Prod.*, vol. 11, no. 2, pp. 87-90.

Singh, D. K., Thakur, M., Raghav, P. R., Varshney, B. C. 1993, "Chemotherapeutic trials with four drugs in crossbred calves experimentally infected with *Theileria annulata*", *Res. Vet. Sci.*, vol. 54, no. 1, pp. 68-71.

Slama-Schwok, A., Zakrzewska, K., Leger, G., Leroux, Y., Takahashi, M., Kas, E., & Debey, P. 2000, "Structural changes induced by binding of the high-mobility group I protein to a mouse satellite DNA sequence", *Biophys. J.*, vol. 78, no. 5, pp. 2543-2559.

Smale, S. T. & Baltimore, D. 1989, "The "initiator" as a transcription control element", *Cell*, vol. 57, no. 1, pp. 103-113.

Smith, H. O. & Birnstiel, M. L. 1976, "A simple method for restriction site mapping", *Nucleic Acids Res.*, vol. 3, pp. 2387-2398.

Soete, M., Camus, D., & Dubremetz, J. F. 1994, "Experimental induction of bradizoite-specific antigen expression and cyst formation by the RH strain of *Toxoplasma gondii* in vitro", *Exp. Parasitol.*, vol. 78, pp. 361-370.

Somerville, R. P., Littlebury, P., Pipano, E., Brown, C. G., Shkap, V., Adamson, R. E., Oliver, R. A., Glass, E. J., & Hall, F. R. 1997, "Phenotypic and genotypic alterations associated with the attenuation of a *Theileria annulata* vaccine cell line from Turkey.", *Vaccine*, vol. 16, no. 569, p. 575.

Somerville, R. P., Adamson, R. E., Brown, C. G., & Hall, F. R. 1998a, "Metastasis of *Theileria annulata* macroschizont-infected cells in scid mice is mediated by matrix metalloproteinases", *Parasitol.*, vol. 116 (Pt. 3), pp. 223-228.

Somerville, R. P., Littlebury, P., Pipano, E., Brown, C. G., Shkap, V., Adamson, R. E., Oliver, R. A., Glass, E. J., & Hall, F. R. 1998b, "Phenotypic and genotypic alterations associated with the attenuation of a *Theileria annulata* vaccine cell line from Turkey", *Vaccine*, vol. 16, no. 6, pp. 569-575.

Southern, E. M. 1975, "Detection of specific sequences among DNA fragments separated by gel electrophoresis", *J. Mol. Biol.*, vol. 98, no. 3, pp. 503-517.

Spooner, R. L., Innes, E. A., Glass, E. J., & Brown, C. G. 1989, "*Theileria annulata* and *T. parva* infect and transform different bovine mononuclear cells", *Immunol.*, vol. 66, no. 2, pp. 284-288.

Spooner, P. R. 1990, "Oxytetracycline inhibition of mitochondrial protein synthesis in bovine lymphocytes infected with *Theileria parva* or stimulated by mitogen", *Parasitol.*, vol. 101 (Pt 3), pp. 387-393.

Srivastava, A. K. & Sharma, D. N. 1981, "Studies on the occurrence, clinical features and clinicopathological and pathomorphological aspects of theileriosis in calves", *Veterinary Research Journal*, vol. 4, pp. 22-29.

Stadler, L. 2000, "Characterisation of TashHN: a parasite gene postulated to play a role in the process of cell proliferation and attenuation of differentiation in *Theileria annulata* in vitro.", Second Year Practical Report, Unpublished Work.

Stagg, D. A., Chasey, D., Young, A. S., Morzaria, S. P., Dolan, T. T. 1980, "Synchronization of the division of *Theileria* macroschizonts and their mammalian host cells" *Ann. Trop. Med. Parasitol.*, vol. 74, pp. 263–265.

Stagg, D. A., Dolan, T. T., Leitch, B. L., & Young, A. S. 1981, "The initial stages of infection of cattle cells with *Theileria parva* sporozoites in vitro", *Parasitol.*, vol. 83, (Pt. 1), pp. 191-197.

Stagg, D. A., Young, A. S., Leitch, B. L., Grootenhuis, J. G., & Dolan, T. T. 1983, "Infection of mammalian cells with *Theileria* species", *Parasitol.*, vol. 86 (Pt. 2), pp. 243-254.

Stambolic, V., Mak, T. W., & Woodgett, J. R. 1999, "Modulation of cellular apoptotic potential: contributions to oncogenesis", *Oncogene*, vol. 18, no. 45, pp. 6094-6103.

Stephens, S. A. & Howard, C. J. 2002, "Infection and transformation of dendritic cells from bovine afferent lymph by *Theileria annulata*", *Parasitol.*, vol. 124, (Pt. 5), pp. 485-493.

Sutherland, I. A., Shiels, B. R., Jackson, L., Brown, D. J., Brown, C. G., & Preston, P. M. 1996, "*Theileria annulata*: altered gene expression and clonal selection during continuous in vitro culture", *Exp. Parasitol.*, vol. 83, no. 1, pp. 125-133.

Swan, D. G., McDonald, F., & Shiels, B. R. 1996, "A stage regulated cyclophilin gene from *Theileria annulata*" *Mol. Biochem. Parasitol.*, vol. 81, issue 1, pp. 111-114.

Swan, D. G. & Phillips, K. 1997, Unpublished Work.

Swan, D. G., Phillips, K., Tait, A., & Shiels, B. R. 1999, "Evidence for localisation of a *Theileria* parasite AT hook DNA-binding protein to the nucleus of immortalised bovine host cells", *Mol. Biochem. Parasitol.*, vol. 101, no. 1-2, pp. 117-129.

Swan, D. G., Stern, R., McKellar, S., Phillips, K., Oura, C. A., Karagenc, T. I., Stadler, L., & Shiels, B. R. 2001a, "Characterisation of a cluster of genes encoding *Theileria annulata* AT hook DNA-binding proteins and evidence for localisation to the host cell nucleus", *J. Cell Sci.*, vol. 114, (Pt. 15), pp. 2747-2754.

Swan, D. G., Phillips, K., McKellar, S., Hamilton, C., & Shiels, B. R. 2001b, "Temporal co-ordination of macroschizont and merozoite gene expression during stage differentiation of *Theileria annulata*", *Mol. Biochem. Parasitol.*, vol. 113, no. 2, pp. 233-239.

Swan, D.G., 2001, Unpublished Work.

Swan, D. G. & Phillips, K. 2001, Unpublished Work.

Tait, A. & Hall, F. R. 1990, "*Theileria annulata*: control measures, diagnosis and the potential use of subunit vaccines", *Rev. Sci. Tech.*, vol. 9, no. 2, pp. 387-403.

Tallini, G. & Dal Cin, P. 1999, "HMGI(Y) and HMGI-C dysregulation: a common occurrence in human tumors", *Adv. Anat. Pathol.*, vol. 6, no. 5, pp. 237-246.

Temple, S. & Raff, M. C. 1986, "Clonal analysis of oligodendrocyte development in culture: evidence for a developmental clock that counts cell divisions", *Cell*, vol. 44, pp. 773-779.

Thanos, D. & Maniatis, T. 1992, "The high mobility group protein HMG I(Y) is required for NF-kappa B-dependent virus induction of the human IFN-beta gene", *Cell*, vol. 71, no. 5, pp. 777-789.

Thomas, S.M., Brugge, J.S. 1997, "Cellular functions regulated by Src family kinases.", *Annu. Rev. Cell Dev. Biol.*, vol. 13, pp. 513-609.

Tkachuk, D. C., Kohler, S., & Cleary, M. L. 1992, "Involvement of a homolog of *Drosophila* trithorax by 11q23 chromosomal translocations in acute leukemias", *Cell*, vol. 71, no. 4, pp. 691-700.

Towbin, H., Staehelin, T., & Gordon, J. 1979, "Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications", *Proc. Natl. Acad. Sci.U.S.A*, vol. 76, no. 9, pp. 4350-4354.

Towler, D. A., Gordon, J. I., Adams, S. P., Glaser, L. 1988, "The biology and enzymology of eukaryotic protein acylation", *Annu. Rev. Biochem.*, vol. 57, pp. 69-99.

Triezenberg, S. J. 1995, "Structure and function of transcriptional activation domains", *Curr. Opin. Genet. Dev.*, vol. 5, pp. 1990-1996.

Tsur, I., & Adler, S. 1963, "Growth of *Theileria annulata* schizonts in monolayer tissue cultures" *J. Protozool.*, Suppl. 10, p. 36.

Tsur, I., Adler, S., Pipano, E., & Senft, Z. 1964, in A. Corradetti, ed."Continuous growth of *Theileria annulata* schizonts in monolayer tissue culture." Pergamon Press, Oxford, pp. 266-267.

Tuazon, P. T. & Traugh, J. A. 1991, "Casein kinase I and II--multipotential serine protein kinases: structure, function, and regulation", *Adv. Second Messenger Phosphoprotein Res.*, vol. 23, pp. 123-164.

Uilenberg, G. 1981, "*Theileria* infection other than East Coast Fever," in M. Ristic & I. McIntyre, eds. *Diseases of Cattle in the Tropics*. Martinus Nijhoff Publishers, The Hague, Boston, London, pp. 411-427.

Uilenberg, G. 1999, "Immunization against diseases caused by *Theileria parva*: a review", *Trop. Med. Int. Health*, vol.4 (9), pp. A12-20.

Ulich, T. R., del Castillo, J., Keys, M., Granger, G. A., & Ni, R. X. 1987, "Kinetics and mechanisms of recombinant human interleukin 1 and tumor necrosis factor-alpha-induced changes in circulating numbers of neutrophils and lymphocytes", *J. Immunol.*, vol. 139, no. 10, pp. 3406-3415.

Urquhart, G. M., Armour, J., Duncan, J. L., Dunn, A. M., & Jennings, F. W. 1987, "*Veterinary Parasitology*", Longman Scientific and Technical, New York.

Van der Ploeg, L. H. T., Gianni, S. H., & Cantor, C. R. 1985, "Heat shock genes: regulatory role for differentiation in parasitic protozoa", *Science*, vol. 228, pp. 1443-1446.

van Lin, L. H., Janse, C. J., & Waters, A. P. 2000, "The conserved genome organisation of non-falciparum malaria species: the need to know more", *Int. J. Parasitol.*, vol. 30, no. 4, pp. 357-370.

van Vorstenbosch, C. J., Uilenberg, G., & van Dijk, J. E. 1978, "Erythrocytic forms of *Theileria velifera*", *Res. Vet. Sci.*, vol. 24, no. 2, pp. 214-221.

Vickerman, K. & Irvin, A. D. 1981, "Association of the parasite of East Coast fever (*Theileria parva*) with spindle microtubules of the bovine lymphoblast", *Trans. R. Soc. Trop. Med. Hyg.*, vol. 75, p. 329.

Viseras, J., Garcia-Fernandez, P., & Adroher, F. J. 1997, "Field trial of immunization with an experimental vaccine against Mediterranean theileriosis in Spain", *Vet. Res.*, vol. 28, no. 4, pp. 397-403.

Visser, A. E., Abraham, A., Sakyi, L. J., Brown, C. G., & Preston, P. M. 1995, "Nitric oxide inhibits establishment of macroschizont-infected cell lines and is produced by macrophages of calves undergoing bovine tropical theileriosis or East Coast fever", *Parasite Immunol.*, vol. 17, no. 2, pp. 91-102.

Voller, A., Bidwell, D. E., Bartlett, A., Fleck, D. G., Perkins, M., & Oladehin, B. 1976, "A microplate enzyme-immunoassay for *Toxoplasma* antibody", *J. Clin. Pathol.*, vol. 29, no. 2, pp. 150-153.

Wagner, G. G., Duffus, W. P., & Burridge, M. J. 1974, "The specific immunoglobulin response in cattle immunised with isolated *Theileria parva* antigens", *Parasitol.*, vol. 69, pp. 43-53.

Walker, A. R. & McKellar, S. B. 1983, "Observations on the separation of *Theileria* sporozoites from ticks", *Int. J. Parasitol.*, vol. 13, no. 3, pp. 313-318.

Walker, A. R. 1990, "Parasitic adaptations in the transmission of *Theileria* by ticks-a review", *Trop. Anim. Health Prod.*, vol. 22, no. 1, pp. 23-33.

Wang, D. Z., Ray, P., & Boothby, M. 1995, "Interleukin 4-inducible phosphorylation of HMG-I(Y) is inhibited by rapamycin", *J. Biol. Chem.*, vol. 270, no. 39, pp. 22924-22932.

Wang, D. Z., Zamorano, J., Keegan, A. D., & Boothby, M. 1997, "HMG-I(Y) phosphorylation status as a nuclear target regulated through insulin receptor substrate-1 and the 14R motif of the interleuin-4 receptor.", *J. Biol. Chem.*, vol. 272, pp. 25083-25090.

Waring, P. M. & Cleary, M. L. 1997, "Disruption of a homolog of trithorax by 11q23 translocations: leukemogenic and transcriptional implications", *Curr. Top. Microbiol. Immunol.*, vol. 220, pp. 1-23.

Waters, A. P., Syin, C., & McCutchan, T. F. 1989, "Developmental regulation of stage-specific ribosome populations in *Plasmodium*", *Nature*, vol. 342, no. 6248, pp. 438-440.

Wathanga, J. M., Jones, T. W., & Brown, C. G. 1986, "Cryopreservation of *Theileria*-infected lymphoblastoid cells with functional assessment of viability", *Trop. Anim. Health Prod.*, vol. 18, no. 4, pp. 191-197.

Weiss, L. M., Ma, Y. F., Takvorian, P. M., Tanowitz, H. B., Wittner, M. 1998, "Bradyzoite development in *Toxoplasma gondii* and the hsp70 stress response" *Infect. Immun.*, vol. 66, no.7, pp. 3295-3302.

Wells, C. W. & McKeever, D. J. 1998, "The entry and development of *Theileria parva* sporozoites within bovine dendritic cells from the afferent lymph", in *Biology*, Comm. Eur. Soc. Microsc., Brussels, pp. 597-598.

Wenyon, C. M. 1926, "Protozoology: A Manual for Medical Men, Veterinarians and Zoologists" Balliere, Tindall and Cox, London, pp. 1029-1039.

Wenzel, R. & Herrmann, R. 1996, "Cosmid Cloning with Small Genomes," in B. Birren & E. Lai, eds. *Nonmammalian Genomic Analysis: A Practical Guide*. Academic Press, San Diego, London, pp. 197-222.

Whiteside, S. T. & Goodbourn, S. 1993, "Signal transduction and nuclear targeting: regulation of transcription factor activity by subcellular localisation", *J. Cell Sci.*, vol. 104 (Pt 4), pp. 949-955.

Wiesgigl, M. & Clos, J. 2001, "Heat shock protein 90 homeostasis controls stage differentiation in *Leishmania donovani*" *Mol. Biol. Cell* vol.12, issue 11, pp. 3307-3316.

Williams, R. O. & Dobbelaere, D. A. 1993, "The molecular basis of transformation of lymphocytes by *Theileria parva* infection", *Semin. Cell Biol.*, vol. 4, no. 5, pp. 363-371.

Williamson, S. 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., Walker, A., Beck, P., Shiels, B., Fletcher, J., & Hall, R. 1989, "*Theileria annulata* sporozoite surface antigen expressed in Escherichia coli elicits neutralizing antibody", *Proc. Natl. Acad. Sci. U.S.A*, vol. 86, no. 12, pp. 4639-4643.

Wootton, J. C. & Federhen, S. 1996, "Analysis of compositionally biased regions in sequence databases", *Methods Enzymol.*, vol. 266, pp. 554-571.

Xiao, D. M., Pak, J. H., Wang, X., Sato, T., Huang, F. L., Chen, H. C., & Huang, K. P. 2000, "Phosphorylation of HMG-I by protein kinase C attenuates its binding affinity to the promoter regions of protein kinase C gamma and neurogranin/RC3 genes", *J. Neurochem.*, vol. 74, no. 1, pp. 392-399.

Yanisch-Perron, C., Vieira, J., & Messing, J. 1985, "Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors", *Gene*, vol. 33, no. 1, pp. 103-119.

Young, A. S. 1977, "*Theileria mutans* the infectivity for cattle of parasites derived from prefed *Amblyomma variegatum* nymphs", *Tropenmed. Parasitol.*, vol. 28, no. 4, pp. 521-527.

Young, A. S., Grootenhuis, J. G., Smith, K., Flowers, M. J., Dolan, T. T., & Brocklesby, D. W. 1978, "Structures associated with *Theileria* parasites in eland erythrocytes", *Ann. Trop. Med. Parasitol.*, vol. 72, no. 5, pp. 443-454.

Young, A. S. & Leitch, B. L. 1981, "Epidemiology of East Coast fever: some effects of temperature on the development of *Theileria parva* in the tick vector, *Rhipicephalus appendiculatus*", *Parasitol.*, vol. 83, (Pt. 1), pp. 199-211.

Young, A. S., Leitch, B. L., Dolan, T. T., Newson, R. M., Ngumi, P. N., & Omwoyo, P. L. 1983, "Transmission of *Theileria parva* by a population of Rhipicephalus appendiculatus under simulated natural conditions", *Parasitol.*, vol. 86, (Pt. 2), pp. 255-67.

Zablotsky, V. T. 1983, "Duration of post-vaccinal immunity in bovine theileriosis", *Tr. Vses. Inst. Eksp. Vet.*, vol. 57, pp. 56-61.

Zalut, C., Henzel, W. J., & Harris, H. W. J. 1980, "Micro quantitative Edman manual sequencing", *J. Biochem. and Biophys. Methods*, vol. 3, no. 1, pp. 11-30.

Zentner, M. D., Lin, H. H., Deng, H. T., Kim, K. J., Shih, H. M., & Ann, D. K. 2001, "Requirement for high mobility group protein HMGI-C interaction with STAT3 inhibitor PIAS3 in repression of alpha-subunit of epithelial Na+ channel (alpha-ENaC) transcription by Ras activation in salivary epithelial cells", *J. Biol. Chem.*, vol. 276, no. 32, pp. 29805-29814.

Zilberstein, D., Blumenfeld, N., Liveanu, V., Gepstein, A., & Jaffe, C. L. 1991, "Growth at acidic pH induces an amastigote stage-specific protein in *Leishmania* promastigotes.", *Mol. Biochem. Parasitol.*, vol. 45, pp. 175-178.

Appendix A Sequence of Primers

Primers for p600 synthesis

HMG-1: 5′ TATTGGATCCCAAACAATTGGGAATTT 3′

HMG-2: 5′ CATCGAATTCATCTCGATCTTCACAAT 3′

5' RACE Primers

rsp1: 3′ CCCGCTCACGTAAATAA 5′

rsp2: 3′ CGGACGATTCAAATGAA 5′

rsp3: 3′ CGCTTCCCCAAATTAGT 5′

Primers for TashAT1-N synthesis

1 5′ GAAG**GAATTC**TGGTGGAATTTTAATAAA 3′

2 5′ TTTA**GGATCC**GTAAAATTTGCTTCTTCC 3′

Sequence in bold type indicates the EcoRI and BamHI restriction sites of primers 1 and 2, respectively.

