# Molecular characterisation of novel functionally important molecules of the model parasitic nematode, *Toxocara canis*.

Thesis submitted for the degree of
Doctor of Philosophy of the
University of Edinburgh

by

Kevin Kwaku Adjei Tetteh B.Sc (Hons)

Institute of Cell, Animal and Population Biology
University of Edinburgh
Kings' Buildings
West Mains Road
Edinburgh
EH9 3JT

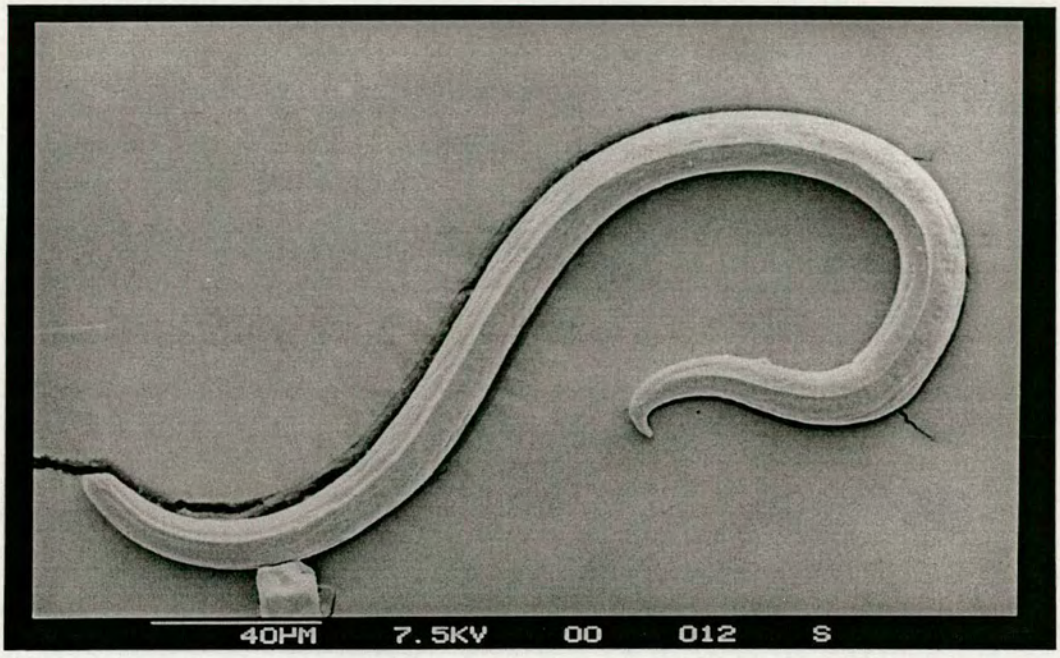
July 1999

# Abstract

The gastrointestinal parasite of dogs and their related species, *Toxocara canis*, is a prime example of a zoonotic parasite. It is the principal agent of visceral larva migrans and also a cause of ocular larval migrans. As a result of the close association that dogs have with man, the parasite enjoys a world-wide distribution. Infections can be contracted from contaminated soil and handling of infected dogs, and is particularly prevalent amongst professions that have a close association with dogs, such as hydatid control officers in New Zealand. Once inside the non-canid host, the parasite enters a state of arrested development, in which it neither grows nor differentiates. In this state the parasite releases up to 1% of its body weight in excretory/secretory antigens per day. Assuming that this high protein production was in some way linked to immune evasion, a modest EST project was undertaken using a cDNA library generated from infective larvae. The hypothesis behind this approach was that the high protein output demonstrated by these parasites would be mirrored at the mRNA level.

In total 266 clones were sequenced, the majority of which were from the both the 5' and 3' ends of the transcripts. Homologues for these genes were sought by similarity searching against the GenBank protein and the dbEST nucleotide databases. Cluster analysis of the clones identified 129 distinct gene products, all but three of which represented new genes. The majority of the genes (96) were represented by single clones, although 8 transcripts were present at high frequencies, each composing >2% of all the clones sequenced. These high abundance transcripts include a mucin and a novel C-type lectin, which together comprise the two major excretory/secretory antigens released by the parasite.

Four highly expressed novel transcripts were found, termed *ant* genes (*a*bundant *n*ovel *t*ranscripts). Together these genes represented 18% of all the cDNA clones isolated, but no similar sequences occur in the *C. elegans* genome. While the coding regions of these genes are dissimilar, they exhibit a remarkable level of similarity in their 3' UTR at the nucleotide level. The discovery of these abundant, parasite specific genes, of newly-identified lectins and mucins, as well as a range of conserved and novel proteins, provide defined candidates for future analysis of the molecular basis of immune evasion by *Toxocara*.




Electron micrograph showing an infective larva of *Toxocara canis*

# Table of Contents

# List of Figures

## Chapter 1



## Chapter 3

## Chapter 4

## Chapter 5

# List of Tables

## Chapter 1



## Chapter 3



## Chapter 4



## Chapter 5

# Acknowledgements

Finding the right words to properly express my thanks and appreciation for the help that I received over the course of my PhD was never going to be an easy task. As much as this is a pleasure it is also something of a necessity.

I wish to begin by expressing my thanks to my supervisor Prof. Rick Maizels. Who not only provided me with the opportunity to study for my doctorate, but whose input during its progress proved invaluable. It has been a pleasure, thanks especially for those 'canny' moments.

To my second supervisor Dr. David Arnot, for the words of wisdom and encouragement, thank you.

Dr. Alex Loukas. I just don't have the words. For all your help, the 70's experience and the many moments to numerous to mention. Thanks.

Bill Gregory, simple words are not enough, and yet, will have to suffice. You have my gratitude for all the help you have given me selflessly over the course of my project.

Also thanks to Dr. Judi Allen. My introduction to research could never have been better.

To Andrew, post-drinking mornings at work no longer hold the same simple pleasures for me.

Thanks also to Xing-xing, Janice, Martin, Natalia, P'ng, Franco and Letitia. It has been a pleasure.

# Chapter 1 Introduction

## 1.0 Introduction

It has been over forty years since the syndrome of visceral larva migrans (VLM) was first described by Paul Beaver and colleagues in 1952 (Soulsby, 1993). This was to be the starting point which paved the way to the vast body of literature that has since appeared on many aspects of the parasite, including the periodicity and range of pathology caused by *Toxocara* species.

### 1.0.1 *Lifecycle of* T. canis *in the dog*

The ascarid gastrointestinal parasitic nematode of canids, *Toxocara canis*, was first described by Werner in 1782 (Morgan and Hawkins, 1953). Later, in 1905, Stiles began the taxonomic classification of the parasite (Yorke and Maplestone, 1926), but it was not until 1958 that the complete lifecycle was finally defined by Sprent (Gillespie, 1987) (see fig 1 below).

The life cycle of *T. canis* is typically complex of ascarid infections, involving both intermediate and paratenic hosts (Gillespie, 1987). Mature dogs become infected by predation of infected hosts, from contact with other infected dogs or from contact with a contaminated environment (fig 1). Once within the host the parasite undergoes a period of somatic migration following a defined route which includes passage through such organs as the lungs, liver and heart. A high infecting dose can result in increasing levels of mobidity and mortality to the host. This period of somatic migration culminates with the larvae reaching full patency in the gut of the canid. Once the parasite reaches adulthood the adult female is capable of producing up to 20,000 eggs per day, up to a period of 6 months. Although surrounded by a tough resilient coating the eggs are susceptible to desiccation, sunlight and temperatures below -15°C. The eggs are very long lived but require suitable conditions in terms of temperature and humidity before reaching complete development after a duration of 2-6 weeks. (Gillespie, 1988).

Although not all the migrating larvae will go on to reach full patency, some will deviate from their migration and enter a state of arrested development within muscle and organs of the host. This state of arrested development not only provides a means for the

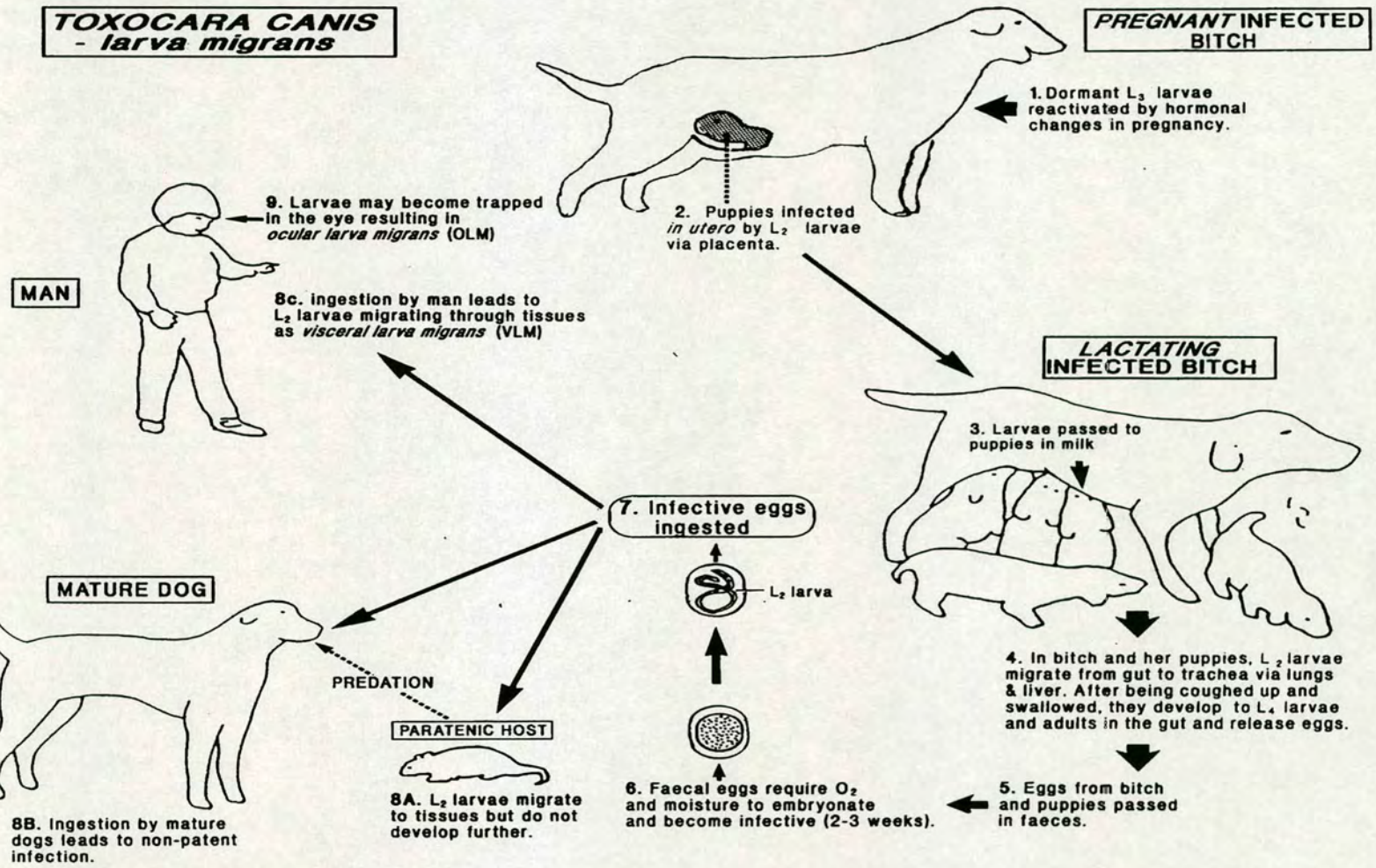



Figure 1.1. Diagramatic representation of the *Toxocara canis* lifecycle (Smyth, 1994).

parasites continued existence via predation of an infected host, but also a source of infection of subsequent generations through the bitch. Arrested larvae within the pregnant bitch can become activated upon receiving hormonal cues released during pregnancy. Once activated, the larvae undergo a period of somatic migration whereby the developing embryos become infected *in utero,* via the placenta. Puppies can also become infected via the transmammary route, resulting from the ingestion of larvae in the milk of the infected bitch. This two pronged attack in infection of the puppies is highly effective resulting in a 100% infection rate of the puppies. The bitch and puppies then enter a cycle of re-infection during weaning, via the faeces and vomitus. Infections can be eliminated with the use of anthelmintics (Gillespie, 1988).

### 1.0.2 *Lifecycle in humans*

The first incidence of *T. canis* in humans was diagnosed in a human eye in 1952 by Paul Beaver. Since that time, *Toxocara* infections have been found not only in the eye but in the brain, liver, lungs and to lesser extents in other regions of the body . It has always been very difficult to diagnose *T. canis* infections and as a result *Toxocara* related illness have often been misdiagnosed. The unnecessary enucleation of the eye due to the incorrect diagnosis of retinoblastoma, a type of cancer of the eye, led to increased awareness of this parasitic zoonosis.

Human infections are mainly contracted by children resulting from the ingestion of contaminated soil (pica) (e.g. sandpits, gardens) (Dent et al., 1956), or from handling of infected dogs (Fig. 1). In this way other helminth eggs can be ingested, and infections contracted. Hence in some cases *T. canis* infection have often been associated with *Trichuris* and *Ascaris* infections. Although the main cause of VLM is due to *T. canis* other ascarids are also known to cause this syndrome, such as *Baylisascaris procyonis* and to a lesser degree, *Lagochilascaris minor* and *T. cati* (Gillespie, 1988).

The pathology associated with these zoonoses tends to arise from the aberrant migratory behaviour of the parasite brought on by the parasite being in the atypical host and therefore receiving inappropriate environmental cues.

### 1.0.3 *Epidemiology of* Toxocara *infections*

*Toxocara canis* is regarded as one of the most widely distributed nematode parasites of mammals, sharing a high prevalence in taxa as dissimilar as the primates and cetaceans (Maizels and Robertson, 1991; Page et al., 1991). The domestic dog represents a large reservoir of *T. canis* throughout the world. In 1993 it was reported that there were approximately 55 million dogs residing in 40% of households in the US, and 7.4 million in 26% of households in the UK (Glickman, 1993). Comparable rates of dog ownership were also reported for other developed countries, including France, Germany and Japan. Published data indicates that of 42,000 dogs 15% are infected with the parasite. These figures would be higher in underdeveloped countries. Uga *et al* (Uga et al., 1996) showed that the incidence of environmental contamination with Toxocara eggs was in the region of 13 - 69% dependant on the district. The reason attributable for the differences was thought due to the fact that there were fewer places for animals to defecate, particularly with reference to the defecation habits of cats (Uga et al., 1996).