Appendix B Deletion Clones Derived from λTa1

Clone	Sense strand	size/kb	clone	Antisense strand	size/kb
1	gd1sp6	3.6	1	3gd1sp6	3.6
3	gd3s	2.9		6903gd1	
5	690gd5	2.8	2	3gd2sp6	3.2
	gd5s1		3	3gd3sp6	2.8
	gd5s		4	3gd4sp6	2.7
	gd5sp6			6903gd4	
	ngd5		6	3gd6s	2.2
7	gd7sp6	2.6	7	3gd7sp6	2.1
	gd7s		8	3gd8sp6	2
9	gd9sp6	1.8	9	3gd9sp6	1.9
	gd9			n3gd9	
10	gd210sp6	1.6	10	6903gd10	1.8
11	gd11sp6	1.4		n3gd10	
				3gd10	
13	gd13sp6	1.2	11	3gd11s	1.7
14	gd14sp6	1.2		6903gd11	
15	gd15sp6	1.1	12	3gd12sp6	1.4
	ngd15		13	n3gd13	1.3
16	gd16sp6	0.6		3gd13sp6	
18	gd18sp6	0.8	14	3gd14sp6	1.2
	ngd18		15	3gd15sp6	1.0
20	gd20sp6	0.6		3gd15s1	
			16	6903gd16	0.9
				3gd16sp6	
				3gd16s	
				3gd16	
			19	3gd19sp6	0.42
				3gd19s	
				n3gd19	
				3gd19	