### 1.0.4 *Toxocariasis*

Most of the existing data on toxocariasis is focused on *T. canis*, primarily because of the close association shared between humans and dogs; this makes the probability of transmission to humans a very real risk. Although cats also share a close bonds with humans, the incidences of *T. cati* infections are far less frequent.

Human infections with *Toxocara* can lead to a variety of syndromes which collectively are referred to as *Toxocariasis*. The pathological syndromes are dependant on the level and frequency of the infecting dose. Most of the pathologies associated with Toxocara infections result from the particular tropism the parasites appear to have for neurological and ocular tissue.

There are a number of syndrome associated with Toxocara the best characterised of which are, ocular larva migrans (OLM) or ocular toxocariasis (OT) and visceral larval migrans (VLM), There is also a third form of the disease known as covert toxocariasis. All of these result from the prolonged migration of the parasite within the host. It is generally accepted that the syndromes of VLM and OLM are caused by a large infecting dose or a small infecting dose respectively. The pathology resulting from the

two syndromes is very different, and results primarily from the infecting dose as opposed to the response of the host immune system. However, it is possible for OLM to occur in conjunction with VLM where the infecting dose is high.

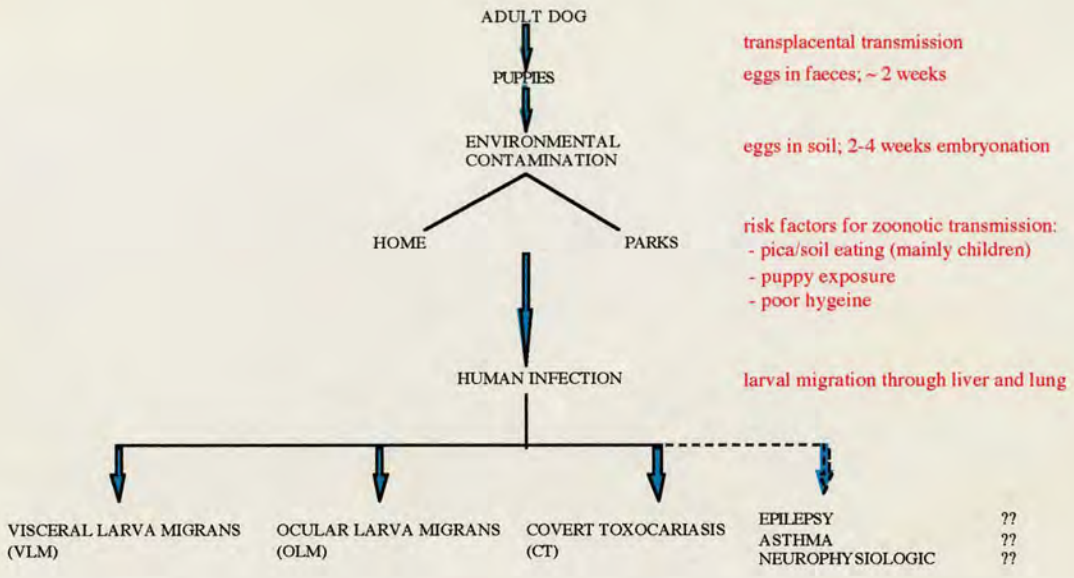



**Figure 1.2** . Schematic highlighting the epidemiological features associated with toxocariasis. The schematic also highlight potential routes by which preventative measures could be implemented in order to reduce the impact on humans.

The syndrome of VLM is associated with the migration of larvae within the host. Generally these infections are associated with children, particularly those between the ages of 2 - 7 years where there is a history of contact with dogs and pica (Glickman, 1993). Although adults, particularly those with a close association with dogs professionally (veterinarians and dog breeders) have shown signs of infection. One study showed that 15.7% of dog breeders in a kennel were seropositive for Toxocara. A similar study based around hydatid disease control officers in New Zealand showed that 25.6% were seropositive compared to 3.3% of the control population (Gillespie, 1988). The resultant clinical signs of VLM are directly linked with the level of the infecting dose and also the level of exposure. Symptoms associated with VLM are abdominal pain, fever, coughing wheezing and asthma (Glickman, 1993). Treatment of VLM involves antihistamines, corticosteroids and bronchodilators. Although effective in the treatment of dogs the use of anthelmintics in humans is somewhat controversial with few clinical trials to evaluate their efficacy (Glickman, 1993).

OLM results from as little as one migrating larva into the region of the eye. The resulting eosinophilc inflammation can lead to damage to the retina or optic nerve. In severe cases this aggressive eosinophilc recruitment leads to partial or full loss of sight. It is often the case that partial sight loss often goes undetected until discovered through a routine medical examination. If the sight of the inflammation is not in an essential area and the level of inflammation is mild then visual defects may go undetected (Gillespie, 1988; Glickman, 1993). Although an infection of both children and adults it was found that OLM tended to occur more in adults than in children and is also more common in males (2:1)(van Knapen and Buijs, 1993). Treatment of OLM is similar to that of VLM (Glickman, 1993).

The third classification of toxocariaisis has less obvious clinical symptoms which include coughing, headaches and behavioural disturbances. This syndrome is termed covert toxocariasis (CT) due to the fact that the symptoms presented tend to be common, non-specific complaints (Taylor, 1993). Although it has been suggested that this syndrome is in fact the more common of the characterised syndromes associated with toxocariasis (Taylor, 1993). Serological studies revealed a statistically valid link between high serum antibody titres to *T. canis* and one or more of the associated symptoms (Glickman, 1993).

A number of other syndromes have been linked with *Toxocara* infections of which include epilepsy, asthma and neurological deficits. Although there has been conflicting evidence for the link between asthma and epilepsy (Glickman, 1993).

### 1.0.5 *Ultrastructure of the nematode cuticle*

*T. canis* belongs to the phylum Nematoda, one of the most abundant groups of animals on the earth, second only to the insects. Both these groups owe much of their success to the possession of a resilient extracellular layer, termed the cuticle (Page, 1991a). Much of the understanding of the nematode cuticle has come about from research on the well characterised free-living nematode *Caenorhabditis elegans* (Blaxter and Robertson, 1998). The lack of a suitable parasitic nematode model species has meant that to date the knowledge obtained about nematode cuticles, in fact nematodes as a whole, is primarily from *C. elegans* with a small proportion of the information obtained from studies on adenophorean nematodes and plant parasites (Blaxter and Robertson, 1998).

The cuticle of *T. canis* is completely surrounded by a 10-20 nm electron dense layer that lies 10 nm distal to the epicuticular surface. This structure, termed the surface coat or glycocalyx is a negatively charged structure that binds the compounds ruthenium red, cationised ferritin (Himmelhoch et al., 1977; Himmelhoch and Zuckerman, 1983; Jansson et al., 1986; Page et al., 1992a; Zuckerman et al., 1979) and wheat germ agglutinin (WGA) (Page et al., 1992d). The surface coat has been implicated in the evasion of the host immune mechanisms (Badley et al., 1987; Page et al., 1992c).

Figure 1.3 shows an oblique freeze fracture section through a *T. canis* larva. The figure shows the internal surface of the surface coat with the invaginations of this structure into the troughs of the cuticle. This demonstrates the complete coverage of the parasite effected by this structure.

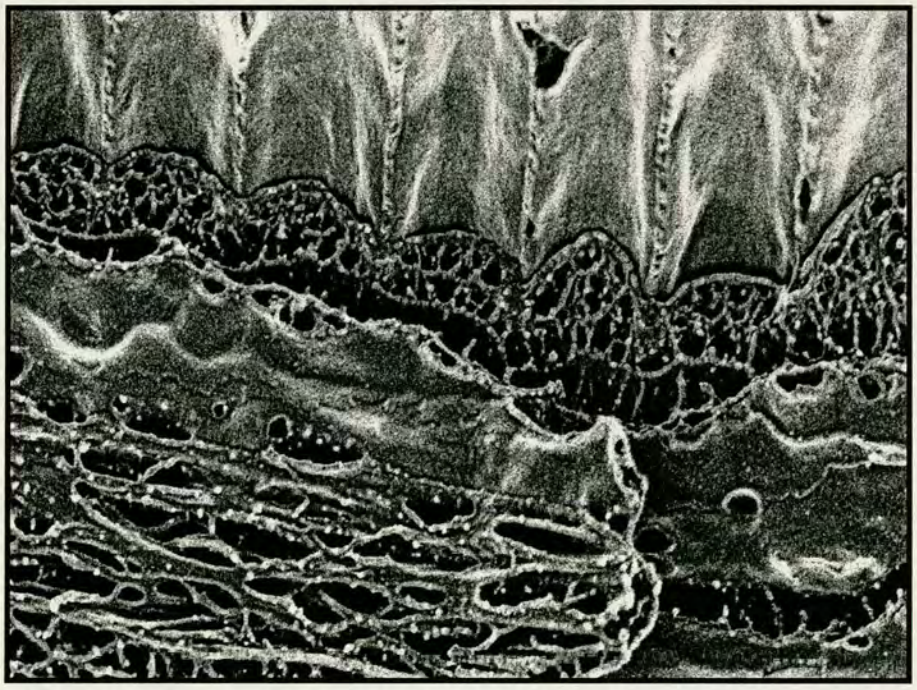


Figure 1.3 Electron micrograph showing an oblique freeze fracture section through the surface of an infective *T.canis* larva. From the figure the internal layer of the surface coat can be seen as well as the invaginations into the grooves of the epicuticle.

1.0.6 Toxcara canis *excretory/secretory (TES) antigens*

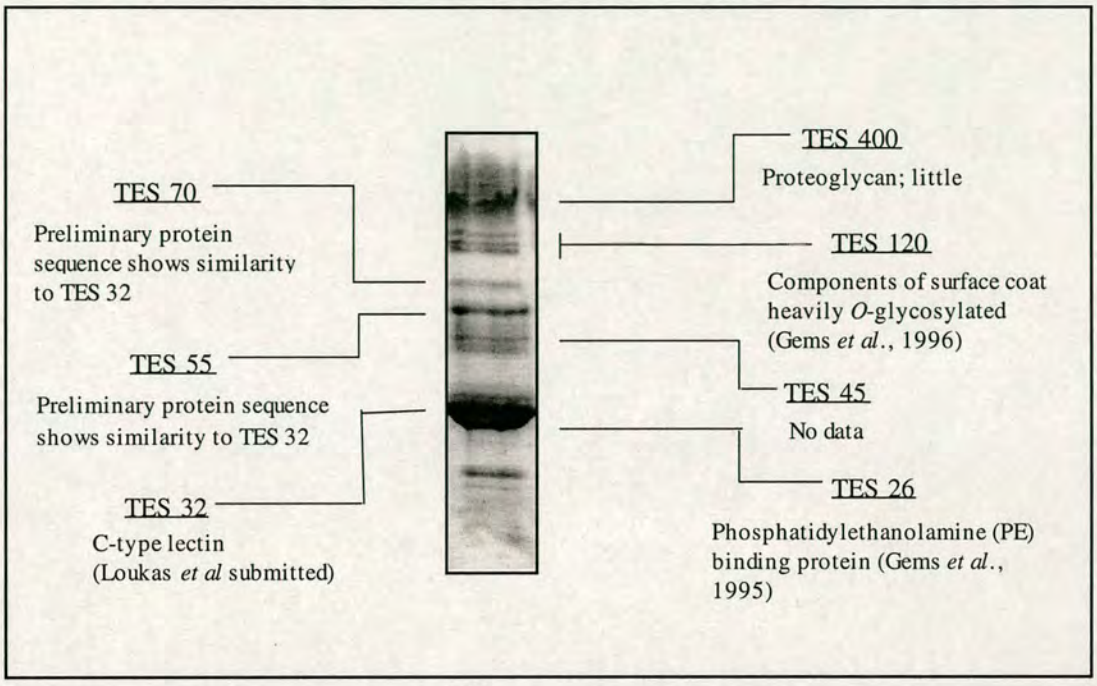



Figure 1.4 Antigen profile of *T. canis* TES. Information relating to the previously characterised bands is indicated.

The importance of TES products in *Toxocara* was first documented by de Savigny in 1975 (de Savigny, 1975). In his report he stated that *T. canis* infective larvae cultured in serum -free medium remained metabolically active for a period of many months. During this time the parasite was shown to produce copious quantities of proteins which proved to be immunologically reactive in serodiagnostic assays. This observation proved to be a milestone in terms of the immunological studies of *Toxocara* and other nematodes.

The vast majority of helminth species secrete varying quantities of what basically amounts to a cocktail of proteins. The components of these parasite cocktails have been investigated as potential targets of the immune response and also for their ability to interfere with host immunity (Maizels et al., 1993).

There are seven major bands in *T. canis* excretory/secretory (TES) antigen that have been characterised to varying degrees. Both TES 26 and TES 120 (Figure 1.4) were first obtained as part of a nematode spliced leader PCR using cDNA prepared from the infective larvae (Gems et al., 1995). TES 26 was shown to be a member of a family of

genes some of which bound the lipid phosphatidylethanolamine with a high degree of specificity. TES 120 was shown to be hyperabundant at the mRNA level (Gems et al., 1995) and from previous experiments to be an abundant component of the parasite surface coat by a combination of techniques which included surface iodination and electron microscopy (Gems and Maizels, 1996). Gel profiles of TES 120 showed it to migrate as multiple bands (Figure 1.4). Further characterisation of this molecule showed it to be a heavily *O*-glycosylated mucin-like protein. Recently, a number of clones were isolated from a *T. canis* expressed sequence tag (EST) project, characterisation of which has shown that they form part of a putative family of mucins (Tetteh et al, manuscript submitted) (Loukas et al, manuscript in preparation). At present five mucin genes have been identified and designated MUC-1 to -5 (see Chapters 3 and 5).

TES 32 is an N-linked glycoprotein found in the excretory/secretory antigen as well as being a component of the parasite surface associated with the cuticle beneath the surface coat. This antigen also represent the major TES product secreted by cultured infective larvae (Gems and Maizels, 1996; Page and Maizels, 1992). Sequence relating to TES 32 was isolated from the *T. canis* EST project (Tetteh et al manuscript submitted) and latter shown to belong to a novel family of calcium dependent lectins described as the C-type lectins or CTLs (Loukas et al manuscript submitted). The key feature of C-type lectins is the presence of a carbohydrate recognition domain (CRD). This group can be classified into two groups depending on the sugars they bind. Those that bind galactose and its derivatives (Gal-type) and those that bind mannose, glucose and their derivatives (Man-type). To date the TES 32 C-type lectin, *Tc-ctl-1*, is unique in its ability to bind both GalNac and D-mannose (Loukas et al manuscript submitted).

Of the remaining five antigens no information presently exists for TES 45 and the only information ascertained so far for TES 400 is that it is a heavily glycosylated molecule with a small peptide backbone. This leads to the conclusion that TES 400 may be a proteoglycan (Meghji and Maizels, 1986). Both TES 55 and TES 70 have shown at the sequence level to bear similarity to TES 32, the novel C-type lectin (Loukas et al manuscript submitted).

The significance of investigating the excretory/secretory material of nematode species is supported by the success obtained by a number of groups. The acetylcholinesterase from a number of nematode systems have been investigated by many groups as a

1991); two putatively protective excretory-secretory (ES) antigens were isolated from the nematode parasite of ruminants, *Haemonchus contortus* (Schallig et al., 1997). Initial studies with these low molecular weight products afforded a significant degree of protection against *H. contortus* in sheep (Schallig and van Leeuwen, 1997).

### 1.0.7 *Proteases*

Cysteine peptidases can be classified into 35 families (Barrett and Rawlings, 1996). Barrett and Rawlings (Barrett and Rawlings, 1996) arranged these 35 families into five clans each of which was thought to have originated from separate evolutionary origins. The clans were designated CA, CB, CC, CD and CE of which clan CA was the largest and also contained the papain (C1) family of peptidases (Barrett and Rawlings, 1996) (see Table 1).

Proteolytic enzymes have long attracted the interest of investigators in terms of their possible applications as tools in protein sequence analysis (Abe et al., 1993) and their ability to regulate a variety of biological processes, and subsequently their potential as therapeutic targets (Barrett and Rawlings, 1996). Proteases, in particular cysteine peptidases, are present in virtually all living organisms from DNA and RNA viruses to mammalian systems (Barrett and Rawlings, 1996). The cysteine peptidase group of enzymes represents a large group classified into 35 families, of which the papain sub-group is the largest. Papain was also the first cysteine protease to be identified (Light et al 1964), and represents the achetypical cysteine protease. Of the cysteine peptidases identified to date, the majority are classed as endopeptidases with some acting additionally or exclusively as exopeptidases (Barrett and Rawlings, 1996).

Table 1.1

| CLAN | FAMILY | NAME | EC |
|---|---|---|---|
| **CA** | C1 | Papain | 3.4.22.2 |
| | | Chymopapain | 3.4.22.6 |
| | | Caricain | 3.4.22.30 |
| | | Glycyl endopeptidase | 3.4.22.25 |
| | | Stem bromelain | 3.4.22.32 |
| | | Ficain | 3.4.22.3 |
| | | Actinidain | 3.4.22.14 |
| | | Asclepain | 3.4.22.7 |
| | | Vignain | - |
| | | Calotropin | - |
| | | Ananain | 3.4.22.31 |
| | | Comosain | - |
| | | Decapod digestive endopeptidase | - |
| | | Cathepsin L | 3.4.22.27 |
| | | *Fasciola* cysteine peptidase | - |
| | | Cathepsin S | 3.4.22.27 |
| | | Cathepsin O | - |
| | | Cathepsin K | 3.4.22.38 |
| | | Cathepsin H | 3.4.22.16 |
| | | Aleurian | - |
| | | Histolysain | 3.4.22.35 |
| | | Nematode cysteine endopeptidase | - |
| | | Cathepsin B | 3.4.22.1 |
| | | *Schistosoma* cathepsin B | - |
| | | Dipetidylpeptidase I | 3.4.14.1 |
| | | House dust mite cysteine endopeptidase | - |
| | | Cruzain | - |
| | | *Leishmainia* cysteine endopeptidase | - |
| | | *Plasmodium* cysteine endopeptidase | - |
| | | *Theileria* cysteine endopeptidase | - |
| | | Dictyostelium cysteine endopeptidase | - |
| | | Baculovirus cysteine endopeptidase | - |
| | | Yeast bleomycin hydrolase | - |
| | | *Lactococcus* PepC | - |
| | | *Streptococcus* aminopeptidase | - |
| | | *Porphyromonas* tpr proteinsase | - |
| | C2 | μ-Calpain | 3.4.22.17 |
| | | m-Calpain | 3.4.22.17 |
| | | *Drosophila* sol protein | - |
| | C10 | *Streptococcus* streptopain | 3.4.22.10 |
| | | *Porphyromonas* streptopain | - |
| **CB** | C3 | Polio picornavirus piconain 3C | 3.4.22.28 |
| | | Polio piconavirus picornain 2A | 3.4.22.29 |
| | | Cowpea mosaic comovirus 24 kDa endopeptidase | - |
| | | Grapevine fanleaf nepovirus endopeptidase | - |
| | | Hepatitis A virus picornain 3C | - |
| | C4 | Tobacco etch potyvirus L1a proteinase | - |
| | C24 | Feline calicivirus 2C endopeptidase | - |
| | C30 | Murine hepatitis coronavirus p80 endopeptidase | - |
| | C37 | Southhampton virus endopeptidase | - |
| | C38 | Parsnip yellow fleck virus endopeptidase | - |

Table 1.1. (continued)

| CLAN | FAMILY | NAME | EC |
|---|---|---|---|
| **CC** | C6 | Tobacco etch potyvirus helper component proteinase | - |
| | C7 | Chestnut blight fungus virus p29 endopeptidase | - |
| | C9 | Sindbis togavirus nsP2 proteinase | - |
| | C16 | Mouse hepatitis coronavirus papain-like endopeptidase 1 | - |
| | C21 | Tymovirus endopeptidase | - |
| | C23 | Blueberry scorch carlavirus endopeptidase | - |
| | C27 | Rubella rubivirus putative endopeptidase | - |
| | C28 | Foot-and-mouth disease virus leader proteinase | - |
| | C29 | Murine hepatitis coronavirus papain-like endopeptidase 2 | - |
| | C31 | Porcine respiratory and reproductive syndrome arterivirus PCP α endopeptidase | - |
| | | Equine arterivirus PCP α endopeptidase | - |
| | C32 | Equine arterivirus PCP β endopeptidase | - |
| | C33 | Equine arterivirus Nsp2 endopeptidase | - |
| | C34 | Apple chlorotic leaf spot closterovirus putative papain-like endopeptidase | - |
| | C35 | Apple stem grooving capillovirus putative papain-like endopeptidase | - |
| | C36 | Beet necrotic yellow vein furovirus putative papain-like endopeptidase | - |
| **CD** | C14 | Interleukin-1β-converting enzyme | 3.4.22.36 |
| | | CED3 protein | - |
| | | Apopain | - |
| **CE** | C5 | Adenovirus endopeptidase | - |
| - | C11 | Clostropain | 3.4.22.8 |
| - | C12 | Ubiquitin C-terminal hydrolase PGP 9.5 | - |
| | | Ubiquitin conjugate hydrolase YUH1 | - |
| | | Ubiquitin conjugate hydrolase L3 | - |
| - | C13 | *Schistosoma* haemoglobinase | - |
| | | α-legumain | - |
| | | β-legumain | 3.4.22.34 |
| - | C15 | Pyroglutamylpeptidase I | 3.4.19.3 |
| - | C19 | Isopeptidase T | - |
| | | Ubp1 ubiquitin peptidase | - |
| | | Ubp2 ubiquitin peptidase | - |
| | | Ubp3 ubiquitin peptidase | - |
| | | Ubp4 ubiquitin peptidase | - |
| | | Doa4 ubiquitin peptidase | - |
| | | Ubp5 ubiquitin peptidase | - |
| - | C22 | *Trichomonas* cysteine proteinase | - |
| - | C25 | Gingipain R | 3.4.22.37 |

Tablel 1.1 Shows the characterisation of the cysteine peptidases into clans and families as demonstrated by Barret and Rawlings (Barrett and Rawlings, 1996). The table was originally composed based on the comparison of available amino acids and is arranged in terms of common ancestory.

The biomedical importance of these enzymes is undisputed. Cysteine peptidases are thought to be involved in the pathophysiology of mammals, including man; in both bacterial and parasitic pathogens, in the physiology and pathology of plants and are also involved in the replication of viruses (Barrett and Rawlings, 1996).

Cysteine peptidases are so defined when the thiol group acts as the nucleophile in enzymatic reactions (Barrett and Rawlings, 1996). To date the activity of cysteine peptidases is thought to rest on that of a catalyitc triad of amino acids at the reactive site, with the second member, histidine, serving as a general base. The arrangement of the cysteine and histidine within the triad (Cys/His, His and Cys) in the linear sequence is taken as evidence that the cysteine peptidases may have been derived from many evolutionary origins (Barrett and Rawlings, 1996).

Proteases are utilised by parasitic worms for a range of functions. These include, among other things, hydrolysis of haemoglobin and manipulation of the host immune response (such as cleavage of immunoglobulin and complement components) (Dowd et al., 1994). Figure 1.5 shows a diagrammatic representation of the interaction between a cathepsin L peptidase and it's substrate (Dalton et al., 1995a).

Peptidases have been reported from many helminths including *Haemonchus contortus* (Gamble et al., 1996; Karanu et al., 1993; Rhoads and Fetterer, 1995), *Schistosoma mansoni* (Ghendler et al., 1996), *Brugia pahangi* (Hong et al., 1993), *Toxocara canis* (Robertson et al., 1989) and *Ancylostoma caninum* (Dowd et al., 1994; Hotez and Cerami, 1983).

Identification and localisation of proteases could enable immunological and chemotherapeutical strategies to be employed against the parasite (Gamble et al., 1996; Wasilewski et al., 1996). The identification of such functionally important proteases in *T. canis* would further expand the usefulness of this parasite as a model system by which to study parasite survival in the host.

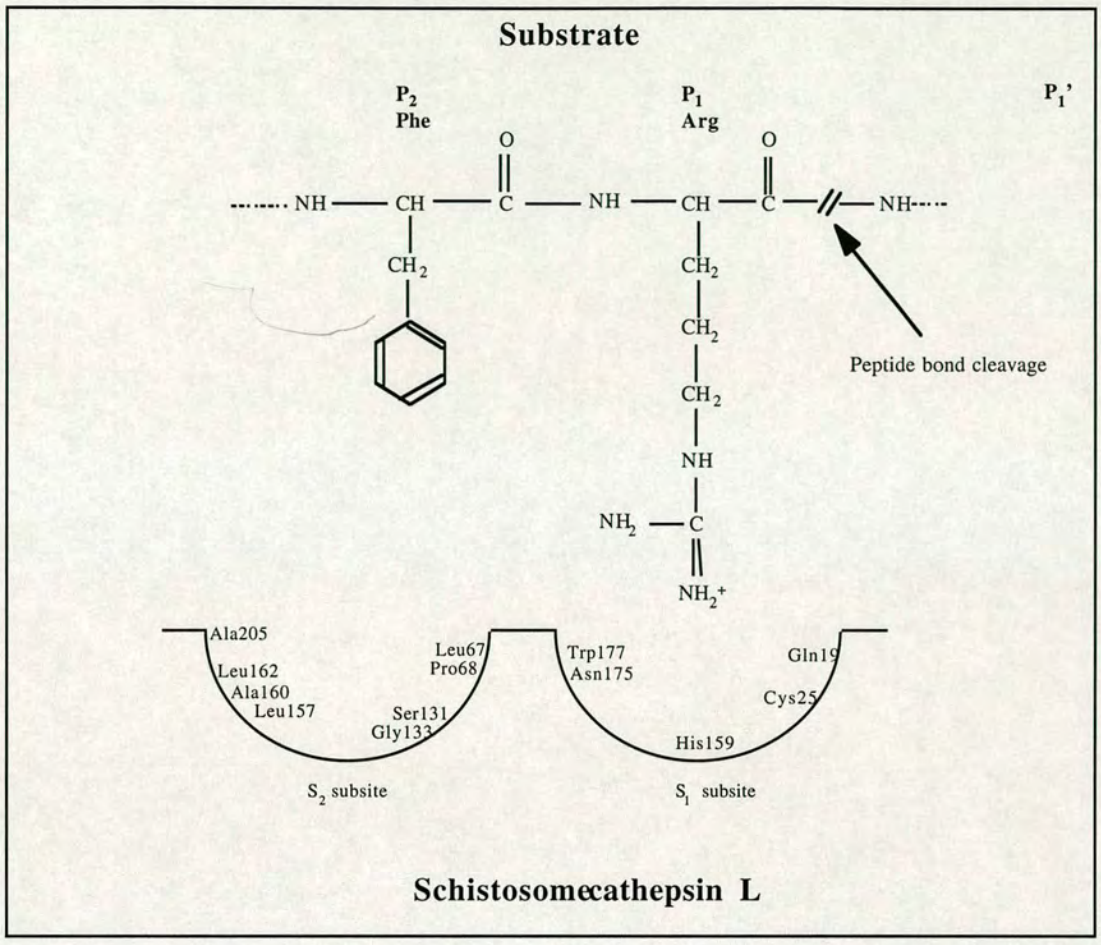



**Figure 1.5** Schematic representation highlighting the key interactions between a cathepsin L peptidase - in this example the S. mansoni cathepsin L - and it's substrate. The amino acids of the $S_1$ subsite of the cathepsin L peptidase (indicated in figure) interact with the $P_1$ residue situated on the N-terminal side of the scissle peptide bond and are involved in the catalytic cleavage of the bond (indicated by arrow). The interactions between the amino acids of the enzyme $S_2$ subsite and the residue at the $P_2$ position (in this example, Phenylalanine) of the substrate determine the specificity of the cysteine peptidases. In this diagrammatic example the cathepsin L-like proteinases have a preference for hydrophobic residues, such as the phenylalanine at position $P_2$.

### 1.0.8 *Mucins*

Mucus glycoproteins or mucins are a family of heavily glycosylated, high molecular weight proteins (400 to > 1000 kDa) found on the surface of a variety of epithelial cells (Devine and McKenzie, 1992; Kim et al., 1996). Many of these include mammary glands, digestive tract, kidney and testes (Devine and McKenzie, 1992). Mucins can be broadly divided into two groups, those that are membrane associated and those that are secreted. The secreted forms can be further subdivided into those likely to form gels and the nature of the gel formed (Kim et al., 1996).