Appendix C Contiguous DNA Sequence Map of λTa1

> Sense strand
< Antisense strand

```
6903gd1     >                                                    CCTCGAGTCT 10
3gd1sp6     > GAGCtcTCCatcRTCCAAGCtTATCGATTTCgAACCCggAATACCGAACTCCTCGAGTCT 60
CONSENSUS   > GAGCTCTCCATTrTCCAAGCTTATCGATTTCGAACCCGGAATACCGAACTCCTCGAGTCT 60

                .........+.........+.........+.........+.........+.........+

387-6       >                                ATTATTACTACTCAATCCTAATATATTAACTTAATA 36
6903gd1     > AGTTGCcCAATCtTTAATCTGGTCATTATTACTACTCAATCCTAATATATTAACTTAATA 70
3gd1sp6     > AGTTGCCCAATCTTTAATcTgGTCATTATTACTACTCAATCCTAATATATTAACTTAATA 120
CONSENSUS   > AGTTGCCCAATCTTTAATCTGGTCATTATTACTACTCAATCCTAATATATTAACTTAATA 120

                .........+.........+.........+.........+.........+.........+

396-6       >                                                    GACtTCAGTA 10
387-6       > AATTGTTCCTATAATTGATGGGGAAATAATTAATAGATTATATCTGAKCAGACTTCAGTA 96
6903gd1     > AATTGTTCCTATAATTGATGGGGAAATAATTAATAGATTATATCTGATCAGACTTCAGTA 130
3gd1sp6     > AATTGTTCCTATAATTGATGGGGAAATAATTAATAGATTATATCTgAtCAGACTTCAGTA 180
CONSENSUS   > AATTGTTCCTATAATTGATGGGGAAATAATTAATAGATTATATCTGATCAGACTTCAGTA 180

                .........+.........+.........+.........+.........+.........+

396-6       > TGTAACATATATGTGTACATGTACCTATGATATAGTAACATTAtAgGGTTTATAATACAA 70
387-6       > TGTAACATATATGTGTACATGTACCTATGATATAGTAACATTATAGGGTTTATAATACAA 156
6903gd1     > TGTAACATATATGTgTACATGTACCTATgATATAGTAACATTATAGGGTTTATAATACAa 190
3gd1sp6     > TGTAACATATATGTGTACATGTACCTATGATATAGTAACATTATAGGGTTTATAATACAA 240
CONSENSUS   > TGTAACATATATGTGTACATGTACCTATGATATAGTAACATTATAGGGTTTATAATACAA 240

                .........+.........+.........+.........+.........+.........+

gd16sp6     <                                        aACAtCTGttTAACcCGTTAttCtt 25
396-6       > CTTTACATCTATTATAAAGCCtAATTATATATAATTAACATCTGTTTAACCCGTTAttCtT 130
387-6       > CttTACATCTATTATAAAGCCTAATTATATATAATTAACATCTGTgTAACCCGTTAttCTT 216
6903gd1     > CTTTACATCTATTATAAAGCCTAATTATATATAATTAACATCTGTgTAACCCGTTAttCTT 250
3gd1sp6     > CTTTACATCTATTATAAAGCCTAATTATATATAATTAACATCTGTgTAACCCGTTAttCtt 300
CONSENSUS   > CTTTACATCTATTATAAAGCCTAATTATATATAATTAACATCTGTkTAACCCGTTAttCTT 300

                .........+.........+.........-.........+.........+.........+

gd16sp6     < CTTTcTATTCAGATCTAATAATACTTGTTTAGAtCTGGTACGtCCAATATTTAGAtCCAA 85
396-6       > CTTTTTATTCAGaTCTAATAATACTTGTTTaAGAtCtGGTACGtCCAATATTTAGATCCAA 190
387-6       > CTtTcTATTSAGAtCTAATAATACTTGTCTAGATCTGGTACGTCCAATATTTAGATCCAA 276
6903gd1     > CTTTTTATTCAGAtCTAATAATACTTGTYTAGATCTgGTACGTCCAATATTTAGATCCAA 310
3gd1sp6     > CTtTTTATTBAGaGCTAATAATAC=TGtgTAGATCTGGTACGtCCAATATtTAGAtCCAA 360
CONSENSUS   > CTTTTTATTCAGATCTAATAATACTTGTTTAGATCTGGTACGTCCAATATTTAGATCCAA 360

                .........+.........+.........+.........+.........+.........+

390-1       >                    GATA.TTTGACCTTATTATCCCACATATAttGATTTATTTAAG 43
3gd2sp6     >      TTTGAATtcttgAGATAtTT.GACCTTATTATCCCACATATAttGATTTATTTAAG 56
gd20sp6     <   AATTcGAATtcTTGAGATAttTTGACCTTATTAtCCaaCATATATTGATTTATTTAAG 58
gd16sp6     < TTAATTTGAAttCTTGAGATATTTTGACCTTATTAtCCCACATATATTGATTTATTTAAG 145
396-6       > TTAATTTGAATTCtTGAGATATTT.GACCTTATTAT.CCACATATATTGATTTATTTAAG 250
387-6       > TTAATTTGAATTCTTGAGATATTT.GACCTTATTAT.CCACATATATTGATTTATTTAAG 336
6903gd1     > TTAATTTGAATTCTTGAGATATTT.GACCTTATTAT.CCACATATATTGATTTATTTAAG 370
3gd1sp6     > TTAATtTGAATTCTTGAGATATTTaGACCTtATTATCCCACATATATTGATTTATTTAAG 420
CONSENSUS   > TTAATTTGAATTCTTGAGATATTTTGACCTTATTATCCCACATATATTGATTTATTTAAG 420

                .........-.........-.........+.........-.........+.........+

390-1       > GATAATTTTCCATTAAAAGTACAAATAAATATAGATTATTTCTCTATTATTTATAGTTAA 103
3gd2sp6     > GATAATTTTCCATTAAAAGTACAAATAAATATAGATTATtTCTCTATTATTTATAGTTAA 116
gd20sp6     < GATAATtTTCcATTAAAAGTACAAATAAATATAGATTATTTCtCTATTATTTATAGTTAA 118
gd16sp6     < GATAATTTTCCATTAAAAGTACAAATAAATATAGATTATTTCTCTATTATTTATAGTTAA 205
396-6       > GATAATTTTCCATTAAAAGtaCAAATAAATATAGATTATTTCTCTATTATTTATAGTTAA 310
387-6       > GATAATTTTCCATTAAAAGTACAAATAAATATAGATTATtTCTCTATTATTTATAGTTAA 396
6903gd1     > GATAaTTTTCCATTAAAAGTACAAATAAATATAGATTATtTCTCTATTATTTATAGTTAA 430
3gd1sp6     > GATAATttttCCATTAAAAGTACAAATAAATATAGATTATtTCTCTATTATTtATAGTTAA 480
CONSENSUS   > GATAATTTTCCATTAAAAGTACAAATAAATATAGATTATTTCTCTATTATTTATAGTTAA 480
```

```
                 .........+.........+.........+.........+.........+.........+

390-1          > CACTCAGAATAAGAGTGTGTAAGTgGGTTATTGAGGTAGTTATGTTGATTATGTGTCTGA 163
3gd2sp6        > cACTCAGAATAAGAGTGTGTAAGTNCGTTATTGACGTACTTATGTtGATTATGtGTCTgA 176
gd20sp6        < cACtCAGAATAAgAGTGTGTAAGTGGGTTATTGAGGTAGTTATGTTGATTATGTGTCTGA 178
gd16sp6        < CACTCAGAATAAGAGTGTGTAAGTGGGTTATTGAGGTAGTTATGTTGATTATGTGTCTGA 265
396 5          > cACTcAGAATAAGAGTGTGTAAGTGGGTTATTGAGGTAGTTATGTTGATTATGTGTCTGA 370
387-6          > CACTCAGAATAAGAGTGTGTAAGTGGGTTATTGAGGTAGTTATGTTGATTATGTGTCTGA 456
6903gd1        > CACTCAGAATAAGAGTGTGTAAGTNGGtTATTGAGGTAGTTATGTTGATTATGTGTCTGA 490
3gd1sp6        > CAcTCAGAATAAGAGTgTgTAAGTcGGTTATTgAGGtAGTTATGTTGATtATGtGtCTgA 540
CONSENSUS      > CACTCAGAATAAGAGTGTGTAAGTGGGTTATTGAGGTAGTTATGTTGATTATGTGTCTGA 540

                 .........+.........+.........+.........+.........+.........+

390-1          > TGAAGGTTCAGgTTTTTGTTTTCTTGGTCTACCTCGTTTACGTTCAGGTTGTTCAGGTTC 223
3gd2sp6        > TGAAG.TTCAGGTTTTTGTTTTcTTGGTCTACCTCGTTTACGTTTAGGTTGTTCAGGTTC 236
gd20sp6        < TGAAGGTTCAGGTTTTTGTTTTCTTGGTCTACCtCGTTTACGTTTAGGTTGTTCAGGTTC 238
gd16sp6        < TGAAGGTTCAGGTTTTTGTTTTCTTGGTCTACCTCGTTTACGTTTAGGTTGTTCAGGTTC 325
396-6          > TGAAGGTTCAGGTTTTTGTTTTCTTGGTCTACCTCGTTTACGTTTAGGTTGTTCAGGTTC 430
387-6          > TGAAGGTTCAGGTTTTTGTTTTCTTGGTCTACCTCGTTTACGTTTAGGTTGTTCAGGTTC 516
6903gd1        > TGAAGGTTCACG                                                  502
3gd1sp6        > TGAA.GGtTCA..GGTTGTTtGT                                      562
CONSENSUS      > TGAAGGTTCACGTTTTTGTTTTCTTGGTCTACCTCGTTTACGTTTAGGTTGTTCAGGTTC 600

                 .........+.........+.........+.........+.........+.........+

390-1          > AGGTTTATGTtTTCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCAGTTTTCATGTTTCT 283
3gd2sp6        > AGGTTTATGTTTtCTtGGtCTAAGTAACCAAGTtTTTTTAGTTTCAGTTTtCATGTTTCT 296
gd20sp6        < AGGTTTATGTTTTCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCAGTTTTCATGTTTCT 298
gd16sp6        < AGGTTTATGTTTTCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCAGTTTTCATGTTTCT 385
396-6          > AGGTTTATGTTTtCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCAGTTTTCATGTTTCT 490
387-6          > AGGTTTATGTTTTCTTGGTCTAAGTAACCAAGT                             549
CONSENSUS      > AGGTTTATGTTTTCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCAGTTTTCATGTTTCT 660

                 .........+.........+.........+.........+.........+.........+

gd16sp6        <                                                          T 1
gd15sp6        <                                              CATAttTCTGtTTTCTAGGCCTACCCCGTTT 31
390-1          > TGGTCTAAGTAACCAAGTTTTTTTAGTTTCATATTTCTGTTTTCTAGGCCTACCCCGTTT 343
3gd2sp6        > TGGTCTAAGTAACCAAGtTTTTTTTAGTTtCAtATTTCtgtTTTCTAGGCCTAcCcCgTTT 356
gd20sp6        < TGGTCTAAGTAACCaAAGTTTTTTTAGTTTCATATTTCTGTTTTCTAGGCCTACCCcGTTT 358
gd16sp6        < TGGTCTAAGTAACCaAAGTTTTTTTAGTTTCATATTTCTGTTTTCT                 430
396 6         > TGGTCTAAgtAACNAAGTTTtttTTAGTTTCATATTTctGTTTTCT                 535
CONSENSUS      > TGGTCTAAGTAACCAAGTTTTTTTTAGTTTCATATTTCTGTTTTCTAGGCCTACCCCGTTT 720

                 .........+.........+.........+.........+.........+.........+

gd14sp6        <                                                        TTc 3
gd18sp6        < ACGtTNAGGTTGttMAGGatCAGGtTTCTGTTcTCttGGTCTAAGTAACCAAGTYTTTTT 61
gd15sp6        < ACGtTTAGGTTGtTCAGGNTCACGTTTCTGTTTTCTTGGtCTAAGTAACCAAGTTTTTTT 91
390-1          > ACGTTTAGGTTGTTCAGGTTCAGGTTTCTGTTTTCTTGGTCTAAGTAACCAAGTTTTTTT 403
3gd2sp6        > ACG                                                           359
gd20sp6        < ACGTTTAGGTTGtttaGGTTCAGGTTTCTGTTTTTTTGGTCTAAGTAAC             407
CONSENSUS      > ACGTTTAGGTTGTTCAGGTTCAGGTTTCTGTTTTCTTGGTCTAAGTAACCAAGTTTTTTT 780

                 .........+.........+.........+.........+.........+.........+

gd14sp6        < ACtttCAGTTttCATGTTTCttGGcCTAAGTAACCAAGTTTTTTTAGTtTCATATTTCTG 63
gd18sp6        < AGTTTCAGTTWTCATGTCTCTTGGTCTAAGTAACCAAGTttTTTtAGTtTCAtATtTCTG 121
gd15sp6        < AGTTTCAGTTTTCATGTttCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCATATTTCTG 151
390-1          > AGTTTCAGTTTTCATGTTTCTTGGTCTAAGTAACCAAGTtTTTTTAGTTTCATATTTCTG 463
CONSENSUS      > AGTTTCAGTTTTCATGTTTCTTGGTCTAAGTAACCAAGTTTTTTTAGTTTCATATTTCTG 840

                 .........+.........+.........+.........+.........+.........+

ngd18          <                                                      CTTGG 5
ngd15          <                                                      CTTGC 5
gd1?sp6        <                                                AGGTTTcTGTTTTcTtGG 18
3gd3sp6        >   tTCTAGGCCTACCCCGtTTACGTTTAGGTTGTTCAGGTTCAGGTTTCTGTtTTCTTGG 58
gd14sp6        < TTTTCTAGGCCTACCCCGTTTACGTTTAGGTTGTTCAGGaTCAGGTHTCTGTTTTCTTGG 123
gd18sp6        < TtTTCTAGGCCTACCCCGTTTACGTTTAGGTTGTTCAGGTTCAGGTTTCTGTTcTCTTGG 181
gd15sp6        < TTTTCTAGGCCTACCCCGTTTACGTTTAGGTTGTTCAGGTTCAGGTTTCTGTTTTCTTGG 211
390-1          > TtTTCTAGGCCTAcCCCGTTTACGtTTAGGTTGTTCAGGGTCACGTTTCTGTTTTCTTGG 523
CONSENSUS      > TTTTCTAGGCCTACCCCGTTTACGTTTAGGTTGTTCAGGTTCAGGTTTCTGTTTTCTTGG 900

                 .........+.........+.........+.........+.........+.........+

gd13sp6        <                                                  ACTTCTGTTAT 11
```

```
390-2       >                                                GTTTCTTGGTTACTTCTGTTAT  22
ngd18       <  TCtACCTGGTCTACGTTTAGGTACtTCTGGtTGYTCATGTTTCTTGGTTACTTCTGTTAT  65
ngd15       <  TCTACCTGGTCTACGTTTAGGTACTTCTGGTTGNTCATGTTTCTTGGTTACTTCTGTTAT  65
gd11sp6     <  TcTACcTGGTCtAcGTTgAGGTACTTCTGGTtGTTCATGTgTCTTgGGTqACTTGTGTtAT  78
3gd3sp6     >  TCTACCTgGTCTACGTtTAGGTACTTCTgGTTGTTCATGTtTCTTqGTTACTTCTqTTAT  118
gd14sp6     <  TCTACCTGGTCTACGTTTAGGTACTTCTGGTTGTTCATGTTTCTTGGTTACtTCTGTTAT  183
gd18sp6     <  TCTACCTGGTCTACGTTTAGGTACTTCTGGTTGTTCATGTcTCTTGGTtACTTCTGTTAT  241
gd15sp6     <  TCTACCTGGTCTACGTTTAGGTACTTCTGGTTGTTCATGTTTCTTGGTTACTTCTGTTAT  271
390-1       >  tCTACCTGGtCTACGttTAGGTACTTCTGGtTGTTCAtGtTtCTtGGtTACTTCTG.TAT  583
CONSENSUS   >  TCTACCTGGTCTACGTTTAGGTACTTCTGGTTGTTCATGTTTCTTGGTTACTTCTGTTAT  960

                ........+.........+.........+.........+.........+.........+

390-3       >          AAAATTCCCAATTGTTTGAATACCAataTTTCTAATTCTTGAATGTCTGTttG  54
gd13sp6     <  AtCAGAAAAATtCCCAATTGTtTGAATACCAACATTtTCTAATTCTTCAATGtCTGTTTC  71
390-2       >  ATCAgaAAAATTCCCAATTGTTTGAATACCAACATTTTCTAATTCTTGAACGTCTGTTTG  82
ngd18       <  ATCAGAAAAATTCCCAATTGTTTGAATACCAATATTTTCTAATTCTTGAATGTCTGTTTG  125
ngd15       <  ATCAGAAAAATTCCCAATTGTTTGAATACCAATATTTTCTAATTCTTGAACGTCTGTTTG  125
gd11sp6     <  ATcAGAAAAATTCCcAATTGTTTGAATACcAATATTTTcTAATTcTLGAACGTCTGTTTG  138
3gd3sp6     >  ATcAgAAAAATTCCCAATTGTTTgAATACCAATATTTTCTAATTCTTGAACGTCTGttTG  178
gd14sp6     <  ATCAGAAAAATTCCCAATTGTTTGAATACCAATATTTTCTAATTCTTGAATGTCTGTTTG  243
gd18sp6     <  ATCAGAAAAATTCCCAATTGTTTGAATACCAATATTTTCTAATTCTTGAATGTCTGTTTG  301
gd15sp6     <  ATCAGAAAAATTCCCAATTGTTTGAATACCAATATTTTCTAATTCTTGAATGTCTGTTTG  331
390-1       >  ATCAG                                                        588
CONSENSUS   >  ATCAGAAAAATTCCCAATTGTTTGAATACCAATATTTTCTAATTCTTGAATGTCTGTTTG  1020

                ........+.........+.........+.........+.........+.........+

390-3       >  TGTATCTATATCTTCTACTTCTTGAATGtcTGTTTgAaTCGAGCTTTCTTTAGTTTGAAT  114
gd13sp6     <  TGTATCTATATCTTCTACTTCTTGAATGTCTGTVtGAATCGAGCTttCTTTAGTtTGAAT  131
390-2       >  TGTATCTATATCTTCTACTTCTTGAATGTcTGTTTGAATCGAGCTTTCTTTAGTTTGAAT  142
ngd18       <  TGTATCTATATCTTCTACtTCTTGAATGTCTGTTTGAATAGAGCTTTCTTTAGT        179
ngd15       <  TGTATCTATATCTTCTACTTCTTGAATGTCTGTTTGAATCGAGCTTTCTTTAGTTTGAAT  185
gd11sp6     <  TGTATCTATATCTTCTACTTcTTGAATGTCTGTtTGAATCGAGCTTTCTTTAGTTTGAAT  198
3gd3sp6     >  TGTATCTATATCTTCTACTTCTTGAATGTCTGTYTgAATCGAGCTTTCTTTAGTTTGAAT  238
gd14sp6     <  TGTATCTATATCTTCTACTTCTTGAATGTCTGTTTGAATCGAGCTTTCTTTAGTTTGAAT  303
gd18sp6     <  TGTATCTATATCTTCTACTTCTTGAATGTCTGTYTGAatCGAGCTTTC              349
gd15sp6     <  TGTATCTATATCTTCTACTTCTTGAATGTCTGTTTGAATCGAGCTTTCTTTAGTTTGAAT  391
CONSENSUS   >  TGTATCTATATCTTCTACTTCTTGAATGTCTGTTTGAATCGAGCTTTCTTTAGTTTGAAT  1080

                ........+.........+.........+.........+.........+.........+

n3gd13      <                                          ATGAATTTCTGTTTGAAYYCcAATATCTTC  30
390-3       >  ATCTGTTTGTGTTACAATATTTTCTAATTCATGAATTTCTGTTTGAATTCCAATATCTTC  174
gd13sp6     <  ATCTGTTTGTGTtACAATATTTtTCTAATTCATGAATtCTGTttGAATtCCAATATCTTC  191
390-2       >  ATCTGTTTGTGTTACAATATTTTCTAATTCATGAATTCCTGTTTGAATTCCAATATCTTC  202
ngd15       <  ATCTGTTTCTGTTACAATATTTTCTaATTCATGAATTCclGTTT                  229
gd11sp6     <  ATCTGTTTGTGTtAcAATATTTTcTAATTcATGAATTCcTGTTTGAATtCcAATATCTTc  258
3gd3sp6     >  ATCTGTtTGTGTTACAATATTTTCTAATTCATGAATTCCTGTtTgAATTCCAATATCTTC  298
gd14sp6     <  ATCTGTTTGTGTTACAATATTTTCTAATTCATGAATTCCTGTTTGAATTCCAATATCTTC  363
gd15sp6     <  ATCTGTTTGTGTTACAATATTTTCTAATTCATGAATTTCTGTTTGAATTCCAATATCTTC  442
CONSENSUS   >  ATCTGTTTGTGTTACAATATTTTCTAATTCATGAATTTCTGTTTGAATTCCAATATCTTC  1140

                ........+.........+.........+.........+.........+.........+

3gd4sp6      >                                    TATCcATTcTGAGTTAATATTCCATCTGATTCtat  35
n3gd13      <  TATTTCTTGGATANCTGtTTGTGTGTATCTATTCTGAGTTAATATTCCATCTGAttCTTC  96
390-3       >  TATTTCTTGAATATCTGTTTGTGTGTATCTATTCTGAGTTAATATTCCATCTGATTCTTC  234
gd13sp6     <  TATtTCTTGAATATCTGTTTGTGTGTATCTATTcTGAGTtAATAT                 236
390-2       >  TATTTCTTGAATATCTGTTTGTGTGTATCTATTCTGAGTTAATATTCCATCTGATTCTTT  262
gd11sp6     <  TATtTCTTGAATATcTGTTTGTGTGTATcTATTcATGAGTTAATATTcCaTGTGATTcTTT  318
3gd3sp6     >  TATTTCTTGAATATCTGTtTGTGTGTATCTATTCTgAGTGAATATTCCATCTgAtTCTTT  358
gd14sp6     <  TATTTCTTGAATATCTGTTTGTGTGTATCTATTCTGAGTTAATATTCCATCTGATTCTTT  423
CONSENSUS   >  TATTTCTTGAATATCTGTTTGTGTGTATCTATTCTGAGTTAATATTCCATCTGATTCTTT  1200

                ........+.........+.........+.........+.........+.........+

3gd6s       >                                          TGAtTCAAGTTCCACTGGAAT  21
6903gd4     >                  ATTAGATTCATCAATTTCTTCATCATCTGATTCAAGTTCCACTGGAAT  48
gd210sp6    <          TGATACATGAGATTCATaAATRTCTTCATcATCTGATTCAAGTTcTACTGGAAT  54
3gd4sp6     >  AGGWWTTGAtACATTAgATTCATCAATTTCTTCATCATCTGATTCAAGTTCCACTGGAAT  95
n3gd13      <  AGGtTTTGATAcATTAGATTcATcaATTTCTTCATCATC                       129
390-3       >  AGGTTTTGATACATTAGATTCATCAATTTCTTCATCATCTGATTCAAGTTCCACTGGAAT  294
390-2       >  AGGTTTTGATACATTAGATTCATCAATTTCTTCATCATCTGATTCAAGTTCCACTGGAAT  322
gd11sp6     <  AGGTTTTGATAcATTaAGATTCATCAATTTCTTCATCATgTGATTcAACTTCTtAcTGGAAt  378
3gd3sp6     >  AGGtTTTgATACATTAGATTCATCAATTTCTTCATCATCTgATTCAAGTTCCACTGGAAT  418
gd14sp6     <  AGGTTTTGATACATTAGATTC                                         444
CONSENSUS   >  AGGTTTTGATACATTAGATTCATCAATTTCTTCATCATCTGATTCAAGTTCCACTGGAAT  1260
```

```
                 .........+.........+.........+.........+.........+.........+
3gd6s          > AGTtTCTGGTTCTAACTCTGTAGGTTCTTgTtCTYTTTcTGTtCGTTGAGTAGGTTCTTC  81
6903gd4        > AGTTTCTGGTTCTAACTCTGTAGGTTCTTGTTGTTTTTCTGTTCGTTGAGTAGGTTCTTC 108
gd210sp6       < AGtgTCTGGtTCTAACTCTGtAGGttCTTGTTGTTTTtCTGTTcGTTGAGtAGGtTCTTc 114
3gd4sp6        > AGTTTCTGGTTCTAACTCTGTAGGTTCTTGTTGTTTTCACTGTTTGTTGAGTAGGTTCTTC 155
390-3          > AGTTTCTGGTTCTAACTCTGTAGGTTCTTGTTGTttttCCGTYTGTTGAGTAGGGTCTTC 354
390-2          > ACTtTCTGGTTCTAACTCTGTAGGTTCTTGTTGTTTTTCTGTTTGTTGAGTAGGTTCTTC 382
gd11sp6        < AGTTTGTGGTTCtAA                                               393
3gd3sp6        > AGTtTCTGGTTCTAACTC.gTAGGTTCTTgTTGTcTTTCTGTCTcTTGAGTAGGKTCTTC 478
CONSENSUS      > AGTTTCTGGTTCTAACTCTGTAGGTTCTTGTTGTTTTTCTGTTTGTTGAGTAGGTTCTTC 1320

                 .........+.........+.........+.........+.........+.........+

3gd8sp6        >                                                  aTTgtGaaaCTAT  13
3gd7sp6        >                    TTTCTTTTcTAgAtcaATtCCTTCCTcAqCTAATgATTgTaAAACTAT  48
3gd6s          > TCTgCCAACGATTTTcTTTTCTAGAtCAATTCCTTCCTCAGCTAATGATTCTAAAACTAT 141