The structure of mucins consists of a non-globular protein backbone, with carbohydrate side chains of varying lengths, sequence, composition and anomeric linkages attached via many *O*-glycosidic linkages. A major feature of mucins is the presence of repeat peptide sequences, of different amino acid composition and length, flanked on either side by non-repetitive domains; these repeat domains have high levels of serine and/or threonine present, potential O-glycosylation sites (Kim et al., 1996). The core proteins of mucins are composed of high levels of serine, threonine, alanine, proline and glycine with low amounts of aromatic and sulphur containing amino acids (Devine and McKenzie, 1992; Kim et al., 1996). The glycosylation of mucins accounts for 50 - 85% of the total molecular weight of mucins giving them a much higher buoyant density than non-glycosylated proteins. This high level of glycoslation differs between species, for example, whereas human mucins are highly glycosylated, mucins from sheep possess one carbohydrate molecules per mucin (Devine and McKenzie, 1992).

Ilg *et al* (1996) purified a secreted filamentous mucin-like proteophosphoglycan from the protozoan parasite *Leishmania major*. The secretion of novel, complex glycoconjugates is thought to be important in the survival and development of the parasite within the sandfly vector and its mammalian host (Ilg et al., 1996; McConville and Bacic, 1990)

### 1.0.9 *Diagnosis of* Toxocara canis

The initial attempts at diagnosis of *Toxocara* infections was based predominately on assays that detected *Toxocara* antibodies. Before that, diagnosis was based on morphological identification carried out on specimens obtained by biopsy or from

necropsy (Nichols, 1956). Apart from the technical difficulty involved, classification of the isolated organisms also proved to be a problem. Such as the accurate determination of the parasites genus. The problem with the majority of the techniques employed was that they lacked specificity; failing to distinguish toxocaral infections from those of ascariasis, filariasis and trichinellosis. Among the tools employed in the diagnosis of VLM infections were the skin sensitivity test, where reactions to *Toxocara* antigens were observed e.g. larval precipitation, complement fixation and immunofluorescent testing (van Knapen and Buijs, 1993). To compound the problems already plaguing these techniques false positive results were arising from crossreactivity between the antigens beings used and the heterophile antibodies or isohaemaglutinins (Cypess et al., 1977). Often the antigens used in these earlier assays were derived from adult material, a stage of the parasite not present in humans as the parasite does not develop beyond the infective larvae.

de Savigny (1975) described the successful culturing of the infective larva of *Toxocara canis*. While in culture the parasites remained in a developmentally arrested yet metabolically active state releasing approximately 1% of the body weight in excretory-secretory antigen (TES) per day (Meghji and Maizels, 1986). These antigens are composed of material found in the cuticle and also a fuzzy outermost layer of the parasite termed the surface coat (Page et al., 1992a; Page et al., 1992b). These antigens used in conjunction with an enzyme linked immunosorbent assay (ELISA) (van Knapen and Buijs, 1993) and also immunofluoresence (Gillespie et al., 1993) provided investigators with suitable diagnostic tools in the diagnosis of toxocaral infections. The first assay system which later became commercially available was based on the detection of human anti-*Toxocara* antibodies (Jacquier et al., 1991). Gillespie *et al* later described an antigen capture method which detected repeat elements found the TES antigens of *T. canis* (Gillespie et al., 1993).

Despite the fact that an ELISA kit has been established since 1991 (Jacquier et al., 1991), the system is still not ideal in that it cannot distinguish between active infections and dormant infections (Gillespie et al., 1993). The problems of species cross-reactivity still appear to be evident (Gillespie et al., 1993; Lynch et al., 1988). So it would appear that the search is still on for a *Toxocara* specific diagnostic antigen. In a report by Loukas et al., (1998) the identification of a cathepsin L-like cysteine protease was described. In a preliminary experiment using a bank of human serum, patients previously exposed to *T. canis* recognised the cathepsin whereas the control sera did not. Further experiments are presently underway to investigate the likelihood of r*Tc*-

cpl-1 being an effective diagnostic antigen for the diagnosis of human toxocariasis (Loukas et al., 1998).

### 1.0.10 *Arrested larval Development*

The phenomenon of arrested development, also refered to as hypobiosis, is an evolutionary adaptation designed to protect the organism from harsh, less favourable times. In such situations the larvae arrest their developmental cycle for a period of several months to years. In some species hyperbiotic rates of up to 100% were observed. In such cases it is thought unlikely that environmental cues alone could be responsible. Jacquiet *et al* (1996) hypothesise that factors in the parasite were genetically predetermined to promote the onset of hypobiosis (Jacquiet et al., 1996).

Although *T. canis* is a ubiquitous parasite, its level of incidence in humans both in developing and developed countries is minor (Page, 1991). The ability to collect the excretory/secretory antigens provides the investigator with the opportunity to investigate a biochemically conserved set of nematode antigens derived from an important parasite compartment. The state of arrested development of the larvae *in vivo* is mirrored *in vitro*, where larvae cultured in serum-free culture media are capable of surviving for in excess of 18 months (de Savigny, 1975). This may prove to be of value in the diagnosis or prophylaxis of a complicated zoonosis (Meghji and Maizels, 1986), providing information that could prove to be of value in the study of other less hardy parasite systems. *T. canis* larvae provide a unique opportunity to examine the release of antigenic molecules by a pathogenic organism (Maizels et al., 1987), primarily because they remain in this arrested state for many years, lying dormant in the musculature or the nervous system. In this state the parasite is developmentally and reproductivley arrested but is metabolically active releasing a steady 1% of its body weight in ES antigens per day (Gems and Maizels, 1996; Meghji and Maizels, 1986).

The unique ability of *T. canis* to survive for extended periods in artificial culture media provides an excellent system by which to study many facets of the parasites biology. This includes the nature of the surface coat and the nature and function of the parasites antigens.

*1.0.11 Immunomodulation of the host immune responses*

Helminthic parasites employ a variety of immune evasive strategies to be able to avoid the defensive measures generated by the host immune system (see section 1.0.5 and 1.0.6) (Finkelman et al., 1991). The characteristics features associated with multicellular parasite infections are classically the induction of the Th2 immune response featuring production of IgE, eosinophilia and mastocytosis. Due to the persistant nature of *T. canis* infections the infective larvae elicit strong, persistant immune responses demonstrated in particular by IgM, IgG and IgE antibodies to *Toxocara* excretory/secretory antigens (TES)(see section 1.0.6). Although it has recently been reported that it is now possible to distinguish, serologically, between the different syndromes presented during toxocaral infections (Obwaller et al., 1998) it still remains unanswered as to how exactly these responses are being induced and by what. In order to ba able to devise an approapriate chemotherapeutic strategy or vaccine approach a number of fundamental questions must be tackled.

i) Diagnosis: Although constantly improving, existing methods remain unsatisfactory, particularly in terms of sensitivity and treatment surveilance (Zhong-Jin et al., 1999).

ii) Immunopathology: A large proportion of the pathology associated with *T. canis* is associated with immune-mediated inflammatory responses. Although there is some evidence that some damage arises from migrating larvae (Smith, 1991). It has also been suggested (Obwaller et al., 1998) that the pathology seen in *T. canis* infections is the result of an inappropriate immune response by the host, in terms of parasite clearance or hypersensitivity (see section 1.0.4). This point is of particular interest as it could be a deliberate strategy by the invading parasites to misdirect the host immune response

iii) Parasite survival/persistance within the host.

### 1.2.1 Aims of project

*Toxocara* serves well as a model system by which to study some of the many facets of parasitic nematode biology. The longevity exhibited by the parasites in culture together with the vast amount of excreted antigen released by the parasite enables investigators to study an important parasite compartment. It also allows investigators to characterise a biochemically conserved set of nematode antigens derived from an important parasite of some medical concern.

At the outset of the project it was proposed that the purpose of this proposal was to provide information that would serve to expand existing knowledge about the parasite as a whole. The purpose of this was to obtain data that could be applied to the treatment and/or control of Toxocariasis and also applied to the treatment and control of other more important zoonoses of medical concern.

To tackle the task ahead it was determined that in order to achieve the somewhat ambitious goals of the project the most significant approach would be the initiation of an expressed sequence tag (EST) project (Chapter 3). The significance of this approach was that since the parasite stage being investigated was developmentally arrested, reproductively immature yet metabolically active, the focus of its energies would be directed solely at being able to survive the malign environment of the host immune system.

By these criteria the goals of the project have been met.

# Chapter 2 Materials and Methods

## 2.0 Materials and Methods

### 2.0.1. *Collection and culture of* T. canis *adults.*

Adult parasites were collected from The Royal Veterinary College, Potters Bars, North London from postmortemed dogs suspected of having a *Toxocara canis* infection. The gut of the dog was removed and washed over a sieve using warm physiological saline. After being washed free of mucous and other debris the worms were subjected to an extensive washing protocol comprised of 5 washes in physiological saline followed by 5 washes in supplemented RPMI (see appendix) to remove as much of the contaminating material as possible. The parasites were left in supplemented RPMI for approximately 5 hours at room temperature (r.t.) before being washed a further 5 times in supplemented RPMI. Adult worms were cultured in 150 $cm^2$ vented culture flasks (Corning Costar) at 37°C and under 5% $CO_2$. The worms were maintained in this state for approximately 7 days; significant deterioration was observed after this period. The medium was changed on a daily basis to collect the adult excretory/secretory (ES) material and to harvest the eggs. After the 7 days the worms were washed free of RPMI and the males stored at -70°C. The females were stored in a 5% formalin solution at room temperature to encourage further embryonation of eggs before subsequent bifurcation of the adults.

### 2.0.2. *In vitro culture of* T. canis *larvae*

Infective larvae were hatched from embryonated ova obtained from adult female worms as described above, and more recently from the Center for Parasitology, The University of Texas. Hatching was performed following a modification of techniques described in de Savigny (1975), Oaks and Kayes (1979), Smith (1989) and Maizels (1984). Culturing was performed following a modification of the Oaks and Kayes (1979) protocol. Briefly, embryonated ova were aliquoted into siliconised 50 ml centrifuge tubes and washed free of formalin with sterile water by centrifuging at 1000*g* for 5 mins. The pellet was then resuspended to a 10% v/v suspension in sodium hypochlorite and incubated for 10 minutes at 37°C, or until the tough pitted outer shell had been dissolved. During the incubation period the tube was agitated at regular intervals to disperse the ova and prevent the build up of froth formed from the dissolved egg shell proteins. A Swinnex 47 filter unit (Millipore) was assembled with

a 10 $\mu$m nylon filter and supported in a clamp stand above a large beaker. The barrel of a 50 ml syringe was attached to the top of the filter unit to function as a funnel. On completion of the incubation period the egg suspension was immediately passed through the Swinnex unit, followed by approximately 500 mls of sterile water. On completion the unit was reversed and the syringe barrel re-attached to the upturned unit. Supplemented RPMI was the forced through the filter unit using the attached syringe to flush the eggs of the filter membrane. Eggs were collected in a 50 ml siliconised centrifuge tube; a steady stream of $CO_2$ was bubbled through the suspension for 4-5 mins to initiate the hatching of the larvae. The suspension was then aliquoted into a modified Baermann apparatus (see below) and incubated overnight at 37°C, 5% $CO_2$. The hatched larvae were collected over a period of several days as they emerged, and were cultured in 25 $cm^2$ vent cap culture flasks at a density of 25,000 - 35, 000 larvae per 10 mls of supplemented RPMI at 37°C, 5% $CO_2$. The culture supernatent from the larvae was replaced once a week with fresh media; the culture supernatant was centrifuged at 1500 *g* for 10 min and stored at -80°C.

### 2.0.3 *Construction of Baermann apparatus*

The construction of the Baermann apparatus was performed following a modification of the method outlined in Oaks (1979). Essentially, the elongated tip of a Pasteur pipette was snapped off and flame polished to remove any sharp edges. The pipettes were then siliconised in dimethlydichlorosilane (DMCS) solution, followed by extensive washes, once with 70% alcohol, followed by two with water then completed with a final wash of 70% alcohol . The dried pipettes were loosely plugged with cotton wool (approximately 1 inch) and autoclaved before use.

### 2.0.4 *Preparation of larval* T. canis *excretory/secretory (TES) products*

Approximately 2 litres of culture supernatent was thawed from -70°C storage and concentrated under a nitrogen induced pressure of 30 lbs/$in^2$ over a 10,000 molecular weight cut-off membrane using an Amicon ultraconcentration unit. The concentrated TES material was washed with two volumes of sterile phosphate buffered saline

(PBS), and concentrated to a final volume of 3 mls; storage was in 0.5 ml aliquots at -70°C.

### 2.0.5 $^{125}$ *I labelling of soluble proteins*

TES products (10 μg) were labelled as described in (Maizels et al., 1991). The labelled protein were loaded onto a G-25 column (Pharmacia) and washed with 5 mls of PBS, and a succession of 250 μl aliquots collected. 1 μl of each tube was tested for total activity and percentage tricholoacetic acid (TCA) precipitability. The fractions with >80% TCA precipitated radioactivity (usually the first peak) were retained. The labelled aliquots were stored at -20°C.

### 2.0.6 *Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS PAGE)*

Preparation of the separating portion of the gel was performed following the details outlined below:

Separating gel: (~ 20mls)

| | 7.5% | 10% | 12% | 15% |
|---|---|---|---|---|
| Acrylamide/BIS 39 : 0.5 (40% stock) | 3.75 ml | 5.0 ml | 6 ml | 7.5 ml |
| 1.5 M Tris Base pH 8.8 (90.86g/500ml) | 5.0 ml | 5.0 ml | 5.0 ml | 5.0 ml |
| 10% SDS | 200 μl | 200 μl | 200 μl | 200 μl |
| Distilled water | 11.0 ml | 9.7 ml | 8.7 ml | 7.2 ml |
| 10% APS | 100 μl | 100 μl | 100 μl | 100μl |
| TEMED | 8 μl | 8 μl | 8 μl | 8 μl |

The water, Tris, acrylamide and sodium dodecyl sulphate (SDS) were properly mixed together, before addition of the ammonium persulphate (APS) and the TEMED.

The Stacking portion of the gel was followed as outlined overleaf:

Stacking gel (5%): (~ 10 ml)

| | |
|---|---|
| Acrylamide/BIS 39 : 0.5 (40% stock) | 1.25 ml |
| 1.0 M Tris Base pH 6.8 | 1.25 ml |
| Distilled Water | 7.45 ml |
| 10% SDS | 100 µl |
| 10% APS | 50 µl |
| TEMED | 10 µl |

2.0.7 *5 - 25% SDS-PAGE Gradient gels*

The SDS PAGE gradient gels were set up using the Hoeffer SE 600 gel system. The gel components were prepared as outlined in (Maizels lab manual). Essentially the gels were prepared as detailed below:

Separating gel: 67 ml

| 5% | | 25% | |
|---|---|---|---|
| 4.2 ml | Acrylamide | 21 ml | Acrylamide |
| 4.2 ml | Solution M* | 4.2 ml | Solution M* |
| 25.2 ml | Distilled water | 8.4 ml | Distilled water |
| 10 µl | TEMED | 10 µl | TEMED |
| 150 µl | APS | 100 µl | APS |

* solution M - see appendix

The stacking portion of the gel was prepared as outlined below:

Stack: ~ 20 ml

| | |
|---|---|
| 2.5 ml | Acrylamide |
| 2.5 ml | Solution S* |
| 15 ml | Distilled water |
| 10 µl | TEMED |
| 200 µl | APS |

* solution S - see appendix

2.0.8 *RNA preparation*

RNA was extracted from 40,000 *T. canis* infective larva using the TRISOLV™ RNA isolation reagent (Gibco BRL). The procedure was outlined in the protocol (as supplied with reagent) but essentially the 40,000 infective larva were homogenised in 100 μl of the TRISOLV™ reagent using a microhomogeniser (Biomedix). The volume of the homogenate was increased to 1 ml with the TRISOLV reagent and incubated at room temperature for 5 minutes before proceeding with the extraction. On completion of the extraction, the RNA was resuspended in 100 μl water to which was added 100 mM EDTA, pH 8.0. Due to the instability of the RNA, 10μl of 3M sodium acetate (pH 7) and 250 μl of 100% ethanol was added and the tube stored at -70°C.

2.0.9 *Construction of* T. canis *infective larva cDNA library*

The infective larva cDNA library was a kind gift from Cindy Tripp and Robert Grieve (Heska, Fort Collins). The library had been constructed from 200,000 *T. canis* infective larvae following a single step guanadine-phenol-chloroform extraction protocol. cDNA synthesised from the mRNA obtained was unidirectionally cloned into the Uni-Zap XR phage vector (Stratagene), using packaging extracts from Stratagene. The amplified library contained $1.9 \times 10^9$ phage/ml with 91% recombinants. The possibility of host contamination was essentially eliminated because eggs were first incubated *in vitro* in formalin, and subsequently larvae collected through Baermann filters were cultured in serum-free RPMI 1640 medium for extended periods.

2.0.10 *Picking of individual clones*

Single clones were identified as serendipitously picked phage from the plated library. The methodology used was that outlined in the Stratagene protocol. Essentially, 1 μl of serially diluted cDNA library was incubated with 200 μl of XL1-Blue *E. coli* for 15 mins at 37°C. The cells were then mixed with 3 mls Lauria-Burtani (LB -see appendix) top agar (48°C), poured onto previously prepared 82 mm agar plates supplemented with tetracycline (12.5 μg/ml) and allowed to set. The plates were then incubated at 37°C overnight. Single plaques were cored out from semi-confluent plates using sterile Pasteur pipettes and the agar plug ejected into 1.5 ml centrifuge tubes containing

500 µl of SM buffer, with 20 µl chloroform as a preservative. The cored plaques were briefly vortexed and left at 4°C overnight (o/n) to allow the phage to elute from the agar plug.

2.0.11 *Excision of phage insert*

The methodology for this procedure was as described in the Stratagene cDNA synthesis kit (Stratagene) manual. SOLR and XL1-Blue bacteria were streaked out onto kanamycin and tetracyline plates, respectively, from frozen stocks and incubated at 37°C o/n. The following day single colonies were picked from each plate and used to inoculate 10 mls of LB broth containing the appropriate antibiotic and this was allowed to grow o/n at 37°C in a shaking incubator. A 0.5 ml aliquot was taken from each culture and used to inoculate 50 mls of LB-kan (SOLR) and LB-tet (XL1-Blue) broth in conical flasks. The flask were then incubated at 37°C in a shaking incubator until mid log phase ($OD_{600}$ of 0.2 - 0.5; approximately 3 hours. XL1-Blue bacteria were centrifuged and resuspended in 10 mM $MgSO_4$ to an $OD_{600}$ of 1.0; the SOLR bacteria were allowed to continue growth at room temperature (rt) or stored at 4°C o/n. Aliquots of 200 µl of the XL1-Blue bacteria were combined in a 15 ml centrifuge tube with 250 µl of the eluted phage stock plus 1 µl of the ExAssist helper phage (Stratagene). The reaction was incubated for 15 mins at 37°C. 3 mls of LB broth was added and the tube incubated in a shaking incubator for 2-2$^1/_2$ at 37°C. Once completed the tube was incubated at 70°C for 15 mins and centrifuged for 15 mins. The supernatants, containing the excised phagemid, were aliquoted into two 1.5 ml tubes and stored at -80°C.

2.0.12 *Preparation of phagemid DNA*

Preparation of phagemid DNA from excised phagemid stocks was performed as outlined, in the Strategene cDNA synthesis kit manual (Strategene). Briefly, 10 and 100 µl of each phagemid was aliquoted into two 1.5 ml centrifuge tubes respectively, each with 200 µl of SOLR cells ($OD_{600}$ = 1.0). The tubes were incubated at 37 °C for 15 mins; 100 µl of each was spread onto LB-ampicillin (50 µg/ml) plates and incubated o/n at 37 °C. A single colony from each phagemid was used to inoculate 10 mls of LB-amp and incubated o/n at 37 °C in a shaking incubator.

Phagemid DNA was prepared from the 10 ml o/n culture using the Qiagen Spin Miniprep plasmid kit (Qiagen, Germany). The procedure used was according to the Qiagen manual. The purified phagemid was eluted into 50 μl of water and stored at -20°C.

2.0.13 *Dye Terminator cycle Sequencing Reaction*

The phagemid templates previously prepared were sequenced using the dye terminator (dt) cycle sequencing chemistry with Amplitaq™ DNA polymerase (FS enzyme - Perkin Elmer). Two microlitres of each plasmid template was used in a dt reaction as standard (see table below):

| | |
|---|---|
| Template | 2.0 μl |
| Dye terminator mix | 4.0 μl |
| Primer (20 ng/μl) | 0.5 μl |
| Water | 3.5 μl |

The reaction was carried out as outlined in the Perkin Elmer manual. Using the Bluescript primers M13 Forward & M13 Reverse, T3 and T7.
Samples were completed for sequencing on the ABI, following a precipitation step designed to remove unincorporated primer and reaction mix. This protocol was outlined in the Perkin Elmer manual.

The dt reactions were analysed using an ABI 377 automated sequencer (Applied Biosystems, Foster City, CA). All clones were sequenced from the 5' end, and most were subjected to 3' sequencing except where 5' sequences gave high quality sequence through the poly-A+ tail, or where the 5' sequence showed unequivocal identity with a previously characterised *T. canis* clone. Most of the clones were sequenced multiple times to reduce eliminate the incidence of error in the sequence data.

2.0.14 *Analysis of sequence data*

Sequence data was edited in the first instance using the program SeqEd v1.0.3 (Applied Biosystems) which allowed comparison between the electropheretic trace and the nucleotide sequence. This was used to identify and remove vector sequence and

flanking sequences. Sequences were aligned using AsemblyLign and the MacVector 6.0 (Oxford Molecular). It was possible to perform nonredundant database searches using BLAST (Basic Local Alignment Search Tool) (Altschul et al 1990) on the National Center for Biotechnology Information (NCBI) server.

# Chapter 3 *Toxocara canis* Expressed Sequence Tag (EST) Project

## 3.1 Introduction

Genome sequencing is increasing at an exponential rate. Since the pioneering studies on *Ascaris* more than a century ago, the genomes of parasitic nematodes have been under intense investigation. At present a vast body of data derived from genome sequencing projects is rapidly accumulating in public databases and it is predicted that investigators would soon have the necessary foundation with which to determine biological function from sequence data (Miklos and Rubin, 1996). A prediction already close to being realised.

One of the main functions of genome sequencing projects such as whole genome or expressed sequence tag (EST) sequencing is to provide a database of sequence information, preferably from a model species (Miklos and Rubin, 1996). With this information it would be possible to answer fundamental questions relating to the elucidation of biochemical pathways (Blattner et al., 1997; Bult et al., 1996), identifying homologues which determine the function in different species including humans (McCombie et al., 1992; Sonnhammer and Durbin, 1997) and ascertain virulence factors.

As a gene discovery approach, whole genome sequencing is not a new idea. In comparison to other animal genomes those of nematodes are unusually small - the genome size of ascarids is in the order of 0.51 - 2.4 x $10^8$ bp with repeated sequences accounting for 13 - 32% (Hammond and Bianco, 1992). Nonetheless, it is an uneconomical and time-consuming option when compared to the EST approach. ESTs are generally generated from single reads from one or both ends of a clone. This is an efficient means of identifying new genes and characterising the expressed gene content of cells and tissues (Adams et al., 1992; Adams et al., 1991; Hoog, 1991). EST sequencing is a versatile approach that can be scaled large or small depending on the life stage under investigation can be biased towards genes of importance to the parasite against which chemotherapeutic targets can be designed; those expressed at medium or high abundance during the lifecycle stage.

In contrast to this, however, El-Sayed and Donelson (1997) showed in their survey of the *Trypanosoma brucei rhodesiense* genome that this was not always the case. Using a single pass sequencing approach on both ESTs and genome survey sequences (GSSs) they showed very little difference in the novel gene discovery potential.

The arrested stage of *T. canis* is remarkable for surviving without development for as long as 9 years *in vivo* (Beaver, 1966), without reproduction or differentiation, and without succumbing to attack by the host immune system. In the absence of cell division, tissue growth or gametogenesis, we hypothesised that a significant proportion of *T. canis* larval protein production is likely to be directed at immune evasion, and that this may be reflected in the spectrum of mRNAs being synthesised by the arrested larvae. We therefore adopted the Expressed Sequence Tag (EST) approach and now report on the discovery of 126 new genes representing approximately 1% of the organisms predicted gene complement.

Table 3.1.

| dbEST rank | Organism Description | EST entries |
|---|---|---|
| 1 | *Homo sapiens* (human) | 847,851 |
| 3 | *Caenorhabditis elegans* (nematode) | 72,521 |
| 4 | *Arabidopsis thaliana* (thale cress) | 33,932 |
| 6 | *Drosophila melanogaster* (fruit fly) | 16,859 |
| 7 | *Brugia malayi* (parasitic nematode) | 10,806 |
| 8 | *Toxoplasma gondii* (protozoan) | 10,671 |
| 10 | *Schistosoma mansoni* (blood fluke) | 3,441 |
| 12 | *Saccharomyces cerevisiae* (baker's yeast) | 3,042 |
| 13 | *Plasmodium falciparum* (malaria parasite) | 2,872 |
| 14 | *Trypanosoma brucei rhodesiense* (protozoan) | 2,732 |
| 16 | *Zea mays* (maize) | 1,757 |
| 17 | *Onchocerca volvulus* (nematode) | 1,765 |
| 19 | *Sus scrofa* (pig) | 1,323 |
| 21 | *Dictyostelium discoideum* | 795 |
| 25 | *Trypanosoma cruzi* (protozoan) | 575 |
| 28 | *Schistosoma japonicum* (blood fluke) | 538 |
| 29 | *Leishmania major* (protozoan) | 535 |
| 36 | *Gallu gallus* (chicken) | 216 |
| 40 | *Trypanosoma brucei brucei* (protozoan) | 143 |
| 44 | *Canis familiaris* (dog) | 107 |
| 51 | *Toxocara canis* (parasitic nematode) | 86 |
| 58 | *Strongyloides stercoralis* (nematode) | 57 |
| 70 | *Meloidogyne javanica* (root-knot nematode) | 22 |
| 80 | *Urechis caupo* (carp) | 10 |
| 82 | *Trichostrongylus vitrinus* (parasitic nematode) | 9 |
| 91 | *Dictyocaulus viviparus* (bovine lungworm) | 2 |
| 94 | *Hydra vulgaris* | 1 |
| 96 | Total (organisms) | |
| | Total (EST entries) | 1,303,930 |

Table 3.1 Summary of the EST sequences entered into the dbEST database showing the number of EST's deposited into the public database.

Note: The above information was compiled from the dbEST Summary of Organisms (http://biology.bjmu.edu.cn/biology/ncbihtm/3458cd57.HTM) which was last updated on 31.10.97.

## 3.2 Materials and Methods:

### 3.2.1 T. canis *infective larvae cDNA Library*

A description of the construction of the cDNA library by Dr. Tripp (Heska, Fort Collins) is given in Chapter 2: Materials and Methods.

### 3.2.2 *Selection of individual clones*

XL1-Blue *E. coli* bacteria were streaked out from frozen stocks onto Lauria-Bertani (LB) agar plates containing 12.5 μg/ml tetracycline (tet) and grown overnight at 37$^0$C. A 10 ml culture of XL1-blue *E. coli* cells was prepared from a single colony and allowed to grow for approximately 16 hours or overnight in a 37$^0$C shaking incubator. A 0.5 ml aliquot of this culture was used to inoculate 50 mls of LB tet in a conical flask This was grown until an $OD_{600}$ of 0.5 was reached, approximately 3 hours. The *T. canis* infective stage cDNA library was titrated in SM buffer (Chapter 2: Materials and Methods) to yield $10^{-0}$ - $10^{-9}$ plaque forming units (pfu) per ml. One microlitre of each phage dilution was added to 200 μl of XL1-Blue *E. coli* bacteria and placed at 37$^0$C for 15 minutes to allow attachment of the phage to the cells. After this period the cells were mixed with 3 ml of top agar (Chapter 2: Materials and Methods), which had been previously melted and cooled to 50$^0$C, and plated out onto agar plates and allowed to set. Once set the plates were incubated at 37$^0$C overnight. The following day plaques were cored out from agar plates, using a pasteur pipette, where the density allowed single plaques to be picked with no possibility of contamination. This corresponded to the dilutions $10^{-3}$, $10^{-4}$ and $10^{-5}$. The cored single plaques were then dispenced into 500 μl of SM and incubated at 4$^0$C to allow the phage to elute out of the agar plug. Twenty microlitres of chloroform was added to act as a preservative.