6903gd4        > TCTGCCAACGATTTTCTTTTCTAGATCAATTCCTTCCTCAGCTAATGATTGTGAAACTAT 168
gd210sp6       < TCTGCCAACGATTgTCTgGTCTAGATcAATtCCTtCCTCAGCTAATGATTGTGAAACTAT 174
3gd4sp6        > CCTGCCAACGATTTTCTTtTCTAGAtCAATTCCTTCCTCAGCTAATGATTGTGAAACTAT 215
390-3          > TCTGCCAACGATTTCTTTTCTAGATCAATTCCTTCCTCAGCTAATGATTGTGAAACTAT 414
390-2          > CCTGCCAACGATTtTCTtTTCTAGATCAATTCCTTCCTCAGCTAATGATTGTGAAACTAT 442
3gd3sp6        > TCT                                                           481
CONSENSUS      > TCTGCCAACGATTTTCTTTTCTAGATCAATTCCTTCCTCAGCTAATGATTGTGAAACTAT 1380

                 .........+.........+.........+.........+.........+.........+

3gd8sp6        > AtcaACATTTTTATcATgTTgTTCACTTTGtGGTTTATCTtTtCTTGTCTATCTCGTTT 73
3gd7sp6        > ATCAACATTTTTATCATgTTgTTcACTTTgtggtTTATcTTNTCTTgTCTATCTCGTTT 108
3gd6s          > ATCAACATTTTTATCATGTTGTTCACTTTGTGGTTTATCTtTCCTtTGTCTATCTCGTTT 201
6903gd4        > ATCAACATTTTTATCATGTTGTTCACTTTGTGGTTTATCTTTTCTTTGTCTATCTCGTTT 228
gd210sp6       < ATCAACATTttTATCATGtTGtTCACTTTGTGGTTTATCTtgTCTTTGTcTATCTCGTTT 234
3gd4sp6        > ATCAACATTTTTATCATGTTGTTCACTTTGTGGTTTATCTatTCTTTGTCTATCTCGTTT 275
390-3          > ATCAACATTTTTATCATGTTGTTCACTTTGTGGTTTATCTccACTAtGtCTATCTCGttT 474
390-2          > ATCAACATTTCTATCATGTTGTTCACTTTGTGGTTTATCTTTTCTTTGTCTATCTCGTTT 502
CONSENSUS      > ATCAACATTTTTATCATGTTGTTCACTTTGTGGTTTATCTTTTCTTTGTCTATCTCGTTT 1440

                 .........+.........+.........+.........+.........+.........+

3gd9sp6        >                                                        AAG   3
n3gd9          >                                                       TGGGAAG  7
3gd8sp6        > TAGTTTACGTTTATCAACTTTKTCTGTTTTTACATTTTTtTTcAGKcTcGGAATTgGAAG 133
3gd7sp6        > TAGTTTACGTTTATCAACtTTtTCTaTTTTTACATTTTtATTCAGTCTCGGAATTgGAAG 168
3gd6s          > TAGTTTACGTTTATCAACtTTtTCTGTKTTTTACATtTTTtTTCAGTCTCGGAATTGGAAG 261
6903gd4        > TAGTTTACGTTTATCAACTTTTTTCTGTTTTTACATTTTTttTCAGTCTCGCAATTCGAAG 288
gd210sp6       < TAGTTtACGTTTATCAACTTTtTCTGTTttTACATKKgtTGTCAGTCTCGGAATtGGAAG 294
3gd4sp6        > TAGTTTACGTTTATCAACTTTAtCtGtttttACAttttttttCAGtCTCGGAATTGGAAG 335
390 3          > TAGTTTACGTTTATCAAC.TTGcCTG.TTTTACATTTTT.TTCAGTCTCGGAATTGGAAG 534
390-2          > TAGTTTACGTTTATCAACTCTTTCTG.TTTCACACTTTTKTTCAGTCTCGGAATtGGAAG 562
CONSENSUS      > TAGTTTACGTTTATCAACTCTTTCTCTGTTTTTACATTTCTTTTCAGTCTCGGAATTGGAAG 1500

                 .........+.........+.........+.........+.........+.........+

3gd9sp6        > CTTACCTWGTGGaaTCTTAATAAATACtGAACAAaAACTTTTgTAATAATCCTCAAACTC  63
n3gd9          > CTTACCTGGTGGAATTTTAATAaATACTGAACAAAAACTTTTGTAATAATCCTCAAACTC  67
3gd8sp6        > CTTACCTgGTGGAATTTTAATAAATACTGAACAAAAACTTTTGTAATAATCCTCAAACTC 193
3gd7sp6        > CtTACCtAGTGGAATTTTAATAAATACTgAACAAAAACTTttgtAATAATCCTCAAACTC 228
3gd6s          > CtTACCTGGtGGAATTTTAATAAATACTGAACAAAAACtTTTGTAATAATCCTCAAACTC 321
6903gd4        > CTTACCTGGGGGAATTTTAATAAATACTGAACAAAAACTtTTGTAATAATCCTCAAACTC 348
gd210sp6       < CTTACCTGGTGGAATtTTAATAAATACTGAACAAAAACTttTGTaATAATCCTCAAACTc 354
3gd4sp6        > CTTACCTgGTGGAATTTTAATAAATACTGAACAAAAACTTTTGTAATAATCCTCAAACTC 395
390-3          > CTTACCTGGTGGAAtTTTAATAAATACTGAACaAAAAcTTTTGTAATAATCCTCAAACTC 594
390-2          > CTTACCtGGtGGAAtTTtAATaAATACTGAACaAAAACTTTtGT                  606
CONSENSUS      > CTTACCTGGTGGAATTTTAATAAATACTGAACAAAAACTTTTGTAATAATCCTCAAACTC 1560

                 .........+.........+.........+.........+.........+.........+

n3gd10         >                                AACTTCCCSCTCACGTAAATAAATTTCATAGCTTCATC  38
3gd9sp6        > ACTCTTgTTGATATATTTATAGAAGTTCCCGCTCACGTAAATAAATTTCATAGCTTCATC 123
n3gd9          > ACTCTTGTTGATATATTTATAGAAGTTCCCGCTCACGTAAATAAATTTCATAGCTTCATC 127
3gd8sp6        > ACTCTTGTTGATATATTTATAGAAGTTCCCGCTCACGTAAATAAATTTCATAGCTTCATC 253
3gd7sp6        > ACTCTagTTGATATATTTATAGAAGTTCCCgCTCACGTAAATAAATTTCATAGCTTCATC 288
3gd6s          > ACTCTTGTTGATATATTTATAGAACtTCcCCCTCACGTAAATAAATTTCATAGCtTCATC 381
6903gd4        > ACTCTTGTTGATATATTTATAGAAGTTCCCGCTCACGTAAATAAATTTCATAGCTTCATC 408
gd210sp6       < ACTCTTGTTGATATATTKATAGAAGTTCCTGCTcACGTAAATAAATTtcATAGCTTCATc 414
3gd4sp6        > ACTCTTGTtGATATATtTATAGAAGTTCCCGCTCACGTAAATAAATTTCATAGCTtCATC 455
390-3          > ACTC                                                          598
CONSENSUS      > ACTCTTGTTGATATATTTATAGAAGTTCCCGCTCACGTAAATAAATTTCATAGCTTCATC 1620
```

```
                 .........+.........+.........+.........+.........+.........+
6903gd10      >  ACGGCATTTCCAATTTCAATGGTAATTAATGTCTTATCGGACGATTCAAATGAAAATCT 59
n3gd10        > TACGgCATTTCCAATTTCAATGGTAATTAATGTCTTATCGGACGATTCAAATGAAAATCT 98
3gd9sp6       > TACGGCATTTCCAATTTCAATgGTAATTAATGTCTTATCGGACGATTCAAATgAAAATCT 183
n3gd9         > TACGGCATTTCCAATTTCAATGGTAATTAATGTCTTATCGGACGATTCAAATGAAAATCT 187
3gd8sp6       > TACGGCATTTCCAATTTCAATGGTAATTAATGTCTTATCGGACGATTCAAATGAAAATCT 313
3gd7sp6       > TACGGCATTtCCAAtTtCAAtgGtAATTAATgTCTTATCGgACGATTCAAAtgAAAAtCT 348
3gd6s         > TACGGCATTTCCAATTTCAATGGTAATTAATGTCTTATCGGACGATTCAAATgAAAATCT 441
6903gd4       > TACGGCATTTCCAATTTCAATGGTAATTAATGTCTtATCGGACGATTCAAATGAAAATCT 468
gd210sp6      < TACGGCATTtCCAATTtcAATGGTAATTAATGTCTTATCGGACGATTCAAATGAAAATCT 474
3gd4sp6       > TACGGCATTTCCAATTtCAATGGtAAtTAATGTCTTATCGGACGATTCAAATGAAAAtCT 515
CONSENSUS     > TACGGCATTTCCAATTTCAATGGTAATTAATGTCTTATCGGACGATTCAAATGAAAATCT 1680

                 .........+.........+.........+.........+.........+.........+

rfs2          <                                        CGCTTCCCCAAATTAGTTTCCTtTTAGAACG 31
6903gd10      > AGTAAAACATTTAACATATTCACCTCGATCGCtTCCCCAAATTAGTTTCCTTTTAGAACG 119
n3gd10        > AGTAAAACATTTAACATATTCACCTCGATCGCTTCCCCAAATTAGTTTCCTTTTAGAACG 158
3gd9sp6       > AGTAAAACATTTAACATATTCACCTCGATCGCTTCCCCAAATTAGTTTCCTTTTAGAACG 243
n3gd9         > AGTAAAACATTTAACATATTCACCTCGATCGCTTCCCCAAATTAGTTTCCTTTTAGAACG 247
3gd8sp6       > AGTAAAACATTTAACATATTCACCTCGATCGMTTCCCCAAATTAGTTTCCTTTTAGAACG 373
3gd7sp6       > AGTAAAACAtgTAACATATTCACCtCGA                                  376
3gd6s         > AGTAAAACATTTAACATATTCACCTCGAGcGCgTCCCCAAATtAGTTTCCT           492
6903gd4       > AGTAAAACATTTAACATATTCACCTCGAGCGCTTCCcCAAATTAGTTTCCTTTTAGAACG 528
gd210sp6      < AGTAAAACATKTAACATATTCACcTCGATCGCTTCCCCAAATtAGTTTCCTcTTAGAACG 524
3gd4sp6       > AGTAAAACATTTAACATATTCACCTCGAtCGMTTCCCCAAAATAG                 560
CONSENSUS     > AGTAAAACATTTAACATATTCACCTCGATCGCTTCCCCAAATTAGTTTCCTTTTAGAACG 1740

                 .........+.........+.........+.........+.........+.........+

rfs2          < TATTTCAGTTATCTTATCTTCTGGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 91
6903gd10      > TATTTCAGTTATCTTATCTTCTGGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 179
n3gd10        > TATTTCAGTTATCTTATCTTCTGGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 218
3gd9sp6       > TATTTCAGTTATCTTATCTTCTTGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 303
n3gd9         > TATTTCAGTTATCTTATCTTCTGGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 307
3gd8sp6       > TATTTCAGTTATCTTATCTTCTGGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 433
6903gd4       > TAT.TCAGtTATCT                                               542
gd210sp6      < TATVTCAGTtATCTTATCTTcTGGTGTAGA                               564
CONSENSUS     > TATTTCAGTTATCTTATCTTCTGGTGTAGATAATATCATAATCTTAGTTAATCTGTCTTC 1800

                 .........+.........+.........+.........+.........+.........+

rfs2          < AACAATTTTTTATTGTATAAAAATTAGGATTATCTAAATTATCCAAATATAACATCTCGGA 151
6903gd10      > AACAATTTTTATTGTATAAAAaTTAGGATTATCTAAATTATCCAAATATAACATCTCGGA 239
n3gd10        > AACAATTTTTATTGTATAAAAATTAGGATTATcTAAATTATCCAAATATAACATCTCGGA 278
3gd9sp6       > AACAATTTTTATTGTATAAAAATTAGGATTATCTAAATTATCCAAATATAATATTTCGGA 363
n3gd9         > AACAAT                                                       313
3gd8sp6       > AACAATTTTCATTGTATAAAAATTAGGATTATCTAAATTATCCAAATATAACATTTCGGA 493
CONSENSUS     > AACAATTTTCATTGTATAAAAATTAGGATTATCTAAATTATCCAAATATAACATTTCGGA 1860

                 .........+.........+.........+.........+.........+.........+

rfs2          < AGAaGCAAATTTTACGCGGTATAAAAATAATGTAAATATTATGTGAGAGAGTTTCAATAC 211
6903gd10      > AGAAGCAAAtTTTACGCGGtATAAAAATAATGTAAATATTATGTGAGAcAGggTCAATAC 299
n3gd10        > AGAAGCAAATTTTACGCGGTATAAAAATAATGTAAATATTATGTGAGAGAGTTTCAATAC 338
3gd9sp6       > ACAACCAAATTTTACGCgGTATAAAAaTAATGTAAATATTATGTGAGAGAGtgTCAATAC 423
3gd8sp6       > aGaAGCAAATTTTACGCGGTATAAAaATTaTgTaaATATTAT                    535
CONSENSUS     > ACAAGCAAATTTTACGCGGTATAAAAATAATGTAAATATTATGTGAGAGAGTTTCAATAC 1920

                 .........+.........+.........+.........+.........+.........+

gd9sp6        <                                                          AT 2
rfs2          < AACCATCATCAAAcGCTTATACTACTACTATTAAAAATAACACAGTAATTGCTAGATCTGGAT 271
6903gd10      > AACCATCATCAAACGCTTATACTACTACTATTAAAAATAACACAGTAATTgCTAGATCTGGAT 359
n3gd10        > AACCATCATCAAACGCTTATACTACTACTATTAAAAATAACACAGTAATTGCTAGATCTGGAT 398
3gd9sp6       > AACCATCATCAAACGCTTATACTAcCTaTT                                452
CONSENSUS     > AACCATCATCAAACGCTTATACTACTACTATTAAAAATAACACAGTAATTGCTAGATCTGGAT 1980

                 .........+.........+.........+.........+.........+.........+

3gd11s        >                                                  AACATGGATGtTTGT 15
3gd10         >                                        TATTGAAAATATTACTATAACAtGGATgtTtGT 33
3gd12sp6      >                                  ATTAATCTTATTGAAAATATTACTATAACatGGATGTTTGT 41
gd9sp6        < GGGGAAATAAAATATTTAGATYAATCTTATTGAAAATATTACTATAACATGGATGTTTGT 62
rfs2          < GGGGAAATAAAATATTTAGATTAATCTTATTGAAAATATTACTATAACATGGATGTTTGT 331
6903gd10      > GGGGAAATAAAATATTTAGATTAATCTTATTGAAAATATTACTATAACATGGATGTTTGT 419
CONSENSUS     > GGGGAAATAAAATATTTAGATTAATCTTATTGAAAATATTACTATAACATGGATGTTTGT 2040
```

```
                .........*.........+.........+.........+.........+.........+
gd7sp6      <                                                           A 1
690gd5      <                                                    TGATCTACAAAATtTGTTTtAGA 23
gd5s1       <                                                   ATTATGATCTACAAAATTTGTctcAGA 27
6903gd11    >     ATATGTAATAKATAT.AtCATTAATATTTATTATGaBCTACAAAVtTTGTtTTAGA 56
gd5sp6      < AtAAAtATGtAAtAGAtATGATCAttAATAtTTAtTATGATCTACAAAAtTTGtTTtAGA 60
3gd11s      > ATAAATATGTAATAGATATGATCATTAATATTTATTATGAtCTACAAAATTTgTgTTAGA 75
3gd10       > ATAAATATGTAATAGATATGATCATTAATATTTATTAtGAtCTACAAaAtTtGtKTTAGA 93
3gd12sp6    > ATAAATATGTAATAGATATGATCATTAATATTTATTAWGAtCTACAAAATTTGtTTTAGA 101
gd9sp6      < ATAAATATGTAATAGATATGATCATTAATATTTATTATGATCTACAAAATTTGTcTTAGA 122
rfs2        < ATAAATATGTAATAGATATGATCATTAATATTTATTATGATCTACAAAATTTGTTTTAGA 391
6903gd10    > ATAAATATGTAAATAgATATgACCATTAATATTTATTAtgACCTACA              466
CONSENSUS   > ATAAATATGTAATAGATATGATCATTAATATTTATTATGATCTACAAAATTTGTTTTAGA 2100

                .........+.........+.........+.........+.........+.........+

3gd13sp6    >                   ACAAGATTTTCAATTTTTTGCATCTCAgSVKaTAATAAGTTTAGA 45
gd3s        <                   AACAACATTTTgAATCgTCYGCATCTcAGTATATAATAAGTTVAGA 46
gd7sp6      < TCAAttCCCATTTAAACAAGATTTCAATCTTTTGCATCTCAGTATATAATAAGTTtAGA 61
690gd5      < tCAAttCCCATTTAAACAAGATTcTCAATCTTTTGCATCTCAGTATATAATAAGTTYAGA 83
gd5s1       < TCAATtCCCATtYAAACAAgATtCTCAATCCTTtGCATCTCAGTATATAATAAGTtCAGA 87
6903gd11    > TCAATTCCCATTTAAACAAGAtTTTCAATCTTYTGCATCTCAGTATATAATAAGTtTAGA 116
gd5sp6      < tCAAttCCCATTTAAACAAGAtTCTCAATcTTTTGCATCtCAGTATATAATAAGtTCAGA 120
3gd11s      > tCAATTCCCATTTAAACAAGATTTTCAATCTTBBTGCATCTCAGTATATAATAAGTtTAGA 139
3gd10       > TCAATTCCCANTTAAACAAGaTTTTCAaTtTKtTGCATCTCAGTATATAATAAGTtTAGA 153
3gd12sp6    > ccAATTCCCATTTAAACAAGATTTTCAATTTKKTGCATCTCAGTATATAATAAGTTTAGA 161
gd9sp6      < TCAATTCCCATTTAAACAAGATTtTCAATctTTTGCATCTCAGTATATAATAAGTTYAGA 182
rfs2        < TCaATTCCCATTT                                               404
CONSENSUS   > TCAATTCCCATTTAAACAAGATCTTCAATTTTTTGCATCTCAGTATACAATAAGTTTAGA 2160

                .........+.........+.........+.........+.........+.........+

gd7s        <                   AACCATAcTGAGTTAtcCAtCttTTtAAAtcCTAAtCATTTGAAA 45
3gd13sp6    > CATATATCATTCAGCAACCATACTGAGTTATTCATCTTTTTAAATTCTAATCATTTGAAA 105
gd3s        < TATATATcATTCAGcAACCATAcTGAGTTATTCATCTTTTTAAATYCTAATCATTTGAAA 106
gd7sp6      < TATATATCATTCAGCAACCATACTGAGTTATTCATCTTTTTAAATtCTAATCATTTGAAA 121
690gd5      < TATATATCATTCAGCAACCATACTGAGTTATTCATCTttTTAAATtCTAATCATTTCAAA 143
gd5s1       < TATATATCATTCAGCAACCATACTGAGTYATTCATCTcctcAAATcCTAATCAtcTGAAA 147
6903gd11    > TATATATCATTCAGCAACCATACTgAGTTATTCATCTTTTTAAATTCTAATCATTTGAAA 176
gd5sp6      < TATATAtCATtCAGCAACCATACTGAGttATtCATCTTTTTAAATYCTAATCATTTGAAA 180
3gd11s      > TACATATCAcTCAGCAACCATACTGAGTTATTCATCTBTBTAAATTCTAATCATTTGAaA 195
3gd10       > TACATATCATTCAGCAACCATACTGAGTTATTCATCTKTTTAAATTCTAATCATTTGAAA 213
3gd12sp6    > TACATATCATTCAGCAACCATACTGAGTTATTCATCTtTTTTAAATTCTAATCATTTGAAA 221
gd9sp6      < TATATATCATTCAGCAACCATACTGAGTTATTCATCTTTtTtAAATYCTAATCATttGAAA 242
CONSENSUS   > TATATATCATTCAGCAACCATACTGACTTATTCATCTTTTTAAATTCTAATCATTTGAAA 2220

                .........+.........+.........+.........+.........+.........+

gd7s        < TTGGGCTTTtTATTtAMTCATATTTAATATTTATTTGtCGTCTAAATATCCCTAATTTtAC 165
3gd13sp6    > TTGGGCTTTTATTTACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 165
gd3s        < TTGGGcTTTtATTtAYTcATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTHAC 166
gd7sp6      < TTGGGCTTTTATTtACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 181
690gd5      < tTGGGCTTTTATTTACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 203
gd5s1       < TTGGGCTttYATTCACTCATAtctAATATTtATcTGTCGTCTAAATATTCCTaATttYAC 207
6903gd11    > TTGGGCTTTTATTTACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 236
gd5sp6      < TTGGGCTTTTATTtACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 240
3gd11s      > TTGGGCTTtTaTTTACTCATATTTAATATTTATBTGTCGTCTAAaTATTCCTAATTtTAC 255
3gd10       > TTGGGCTTTTATTTACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 273
3gd12sp6    > TTGGGCTTTTATTTACTCATATTTAATATTTATtTGtCGtCTAAATATTCCTAATTGTAC 281
gd9sp6      < TTGGGCTTTtATTtACTCATATTtAATATTTATtTGTCGTCTAAATATTCCTAACTTYAC 302
CONSENSUS   > TTGGGCTTTTATTTACTCATATTTAATATTTATTTGTCGTCTAAATATTCCTAATTTTAC 2280

                .........+.........+.........+.........+.........+.........+

gd5s        <                                   AATGGTGAtAtATGACATCAAATTtCAATAACA 33
gd7s        < TTCTtAGATGTAGTTAATTAGtcCCTtAATGGTGATATATGACATCAAATttCAATAACA 165
3gd13sp6    > TTCTTAGATGTAGTTAATTAGTTCCTTAATGGTGATATATGACATCAAATTTCAATAACA 225
gd3s        < TTcTaAGATGTAGTtAATTAGTTCCTtAATGGTGATATATGACATCAAATtTCAATAAcA 226
gd7sp6      < TTCTTAGATGTAGTTAATTAGTTCCTTAATGGTGATATATGACATCAAATTtCAATAACA 241
690gd5      < TTCTTAGATGTAGTTAATTAGTTCCTTAATGGTGATATATGACATCAAATtTCAATAACA 263
gd5s1       < TTCTYAGATGTAGTYAATtAGTtCCTYAATGGTgATATATGACATCAAATcTCAATAACA 267
6903gd11    > TTCTTAGATGTAGTTAATTAGTTCCtTAATGGTGATATATGACATCAAAYTTCAATAACA 296
gd5sp6      < TTCTtAGATGTAGTTAATTAGTcCCTTAATGGTGATATATGACATCAAATTtCAATAACA 300
3gd11s      > TTCTTAGATGTAGTTAATTAGTTCCTTAATGGTGATATATGACATCAAATTTCAATAACA 315
3gd10       > TTCTTAGATGTAGTTAATTAGTTCCtTAATGGTGATATATGACATCAAATTTCAATAACA 333
3gd12sp6    > TTCTtAGATGTAGTTAATTAGTTCCYTAATGGTGATATATGACATCAAATTTCAATAACA 341
gd9sp6      < TTCTtAGATGTAGTTAATTAGTTCCTtAATGGTGATATATGACATCAAATtTCAATAACA 362
```

```
CONSENSUS   > TTCTTAGATGTAGTTAATTAGTTCCTTAATGGTGATATATGACACCAAATTTCAATAACA 2340

                .........+.........+.........+.........+.........+.........+

ngd5        <                                  TGTGttgAABTATGGTGGCTTGTGATAATACTAATCT  37
gd9         <                TAAATTCATATATTtTTTGTGTTCAATTATGGTGGCTTGTGATAATACTaATCT  53
gd5s        < TcTGATtcAAAgTCATATAtctTTGTGTtcAALTATGGTGGCTTGTGATAATACTAATCT  93
gd7s        < TTTGATTtAAATTCATATATTTTTGTGTTaAATTATGGTGGCTTGTGATAATACTAATCT 225
3gd13sp6    > TTTGATTTAAATtCATATATTtTStGTgTTCAATTATGGTGGCTTGTGATAATACTAATCT 285
gd3s        < TtTGATTNAAATTaATATATtTTTGTGTTcAATTATGGTGGCTTGTGATAATACTAATCT 286
gd7sp6      < TTTGATTTAAATTCATATATTTTTGTGTTCAATTATGGTGGCTTGTGATAATACTAATCT 301
690gd5      < TTTGATTtAAATTCATATATTTTTGTGTTCAATTATGGTGGCTTGTGATAATACTAATCT 323
gd5s1       < TcTgATtcAAAcTCATATAtcYttGTGTTCAATTATGGTGGCTTGTgATAATACTAATCT 327
6903gd11    > TTTGATTTAAATTCATATATtTYtGTgTTCAATTATGGTGGCTTGTcATAATACTAATCT 356
gd5sp6      < TcTGAtTtAAAtTcATATATTTTTGTGTTcAATTATGGTGGCTTGTGATAATACTAATCT 360
3gd11s      > TTTGATTTAAATTCATATTtttGTGTTCAATTATGGTGGCTTGTGATAATACTAATCT 375
3gd10       > TTTGATTTAAATTCATATATTTYTGTGTTCAATTATGGTGGCTTGTGATAATACTAATCT 393
3gd12sp6    > TTTGATTTAAATTCATATATTTgtGTGTCCAATtATGGtGGCTtGTGATAATACTAATCT 401
gd9sp6      < TTTGATTtAAATTCATATATTTTTGTGTTCAATTATGGTGGCTTGTGATAATACTAATCT 422
CONSENSUS   > TTTCATTTAAATTCATATATTTTTGTGTTCAATTATGGTGGCTTGTGATAATACTAATCT 2400

                .........+.........+.........+.........+.........+.........+

3gd15sp6    >                                                       GAGTG   5
3gd15s1     >                                          GGAAGAGaTGAACCATTTGTaTTTTcTCTtTCAGTc  36
ngd5        < GGTTTATATGGtCTTCTCTTAAGTGGAAGAGATGAACCATTTGTATTtTCTGTYTGAGTc  97
gd9         < GGTTTATATGGTCTTTTCTTaAGTGGaAGAGATGAACCATTTGTATTTTCTGTTTCAGTT 133
gd5s        < GGtTTATATGGtCTTTcCCTYAAGTGGAAGAGATGAACCATTTGTATTTTCTGTTcTGAGTc 153
gd7s        < GGTTTATATGGTCTTTcTCTtAAGTGGAAGAGATGAACCATTTGTATTTTCTGTTTGAGTc 285
3gd13sp6    > CGTTTATATGGtCTTTTCTtAAGtGGAACAGATGAACCATTTGTATtTTCTGTTTGAGTT 345
gd3s        < GGTTTATATGGTCTTaTCTtAAGTGGAAGAGATGAACCATTTGTATTTTTTGTTTGAGTN 346
gd7sp6      < CGTTTATATGGTCTTcTCTtAAGTGGAAGAGATGAACCATTTGTATTTTCTGTTTGAGTc 361
690gd5      < GGTTTATATGGTCTTYTCTTAAGTGGAAGAGATGAACCATTTGTATTTTCTGTTTGAGTY 383
gd5s1       < GGTYtATATGGTCTTcTCTTAAGTGGAAGAGATGAACCATtTGTATTYTCTGTtTGAGTc 387
6903gd11    > GGTTTTATATGGTCTttTCTTAAGTGGAAGANAtGAACCAtTTgTATTTTCTGTTTNAGTc 416
gd5sp6      < GGTTTTATATGGTCTTcTCTTAAGTGGAAGAGATGAACCATTTGTATTTTCTGTTTGAGTc 420
3gd11s      > GGTTTTATATGGTCTtgTCTTAAGTGGAAGAGATGAACCATTTGTaTTTTCTCTBTGAGTc 435
3gd10       > GGGTTATATGGTCTTTTCTTAAGTGGAAGA                                 423
3gd12sp6    > CGTCTATATGGGCTTTTCTTAAGTCGAAGAGAtGAACCAtttGTAtTTTCTGTKTGAGTT 461
gd9sp6      < GGTTTATATGGTCTTcTCTtAAGTGGAAGAGATGAACCATTTGTATTTTCTGTTTGAGTc 482
CONSENSUS   > GGTTTATATGGTCTTtTCTTAACTCGAAGAGATGAACCATTTGTATTTTCTGTTTGAGTT 2460

                .........+.........+.........+.........+.........+.........+

3gd15sp6    > TCAATATCTgGCGTTTGaGGGGtATCtGattGGGtAGAGCtTTCTgTAGTtGGAATATCT  65
3gd15s1     > TCAATATCTTGCGTTTGTTGGATATCTGTtTGGGTAGAGCTTTCTTTAGTTGGTATATCT  96
ngd5        < TCAATATCTTGCGTtTGTTGGATATCTGTTTGGGTAGAGCTcTCTTAGTTGGTATATCT 157
gd9         < TCaATaTCTTgCGTTTgTTGGATaTCTGTttgGgtAgagCTTTCtTTAGtTGGT      167
gd5s        < tcAATATCTtGCGTcTGTtGGATATCTGTTYGGGTAGAGCTtTCTTtAGTTGGTATATCT 213
gd7s        < tcAATATCTTGCGTTTGTTGGATATCTGTTTGGGTAGAGCTTTCTTTAGTTGGTATATCT 345
3gd13sp6    > TCAATAtCtCGCGtTTGttGGAtATCTGTgTGGGtAGAGCTttCttAGttGGtATATCT 405
gd3s        < TSAATATCTTGCGTtTGTTGGATATCTGTTtGGGTAGAGCTWTCTTTAGTTGGTATATCT 406
gd7sp6      < TCAATATCTTGCGTTTGTTGGATATCTGTTTGGCTAGAGCTTTCTTTAGTTGGTATATCT 421
690qd5      < TCAATATCTTGCGTTTGTTGGATATCTGTTTGGGTAGAGCTctCTTTAGTTGGTATATCT 443
gd5s1       < TCAATATCTYGCGTYTGTTGGATATCTGTYTGGGTAGAGCTctCTttAGTTGGTATATCT 447
5903gd11    > TCAATATCTTGCGTtTgTTGGATATCTgtcT.GGTAGA.CGTTCTTTAGTTGGTATATCT 476
gd5sp6      < TCAATATCTTGCGTTTGTTGGATATCTGTTTGGGTAGAGCTTTCTTTAGTTGGTATATCT 480
3gd11s      > TCAATATCTTGCGTTTGTTGGaTATCTGTtTGGGTAGAGCTTTCTTTAGTTGGTATATCT 495
3gd12sp6    > TCAATAtCttGCGttTGttG                                          481
gd9sp6      < TCAATATCTtGCGT                                                495
CONSENSUS   > TCAATATCTTGCGTTTGTTGGATATCTGTTTGGGTAGAGCTTTCTTTAGTTGGTATATCT 2520

                .........+.........+.........+.........+.........+.........+

6903gd16    >                                                    AACATCCATATCA  13
3gd16sp6    >                                        ATTgTATATaGgAgtgAGTAGGTTCATCAACATCCATATCA  41
3gd14sp6    >                                   CGGATTGTATATGGGAGtgAGTAGGTTcATcAACATCCATATCA  44
3gd16s      >                                ATCGGATTTTATATaggAGTgaGTAGGTTCATcAACATCCATATcA  46
3gd15sp6    > GtTtGAGTAATAGCATCGGAggGKAGAttGGAGTGAGTAGGtTCATCAACATCCATATCA 125
3gd15s1     > GTtTGAGTAATAGCATCGGATTGTATATGGGAGtgAGTAGGTTCATCAACATCCATATCA 156
ngd5        < GTcTGAGTAATAGCAtCGGATTGTATATGGGAGTGAGTAGGTTCATCAACATCCATATCA 217
gd5s        < GTctGAGTAATAGCATCGGATtGTATATGGGAGTGAGTAGGTTCATCAACATCCATATcA 273
gd7s        < GTTTGAGTAATAGCATCGGATTGTATATGGGAGTGAGTAGGTTCATCAACATCCATATCA 405
3gd13sp6    > GTcTgAGTAATAGCATCGGAtTgtATATagSAGtgAGTAGGTtCAtCAACAtCCA       450
gd3s        < GTWtGAGTAATAGCATCGGATtgTATATGGGAGTGAGTAGGTTCATcAACATCCATATcA 466
gd7sp6      < GTTTCAGTAATAGCATCGGATTGTATATGGGAGtGAGTAGGTTCATCAACATCCATATCA 481
690gd5      < GTtTGAGTAATAGCATCGGATTGTATATGGGAGTGAGTAGGTTCATCAACGT          495
gd5s1       < GTctGAGTMATAGCATCGGATtGTATATGGGAGTGAGTAGGTTCATCAACATCCATATCA 507
```

```
6903gd11    > GTCTNAGTAATANCATCgNACTNTATATNGNAGTGAGTAGGGTCATCAACAtCCATATCA 536
gd5sp6      < GTYTGAGTAATAGCATCGGATTGTATATGGGAGtGAGTAGGTTCATCAACATCCATATCa 540
3gd11s      > GTTTGAGTAATAGCATCGGaTTGTATATGGGaGtGAGTaGGTTCATCAaCATCCATATCA 555
CONSENSUS   > GTTTGAGTAATAGCATCGGATTGTATATGGGAGTGAGTAGGTTCATCAACATCCATATCA 2580

                .........+.........+.........+.........+.........+.........+

6903gd16    > CTGTCTgATGATATTTCCAATTCAAAAATTTCAGGtTCTAGTTCTTCCTGATAAACTTGA 73
3gd16sp6    > CTGTCTgATGATATTTcCAATTCAAAAaTTTCAGGTTCTAGTTCTTCCTGATAAACTTGA 101
3gd14sp6    > CTGTCTGATGACATTTCCAATTCAAAAATTTCAGGTTCTAGTTCTTCCTGATAAACTTGA 104
3gd16s      > CTGTCTgATGAtATTTcCAATTCAAAAaTTTCAGGTTcTAGTTCTTCCTgATAAACTTgA 106
3gd15sp6    > CTGtCTGATGACATTKCCAAttCAAAAAtttCAGGtTCTAGTTCTTtCCtGATAAACTtGA 185
3gd15s1     > CTGTCTGATGATATTTCCAATTCAAAAATTTCAGCTTCTAGTTCTTCCTGATAAACTTGA 216
ngd5        < cTGTCTGATGATATTTCCAATTCAAAAA                                  245
gd5s        < CTGTCTGATGATATTTCCAA                                          293
gd7s        < CTGTCTGATGATATtt                                             421
gd3s        < CtKtcTGATGATATTTCCAATTcAAAAATDTcAGGTTATAGTT                   509
gd7sp6      < CTGTCtYWTGATATTTCC                                           499
gd5s1       < CT                                                           509
6903gd11    > C                                                            537
gd5sp6      < MTGTCTGATGATATTTCcAaTT                                       562
3gd11s      > CTGTCTGaTGaTA.TTCC.ATTCAAA                                   581
CONSENSUS   > CTGTCTGATGATATTTCCAATTCAAAAATTTCAGGTTCTAGTTCTTCCTGATAAACTTGA 2640

                .........+.........+.........+.........+.........+.........+

6903gd16    > GTTGATATGTTAGCtTGTCTCCTCTGGGGGtTTTTTARNtcttCGtTCCAGAGGCACCtCT 133
3gd16sp6    > GTTGATATGTTAGCtTGTCTCCTCTGGGGKTTTTAVBtCTCCGTTCCAGAGGCACCtCT 161
3gd14sp6    > GTTGATATGTTAGCTTGTcTCCTCTGGGGTTTTTTAATTCTTCGTTCCACAGGCACCTCT 164
3gd16s      > GTTGATATGTTAGCtTGTMTCCTcTgGGGgTTTTTARtTcTTCGTTCCAgAtGCACctcT 166
3gd15sp6    > GtTGATATGtTAGCTTGtCtCCCCTGGGG                                 214
3gd15s1     > GTTGATATGTTAGCTTGTCTCCTCTGGGGCTTTTTAATTCTTCGTTCCAGAGGCACCTCT 276
CONSENSUS   > GTTGATATGTTAGCTTGTCTCCTCTGGGGCTTTTTAATTCTTCGTTCCAGAGGCACCTCT 2700

                .........+.........+.........+.........+.........+.........+

6903gd16    > ACTTGCAATCCAGATGAACTACTTATTGTATCTaATAATGAACTGGtTGTgTAACCTGTG 193
3gd16sp6    > ACtTGCAATCCAGATGAaCTACTTATTGTATCTKATAATGAACtGGtTGtgTgACCcgTG 221
3gd14sp6    > ACTTGCAATCCAGATGAATTACTTATTGTATCTGATAATGAACTGGTTGTGTGACCTGTG 224
3gd16s      > ACtTgCAATCCAgATGAaTTASTTATTGTATCTtATAATGAACTgGtTGTGTgACCtGTG 226
3gd15s1     > ACTTGCAATCCAGATGAATTACTTATTGTATCTGATAATGAACTGGTTGTGTgACCCGTG 336
CONSENSUS   > ACTTGCAATCCAGATGAATTACTTATTGTATCTGATAATGAACTGGTTGTGTGACCTGTG 2760

                .........+.........+.........+.........+.........+.........+

387-8       <                                   CATTATATCTGAACTTAAGAATtTCTTCAGTTTCT 35
6903gd16    > TGTTCTCTTGAATKAGTTGGTCTTTCATTATATCTGAACtTAAGARTgTCTTCAGTtTCT 253
3gd16sp6    > TGTTCTCTTGAATtAgTTGGTCTtTCATTATATCTgAACtTAAGAVTBTCTTCAGtSTCT 281
3gd14sp6    > TGTTCTCTTGAATGAGTTGGTCTTTCATTATATCTGAACTTAAGAATTTCTTCAGTTTCT 284
3gd16s      > TGTTCTCTTgAATcAGTTGGTCTTTCATTATATCTtAACtTAAGARTtTcTTCAGTtTCT 286
3gd15s1     > TGTTCTCTTGAATGAGTTGGTCTTTCATTATATCTGAACTTAAGAATTTCTTCAGTTTCT 396
CONSENSUS   > TGTTCTCTTGAATGAGTTGGTCTTTCATTATATCTGAACTTAAGAATTTCTTCAGTTTCT 2820

                .........+.........+.........+.........+.........+.........+

387-8       < GAAGAAGTCTCAGTGTCTTCATttGAAGGAGTAACATACTTTCTTCTCTTAtTCTTACCT 95
6903gd16    > GAAGAAGttCAGTGTCTTCATTTGAAGGAGTAACATACTTTCTTCTCTTATTCTTACCT 313
3gd16sp6    > CAAGAAGtBTCAGTGTCTTCATTTgAAGGAGTAACATACTTTcTTCTCTTATTCTTACCt 341
3gd14sp6    > GAAGAAGTTTCAGTGTCTTCATTTGAAGGAGTAACATACTTTCTTCTcTTATTCTTACCT 344
3gd16s      > GAAGAAGttTCAGTgTCTTCATTTgAAGGAGTAACATAStTTcTTCTcTTATTCTTACCT 346
3gd15s1     > GAAGAAGTTTCAGTGTCTTCATTTGAAGGAGTAACATACTtTCTtCTCTNATtCTTACCT 456
CONSENSUS   > GAAGAAGTTTCAGTGTCTTCATTTGAAGGAGTAACATACTTTCTTCTCTTATTCTTACCT 2880

                .........+.........+.........+.........+.........+.........+

387-8       < TTTGGATGTTTTGAAAACCCTYYtTTGTTCCTTTGACGTGTTGAAGGATCGGTGGTATG 155
6903gd16    > TTTGGATGtSTTGAAAACCCttTTTTGTTCCTTTGACGTRTTCAAGGgATCDGtGGTATG 373
3gd16sp6    > TTTGGATGSSTTGAAAACCCgTTTtTGTTCCttGACGTgTTGAAGGgATCgGTGGTATG 401
3gd14sp6    > TTTGGATGTTTTGAAAACCCTTTTTTGTTCCTTTGACGTGTTGAAGGGATCGGTGGTATG 404
3gd16s      > TTTGGATGtTTTgAAAACCStTTTTTGTTCCTTTGACGTRTTGAAGGGATCCGTGGTATg 406
CONSENSUS   > TTTGGATGTTTTGAAAACCCTTTTTTGTTCCTTTGACGTGTTGAAGGGATCGGTGGTATG 2940

                .........+.........+.........+.........+.........+.........+

3gd19sp6    >                                                 TTCTTTTTGTATAGTACTA 19
3gd19s      >                                                 TCTCCTtttGTATAGTACTA 21
n3gd19      >                        GTTTTAGGATGAAAAATCTTTGTATtStCTTCTTTTTGTATAGTACTA 48
3gd16       >                      AGTTttAGGATGAAAAATCTttGTATtGTCTtCTttTTGTATAGTACTA 49
```

```
3gd19       >          TTCAGTTTTAGGATGAAAAATCTTTGTATTgTCTTCTTTTTGTATAGTACTA 52
387-5       <        TATTCAGTTTTAGGATGAAAAATCTTTGTATTGTCTTCTtTTTGTATAGTACTA 54
gd1sp6      <     AGtAtTCAGTtTTAGGATGAAAAAtCTtTGTATTGTCTtCtTTTTGTATAGTACTA 56
387-8       < GCATAGTATTCAGTTTTAGGATGAAAAATCTTTGTATTGTCTTCTtTTGTATAGTACTA 215
6903gd16    > GGATAGTATTCAGTTTTAGGATGAAAAATCTTTGTATTGtCTtCTTtTtGTATAGTACTA 433
3gd16sp6    > GgATAGTATTCAGTtTTAGGATGAAAAATCtTTGTATTGTCTTCTttTTGTATAGtACTA 461
3gd14sp6    > GGATAGTATTCAGTTTTAGGATGAAAAATCTTTGTATTGTCTTCTTTTTGTATAGTACTA 464
3gd16s      > GGATAGTATTCAGTTTTAGGATgAAAAATCTTTgTATTgTCTTCTTTTTGTATAGTACTA 465
CONSENSUS   > GGATAGTATTCAGTTTTAGGATGAAAAATCTTTGTATTGTCTTCTTTTTGTATAGTACTA 3000

                .........-.........+.........-.........+.........+.........-

3gd19sp6    > AGACTTAAaATCTTttCTTGATAAGTtTTTTGAtCAATTTCAAAATAtCATCTTTtCTT 79
3gd19s      > AGACttAAAATCttTTCTTCATAAGtctYTTgAtCAAttTCAAAATATTCATCTttTCTt 81
n3gd19      > AGACTTAAAATCTTTTCTTGATAAGTTTTTTGATCAATTTCAAAATATTCATCTTTTCTT 108
3gd16       > AGACTTAAAATCTTTtCTtGATAAGttTTtTtGAtCAATTTCAAAATATTCATCTtTTCTT 109
3gd19       > AGACTTAAAATCTTTTCTTGATAAGTTTTTTGATCAATTTCAAAATATTCATCTTTTCTT 112
387-5       < AGACTTAAAATCTTTCTTGATAAGTTTTTTGATCAATtTCAAaATATTCATCTTTTCTT 114
gd1sp6      < AGACTTAAAATCTTTTCtTGAtAAGtTTTTTGATCAATTTCAAAATATTCATCTTTTCTT 116
387-8       < AGACTtAAAATCTTTTCTTGATAAGTTTTTTGATCAATtTCAAAATATTCATCTTTTCTT 275
6903gd16    > AGACTTAAAATCTTTTCTTGATAAGtTTTttGAACAATttCAAAATATtCATCT 487
3gd16sp6    > AGACTTAAAATC 473
3gd14sp6    > AGACTTAAAATCTTTTCTTGATAAGTTTTTTGATCAATTTCAAAATATTCATCTTTTCTT 524
3gd16s      > AgACTTAAAATcTTTTCTTgATAAGTTTTTTGATCAATTTCAAAATATTCATcTTTTCTT 526
CONSENSUS   > AGACTTAAAATCTTTTCTTGATAAGTTTTTTGATCAATTTCAAAATATTCATCTTTTCTT 3060

                .........+.........+.........+.........-.........+.........+

3gd19sp6    > TtGCAATATATCTttTtGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 139
3gd19s      > TTGCAATATATCTttTTgTgACGtAGATgCCGgAAAATCAATtTCAATATCGACAATtACt 141
n3gd19      > TTGCAATATATCTttTTGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 168
3gd16       > TTGCAATATATCTTTTTGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 169
3gd19       > TTGCAATATATCTTTTTGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 172
387-5       < TTGCAATATATCTTTTTGTGACGTAGATGCCGGAAAATCAATtTCAATATCGACAATTACT 174
gd1sp6      < TTGCAATATATCTTTTTGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 176
387-8       < TTGCAATATATCTTTTTGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 335
3gd14sp6    > TTGCAATATATCTTTTTGTGACGTAGATGCCGGaAAATCAATTTCAATATCGACAATTACT 584
CONSENSUS   > TTGCAATATATCTTTTTGTGACGTAGATGCCGGAAAATCAATTTCAATATCGACAATTACT 3120

                .........+.........+.........+.........+.........+.........+

3gd19sp6    > TCTcCAGAATATTTAAATttTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGAGCA 199
3gd19s      > TCTCCAgAATATtTAAaTttTgTAATTgTAATCATTCTgCATTTCTCACCTAACAgAgCA 201
n3gd19      > TCTCCAGAATATTTAAaTTTTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGAGCA 228
3gd16       > TCTCCAGAATATTTAAATTTTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGAGCA 229
3gd19       > TCTCCAGAATATTTAAATTTTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGACCA 232
387-5       < TCTcCAGAATATTTAAATTTTGTAATTGTAATCATTCTGCATtTCTCACCTAACAGAGCA 234
gd1sp6      < TCTcCAGAATATTTAAATTTTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGACCA 236
387-8       < TCTcCAGAATATTTAAATTTTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGAGCA 395
3gd14sp6    > TCTcCAGAATATTTA 599
CONSENSUS   > TCTCCAGAATATTTAAATTTTGTAATTGTAATCATTCTGCATTTCTCACCTAACAGACCA 3180

                .........+.........+.........+.........+.........+.........+

3gd19sp6    > GACCATACCAATTtTTCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 259
3gd19s      > GACCAtACCAatttTTCTCCATCATAAACCTTTCTAATTgGtTgGtTATTGGTAGGaTAA 261
n3gd19      > GACCATACCAaTgtGTCTCCATCATaaACCTNTCTAATtGGHTGGTTATTGGtAGGaTAA 288
3gd16       > GACCATACCAATTtTtCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 289
3gd19       > GACCATACCAATTTTTCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 292
387-5       < GaCCATACCAATTTTTCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 294
gd1sp6      < GACCATACCAATTTTTCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 296
387-8       < GACCATACCAATTTTTCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 455
CONSENSUS   > GACCATACCAATTTTTCTCCATCATAAACCTTTCTAATTGGTTGGTTATTGGTAGGATAA 3240

                .........+.........+.........+.........+.........+.........+

3gd19sp6    > TttTGAaTtGTATATATCACTtATATTtTAATAAGtCTGAAGACACAAGATTTACATAAGA 319
3gd19s      > TttTgAaTtgtATATATCACttATATTTTAATAAGTCTgAAGACACAAGatTTACATAAgA 321
n3gd19      > T 289
3gd16       > ATTTTTATCTTAGTCATTCCATTTTCAGCATATTCGACTATCGGTAATTTTGAATTGTAT 349
3gd19       > ATctctaTCTTAGTCATTCCATTTTCA.CATATTCgaCYATCGgTAATTTTGAATTGTAT 352
387-5       < ATcTttATCTTAGTCATTCCATTTTCAGCATATTCGACTATCGGTAATTTTGAATTGTAT 354
gd1sp6      < ATTTttATCTTAGTCATTCCATTTTCAGCATATTCGACTATCGGTAATTTTGAATTGTAT 356
387-8       < ATtTttATCTTAGTCATTCCATTTTCAGCATATTCGACTATCGGTAATTTTGAATTGTAT 515
CONSENSUS   > ATTTTTATCTTAGTCATTCCATTTTCAGCATATTCGACTATCGGTAATTTTGAATTGTAT 3300

                .........+.........+.........+.........+.........+.........+
```

```
3gd19sp6    > GAAAGTTAGTACAAGAACTAGA                                    341
3gd19s      > gAAaGttAGTAC                                              333
3gd16       > ATATCACTTATATTTAATAAGTCTGAACACACAAGATTTACATAAGAGAAAGTTAGTACA 409
3gd19       > ATATCACTTATATTTAATAAGTCTGAAGACACAAGATTTACATAAGAGAAAGtTAGTACA 412
387-5       < ATATCACTTATATTTAATAAGTCTGAAGACACAAGATTTACATAAGAGAAAGTTAGTACA 414
gd1sp6      < ATATCACTTATATTTAATAAGTCTGAAGACACAAGATTTACATAAGAGAAAGTTAGTACA 416
387-8       < ATATCACTTATATTTAATAAGTCTGAAGACACAAGATTTACATAAGAGAAAGTTAGTACA 575
CONSENSUS   > ATATCACTTATATTTAATAAGTCTGAAGACACAAGATTTACATAAGAGAAAGTTAGTACA 3360

                .........+.........+.........+.........+.........+.........+

3gd16       > AGAACTAG                                                  417
3gd19       > AGAaCTAG                                                  420
387-5       < AGAAC                                                     419
gd1sp6      < AGAACTAG                                                  424
387-8       < AGAACTA                                                   582
CONSENSUS   > AGAACTAG                                                  3368

                .........+.........+.........+.........+.........+.........+
```