### 3.2.3 *Estimation of insert sizes*

The insert sizes of the selected phage were determined by PCR using Bluescript primers (M13 Reverse and M13 Forward) with aliquots of the phage stock as template. A number of the clones proved difficult when attempting to determine the insert sizes. In those cases a restriction digest was performed on the purified plasmid (see Chapter 2: Preparation of plasmid DNA) using the enzymes *Xho 1* and *Xba 1* (Promega). Of the

332 clones picked from the library 66 (19.8%) did not have inserts and were not included in any further analysis. Some clones gave an insert size which later turned out to be an underestimate of the actual size of the clone when the clone was subsequently sequenced. Blue/white screening of the colonies was not carried in the selection of recombinant clones as due to the high titre of the library it was deemed somewhat unnecessary.

### 3.2.4 *Excision of phage insert*

Phagemids were excised as outlined in Chapter 2: Materials and Methods.

### 3.2.5 *Preparation of phagemid DNA*

Preparation of the phagemid DNA was as outlined in Chapter 2: Materials and Methods.

### 3.2.6 *Dye Terminator Cycle Sequencing Reaction*

The dye terminator protocol used was according to the Perkin Elmer manual, as detailed in Chapter 2: Materials and Methods.
For specific clones, internal primers were required to obtain further sequence data. In these cases, the primers used and the positions from which they were designed are shown on the nucleotide sequence, in the relevant sections of this chapter.

### 3.2.7 *Analysis of Sequence data*

Sequence analysis of clones was performed as outlined in Chapter 2: Materials and Methods.

### 3.2.8 *Nomenclature*

The genes identified from the EST project were named following the nematode nomenclature proposal as formed by Bird and Riddle (Bird and Riddle, 1994). Thus all genes are given a 3-letter name and a number, and the prefix *Tc* to denote the species (e.g. *Tc* = *Toxocara canis*). Three sets of interim gene names were used where no functional homology exists: *Tc-not* (= novel transcript), *Tc-ant* (abundant novel transcript, where =7 clones/266 were identical), and *Tc-huf* (homologue of unknown function, where similar sequences had been reported in other nematodes such as *C. elegans* or *Onchocerca volvulus*). For these interim gene designations, the number of the EST clone first sequenced was retained. For cDNAs assigned functional names, clones were generally numbered sequentially (e.g. *Tc-ctl-1*, *Tc-ctl-2* for C-type lectins), except where numbering was significant in other organisms, principally for ribosomal proteins (e.g. *Tc-rps-4*, *-5* and *-9* to conform with established nomenclature).

### 3.2.9 *Database Submission*

Sequences were deposited in NCBI dbEST (Banfi et al., 1998; Boguski et al., 1993) with separate entries for 5' and 3' reads. In a few cases the full-length sequence was also deposited.

## 3.3 RESULTS

### 3.3.1 *Sequencing and homology searches*

A total of 332 clones were sequenced from the 5' end, of which 266 (80%) contained inserts ranging from 128 to 2050 base pairs (Table 3.1). Of these, 223 were also sequenced from the 3' end. Five clones were found to be chimaeric, i.e. the 5' and 3' ends were derived from separate transcripts with different derivations. All nucleotide sequences were subjected to BLASTX searches against the nonredundant (nr) database at NCBI (National Centre for Biotechnology Information) GenBank. Where no homology was found, nucleotide sequences were also subjected to BLASTN searches (at both nr and dbEST databases), and where an unambiguous open reading frame could be assigned, the deduced protein sequence was used to search with TBLASTN (also with nr). Most homologies were found, or were stronger, with the 5' sequences, but a significant minority of homology relationships were found only with 3' sequence reads. In general, a probability value of $10^{-6}$ was sought as minimum degree of similarity.

### 3.3.2 *Clustering analysis*

All sequences were held as MacVector files, and the Align to Folder algorithm used to determine similarities between different EST clones. It was noted that in some of the larger clusters, identity of the 3' sequences was critical to correctly group differently truncated clones with non-overlapping 5' sequences. To some extent this reflected the moderate read lengths achieved in this work.

As a result of these analyses, a total of 129 distinct gene products were identified. Of these, only 3 had previously been characterised, namely mucin-1 (*Tc-muc-1*; (Gems and Maizels, 1996)) phosphatidylethanolamine-binding protein (*Tc-peb-1*; (Gems et al., 1995)) and the 60S ribosomal protein L3 (*Tc-rpl-3*; (Moore et al., 1995)). Table 3.2 gives a gross summary of the number of individual genes, and the total clones isolated, which fell into one of four different homology classes. Approximately 50% (65/129) of the total number of genes showed significant similarity to genes of known function from other species; a further 16 clones had database homologues of unknown function; and 48 genes (37%) showed no similarity to known genes: members of this last class were designated novel genes.

and 48 genes (37%) showed no similarity to known genes: members of this last class were designated novel genes.

**Table 3.2.** Gross summary of the EST project.

| | | Genes | Clones |
|---|---|---|---|
| Identified Homologues | (Table 3.6) | 65 * | 140 |
| Unidentified Homologues | (Table 3.5) | 16 | 20 |
| Abundant Novel Transcripts | (Chapter 4) | 4 | 48 |
| Novel Transcripts | | 44 | 53 |
| TOTAL | | 129 | 261 ‡ |

* Includes 3 genes previously characterised, i.e. *Tc-muc-1, Tc-peb-1* and *Tc-rpl-3.*

‡ Excludes 5 cDNAs found to be chimaeric and so not used in further analysis.

### 3.3.3 *Abundant Transcripts*

The clustering analysis produced some surprising results, with a small number of transcripts being heavily represented in the library. Eight transcripts (6.4% of genes) accounted for 106 clones (39.1%) from the cDNA library (Table 3.3), while the 20 most abundant (all those sequenced 3 or more times) accounted for 145 clones (55.1%). Table 3.4 presents the transcripts characterised in order of frequency, with the most highly abundant clone being cytochrome c oxidase subunit II (25/263 = 9.6% of clones), and the second most common clone being a C-type lectin (16/263 = 6.1%) which in published work was shown to correspond to TES 32, the major surface secreted glycoprotein of *T. canis* larvae (Loukas *et al*, 1999).

**Table 3.3.** Abundant clones

| NUMBER OF CLONES | DESCRIPTION |
|---|---|
| 25 | Cytochrome c Oxidase |
| 16 | C-type lectin (TES 32) *ctl-1* |
| 15 | Abundant Novel Transcript (*ant*) 003 |
| 7 | Abundant Novel Transcript (*ant*) 005 |
| 13 | Abundant Novel Transcript (*ant*) 030 |
| 13 | Abundant Novel Transcript (*ant*) 034 |
| 7 | Surface coat mucin (TES 120/*muc*-1) |
| 5 | Phosphatidylethanolamine (PE) Binding Protein (TES 26) |

The 8 most abundant clones identified as part of the EST project. Together accounting for 106 (39.8%) of the total clones sequenced from the library.

**Table 3.4** Frequency of transcripts

| | | |
|---|---|---|
| Cytochrome c oxidase, subunit II (*cox-2*) | 25 | 9.6% |
| C-type lectin, TES-32 (*ctl-1*) | 16 | 6.1% |
| Abundant Novel Transcripts 003 (*ant-003*) | 15 | 5.7% |
| Abundant Novel Transcripts 034 (*ant-034*) | 13 | 5.0% |
| Abundant Novel Transcripts 030 (*ant-030*) | 13 | 5.0% |
| Abundant Novel Transcripts 005 (*ant-005*) | 7 | 2.7% |
| Mucin, TES-120 (*muc-1*) | 7 | 2.7% |
| PE-binding Protein, TES-26 (*peb-1*) | 5 | 1.9% |
| Novel Transcript 095 (*not-095*) | 4 | 1.5% |
| Protein tyrosine phosphate (*ptp-1*) | 4 | 1.5% |
| 60S Ribosomal Protein L19 (*rpl-19* | 4 | 1.5% |
| ADP/ATP translocase (*aat-1*) | 3 | 1.1% |
| ASP Homologue-1/Tc-CRISP (*vah-1*) | 3 | 1.1% |
| Mucin-2 (*muc-2*) | 3 | 1.1% |
| Novel Transcript 018 (*not-018*) | 3 | 1.1% |
| Novel Transcript 120 (*not-120*) | 3 | 1.1% |
| 60S Ribosomal Protein PO (*rpp-0*) | 3 | 1.1% |
| Superoxide dismutase-3 (*sod-3*) | 3 | 1.1% |
| Tubulin alpha-3 chain (*tua-3*) | 3 | 1.1% |
| 14 transcripts present as two clones * | 28 | 10.7% |
| 96 transcripts present as single clones | 96 | 36.7% |
| Total Clones Sequenced | 261 | |

* The 14 genes present as 2 clones are *aki-1, ash-2, cap-1, huf-001, irp-1, muc-3, not-069, not-126, ofm-1, pam-1, pcc-1, rpl-7a, rpl-9* and *sod-4*. Details of these, apart from the *ants* and *nots* are given in Table 3.6

### 3.3.4 *Abundant Novel Transcripts*

Interestingly, the 4 next most abundant clones were all novel, with no similarities in the database or to each other, but occurred with frequencies between 2.7 - 5.7% of all cDNA clones randomly selected. These 4 transcripts, which together account for more than 18% of the library (Figure 3.1), have been designated *ant* (abundant novel transcript) genes, and retain the number of the EST clone from which they were first isolated (*ant*-003, 005, 030, and 034).

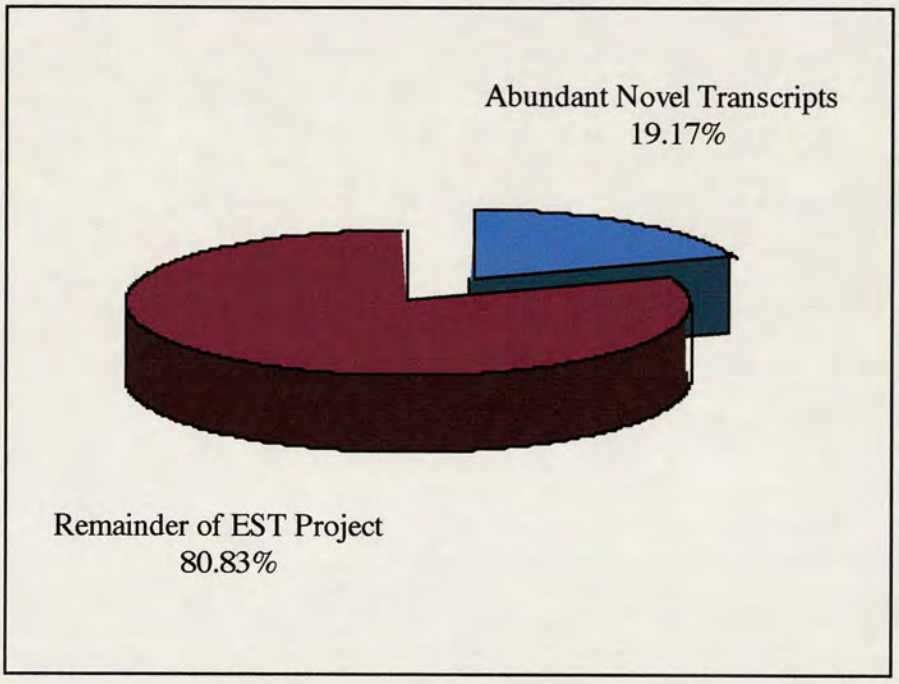



**Figure 3.1.** Pie chart showing the percentage of the EST project devoted to the Abundant Novel Transcripts (*ants*) versus the rest of the clones

### 3.3.5 *Homologues of Genes of Unknown Function* (huf)

Ten clones showed significant similarities to *C. elegans* sequences with no assigned function (Table 3.5, lines 2-12). These were all designated as *huf* genes, retaining the number of the corresponding EST clone. One clone, *Tc-huf-1* contained a tandem array of 3 blocks of 36 amino acids each containing 6 cysteines in identical alignment, and one block containing 5 cysteines. This novel domain, termed NC6 or SXC (nematode cysteine 6 or six cysteine) domains (Blaxter, 1998; Gems et al., 1995), has been found in *T. canis* and *C. elegans* proteins, particularly those associated with nematode surfaces, but no insight into the function of the array has yet been gained.

The pie chart in Figure 3.2 shows the percentage of ESTs that was devoted to these genes of unknown function.

**Figure 3.2.** Pie-chart showing the percentage of the EST project inhabited by the Homologues of Unknown Function (*huf*) and the Venom allergen/*ASP* homologues (*vah*).

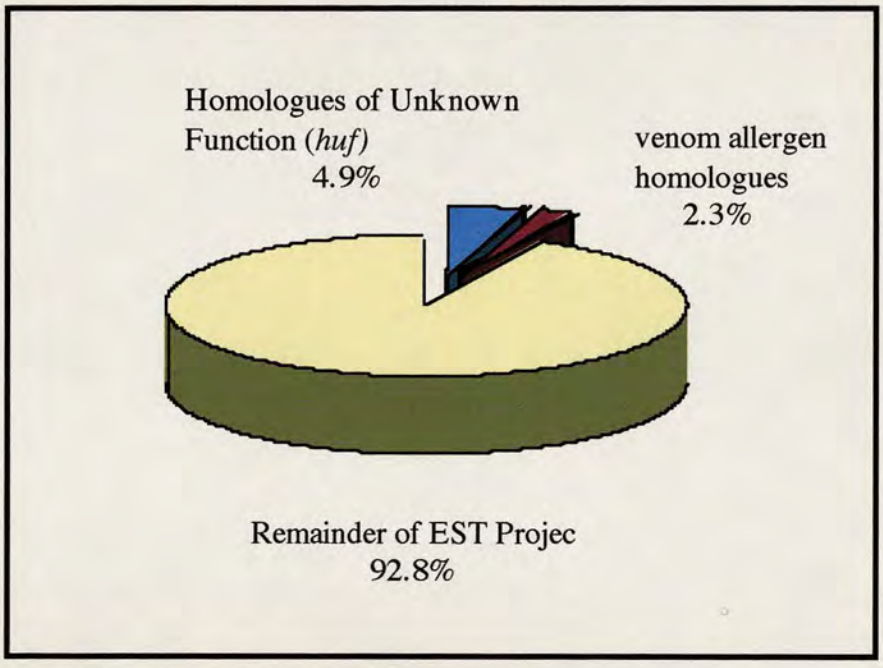


**Table 3.5** *T. canis* ESTs classified as homologues of nematode proteins of Unknown Function (*huf*) (16 genes). All the clones were isolated as single transcript except *vah-1* (3 clones), *vah-2* and *huf-001* (2 clones each).

| | Gene Name | Similarity | p value | EST Name | Insert Size (bp) | 5' Read | 3' Read | 5' Acc. No. | 3' Acc. No. |
|---|---|---|---|---|---|---|---|---|---|
| 1. | *Tc-huf-001* | NC6/SXC motif | | Tc-EST-001 | 800 | 764 | 666 | AA683457 | AA683458 |
| 2. | *Tc-huf-012* | *Caenorhabditis elegans* R53.5 gene | 3.8E-06 | Tc-EST-012 | 850 | 599 | 381 | AA563573 | |
| 3. | *Tc-huf-052* | *C. elegans* yk284h6 cDNA | 5.0E-18 | Tc-EST-052 | 420 | 395 | 602 | AA610968 | AA610969 |
| 4. | *Tc-huf-053* | *C. elegans* C13A10 gene | 2.8E-38 | Tc-EST-053 | 1000 | 589 | 521 | AA610971 | AA610972 |
| 5. | *Tc-huf-062* | *C. elegans* F41D9.1 gene | 4.3E-42 | Tc-EST-062 | 760 | 482 | 288 | AA612566 | AA907959 |
| 6. | *Tc-huf-159* | *C. elegans* C56E6.5 gene | 7.6E-15 | Tc-EST-159 | 1850 | 387 | 393 | AA836705 | AA836706 |
| 7. | *Tc-huf-169* | *C. elegans* K02E10.6 gene | 6.2E-17 | Tc-EST-169 | 1050 | 466 | 558 | AA874757 | AA874758 |
| 8. | *Tc-huf-252* | *C. elegans* Y57G11C.12 gene | 7.7E-30 | Tc-EST-252 | 1150 | 547 | 444 | AA684533 | AA684534 |
| 9. | *Tc-huf-264* | *C. elegans* B0034.3 gene | 1.4E-25 | Tc-EST-264 | 1420 | 589 | 416 | AA684531 | AA684532 |
| 10. | *Tc-huf-287* | *C. elegans* K12H4.5 gene | 5.3E-23 | Tc-EST-287 | 452 | 429 | 491 | AA875794 | AA979711 |
| 11. | *Tc-huf-296* | *C. elegans* T24F1.6 gene | 1.9E-11 | Tc-EST-296 | 2300 | 259 | 578 | AA875806 | AA875807 |
| 12. | *Tc-huf-316* | *C. elegans* T24D8.5 gene | 3.7E-21 | Tc-EST-316 | 720 | 542 | 351 | AA874711 | AA874712 |
| 13. | *Tc-huf-325* | *Onchocerca volvulus* L3CAN05C08SK | 7.4E-47 | Tc-EST-325 | 1250 | 370 | 590 | AA879368 | AA879369 |
| 14. | *Tc-vah-1* | *C. elegans* T05A10.5 venom allergen/ *Ancylostoma* secreted protein (ASP) homologue | 4.6E-15 | Tc-EST-269 | 1870 | 459 | 699 | AI 083044 | AI 083045 |
| 15. | *Tc-vah-2* | *Brugia malayi* Venom allergen/ASP homologue | 1.0E-13 | Tc-EST-303 | 1600 | 550 | 512 | AI 083048 | AI 083049 |
| 16. | *Tc-vah-3* | *Brugia malayi* Venom allergen/ASP homologue | 4.7E-16 | Tc-EST-249 | 650 | 507 | | AI 083050 | AI 083051 |

### 3.3.6 *Venom Allergen Homologues (vah)*

In addition to the aforementioned *huf* genes, three genes were found to show similarity to *Ancylostoma* secreted protein (ASP), which is associated with larval reactivation and development in other nematode species (Hawdon et al., 1996; Schallig et al., 1997). This gene family shows similarity to hymenopteran venom allergens, and in common with members reported from *B. malayi* and *O. volvulus*, has been designated *vah* (venom allergen homologue). Figure 3.3(a) shows an alignment of predicted proteins of *vah-1* and *-2* together with partial predicted sequence for *vah-3*. The full-length sequence for *vah-1* was obtained as part of the EST project, whereas *vah-2* was obtained by Janice Murray (pers. comm.). Identical residues are shaded in dark grey, while similar residues are in lighter grey boxes. Each *vah* gene has one *N*-linked glycosylation site - this is an assumption with *vah-3* as the full-length sequence has not yet been obtained - which is highlighted on the figure in green. *Tc-vah-1* has previously been reported as Tc-CRISP (Cysteine-rich secreted protein) by Leo Liu (pers. comm.), and is remarkable for containing four NC6/SXC domains in tandem with a *vah*-homology domain (Figure 3.3 (a)). *vah-2* does not possess any NC6/SXC domains, a feature which is as yet unknown for *vah-3*. Figure 3.3(b) shows a schematic of the *vah* genes and highlights their relation to each other. The conservation that exists between the NC6 domains in *vah-1* is detailed in Figure 3.3 (c). Both Figures 3.3(a & b) show that these cysteine rich domains are separated by a conserved 13 amino acid gap showing a significant degree of homology with each other. This homology is highlighted in Figure 3.3 (d). Also of interest, is the observation that NC6 domains 1 and 3 of *vah-1* are shorter by one amino acid than the NC6 domains of domains 2 and 4. The difference between the two domain pairs extends further in that there are amino acid differences common to domains 1 and 3 but not to domains 2 and 4 (Figure 3.3 (c)). NC6 domains 1 and 3 show a 50% difference to domains 2 and 4. Domains 2 and 4 in turn show a 36.4% amino acid divergence to domains 1 and 3. These differences are emphasised in Figure 3.3(c) in green for domains 1 and 3 and pink for domains 2 and 4.

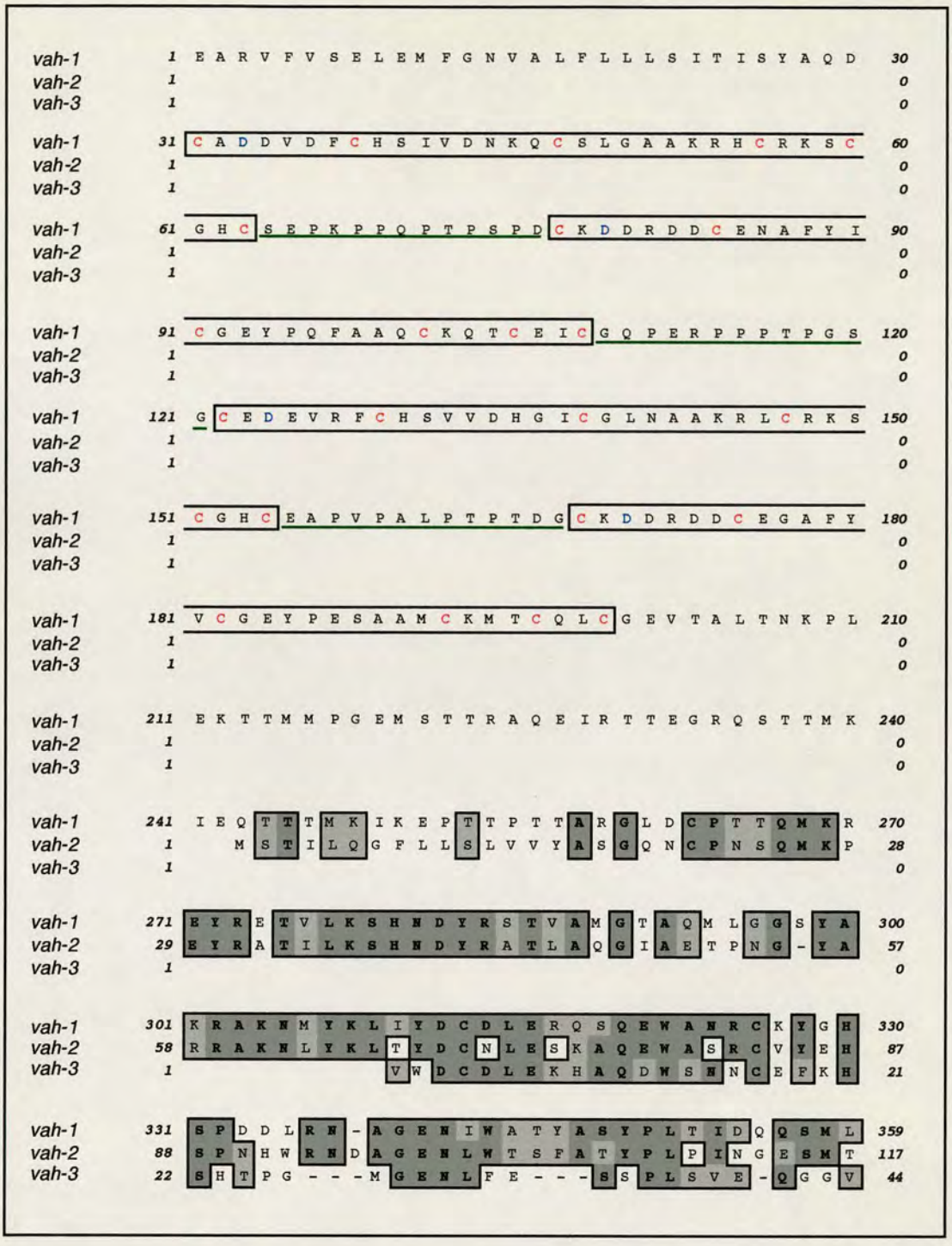



**Figure 3.3 (a).** Amino acid pileup of the *vah* genes. The boxed sections outline the NC6/SXC domains of *vah-1*. The dark green underlined sections show the conserved 13 amino acid gap that separates the NC6/SXC domains.

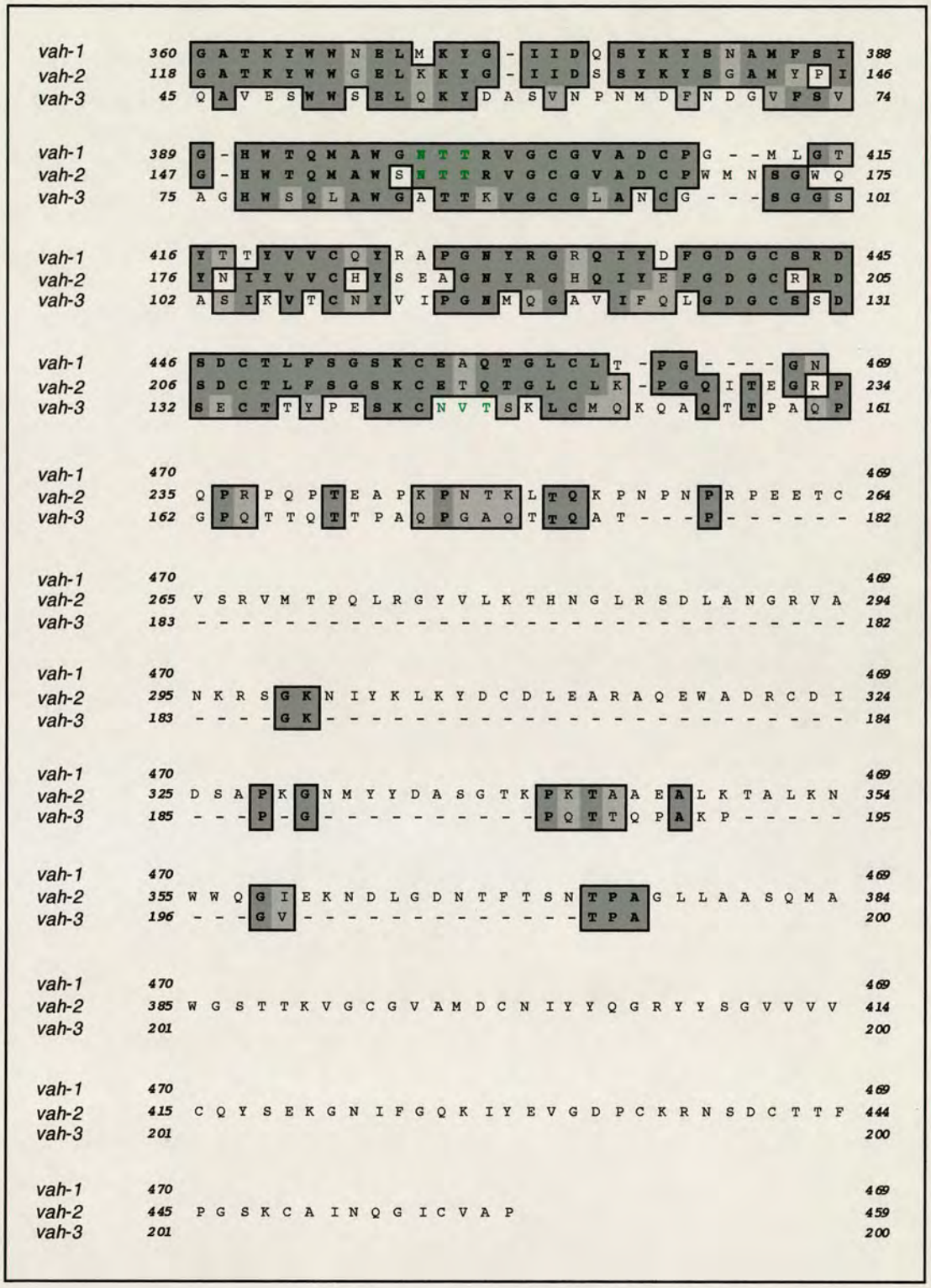



**Figure 3.3 (a) continued** *N*-glycosylation sites are highlighted in bright green.

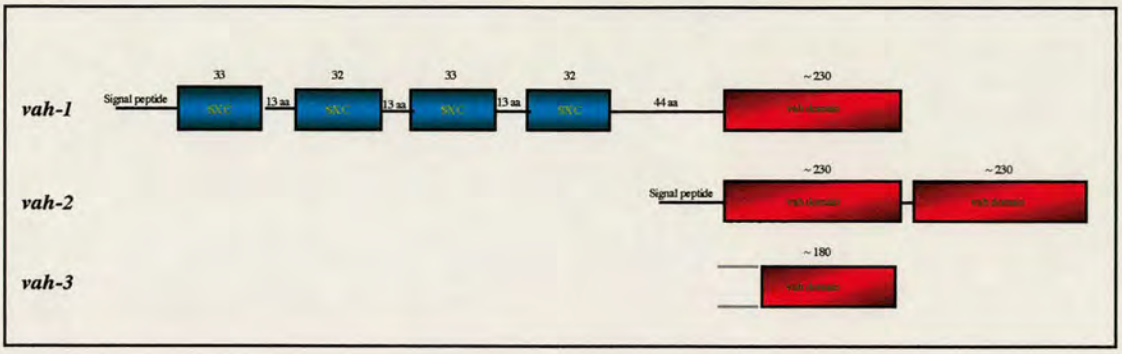



**Figure 3.3 (b).** Schematic representation of the venom allergen homologue (*vah*) genes isolated from the EST project. The grey portion of *vah-3* represents the fact that this conception is derived from a partial sequence.