Appendix D PSORTII Analysis of the Predicted TashAT1 Polypeptide

Input Sequence

```
QUERY (466 aa)

MMVVLKLSHI IFTLFLYRVK FASSEILYLD NLDNPNFYTI KIVEDRLTKI
MILSTPEDKT TEIRSKRKLI WGSDRGEYVK CFTRFSFESS DKTLITIEIG
NAVDEAMKFI YVSGNFYKYI NKSEFEDYYK SFCSVFIKIP PGKLPTPRLK
KNVKTEKVDK RKLKRDRQRK DKPQSEQHDK NVDIVSQSLA EEGIDLEKKT
VGREEPTQQT EKQQEPTELE PETIPVELRG DDERIDRSNV SKPKESDGIL
TQNRYTQTDI QEIEDIGIQT RIIELENIVT QTDIQTKESS IQTDIQEVED
IDTQTDIQEL ENIGIQTIGN FSDTTEVTKK HEQPEVPKRR PGRPRKQKPE
PEQPKRKRGR PRKQKYETKK TWLLRPRNMK TETKKTWLLR PRKQKPEPEQ
PKRKRGRPRK QKYETKKTWL LRPRNMKTET KKTWLLRPRK HKPEPEQPKR
KRGRPRKQKP EPSSDT
```

Results of Subprograms

```
PSG:  a new signal peptide prediction method
      N-region:  length 6;  pos.chg 1;  neg.chg 0
      H-region:  length 11;  peak value  10.31
      PSG score:   5.91

GvH:  von Heijne's method for signal seq. recognition
      GvH score (threshold: -2.1):  -5.36
      possible cleavage site: between 24 and 25

>>> Seems to have no N-terminal signal peptide

ALOM: Klein et al's method for TM region allocation
      Init position for calculation: 1
      Tentative number of TMS(s) for the threshold  0.5:   1
      Number of TMS(s) for threshold  0.5:   1
      INTEGRAL    Likelihood = -2.76   Transmembrane    1 -  17
      PERIPHERAL  Likelihood =  6.15 (at 94)
      ALOM score:  -2.76  (number of TMSs: 1)

MTOP: Prediction of membrane topology (Hartmann et al.)
      Center position for calculation: 8
      Charge difference:  0.0   C( 2.5) - N( 2.5)
      N >= C: N-terminal side will be inside

>>> membrane topology: type 2   (cytoplasmic tail 1 to 1)

MITDISC: discrimination of mitochondrial targetting seq
      R content:        1       Hyd Moment(75):  3.96
      Hyd Moment(95):  5.28     G content:       0
      D/E content:     1        S/T content:     4
      Score: -3.27

Gavel: prediction of cleavage sites for mitochondrial preseq
      R-2 motif at 28  YRV|KF

NUCDISC: discrimination of nuclear localisation signals
      pat4: PKRR (4) at  337
      pat4: KRRP (4) at  338
      pat4: RPRK (4) at  343
      pat4: PKRK (4) at  354
      pat4: KRKR (5) at  355
      pat4: RPRK (4) at  360
      pat4: RPRK (4) at  390
      pat4: PKRK (4) at  401
      pat4: KRKR (5) at  402
      pat4: RPRK (4) at  407
      pat4: RPRK (4) at  437
      pat4: RKHK (3) at  439
      pat4: PKRK (4) at  448
      pat4: KRKR (5) at  449
```

```
      pat4: RPRK (4) at  454
      pat7: PIPRLKK (3) at  145
      pat7: PRLKKNV (5) at  147
      pat7: PEVPKRR (3) at  334
      pat7: PKRRPGR (5) at  337
      pat7: PGRPRKQ (4) at  341
      pat7: PRKQKPE (5) at  344
      pat7: PKQPKRK (3) at  351
      pat7: PKRKRGR (5) at  354
      pat7: PRKQKYE (5) at  361
      pat7: PRKQKPE (5) at  391
      pat7: PEQPKRK (3) at  398
      pat7: PKRKRGR (5) at  401
      pat7: PRKQKYE (5) at  408
      pat7: PRKHKPE (5) at  438
      pat7: PKQPKRK (3) at  445
      pat7: PKRKRGR (5) at  448
      pat7: PRKQKPE (5) at  455
      bipartite: KKNVKTEKVDKRKLKRD at  150
      bipartite: RKQKPEPRQPKRKRGRP at  345
      bipartite: RKQKPEPEQPKRKRGRP at  392
      bipartite: RKHKPEPEQPKRKRGRP at  439
      content of basic residues:  21.2%
      NLS Score: 12.71

KDEL: ER retention motif in the C-terminus: none

ER Membrane Retention Signals: none

SKL: peroxisomal targeting signal in the C-terminus: none

SKL2: 2nd peroxisomal targeting signal:  none

VAC: possible vacuolar targeting motif: found
      KLPI at 143

RNA-binding motif: none

Actinin-type actin-binding motif:
      type 1: none
      type 2: none

NMYR: N-myristoylation pattern : none

Prenylation motif: none

memYQRL: transport motif from cell surface to Golgi: none

Tyrosines in the tail: none

Dileucine motif in the tail: none

checking 63 PROSITE DNA binding motifs: none

checking 71 PROSITE ribosomal protein motifs:  none

checking 33 PROSITE prokaryotic DNA binding motifs: none


NNCN: Reinhardt's ,method for Cytoplasmic/Nuclear discrimination
      Prediction: nuclear
      Reliability: 94.1

COIL: Lupas's algorithm to detect coiled-coil regions
total: 0 residues
```

Results of the *k*-NN Prediction

$k = 9/23$

```
          39.1 %: nuclear
          21.7 %: cytoplasmic
          17.4 %: mitochondrial
           8.7 %: cytoskeletal
           8.7 %: plasma membrane
           4.3 %: vesicles of secretory system

>> prediction for QUERY is nuc (k=23)
```

Appendix E SignalP Analysis of the TashAT1 Predicted Polypeptide

SignalP predictions using networks trained on eukaryotic data

TashAT1 length = 70

# pos	aa	C	S	Y
1	M	0.012	0.907	0.011
2	M	0.012	0.917	0.013
3	V	0.012	0.948	0.018
4	V	0.012	0.954	0.021
5	L	0.012	0.949	0.024
6	K	0.012	0.959	0.028
7	L	0.012	0.947	0.033
8	S	0.012	0.958	0.040
9	H	0.014	0.953	0.050
10	I	0.015	0.953	0.059
11	I	0.035	0.918	0.099
12	F	0.013	0.937	0.066
13	T	0.015	0.957	0.075
14	L	0.051	0.918	0.149
15	F	0.017	0.892	0.090
16	L	0.063	0.600	0.183
17	Y	0.050	0.572	0.166
18	R	0.019	0.831	0.103
19	V	0.092	0.730	0.235
20	K	0.042	0.787	0.162
21	F	0.023	0.853	0.125
22	A	0.277	0.685	0.443
23	S	0.395	0.499	0.538
24	S	0.030	0.363	0.147
25	E	0.644	0.188	0.680
26	I	0.057	0.202	0.198
27	L	0.060	0.125	0.197
28	Y	0.029	0.091	0.131
29	L	0.028	0.111	0.125
30	D	0.118	0.072	0.246
31	N	0.018	0.071	0.091
32	L	0.045	0.045	0.137
33	D	0.031	0.029	0.106
34	N	0.020	0.047	0.081
35	P	0.027	0.032	0.091
36	N	0.020	0.030	0.071
37	F	0.015	0.031	0.056
38	Y	0.025	0.025	0.066
39	T	0.018	0.033	0.047
40	T	0.017	0.042	0.038
41	K	0.036	0.032	0.046
42	I	0.018	0.036	0.025
43	V	0.016	0.029	0.020
44	E	0.023	0.032	0.017
45	D	0.013	0.034	0.007
46	R	0.019	0.033	0.000
47	L	0.017	0.026	0.000
48	T	0.013	0.032	0.000
49	K	0.018	0.035	0.000
50	I	0.015	0.035	0.000
51	M	0.027	0.028	0.000
52	I	0.012	0.027	0.000
53	L	0.014	0.030	0.000
54	S	0.039	0.031	0.000
55	T	0.023	0.032	0.000
56	P	0.016	0.033	0.000
57	E	0.020	0.031	0.000
58	D	0.017	0.058	0.000
59	K	0.022	0.100	0.000
60	I	0.024	0.082	0.000
61	T	0.014	0.075	0.005
62	E	0.064	0.059	0.021

```
   63    I    0.033    0.059    0.018
   64    R    0.026    0.062    0.019
   65    S    0.017    0.039    0.017
   66    K    0.026    0.043    0.022
   67    R    0.015    0.027    0.017
   68    K    0.022    0.025    0.021
   69    L    0.014    0.028    0.016
   70    I    0.012    0.029    0.015

< Is the sequence a signal peptide?
# Measure  Position  Value  Cutoff  Conclusion
  max. C     25       0.644   0.37   YES
  max. Y     25       0.680   0.34   YES
  max. S      6       0.959   0.88   YES
  mean S     1-24     0.833   0.48   YES
# Most likely cleavage site between pos. 24 and 25: ASS-ET
```

OUTPUT INTERPRETATION (Taken from SignalP analysis program)

The **SignalP** WWW server will return three scores between 0 and 1 for each position in your sequence:

C-score (raw cleavage site score)

The output score from networks trained to recognise cleavage sites *vs.* other sequence positions. Trained to be:

high at position +1 (immediately *after* the cleavage site)
low at all other positions.

S-score (signal peptide score)

The output score from networks trained to recognise signal peptide *vs.* non-signal-peptide positions. Trained to be:

high at all positions before the cleavage site
low at 30 positions after the cleavage site and
in the N-terminals of non-secretory proteins.

Y-score (combined cleavage site score)

The prediction of cleavage site location is optimised by observing where the C-score is high *and* the S-score changes from a high to a low value. The Y-score formalises this by combining the height of the C-score with the slope of the S-score.

Specifically, the Y-score is a geometric average between the C-score and a smoothed derivative of the S-score (*i.e.*, the difference between the mean S-score over *d* positions before and *d* positions after the current position, where *d* varies with the chosen network ensemble).

All three scores are averages of five networks trained on different partitions of the data.

For each sequence, **SignalP** will report the maximal C-, S-, and Y-scores, and the mean S-score between the N-terminal and the predicted cleavage site. These values are used to distinguish between signal peptides and non-signal peptides. If the your sequence is predicted to have a signal peptide, the cleavage site is predicted to be immediately before the position with the maximal Y-score.

Appendix F TargetP Analysis of the Predicted TashAT1 Polypeptide

```
T A R G E T P  1.0  prediction results

Number of input sequences:  1
Cleavage site predictions not included.
Using NON-PLANT networks.
```

Name	Length	mTP	SP	other	Loc.	RC
TashAT1	70	0.047	0.877	0.085	S	2
cutoff		0.00	0.00	0.00		

INTERPRETATION (Taken from TargetP program)

COLUMNS:

Name
Sequence name as annotated in fasta file (without initial ">") or on TargetP input page. The name may be of any length, but only 30 characters will be preserved throughout the prediction.

Length
Sequence length. Only the 130 N-terminal amino acids are used in the prediction; submitting sequences longer than 130 residues does not improve the prediction
(but it does slow down the prediction).

cTP/mTP/SP/other
The neural network output score for each of the possible categories. If non-plant version is chosen, cTP is not included as a possible location. The scores are NOT probabilities, and they do NOT necessarily add to one (1). However, the location with the highest score is the most likely one according to TargetP, and the relation between the scores may be an indication of how certain the prediction is (see column RC).

Loc.
The subcellular localisation predicted by TargetP:
C : Chloroplast, *i.e. the sequence contains a chloroplast transit peptide, cTP*
M : Mitochondrion, *i.e. the sequence contains a mitochondrial targeting peptide, mTP*
S : Secretory pathway, *i.e. the sequence contains a signal peptide, SP*:
any other location* : "do not know".
This character appears if cutoff restrictions were demanded and the winning network output score for a sequence was BELOW the requested cutoff for that category. The asterisk shows that no prediction was done by TargetP (although the output scores and RCs are presented also for these sequences).

RC
Reliability Class: a measure of the size of the difference (diff) between the highest (winning) and the second highest output scores. There are 5 reliability classes, defined as follow:

RC 1: diff > 0.800
RC 2: 0.800 > diff > 0.600
RC 3: 0.600 > diff > 0.400
RC 4: 0.400 > diff > 0.200
RC 5: 0.200 > diff

TPlen
For sequences predicted to contain a cTP/mTP/SP, this is the predicted *length* of the presequence. For SPs, SignalP is used in this prediction, and for cTPs, ChloroP is used.

Appendix G Pestfind Analysis of TashAT1 for PEST Sequences

```
ANALYSED SEQUENCE:
     ------------------------------------------------------------

YKRLMMVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDR 50
                          ----------------------

LTKIMILSTPEDKITEIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLIT 100
                                                  ooooo

IEIGNAVDEAMKFIYVSGNFYKYINKSEFEDYYKSFCSVFIKIPPGKLPI 150
oooooooooo

PRLKKNVKTEKVDKRKLKRDRQRKDKPQSEQHDKNVDIVSQSLAEEGIDL 200
                                    ooooooooooooooooo

EKKIVCREEPTQQTEKQQEPTELEPETIPVELESDDEEIDESNVSKPKES 250
o               ++++++++++++++++++++++++++++++

DGILTQNRYLQTDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQ 300
       oooooooooooooooooooooooooooooooooooooooooooooo

EVEDIDTQTDIQELENIGIQTIGNFSDITEVTKKIIRQPEVPKRRPGRPRK 350
oooooooooooooooooooooooooooooooo

QKPEPEQPKRKRGRPRKQKYETKKTWLLRPRNMKTETKKTWLLRPRKQKP 400


EPEQPKRKRGRPRKQKYETKKTWLLRPRNMKTETKKTWLLRPRKHKPEPE 450


QPKRKRGRPRKQKPEPSSDT 470


++++++    possible PEST sequences
------    poor PEST sequences
oooooo    invalid PEST sequences


POTENTIAL PEST SEQUENCES:
--------------------------------------------------------------------------

216  KQQEPTELEPETIPVELESDDEEIDESNVSK  246

mole fraction of PESST   :  58.93
hydrophobicity index     :  30.60
PEST-FIND score          :  +17.11
-----------------------------------------------

POOR PEST SEQUENCES:
--------------------------------------------------------------------------

24  KFASSEILYLDNLDNPNFYTIK  45
PEST-FIND score          :  -14.64

INVALID PEST SEQUENCES:
--------------------------------------------------------------------------

96  KTLITIEIGNAVDEAMK  112

184  KNVDIVSQSLAEEGIDLEK  202

258  RYTQTDIQEIEDIGIQTEIH  277

277  HELENIVTQTDIQTK  291

291  KESSIQTDIQEVEDIDTQTDIQELENIGIQTIGNFSDITEVTK  333
```

Appendix H Prosite Search of the TashAT1 Predicted Polypeptide

PROSCAN result for : TashAT1

```
PROSITE: Bairoch A., Bucher P. and Hofmann K. The PROSITE database, its
status in 1997 Nucleic Acids Res. (1997) Jan 1;25(1):217-221
PROSCAN : PROSCAN has been developed at IBCP.

Protein TashAT1 using PROSITE.BASE as reference site file


MMVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLIKIMILSTPEDKITEIRSKRKLIWGS
DRGEYVKCFTRFSFESSDKTLITIEICNAVDEAMKFIYVSGNFYKYINKSEFEDYYKSFCSVFIKIPPGKLPI
PRLKKNVKTEKVDKRKLKRDRQRKDKPQSEQHDKNVDIVSQSLAEEGIDLEKKIVGREEPTQQTEKQQEPTEL
EPETIPVELESDDEEIDESNVSKPKESDGILTQNRYTQTDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQ
TDIQEVEDIDTQTDIQELENIGIQTIGNFSDITEVTKKHEQPEVPKRRPGRPRKQKPEPEQPKRKRGRPRKQK
YETKKTWLLRPRNMKTETKKTWLLRPRKQKPEPEQPKRKRGRPRKQKYETKKTWLLRPRNMKTETKKTWLLRP
RKHKPEPEQPKRKRGRPRKQKPEPSSDT


Similarity percentage 100
-------------------------------------------------------------------------
-------
N-glycosylation site.
Prosite access number: PS00001
Prosite documentation access number: PDOC00001
N-{P}-[ST]-{P}.
Randomised probability: 5.138e-03 .

Site :  121 to  124 NKSE.   Identity.
Site :  239 to  242 NVSK.   Identity.
Site :  320 to  323 NFSD.   Identity.
-------------------------------------------------------------------------
-------
Protein kinase C phosphorylation site.
Prosite access number: PS00005
Prosite documentation access number: PDOC00005
[ST]-x-[RK].
Randomised probability: 1.423e-02 .

Site :   39 to   41 TIK.   Identity.
Site :   65 to   67 SKR.   Identity.
Site :   73 to   75 SDR.   Identity.
Site :   90 to   92 SDK.   Identity.
Site :  155 to  157 TEK.   Identity.
Site :  210 to  212 TEK.   Identity.
Site :  328 to  330 TKK.   Identity.
Site :  368 to  370 TKK.   Identity.
Site :  383 to  385 TKK.   Identity.
Site :  415 to  417 TKK.   Identity.
Site :  430 to  432 TKK.   Identity.
-------------------------------------------------------------------------
-------
Casein kinase II phosphorylation site.
Prosite access number: PS00006
Prosite documentation access number: PDOC00006
[ST]-x(2)-[DE].
Randomised probability: 1.482e-02 .

Site :   54 to   57 STPE.   Identity.
Site :   55 to   58 TPED.   Identity.
Site :  123 to  126 SEPE.   Identity.
Site :  188 to  191 SLAE.   Identity.
Site :  217 to  220 TELE.   Identity.
Site :  230 to  233 SDDE.   Identity.
Site :  256 to  259 TQTD.   Identity.
Site :  280 to  283 TQTD.   Identity.
Site :  303 to  306 TQTD.   Identity.
-------------------------------------------------------------------------
-------
```

```
Tyrosine kinase phosphorylation site.
Prosite access number: PS00007
Prosite documentation access number: PDOC00007
[RK]-x(2,3)-[DE]-x(2,3)-Y.
Randomised probability: min = 4.074e-04 max = 4.083e-04

Site :  122 to  129 KSEFEDYY.   Identity.
--------------------------------------------------------------------------------
-------
N-myristoylation site.
Prosite access number: PS00008
Prosite documentation access number: PDOC00008
G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.
Randomised probability: 1.397e-02 .

Site :   72 to   77 GSDRGE.   Identity.
--------------------------------------------------------------------------------
    ----
5 different patterns found.
```

PROSITE result file (text): [PROSITE]

Appendix I Predicted Transcription factors that bind to the *TashAT1* upstream region.

Adapted from a chart by MatInspector analysis of the upstream region of *TashAT1* (see Fig. 3.19). Sequence in capitals are highly conserved within the transcription factor binding site consensus. Sequence in capital letters denote the core sequence, the highest, consecutive conserved positions of the matrix, usually 4bp long (see Fig. 3.19 for explanation of Matrix. Sim scores).

Name of family/matrix	Further Information	Position	Strand	Matrix sim.	Sequence
V$CDXF/CDX2.01	Cdx-2 mammalian caudal related intestinal transcr. Factor	2- 20	(-)	0.837	TcctaatT TTActt Cttag
V$MYT1/MYT1.01	MyT1 zinc finger transcription factor involved in primary neurogenesi	7 - 18	(-)	0.842	CtAATT Ttactt
V$NKXH/NKX25.02	homeo domain factor Nkx-2.5/Csx tinman homolog low affinity sites	12 - 19	(-)	0.895	CcTAAT Tt
V$OCTP/OCT1P.01	octamer-binding factor 1 POU-specific domain	14 - 28	(-)	0.900	CtaaatAT TCctaat
V$MEF2/AMEF2.01	myocyte enhancer factor	15 - 32	(-)	0.824	TcgtcTAA Atattcctaa
V$CART/CART1.01	Cart-1 (cartilage homeoprotein 1)	31 - 48	(-)	0.854	AttTAAT Attatttgtc
V$FKHD/HNF3B.01	Hepatocyte Nuclear Factor 3beta	33 - 47	(+)	0.972	CaaataAA TAttaaa
V$HNF1/HNF1.01	hepatic nuclear factor 1	34 - 48	(-)	0.802	AtTTA Atatttatt
V$NKXH/NKX31.01	prostate-specific homeodomain protein NKX3.1	34 - 46	(+)	0.851	AaatAA ATattaa
V$SATB/SATB1.01	Special AT-rich sequence-binding protein 1 predominantly expressed in thymocytes binds to matrix attachment regions (MARs)	34 - 55	(-)	0.917	Tactcatatt TAATatt Tattt
V$BRNF/BRN2.01	POU factor Brn-2 (N-Oct 3)	38 - 53	(-)	0.939	Ctcatatt TAATattt
V$EVI1/EVI1.04	Ecotropic viral integration site 1 encoded factor	45 - 59	(+)	0.834	AAATAt Gagtaaata
V$FKHD/XFD3.01	Xenopus fork head domain factor 3	50 - 63	(+)	0.921	TgagtaAA TAaaag
V$TBPF/TATA.01	cellular and viral TATA box elements	55 69	(+)	0.880	AaaTAAA Agcccaat
V$BRNF/BRN2.01	POU factor Brn-2 (N-Oct 3)	64 - 79	(-)	0.938	AtcatttgA AATtggg
V$PDX1/ISL1.01	Pancreatic and intestinal lim-homeodomain factor	71 - 90	(-)	0.817	Ttaaaatc TAATc Atttga
V$TBPF/MTATA.01	Muscle TATA box	77 - 93	(-)	0.851	CtttTAA Attctaatc
V$MEF2/MEF2.05	MEF2	83 - 92	(+)	0.969	AattTA AAaa
V$HNF1/HNF1.01	hepatic nuclear factor 1	89 - 103	(-)	0.793	AGTTA Ttcatcttt
V$PIT1/PIT1.01	Pit1 GHF-1 pituitary specific pou domain transcription factor	94 - 103	(-)	0.960	AgttAT TCat
V$MEIS/MEIS1.01	Homeobox protein MEIS1 binding site	95 - 106	(-)	0.781	CTGAG Ttattca
V$AP1F/NFE2.01	NF-E2 p45	98 - 108	(+)	0.861	AtaacTC AGta
V$OCT1/OCT1.05	octamer-binding factor 1	115 - 128	(+)	0.935	Ctgaatga TATATa
V$MEIS/MEIS1.01	Homeobox protein MEIS1 binding site	119 - 130	(+)	0.840	ATGAT Atatatc
F$YNIT/NIT2.01	activator of nitrogen-regulated genes	127 - 132	(+)	1.000	TATCta
V$EVI1/EVI1.04	Ecotropic viral integration site 1 encoded factor	127 - 141	(-)	0.857	TAATA Agtttagata
V$NKXH/NKX31.01	prostate-specific homeodomain protein NKX3.1	129 - 141	(-)	0.898	TaatAA GTttaga
V$MYT1/MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis	130 - 140	(-)	0.890	AatAAG Tttag
V$TBPF/TATA.01	cellular and viral TATA box elements	132 - 146	(-)	0.896	GtaTATA Ataagttt

V$OCT1/OCT1.05	octamer-binding factor 1	142 - 155 (+)	0.905	Tatactga GATGCa
V$OCT1/OCT1.02	octamer-binding factor 1	150 - 159 (+)	0.937	GATGC Aaaaa
V$IRFT/IRF1.01	interferon regulatory factor 1	156 - 168 (+)	0.890	Aaaaatt GAAAat
V$EVI1/EVI1.06	Ecotropic viral integration site 1 encoded factor	165 - 173 (-)	0.854	AcaAG ATtt
V$FKHD/HFH1.01	HNF-3/Fkh Homolog 1	168 - 179 (-)	0.898	AtttAAA Caaga
V$OCT1/OCT1.06	octamer-binding factor 1	175 - 186 (-)	0.861	Taaatgg GAATT
V$IKRS/IK1.01	Ikaros 1 potential regulator of lymphocyte differentiation	176 - 188 (-)	0.916	AaatGG GAattga
V$CLOX/CDPCR3HD.01	cut-like homeodomain protein	183 - 192 (+)	0.958	AattGA TCta
V$GATA/GATA3.02	GATA-binding factor 3	184 - 193 (+)	0.928	AtTGA Tctaa
V$PDX1/ISL1.01	Pancreatic and intestinal lim-homeodomain factor	203 - 222 (+)	0.826	Tgtagat CaTAA Taaatatt
V$GATA/GATA3.02	GATA-binding factor 3	204 - 213 (+)	0.934	GtAGA Tcata
V$BRNF/BRN2.01	POU factor Brn-2 (N-Oct 3)	208 - 223 (+)	0.930	AtcataatA AATatta
V$CART/CART1.01	Cart-1 (cartilage homeoprotein 1)	209 - 226 (+)	0.923	TcaTAA Taaatatt Aatg
V$NKXH/NKX25.02	homeo domain factor Nkx-2.5/Csx tinman homolog low affinity sites	210 - 217 (+)	0.878	CaTAA Taa
V$FKHD/XFD2.01	Xenopus fork head domain factor 2	211 - 224 (+)	0.903	AtaaTAA Atattaa
V$HNF1/HNF1.01	hepatic nuclear factor 1	212 - 226 (-)	0.807	CATTA Atatttatta
V$GATA/GATA3.02	GATA-binding factor 3	223 - 232 (-)	0.933	AaTGA Tcata
V$OCT1/OCT1.06	octamer-binding factor 1	223 - 234 (-)	0.851	Gatatgat CATT
V$EVI1/EVI1.05	Ecotropic viral integration site 1 encoded factor	225 - 235 (-)	0.833	Agatat GATCa
V$GATA/GATA3.02	GATA-binding factor 3	228 - 237 (-)	0.949	AtAGA Tatga
V$SATB/SATB1.01	Special AT-rich sequence-binding protein 1 predominantly expressed in thymocytes binds to matrix attachment regions (MARs)	228 - 249 (-)	0.923	Tataaatat GTAAT Agatatga
F$YNIT/NIT2.01	activator of nitrogen-regulated genes	231 - 236 (+)	1.000	TATCta
V$CREB/E4BP4.01	E4BP4 bZIP domain transcriptional repressor	235 - 246 (-)	0.873	AaatatG TAAta
V$TBPF/TATA.01	cellular and viral TATA box elements	236 - 250 (-)	0.939	GtaTAA Atatgtaat
V$VBPF/VBP.01	PAR-type chicken vitellogenin promoter-binding protein	236 - 245 (+)	0.899	ATTAC Atatt
V$FKHD/FREAC7.01	Fork head RElated Activator-7	238 - 253 (-)	0.942	TttgtaTA AAtatgta
V$NKXH/NKX31.01	prostate-specific homeodomain protein NKX3.1	238 - 250 (-)	0.843	GtatAA ATatgta
V$MEF2/MEF2.05	MEF2	242 - 251 (-)	0.978	TgtaTA AAta
V$PIT1/PIT1.01	Pit1 GHF-1 pituitary specific pou domain transcription factor	243 - 252 (+)	0.907	AtttAT ACaa
V$FKHD/HFH3.01	HNF-3/Fkh Homolog 3 (= Freac-6)	246 - 258 (+)	0.981	TatacAA ACatcc
V$SATB/SATB1.01	Special AT-rich sequence-binding	258 - 279 (-)	0.913	Catgttat

	protein 1 predominantly expressed in thymocytes binds to matrix attachment regions (MARs)			AgTAA Tattttcaa
V$FKHD/HNF3B.01	Hepatocyte Nuclear Factor 3beta	290 - 304	(-)	0.934	AaataaA ATAtttag
V$CDXF/CDX2.01	Cdx-2 mammalian caudal related intestinal transcr. Factor	291 - 309	(+)	0.921	Taaatat TTTA Tttccccа
V$BRNF/BRN3.01	POU transcription factor Brn-3	294 - 307	(+)	0.784	ATATttt Atttccc
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	297 - 312	(-)	0.860	GgatggGG AAataaaa
V$CEBP/CEBPB.01	CCAAT/enhancer binding protein beta	299 - 312	(-)	0.941	Ggatggg GAAAtaa
V$MZF1/MZF1.01	MZF1	304 - 311	(-)	0.974	GatGG GGa
V$PAX1/PAX1.01	Pax1 paired domain protein expressed in the developing vertebral column of mouse embryos	307 - 324	(+)	0.675	CCAT Ccagat Ctagcaat
V$BCL6/BCL6.01	POZ/zinc finger protein transcriptional repressor translocations observed in diffuse large cell lymphoma	313 - 326	(-)	0.756	TaaTTGC Tagatct
V$HOMS/S8.01	Binding site for S8 type homeodomain	315 - 330	(-)	0.975	Atctag CaATT Actgt
V$PDX1/PDX1.01	Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF	317 - 335	(-)	0.744	Ataaca CagTA ATtgctag
V$SATB/SATB1.01	Special AT-rich sequence-binding protein 1 predominantly expressed in thymocytes binds to matrix attachment regions (MARs)	329 - 350	(+)	0.924	Gtgttattt TAATag Tagtat
V$HNF1/HNF1.01	hepatic nuclear factor 1	330 - 344	(+)	0.792	TGTTAt Tttaatag
V$MEF2/MMEF2.01	myocyte enhancer factor	330 - 345	(-)	0.920	ActatTAA Aaataaca
V$CART/XVENT2.01	Xenopus homeodomain factor Xvent-2 early BMP signaling response	338 - 353	(-)	0.822	CTTATac Tactattaa

Appendix J Tblastn Analysis of the Predicted Amino Acid Sequence of TashAT1 with the *T. parva* Genome Database

```
TBLASTN 2.0MP-WashU [27-Aug-2000] [linux-i686 21:46:47 28-Aug-2000]

Copyright (C) 1996-2000 Washington University, Saint Louis, Missouri USA.
All Rights Reserved.

Reference:  Gish, W. (1996-2000) http://blast.wustl.edu

Notice:  statistical significance is estimated under the assumption that the
equivalent of one entire reading frame of the database codes for protein and
that significant alignments will involve only coding reading frames.

Query=  query_sequence
        (466 letters)

Database:  /usr/local/db/ufmg/t_parva
           564 sequences; 8,929,689 total letters.
Searching....10....20....30....40....50....60....70....80....90....100% done
```

Sequences producing High-scoring Segment Pairs:	Reading Frame	High Score	Smallest Sum Probability P(N)	N
443	+1	420	1.5e-47	3
473	+2	91	0.43	2
500	-3	77	0.81	1
450	-1	85	0.97	2


Color Key for Alignment Scores
40
40-50
50-80
80-200
>=200
100
200
300
400

```
>443
        Length = 814,854

  Plus Strand HSPs:

 Score = 244 (91.0 bits), Expect = 2.7e-23, Sum P(3) = 2.7e-23
 Identities = 69/198 (34%), Positives = 99/198 (50%), Frame = +2

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              MV L +  +++  FLY +K   S +L L+NL +     +K VE  ++K MI ST E KIT
Sbjct: 40829 MVRLNILLLLYAGFLYHIKSVYSVLLDLNNLSHSGLQVVKTVERDISKTMIYSTQERKIT 41008

Query:     62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
               I     KLIW   RGRYV+C TR   -  + TLI I+I NAV +   +I++  + Y YI+
Sbjct: 41009 GICKGSKLIWMGYRGEYVRCVTRILSKWCNSTLIIIQINKAVKDDTYYIHIDPSNYVYIS 41188

Query:    122 KSEFEDYYKSFCSVFIKIPPGKLPIPRLKKNVKTEXXXXXXXXXXXXXXXPOSEQHDKN 181
              K EF++       S F +     K     R KK   K +                PQ  Q D+
Sbjct: 41189 KEEFDEVL-DVISKFSQWMYEKYS-KRPKKPEKRKGDIDDTERTKKRSTVEPQP-QSDQP 41359

Query:    182 VDLVSQSLAEEGIDLEKK 199
                V+Q  + E ++ EKK
Sbjct: 41360 EPEVTQVKSIEDVEREKK 41413

 Score = 235 (87.8 bits), Expect = 3.0e-19, Sum P(3) = 3.0e-19
 Identities = 59/151 (39%), Positives = 85/151 (56%), Frame = +2

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              M  L + +-IF +F+Y +X ASS  L L+N     F+T+KIV + + K  I STP+  IT
Sbjct: 45452 MFRLNVLYLIFVVFVYCIKIASSLTLDLNNTSMSEFHTLKIVGNGIIKTTIFSTPDRFIT 45631

Query:     62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
              +I     + IW +  GE  KC T  + E S K L+ IE+ N V   + +-       Y YI
Sbjct: 45632 KICKGCRGIWQALPGESAKCVTYITSELSKKILMIIEVDNPVKHQVYYLNRCRTHYVYIT 45811

Query:    122 KSEFEDYYKSFCSVFIKIPP-GKLPIPRLKK 151
              + +FE   K F  V I  PP GK+PIP+LK+
Sbjct: 45812 REQFR---KEF--VEISYPPSGKVPIPKLKR 45889

 Score = 218 (81.6 bits), Expect = 3.5e-16, Sum P(3) = 3.5e-16
 Identities = 54/160 (33%), Positives = 86/160 (53%), Frame = +2

Query:      1 MMVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKI 60
              +M +L + ++   +    + F  SEIL + NL +  FYTI I+ED ITK MI STP+  I
Sbjct: 35276 IMNILDVKYLGVLIIFNCISFVFSEILNIKNLTDSGFYTIIIIEDGITKTMIYSIPQKII 35455

Query:     61 TEIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYI 120
              TE+R  +++=+W +  GE VKC T ++ + +   ++TIEI N V + M -     + Y Y
Sbjct: 35456 TEVRQCKRVLWTALPGESVKCLTIYTLDWASIRVMTIEINNPVKDGMYYFNRRYSNYVYA 35635

Query:    121 NKSEFEDYYKSFCSV--FI--KIPPG--KLPIPRLKKNVK 154
               K  F+  Y     V  ++  K      K+PIP  K+  K
Sbjct: 35636 TKEMFDFEYAEMARVAKYMHEKYSKSSDKVFIPEQKQPKK 35755

 Score = 215 (80.7 bits), Expect = 9.6e-16, Sum P(3) = 9.6e-16
 Identities = 51/130 (39%), Positives = 75/130 (57%), Frame = +2

Query:     10 IIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKITEIRSKRKL 69
              +IF + +Y   F SS  L L+NL N   F T+KIVE+ - K  I ST +  IT+IR   K-
Sbjct: 50048 LIFVVIVYCTSFVSSLTLDLNNLFNSEFNTLKIVENGIIKTTIFSTADRFITKIRKGSKV 50227

Query:     70 IWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYINKS----EF 125
              IW +  GE  K  T  S + S   ++TIE+ N V++ M +I  + + YKY+ K      EF
Sbjct: 50228 IWQALSGESAKSITHISSKWSKSMVMTIEVDNHVNDDMYYICKTRSDYKYVTKEIFDEEF 50407

Query:    126 EDYYKSFCSV 135
               - YK+F S+
Sbjct: 50408 VELYKTFISM 50437

 Score = 210 (79.0 bits), Expect = 5.0e-15, Sum P(3) = 5.0e-15
 Identities = 45/119 (37%), Positives = 65/119 (54%), Frame = +2

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              MV L +   ++  L LY +K  SS +  + N+ N     ++ VE+ ITK +I STPE KIT
Sbjct: 36614 MVKLNILCLLVILILYHIKIVSSIVFDIKNISNSKVRVVVRTVRNGMIKIVIYSTPERKIT 36793
```

```
Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYI 120
             E+R   K+IW  D GEY+KC   + F      L++++E   N   + M F Y  G  Y  I
Sbjct: 36794 EVRDGYKLIWMADPGEYLKCLNYYEFLWKKNILISVESNNPSKD-MSFFYKRGENYPRI 36967


Score = 188 (71.2 bits), Expect = 3.4e-12, Sum P(3) = 3.4e-12
 Identities = 43/130 (33%), Positives = 69/130 (53%), Frame = +2

Query:     2 VVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKITE 62
             V LK   +IF L  Y +K  SS +L ++N+DN  F  +KIV+  + K+MI  T ++ I E
Sbjct: 42278 VFLKNVCLIFALIFYHIKIVSSALLNINNIDNSKFKVVKIVDGFVIKVMIYPTADNPINE 42457

Query:    63 IRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNF--YKYI 120
             +    K+IW    GE +KC T  + +  D  I+ I++ N   E    TY    +  ++ I
Sbjct: 42458 VCKNSKIIWEVHPGEKIKCVTHITSKLCDAQLMVIDVENP--EKKYTIYFRKYYLHFRVI 42631

Query:   121 NKSEFEDYYK 130
             N++ F+   K
Sbjct: 42632 NENTFDKMLK 42661

 Score = 174 (66.3 bits), Expect = 1.6e-10, Sum P(3) = 1.6e-10
 Identities = 39/123 (31%), Positives = 65/123 (52%), Frame = +2

Query:     4 VLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKITEI 63
             +LK+ ++   L LY +K  S +L L ++ N  F+T+++++  + K MI ST +  I+ +I
Sbjct: 38249 LLKILYLFILTLYHIKIVLSNVLDLRDIKNSEFHTLELLQQGIIKTMIYSTADKPITKI 38428

Query:    64 RSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYINKS 123
               +  IW +  GE  KC T     + +  L+TIEI N     M +IY     Y Y  + S
Sbjct: 38429 CKGSRVIWQALPGESAKCITLIRSKWAPGNILMTIEINNPSKEDMLYIYKRYFHYYYTTEE 38608

Query:   124 EFE 126
              F+
Sbjct: 38609 GFK 38617

 Score = 42 (19.8 bits), Expect = 1.5e-37, Sum P(3) = 1.5e-37
 Identities = 12/30 (40%), Positives = 17/30 (56%), Frame = +3

Query:   284 IQTKESSIQTDIQEVEDIDT--QTDIQELE 311
             + T ES I+ DI+    +DT  +T I  LE
Sbjct: 264897 LNTSESLIKLDIKPFHGLDTSPKTAIINLE 264986

 Score = 148 (57.2 bits), Expect = 2.4e-13, Sum P(3) = 2.4e-13
 Identities = 49/198 (24%), Positives = 94/198 (47%), Frame = +2

Query:     2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
             M  LK ++++  + L  +K   S IL L+NL   +F  I+  ++ - K +I ST E  IT
Sbjct: 25544 MAKLKPTYLVCIVVLCSIKAVLSNILDLNNLTKFSFKIIQYCKENVNKTIIYSTNESPIT 25723

Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
             +I      L++     GE +K  T    F+++ + L+ +EI NA-     K+        +K I+
Sbjct: 25724 QINVGSILLPKPLHGEKIKSVTIMRFKNTKEVLLVLEIENAIASPQKYYSRRKGPFKPID 25903

Query:   122 KSEFEDYYKSFCSVFIKIPPGKLPIP----RLKKNVKTEXXXXXXXXXXXXXXXPQSEQ 177
             +  F    +S  +    K  P K+  P     +-+K+ + +                PQ
Sbjct: 25904 ERTFLLKFESLSNK PKYDPSKIVTPGVDPKVQKSFERKLRKRFRDGFFKEKEITPQDVT 26080

Query:   178 HDKN---VDIVSQSLAEE 192
             +  N    +D1+++S  ++E
Sbjct: 26081 NIANEPPIDILTESFSDE 26134

 Score = 58 (25.5 bits), Expect = 3.2e-39, Sum P(3) = 3.2e-39
 Identities = 23/75 (30%), Positives = 36/75 (48%), Frame = +3

Query:   256 TQTDIQ-EIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQELENIG 314
             T+TD    + +    ++ EI E ++   + +D+    ES+     IQ    D  TQTD   E +N
Sbjct: 52572 TETDSSTKPKSKPLEPEIIESQSS-SDSDMDVDESTGPQVIQS--DATTQTDTHESQNSE 52742

Query:   315 IQTIGNFSDITEVTK 329
              QT+    S     TK
Sbjct: 52743 TQTVIQTSSTETQTK 52787

 Score = 84 (34.6 bits), Expect = 7.2e-07, Sum P(3) = 7.2e-07
 Identities = 20/88 (22%), Positives = 44/88 (50%), Frame = +2

Query:    43 VEDRLTKIMILSTPEDKITEIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNA 102
             + D +   +++ + P    +T +   +  +I  +D+G  +  F+ FS+ +    TLI I    A
Sbjct: 27089 INDGIPTLIVKAKPHKTVTHVVFQGGVIICEADKGSKLLSFSAFSYYNY-YTLIEIIFQTA 27265
```

```
Query:   103 VDEAMKFIYVSGNFYKYINKSEFEDYYK 130
             +  +K+    G  +  ++ + FE YY+
Sbjct: 27266 TNSYIKYFRKHGGDWVEVDITHFEAYYQ 27349


 Score = 80 (33.2 bits), Expect = 1.7e-41, Sum P(3) = 1.7e-41
 Identities = 29/84 (34%), Positives = 46/84 (54%), Frame = +2

Query:   256 TQTDIQEIEDIGIQTEIHELENIVTQTDIQ-TKESSIQTD----IQEVEDIDTQ-TDIQE 309
             TQTD  +      +QTE     +   TQTD + T++S  QTD     IQE + +D + T++ E
Sbjct: 55586 TQTDPIKTRVRSVQTESKRTKIGKTQTDPKKTEDSQTQTDAKIDIQECKSVDEKKTEVSE 55765

Query:   310 L---ENIGIQTIGNFSDITEVTKK 333
             L    ++I  +  G+  D+T + KK
Sbjct: 55766 LRITDSTKAED-GSHGDVTPLEKK 55834

 Score = 73 (30.8 bits), Expect = 8.8e-41, Sum P(3) = 8.8e-41
 Identities = 27/84 (32%), Positives = 47/84 (55%), Frame = +2

Query:   245 ESDGILTQNRYTQTDIQEIEDIGIQTEIHELENI-VTQTD-IQTKESSIQTDIQEVEDID 302
              SD  L++N   Q+D+++ + +G   +  +   I  TQTD I+T+  S+QT+ +  +
Sbjct: 55481 DSDVDLSKNTKIQSDVKKTK-VGSFKKGGKGTKIGKTQTDPIKTRVRSVQTESKRTKIGK 55657

Query:   303 TQTDIQELENIGIQTIGNFSDITE 326
             TQTD ++ L+   QT     DT E
Sbjct: 55658 TQTDPKKTEDSQTQTDAKI-DIQE 55726

 Score = 68 (29.0 bits), Expect = 9.9, Sum P(3) = 1.00
 Identities = 28/147 (19%), Positives = 55/147 (37%), Frame = +2

Query:    10 IIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKITEIRSKRKL 69
             +++ L L  +  +S   + ++NL        +K  ++ + K +I ST    JT +     L
Sbjct: 54575 LLYLLTLCYINLSSGVYIDINNLYGKGLKILKYTKNNVKKTIIYSTTNLPITGLMDNGSL 54754

Query:    70 IWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYINKSEFEDYY 129
             -W     E  +      F  S   ++ + +   +    K         Y  I K  FEDYY
Sbjct: 54755 VWSDYLSENCEKIIVNRFAHSRDLIVEVFLLKPMFTKTKIYLKHKGEYSRITKQIFRNKI 54934

Query:   130 KSFCSVFIKIPPCKLPIFRLKKNVKTE 156
             K   S      P K+  P+++    +E
Sbjct: 54935 KILSEIQNYGPEKIFTPQVETATTSE 55012

 Score = 50 (22.7 bits), Expect = 2.2e-38, Sum P(3) = 2.2e-38
 Identities = 13/40 (32%), Positives = 24/40 (60%), Frame = +2

Query:   258 TDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQE 297
             TD  + ED G   ++  LE  V  TD-  ++SS++ +I++
Sbjct: 55775 TDSIKAED-GSHGDVTPLEKKVFPTDLYHQDSSLEFEIRK 55891

 Score = 269 (99.8 bits), Expect = 7.1e-26, Sum P(3) = 7.1e-26
 Identities = 57/125 (45%), Positives = 75/125 (60%), Frame = +3

Query:     2 MVVIKLSHTIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
             M  LKLSH++FTLFLY +K   S +L LDN+    +  ++  E  +TK+MI ST E KIT
Sbjct: 51735 MATLKLSHVLFTLFLYHIKIVFSNLLDLDNIAGSGYNIVQTSERGITKLMIFSTQEKKIT 51914

Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
              I +K K+ W    GEY KCFT + FE S+  L ++EI N   E +K+        YK IN
Sbjct: 51915 VIYNKEKLEWKGHPGEYTKCFTIYKFEMSNIGLASLEILNPEFENIKYFRSHNLIYKPIN 52094

Query:   122 KSEFE 126
             +  FE
Sbjct: 52095 QETFE 52109

 Score = 68 (29.0 bits), Expect = 2.9e-40, Sum P(3) = 2.9e-40
 Identities = 18/40 (45%), Positives = 22/40 (55%), Frame = +3

Query:   258 TDIQEIE-DIGIQTEIHELENIVTQTDIQTKESSIQTDIQ 296
             T  Q I+ D   QT+ HE +N  TQT IQT  +  QT  Q
Sbjct: 52674 TGPQVIQSDATTQTDTHESQNSETQTVIQTSSTEIQTKTQ 52793

 Score = 49 (22.3 bits), Expect = 6.6e-12, Sum P(3) = 6.6e-12
 Identities = 9/21 (42%), Positives = 14/21 (66%), Frame = +3

Query:   110 IYVSGNFYKYINKSEFEDYYK 130
              +Y+  N YKYIN  -+ + YK
Sbjct: 125118 LYIRINNYKYINNYKYINNYK 125180
```

```
 Score = 42 (19.8 bits), Expect = 1.5e-37, Sum P(3) = 1.5e-37
 Identities = 9/38 (23%), Positives = 20/38 (52%), Frame = +2

Query:    267 GIQTEIHELENIVTQTDIQTKESSIQTDI--QEVEDID 302
              G+Q ++ +L N       ++T E      D+  + V++I+
Sbjct: 410378 GVQEKLKDLSNANISEPLKTSEEQKPLDVKAENVQEIE 410491


 Score = 250 (93.1 bits), Expect = 6.4e-24, Sum P(3) = 6.4e-24
 Identities = 52/130 (40%), Positives = 81/130 (62%), Frame = +1

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              M+ LK + -IF +F+Y +K ASS  L L+N      F+T+K++    + K ++LSTP+ +IT
Sbjct: 48622 MIYLKNTFLIFVVFVYCIKIASSLTLDLNNTSMSEFHTLKLLHQGIAKTIVLSTPDRQIT 48801

Query:     62 EIRSKRKLIWGSDRGEYVKCFTRF-SFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYI 120
              E+R  R+ +W    GE +KC T + SF+ S + L+TIEI N    AM ++    + YKYI
Sbjct: 48802 EVRQGRESVWMGHPGESIKCVTFYISFKWSSEVLMTIEINNFNKTAMYYLRKHHHNYKYI 48981

Query:    121 NKSEFEDYYK 130
              +  EFE  Y+
Sbjct: 48982 SLQEFESRYR 49011

 Score = 138 (53.6 bits), Expect = 1.5e-47, Sum P(3) = 1.5e-47
 Identities = 40/92 (43%), Positives = 49/92 (53%), Frame = +1

Query:    243 PKESD-GILTQNRYTQTDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVEDI 301
             PKE    I TQ RYTQTD  E ED   QT+  + ++  TQ  I T  +  QT I   +
Sbjct: 31288 PKEQQVKILTQIRYTQTDTHESEDTETQTDIQQSKDTETQTVILTDSTETQTLIP-TDST 31464

Query:    302 DTQTDIQELENIGIQTIGNFSDITEV-TKKHE 332
             DTQTD  E +    QT+   +D TE  T  HE
Sbjct: 31465 DTQTDTHESKETETQTVIP-TDSTETQTDTHE 31557

 Score = 216 (81.1 bits), Expect = 7.1e-16, Sum P(3) = 7.1e-16
 Identities = 47/129 (36%), Positives = 73/129 (56%), Frame = +3

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              M+ LK + -IF +F+Y +K ASS  L L+N      F-T+KIVE  +IK  I ST +  IT
Sbjct: 46713 MIYLKNTFLIFVVFVYCIKIASSLTLDLNNTSMSEFHTLKIVEGNVTKTTIFSTADKPIT 46892

Query:     62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
              +-R      IW  + GE  KC + F F+    +TL+ +EI         + + - +   YK +
Sbjct: 46893 KVRQGYIPIWEGNPGESAKCVSYFRFKRPARTLVKVEIEGYNTNFVCYFFKNFTHYKEVI 47072

Query:    122 KSEFEDYYK 130
              K+ FE+ YK
Sbjct: 47073 KNYFENKYK 47099

 Score = 48 (22.0 bits), Expect = 3.5e-38, Sum P(3) = 3.5e-38
 Identities = 11/29 (37%), Positives = 19/29 (65%), Frame = +2

Query:    265 DIGIQTEIHELENIVTQTD----QTKES 289
             D+ I T +HE  N+-T+T+     I T++S
Sbjct: 332336 DLAIDTILHEFLNLLTRTNPNFNINTEDS 332422

 Score = 153 (58.9 bits), Expect = 6.6e-14, Sum P(3) = 6.6e-14
 Identities = 43/160 (26%), Positives = 82/160 (51%), Frame = +3

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              M  L+ ++++  L  Y +   SS+ L +++ N N   ++  E+ +'I I  T E  I
Sbjct: 29064 MARLRFTYLVPVLTFYFINLVGSDKLDINDIYNLNIQRVEYSENGMTTIKIYPTSEQPIR 29243

Query:     62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVS--GNFYKY 119
              ++   + L+W +  GE  K  T   F+ S + ++ ++I   V ++ K IY +  GN Y+
Sbjct: 29244 QVYDGQNLVWSALLGERAKVITVLKFKYSGEVIVKVDIDFPVSKSQK-IYCNRKGN-YQE 29417

Query:    120 INKSEFEDYYKSFCSVFIKIPPGKL--------PIPRLKK 151
              I++   +  +SF S   K  P K+          PIP+++K
Sbjct: 29418 IDRKTCQEKIQSF-STIQKYNPSKTYQPKTERYPIPQMRK 29534

 Score = 205 (77.2 bits), Expect = 2.4e-14, Sum P(3) = 2.4e-14
 Identities = 40/129 (31%), Positives = 73/129 (56%), Frame = +3

Query:      4 VLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKITEI 63
              +LK+ ++   L LY +K   S +L L ++ +  F  +  E  + K ++LS+P+ +ITE+
Sbjct: 43635 LLKILYLFIILTLYHIKIVLSNVLDLRDISSSGFEVAQTHESGMIKTIVLSSPDRQITEV 43814
```

```
Query:    64 RSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYINKS 123
             R  R+L+W    GE VKC T  +F    K L+T+EI N V   + ++Y   Y Y IK
Sbjct: 43815 KQGRELVWMGHPGESVKCLTHTTFMRYKKALVTMEINNPVKHDVFYLYNYFSHYVYTSKD 43994

Query:   124 RFEDYYKSF 132
              +++ ++ +
Sbjct: 43995 IYDEKFRKY 44021

 Score = 46 (21.3 bits), Expect = 5.6e-38, Sum P(3)   5.6e-38
 Identities = 10/27 (37%), Positives = 18/27 (66%), Frame = +2

Query:   279 VTQTDIQTKESSIQTD-IQEVEDIDTQ 304
             +TQT-   ++S IQT+ + E E ID I
Sbjct: 50795 ITQTEPPNEQSEIQTEHVLESEIVDKE 50875

 Score = 45 (20.9 bits), Expect = 7.2e-38, Sum P(3) = 7.2e-38
 Identities = 14/54 (25%), Positives = 25/54 (46%), Frame = +2

Query:   276 ENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQE-LENIGIQCIGNFSDITEVT 328
             + IV  T + T+   I   +++  I    T +Q  L  +G  IG  S I  +
Sbjct: 275420 KGIVVKTGMNTQVGKIAKQLKKASETSKVTPLQRALNRLG-GTIGIISTIVLIS 275578

 Score = 420 (152.9 bits), Expect = 1.5e-47, Sum P(3) = 1.5e-47
 Identities = 88/157 (56%), Positives = 110/157 (70%), Frame = +1

Query:     2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
             M  LKLSHIIFTLFLY++K ASSEILYLDN+    F  IKI+E+R+T+ MI  STP+ +IT
Sbjct: 30538 MATLKLSHIIFTLFLYQIKIASSEILYLDNIVGSGFNIIKIIENRITRTMIYSTPDRQIT 30717

Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
             -+R  RKLIW    GE +KC T FSFESS K LITIEI N   ++++KFIY+  N+++Y+
Sbjct: 30718 QVRQGRKLIWMGYPGESIKCLTIFSFESSSKILITIEIENPAYDSLKFIYMHRNYFRYVT 30897

Query:   122 KSEFEDYY----KSFCSVFIKIPPGKLPIPRLKKNVK 154
             K+ FE  +    K   S   K  PGKLPIPRLKK  K
Sbjct: 30898 KAYFETNFAMQAKPLKSPTSKPIPGKLPIPRLKKPEK 31008

 Score = 179 (68.1 bits), Expect = 4.1e-11, Sum P(3) = 4.1e-11
 Identities = 45/128 (35%), Positives = 64/128 (50%), Frame = +3

Query:     2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
             MV L + +II  L LY VK  SS  L L ++D   F    +VE+ +TK  IL+     I
Sbjct: 53343 MVRLNILYIIVLGLLLYHVKTVSSLTLDLRDIDTSKFDVSSVVENGVTKTTILTKRHIPID 53522

Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
             E+     ++IW    GE V   T +SFE-  K L+ IE+ N-  E + + Y     Y I
Sbjct: 53523 ELYFAGEMIWKGRPGESVNSITHYSFENHHKMLLYIEVDNSAFEDILYFYTRRGIYLDIT 53702


Query:   122 KSEFEDYY 129
             + EF   Y
Sbjct: 53703 EEEFWKLY 53726

 Score = 42 (19.8 bits), Expect = 1.5e-47, Sum P(3) = 1.5e-47
 Identities = 6/16 (37%), Positives = 9/16 (56%), Frame = +2

Query:   419 WLLRPRNMKTETKKTW 434
             W+  PRN  +T  +W
Sbjct: 480626 WICYPRNRLKDTTSSW 480673

 Score = 44 (20.5 bits), Expect = 2.1e-11, Sum P(3) = 2.1e-11
 Identities = 8/22 (36%), Positives = 13/22 (59%), Frame = +1

Query:   109 FIYYVSGNFYKYINKSEFEDYYK 130
             FI+ +    Y Y N++ F D +K
Sbjct: 155497 FIFLKA-VYDYFNQTHFNDIFK 155559

 Score = 154 (59.3 bits), Expect = 3.1e-08, Sum P(3) = 3.1e-08
 Identities = 36/129 (27%), Positives = 65/129 (50%), Frame = +1

Query:     2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
             MV - + ++ F L + R+K  SS +L ++++ N      + V++ +    + S P   IT
Sbjct: 32353 MVRVNILYMSFVLIVCRIKIVSSIVLDINDIVNSGLKVFQRVKNGIITTKVFSRPGMPIT 32532

Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
             +I    + IW    GE V+  T  + E S +T+++IE+ N V E   +++    N YKYI
Sbjct: 32533 QILKGSRPIWNGYPGESVRSLTFITSEWSSETVLSIEVDNPVKEPTLYLHTFYNHYKYIT 32712
```

```
Query:   122 KSEFEDYYK 130
               +    K
Sbjct: 32713 IQSYNQKIK 32739

 Score = 99 (39.9 bits), Expect = 1.7e-43, Sum P(3) = 1.7e-43
 Identities = 31/84 (36%), Positives = 47/84 (55%), Frame = +1

Query:   245 ESDGILTQNRYTQTDIQEIEDIGIQTEI----HELENIV------TQTDI-QTKESSIQT 293
             ++D   +++  TQTD Q+ +D   QT I     F + ++      TQTD  ++KE+  QT
Sbjct: 31333 QTDTHESEDTETQTDTQQSKDIETQIVILTDSTETQTLIPTDSTDTQTDTHESKETETQT 31512

Query:   294 DIQEVEDIDTQTDIQELENIGIQT 317
              I   +  +TQTD  E E+IGIQT
Sbjct: 31513 VIP-TDSTETQTDTHETEDIGIQT 31581


 Score = 61 (26.5 bits), Expect = 2.8e-06, Sum P(3) = 2.8e-06
 Identities = 14/26 (53%), Positives = 16/26 (61%), Frame = +3

Query:   280 TQTDIQTKESSIQTDIQEVEDIDTQT 305
             TQ  IQTK +  QTD Q+ ED   QT
Sbjct: 29946 TQCAIQTKAAETQTDSQQTEDPVVQT 30023

 Score = 46 (21.3 bits), Expect = 7.8e-05, Sum P(3) = 7.8e-05
 Identities = 11/22 (50%), Positives = 14/22 (63%), Frame = +3

Query:   249 ILTQNRYTQTDIQEIEDIGIQT 270
             I T-   TQTD Q+ ED  +QT
Sbjct: 29958 IQTKAAETQTDSQQTEDPVVQT 30023

 Score = 246 (91.7 bits), Expect = 1.7e-23, Sum P(3) = 1.7e-23
 Identities = 57/159 (35%), Positives = 88/159 (55%), Frame = +3

Query:     2 MVVLKLSHTIFTLFLYRVKPASSRILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
             MV L +  +++  FLY +K   S  L +  + +  F+TIK+ E  TKIM+ SF +  IT
Sbjct: 39513 MVRLKTLLLLVAGFLYHIKSVYSNTLNIQIIADSGFFTIKVYENGITKIMVFSTADKFIT 39692

Query:    62 EIRSKRKLIWGSDRGEYVKCFTRFSFESSDKTLITIEICNAVDEAMKFIYVSGNFYKYIN 121
             E+R   K IW S  GE +KC T + F+ S++ L+TIEI N V + M ++++    Y Y
Sbjct: 39693 EVRQGPKSIWDSLPGESIKCLTYYQFKGSNRKLMTIEINNPVKDEMYYLHTHNYNYVYAT 39872

Query:   122 KSEFEDYYKSFCSV--FI--KIPPG--KLPIPRLKKNVK 154
             K  FE  Y     V  ++  K      K+PIP+ K+  K
Sbjct: 39873 KEMFEEEYTEMARVAKYMHEKYSKSSDKVPIPKQKQPKK 39989

 Score = 42 (20.2 bits), Expect = 3.1e-40, Sum P(4) = 3.1e-40
 Identities = 10/34 (29%), Positives = 18/34 (52%), Frame = +2

Query:   294 DIQEVEDIDTQTDIQ-ELENIGIQTIGNFSDITE 326
             D   V + D   +++ +LEN  + ++G  SD  E
Sbjct: 277931 DDLNVSNPDLHPNLRAKLENFDLSSLGGMSDFVE 278032

 Score = 43 (20.2 bits), Expect = 1.2e-37, Sum P(3) = 1.2e-37
 Identities = 10/60 (16%), Positives = 36/60 (60%), Frame = +2

Query:   242 KPKESDGIL-TQNRYTQTDIQ-EIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVE 299
             KP+ ++  +  + + +QT +Q ++ +  +Q ++ E E  + Q ++ +++      +IQ++
Sbjct: 45965 KPETAETTIGNRRKSSQTLMQTQVVETQLQPDLLEPE--IVQVEVESEDDEEGREVQQLQ 46138

 Score = 45 (20.9 bits), Expect = 2.2e-38, Sum P(2) = 2.2e-38
 Identities = 13/60 (21%), Positives = 24/60 (40%), Frame = +3

Query:   250 LTQNRYTQTDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQE 309
             LTQN  T +    + +   T+      N  T+ + Q  +S+ Q   Q  +     T + +
Sbjct: 708783 LTQNIIDTNNEKANKDSLESTTQYDTQPNQSTKDETQPNQSTTQCGTQPNQSTKDDTQLNQ 708962

 Score = 98 (39.6 bits), Expect = 2.2e-43, Sum P(3) = 2.2e-43
 Identities = 25/60 (41%), Positives = 32/60 (53%), Frame = +1

Query:   249 ILTQNRYTQTDIQEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQ 308
             ILT +  TQT I   +    QT+ HE +   TQT I T  +  QTD  E EDI  QT ++
Sbjct: 31414 ILTDSTETQTLIPT-DSTDTQTDTHESKETETQTVIPTDSTETQTDTHETEDIGIQTKLR 31590

 Score = 42 (19.8 bits), Expect = 2.8e-22, Sum P(2) = 2.8e-22
 Identities = 11/22 (50%), Positives = 12/22 (54%), Frame = +2

Query:   294 DIQEVEDIDTQTDIQELENIGI 315
             D   V + D   +++ +LEN  +
Sbjct: 277931 DDLNVSNPDLHPNLRAKLENFDL 277996
```

```
Query:    117 YKYINKSEFEDYYKSFCSVFIK 138
              Y YI  S  + YVK F S F K
Sbjct: 809108 YSYIYNSF*KLVYKWFFSKFAK 809173

 Score = 43 (20.2 bits), Expect = 4.5e-22, Sum P(3) = 4.5e-22
 Identities = 25/93 (26%), Positives = 41/93 (44%), Frame = +1

Query:    113 SGNFYKYINKG-EFEDYYKSFCSVFIKIP-PG--KLPIPRLKK---NVKTEXXXXXXXXX 165
              S  F   +NK  E   Y +S  SV IK P P   + P P + K   +V+ E
Sbjct: 235174 SKEFDDLLNKKLEGLKYEESEASV-IKEPEPSTSREPEPSVIKESESVEIEEPKPSVRRE 235350

Query:    166 XXXXXXXPQSEQHDKNVDIVSQSLAEEGIDLEK 198
                      S + D+  +++S+ L++ G-D +K
Sbjct: 235351 AEPLKYSQTSVKFDRVTEMMSEHLSKSGVDSKK 235449

 Score = 43 (20.2 bits), Expect = 4.5e-22, Sum P(3) = 4.5e-22
 Identities = 8/23 (34%), Positives = 13/23 (56%), Frame = +2

Query:    128 YYKSFCSVFIKIPPGKLPIPRLK 150
              +Y +FC++ +K    KL    LK
Sbjct: 425084 FYSTFCNILLKSTKYKLVTNHLK 425152

 Score = 42 (19.8 bits), Expect = 1.5e-47, Sum P(3) = 1.5e-47
 Identities = 6/16 (37%), Positives = 9/16 (56%), Frame = +2

Query:    372 WLLRPRNMKTETKKIW 387
              W+  PRN   +T  +W
Sbjct: 480626 WICYPRNRLKDCTSSW 480673

 Score = 42 (19.8 bits), Expect = 1.5e-37, Sum P(3) = 1.5e-37
 Identities = 11/44 (25%), Positives = 19/44 (43%), Frame = +2

Query:    261 QEIEDIGIQTEIHELENIVTQTDIQTKESSIQTDIQEVEDIDTQ 304
              Q IED   ++E   LE    Q ++ + E         + +D D +
Sbjct: 42986 QRIEDKSKESESKALEPETIQFEVSSDEEEADEPTSKCDDYDKE 43117

 Score = 43 (20.2 bits), Expect = 3.6e-38, Sum P(2) = 3.6e-38
 Identities = 13/41 (31%), Positives = 21/41 (51%), Frame = +1

Query:    268 IQTEIHE-LENIVTQTDIQTKESSIQTDIQEVEDIDTQTDI 307
              +Q   E I    + +T  +T ++ TQ    EV  I TQT +
Sbjct: 514681 VQNTLEEGLA*FIRETQAETAQNGIQIFELEVTRIWTQTKL 514803

 Score = 228 (85.3 bits), Expect = 7.7e-18, Sum P(3) = 7.7e-18
 Identities = 50/125 (40%), Positives = 72/125 (57%), Frame = +1

Query:      2 MVVLKLSHIIFTLFLYRVKFASSEILYLDNLDNPNFYTIKIVEDRLTKIMILSTPEDKIT 61
              MV +  +  + + L +Y +K A S +L L N+ N  F+ +KIVE  +TK M1 ST +  IT
Sbjct: 33718 MVRVNTLLEPYALCVYCIKTAFSNVLDLKNISNSGFWALKIVEGNITKTMIYSTADRPIT 33897

Query:     62 EIRSKRKLIWGSDRCEYVKCFTRFSPESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYIN 121
              +-R   ++IW     E VKC T +SFE  +K L-TIEI + V   M +        Y YI+
Sbjct: 33898 KVRQGTRVIWEPYSYESVKCVTHYSFELHNKILMTIEISDPVVNDMYYFKRRRTHYVYIS 34077

Query:    122 KSEFE 126
              K EFE
Sbjct: 34078 KKEFE 34092

 Score = 45 (20.9 bits), Expect = 7.2e-38, Sum P(3) = 7.2e-38
 Identities = 7/15 (46%), Positives = 12/15 (80%), Frame = +1

Query:    177 QHDKNVDIVSQSLAE 191
              QHDKN++++ Q+  E
Sbjct: 336256 QHDKNINVMKQAHEE 336300

>473
        Length = 137,969

  Plus Strand HSPs:

 Score = 91 (37.1 bits), Expect = 0.56, Sum P(2) = 0.43
 Identities = 33/124 (26%), Positives = 56/124 (45%), Frame = +2

Query:     12 FTLFLYRVKFASSEILYLDN-LDNPNF----YTIKIVEDRLTKIMILSTPEDKITEIRSK 66
              + L  Y        E+  L++ LD  NF     +++  +-DR+T+ ++L+     ++T +
Sbjct: 56228 YPLTKYHFDVLKDEMESLESTLDLHNFSDSLFSVSHLDDRVTRELLARRLSRVIRVNHN 56407
```

```
Query:    67 RKLIWGSDRGEYVKCFTRFSFESSDK------TLITIEIG-NAVDEAMKFIYVSGNFYKY 119
             L+W S   E  K      S+ + +K       TL T  +  N +DE     YV    Y Y
Sbjct: 56408 GDLVWSSGGTELFKS-AALSYNTMNKLVLVKMTLTTPNMTYNPLDEKSTRYYVDRTLYFY 56584

Query:   120 INKS 123
             NKS
Sbjct: 56585 RNKS 56596

 Score = 46 (21.3 bits), Expect = 0.86, Sum P(2) = 0.58
 Identities = 17/55 (30%), Positives = 30/55 (54%), Frame = +3

Query:   274 ELENIVTQTDIQTKESSIQTDIQEVEDIDTQTDIQELLNIGIQTIGNFSDITEVT 328
             +LEN   +I +Q KES +-T   E+ + ++Q +     N+     G FS+ T +T
Sbjct: 109707 DLEN--RETTLQFKESELRTKELELNEKESQLNNALANNMF----GEFSN-TSLT 109850

 Score = 46 (21.3 bits), Expect = 0.86, Sum P(2) = 0.58
 Identities = 9/22 (40%), Positives = 11/22 (50%), Frame = +2

Query:   119 YINKSEFEDYYKSFCSVFIKIP 140
             Y N  E  DYY SF  +  + P
Sbjct: 66818 YYNDDENNDYYTSFTKLVPRTP 66883

 Score = 48 (22.0 bits), Expect = 0.56, Sum P(2) = 0.43
 Identities = 10/13 (76%), Positives = 11/13 (84%), Frame = -2

Query:   115 NFYKYIN-KSEFE 126
             NFY+YIN KS FE
Sbjct: 83366 NFYQYINQKSSFE 83404

>500
        Length = 12,748

  Minus Strand HSPs:

 Score = 77 (32.2 bits), Expect = 1.6, P = 0.81
 Identities = 27/101 (26%), Positives = 47/101 (46%), Frame = -3

Query:    58 DKITEIRSKRKLIWGSDRGE-----YVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYV 112
             + IT++R   +LI+ +D G       +V+ F +F+ +      + T   G   +   K I +
Sbjct: 11846 EPITQLRFGNELIFPNDTGSNNYNLFVEMFLQFNLKLVSFYVHTPINGLLKEVKRKVISI 11667

Query:   113 SGNFYKYINKSEFEDYYKSFCSVFIKIPPGKLPI-PRLKKN 152
             S + Y YI+   EF   YK     + + +  C  P  P + KN
Sbjct: 11666 SYDTYSYISNEEF- MYKINKPILVSVDIGPKPKHPFVNKN 11550

>450
        Length = 1,309,524

  Minus Strand HSPs:

 Score = 85 (35.0 bits), Expect = 3.6, Sum P(2) = 0.97
 Identities = 30/125 (24%), Positives = 58/125 (46%), Frame = -1

Query:    11 IFTLFLYRVKFASSEILYLDNLDNPNFYTIKI--VEDRLTKIMILSTPEDKITEIRSKRK 68
             I  +F+Y     S+ I+ + +  N NF  I++  V   +T + I+     +KI ++  K++
Sbjct: 300225 IVLIFIYCALVTSNNIIDI-SFPNFNFIAIQLEDVYGNVTMVTIVPNASNKIVKVVDKQR 300049

Query:    69 LIWG-SDRGEYVKCFTRFSFESSDKTLITIEIGNAVDEAMKFIYVSGNFYKYINKSEFED 127
              IW   +  E++      +     D+ I  - + + V + +KF    G    I  S  ED
Sbjct: 300048 QIWNRTSVDEFLSELKLYRLYGEDRLLFIVSVVDNV-KYVKFYIKQG--VPWIETS-MED 299881

Query:   128 YYKSF 132
             +  SF
Sbjct: 299880 FNSSF 299866

 Score = 53 (23.7 bits), Expect = 3.6, Sum P(2) = 0.97
 Identities = 16/68 (23%), Positives = 32/68 (47%), Frame = -2

Query:   253 NRYTQTDIQELE----DIGIQTEIHELEN---IVTQTDIQTKESSIQTDIQEVEDIDTQT 305
             NRYT TD++EI      +    +    EN      +T+  +   +  D+Q  +D+ TQT
Sbjct: 210455 NRYT-TDLKEINLNTHSLIFSNDFSNPENHNKFNFKTEDYSNSHNSHNDLQHTQDLQDSQ 210279

Query:   306 DIQELENI 313
             D+Q+ +++
Sbjct: 210278 DLQDSQDL 210255
```

```
Parameters:
  B=20
  cpus=1
  matrix=blosum62
  E=10
  V=20
  filter=seg

  ctxfactor=6.00

  Query                                  -----  As Used  -----      -----  Computed  ----
  Frame  MatID Matrix name              Lambda     K        H       Lambda     K        H
   +0      0   blosum62                0.315   0.134    0.375      same     same     same
               Q=9,R=2                 0.244   0.0300   0.180       n/a      n/a      n/a

  Query
  Frame  MatID  Length  Eff.length      E     S W   T  X    E2      S2
   +0      0     466       466        10.   65 3   13 22   0.060   38
                                                     35   0.063   42


Statistics:

  Database:  /usr/local/db/ufmg/t_parva
   Title:  /usr/local/db/ufmg/t_parva
   Posted:  4:06:00 PM EDT May 31, 2001
   Format:  BLAST-1.4
   # of letters in database:  8,929,689
   # of sequences in database:  564
   # of database sequences satisfying E:  4
  No. of states in DFA:  490 (52 KB)
  Total size of DFA:  76 KB (128 KB)
  Time to generate neighborhood:  0.00u 0.00s 0.00t  Elapsed: 00:00:00
  No. of threads or processors used:  1
  Search cpu time:  4.72u 0.06s 4.78t  Elapsed: 00:00:06
  Total cpu time:  4.76u 0.06s 4.82t  Elapsed: 00:00:06
```

Appendix K Restriction fragment sizes of λ dash 13 derived from Southern blot analysis of λ dash 13 hybridised to probes AThook1 and Ta369.

PROBE	EcoRI	HindIII	SpeI	XbaI	KpnI/ SpeI	EcoRI/ HindIII	EcoRI/ SpeI	EcoRI/ XbaI	HindIII/ SpeI	HindIII/ XbaI	SpeI/ XbaI
AThook1											
	†1.2	*3.2	*3.2	8	†8	†1.2	†1.2	*0.8	†1.6	*3.2	3.2
	*0.8	†1.6		P5	*3.2	*0.8	*0.8		*1.4	1.2	P3.1
				P3.2						1	*1.3
				1							1
Ta369											
	*2.5	†3.2	*3.2	8	†8	*2.1	*2.1	*2.5	†1.6	3.2	3.2
		*2.8		P7	*3.2			*2.3		*2.8	P2.8
				3.4							*2.0
				3.2							

Approximate sizes of restriction fragments (in kb) mapped to *TashAT1* and the putative *TashAT3* gene (Fig. 3.24) (denoted by * and †, respectively). P : partially digested fragments.