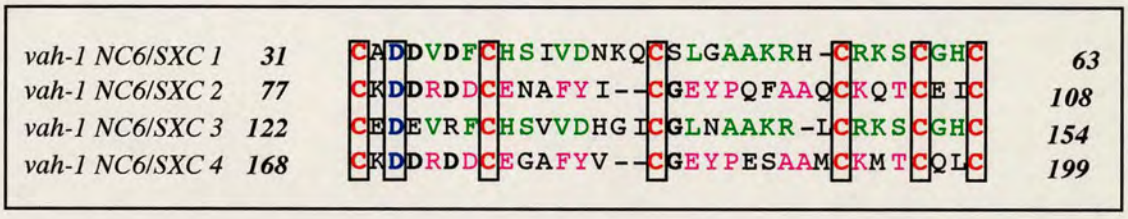



**Figure 3.3 (c).** Amino acid pileup of the four NC6/SXC domains from vah-1. Conserved cysteines are shown in red. The conserved aspartic acid is shown in blue. Residues common to both NC6/SXC 1 and 3 only are shown in green. Whereas those residues common to both NC6/SXC 2 and 4 only are shown in pink.

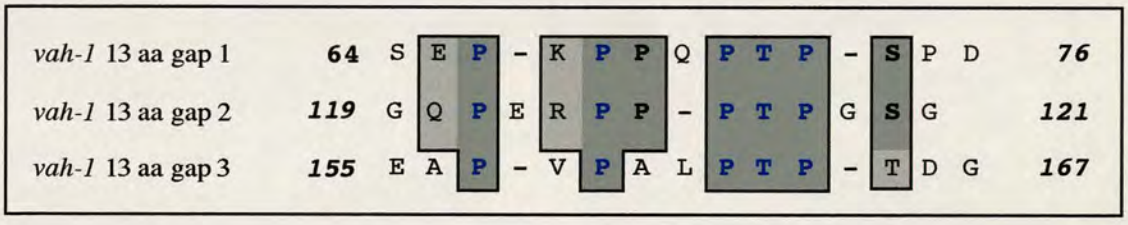



**Figure 3.3 (d).** Pileup of the three 13 amino acid stretch that separates the four NC6/SXC domains of *vah-1*. Residues common to all three domains are highlighted in blue.

### 3.3.7 *Other Novel Genes*

A total of 44 additional transcripts were found to have no significant similarities to any other sequences deposited in the GenBank nonredundant (nr) protein and dbEST databases. These were present at between 1-4 copies in the 261 member data set from *T. canis*, and were designated *not* genes (*no*vel *t*ranscripts) (Figure 3.4). Further studies are ongoing on selected clones, such as *not-018*, for which antibodies to the protein product strongly recognise TES antigens (Chapter 5).

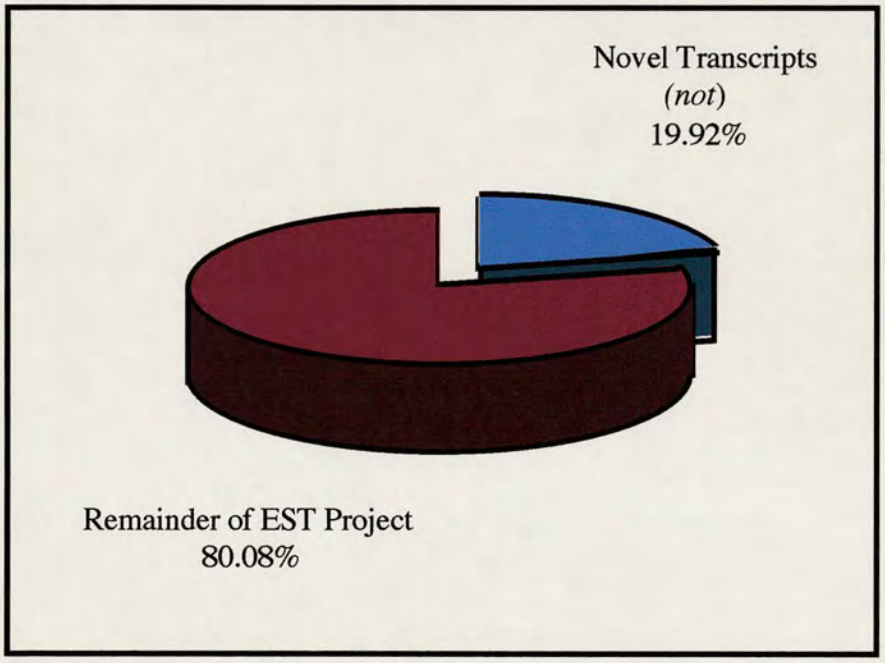



**Figure 3.4.** Pie chart showing the proportion of the EST project devoted to the Novel Transcripts (*nots)* versus the remainder of the project.

**able 3.6** Table of the *no*vel *t*ranscripts (*nots*) identified as part of the *T. canis* EST project

| | Clone number | Gene Name | Identification | Closest Species | Sibling overlaps | Related | Insert size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank | Clone Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Tc-EST-009 | *Tc-not-009* | Novel Transcript 009 | None | None | No | 480 | 554 | 522 | AA563571 | AA683462 | Tc-EST-009 |
| 2 | Tc-EST-018 | *Tc-not-018* | Novel Transcript 018 | None | 018,031,136 | Yes | >1270 | 600 | 671 | AA683463 | AA683464 | Tc-EST-018 |
| 3 | Tc-EST-031 | *Tc-not-031* | Novel Transcript 031 | None | 018,031,136 | Yes | 1000 | 382 | 555 | AA568095 | AA568096 | Tc-EST-031 |
| 4 | Tc-EST-136 | *Tc-not-136* | Novel Transcript 136 | None | 018,031,136 | Yes | 1000 | 620 | 565 | AA835621 | AA835622 | Tc-EST-136 |
| 5 | Tc-EST-028 | *Tc-not-028* | Novel Transcript 028 | None | None | No | 607 | 588 | 600 | AA568091 | AA568092 | Tc-EST-028 |
| 6 | Tc-EST-039 | *Tc-not-039* | Novel Transcript 039 | None | None | No | 450 | 360 | 360 | AA583103 | AA583104 | Tc-EST-039 |
| 7 | Tc-EST-042 | *Tc-not-042* | Novel Transcript 042 | None | None | No | 377 | 377 | | AA603959 | | Tc-EST-042 |
| 8 | Tc-EST-052 | *Tc-not-052* | Novel Transcript 052 | None | None | No | 420 | 395 | 381 | AA610968 | AA610969 | Tc-EST-052 |
| 9 | Tc-EST-056 | *Tc-not-056* | Novel Transcript 056 | None | None | No | 750 | 630 | 400 | AA610962 | AA907597 | Tc-EST-056 |
| 10 | Tc-EST-068 | *Tc-not-068* | Novel Transcript 068 | None | None | No | 370 | 370 | 370 | AA728630 | AA728631 | Tc-EST-068 |
| 11 | Tc-EST-069 | *Tc-not-069* | Novel Transcript 069 | None | 069,071 | Yes | 295 | 294 | 283 | AA888748 | AA888749 | Tc-EST-069 |
| 12 | Tc-EST-071 | *Tc-not-071* | Novel Transcript 071 | None | 069,071 | Yes | 332 | 332 | 332 | AA728647 | AA728648 | Tc-EST-071 |
| 13 | Tc-EST-073 | *Tc-not-073* | Novel Transcript 073 | None | None | No | 400 | 396 | 393 | AA681626 | AA728638 | Tc-EST-073 |
| 14 | Tc-EST-095 | *Tc-not-095* | Novel Transcript 095 | None | 095, 145, 293, 321 | Yes | 350 | 315 | | AA820018 | | Tc-EST-095 |
| 15 | Tc-EST-145 | *Tc-not-145* | Novel Transcript 145 | None | 095, 145, 293, 321 | Yes | 314 | 314 | | AA836694 | | Tc-EST-145 |
| 16 | Tc-EST-293 | *Tc-not-293* | Novel Transcript 293 | None | 095, 145, 293, 321 | Yes | 320 | 311 | 300 | AA875800 | AA875801 | Tc-EST-293 |
| 17 | Tc-EST-321 | *Tc-not-321* | Novel Transcript 321 | None | 095, 145, 293, 321 | Yes | 300 | 300 | 300 | AA879363 | AA883053 | Tc-EST-321 |
| 18 | Tc-EST-098 | *Tc-not-098* | Novel Transcript 098 | None | None | No | 800 | 528 | 501 | AA820022 | AA820023 | Tc-EST-098 |

**Table 3.6 continued**

| | Clone Name | Gene Name | Identification | Closest Species | Sibling overlaps | Related | Insert size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank | Clone Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 19 | Tc-EST-099 | *Tc-not-099* | Novel Transcript 099 | None | None | No | 310 | 250 | 258 | AA820024 | AA820025 | Tc-EST-099 |
| 20 | Tc-EST-103 | *Tc-not-103* | Novel Transcript 103 | None | None | No | 640 | 432 | 538 | AA824218 | AA824219 | Tc-EST-103 |
| 21 | Tc-EST-107 | *Tc-not-107* | Novel Transcript 107 | None | None | No | 1200 | 528 | 516 | AA824226 | AA824227 | Tc-EST-107 |
| 22 | Tc-EST-120 | *Tc-not-120* | Novel Transcript120 | None | 120, 236, 240 | Yes | 630 | 580 | | AA825139 | | Tc-EST-120 |
| 23 | Tc-EST-236 | *Tc-not-236* | Novel Transcript 236 | None | 120, 236, 240 | Yes | 500 | 480 | 463 | AA874742 | AA874743 | Tc-EST-236 |
| 24 | Tc-EST-240 | *Tc-not-240* | Novel Transcript 240 | None | 120, 236, 240 | Yes | 500 | 501 | 481 | AA874744 | AA874745 | Tc-EST-240 |
| 25 | Tc-EST-126 | *Tc-not-126* | Novel Transcript 126 | None | None | No | 570 | 552 | 568 | AA835594 | AA835595 | Tc-EST-126 |
| 26 | Tc-EST-127 | *Tc-not-127* | Novel Transcript 127 | None | None | No | 1000 | 562 | 645 | AA835596 | AA835597 | Tc-EST-127 |
| 27 | Tc-EST-132 | *Tc-not-132* | Novel Transcript 132 | None | None | No | 1430 | 493 | 653 | AA835615 | AA835616 | Tc-EST-132 |
| 28 | Tc-EST-135 | *Tc-not-135* | Novel Transcript 135 | None | None | No | 300 | 292 | 300 | AA835619 | AA835620 | Tc-EST-135 |
| 29 | Tc-EST-138 | *Tc-not-138* | Novel Transcript 138 | None | None | No | 614 | 591 | 590 | AA835625 | AA835626 | Tc-EST-138 |
| 30 | Tc-EST-139 | *Tc-not-139* | Novel Transcript 139 | None | None | No | 605 | 355 | 505 | AA835627 | AA835628 | Tc-EST-139 |
| 31 | Tc-EST-154 | *Tc-not-154* | Novel Transcript 154 | None | None | No | 1300 | 615 | 416 | AA836701 | AA836702 | Tc-EST-154 |
| 32 | Tc-EST-173 | *Tc-not-173* | Novel Transcript 173 | None | None | No | 530 | 328 | 395 | AA873913 | AA873914 | Tc-EST-173 |
| 33 | Tc-EST-195 | *Tc-not-195* | Novel Transcript 195 | None | None | No | 237 | 237 | 218 | AA873902 | AA873903 | Tc-EST-195 |
| 34 | Tc-EST-198 | *Tc-not-198* | Novel Transcript 198 | None | None | No | 498 | 495 | 498 | AA873893 | AA911816 | Tc-EST-198 |
| 35 | Tc-EST-204 | *Tc-not-204* | Novel Transcript 204 | None | None | No | 1000 | 463 | 432 | AA874709 | AI 080931 | Tc-EST-204 |

**Table 3.6 continued**

| | Clone Name | Gene Name | Identification | Closest Species | Sibling overlaps | Related | Insert size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank | Clone Number |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 36 | Tc-EST-216 | *Tc-not-216* | Novel Transcript 216 | None | 218 | No | 314 | 314 | 296 | AA911817 | AA911818 | Tc-EST-216 |
| 37 | Tc-EST-218 | *Tc-not-218* | Novel Transcript 218 | None | 216 | No | 311 | 311 | 302 | AA911819 | AA911820 | Tc-EST-218 |
| 38 | Tc-EST-224 | *Tc-not-224* | Novel Transcript 224 | None | None | No | 300 | 278 | 284 | AA874727 | AA874728 | Tc-EST-224 |
| 39 | Tc-EST-225 | *Tc-not-225* | Novel Transcript 225 | None | None | No | 830 | 520 | 461 | AA911821 | AA911822 | Tc-EST-225 |
| 40 | Tc-EST-230 | *Tc-not-230* | Novel Transcript 230 | None | None | No | 1010 | 436 | 616 | AA874732 | AA874733 | Tc-EST-230 |
| 41 | Tc-EST-231 | *Tc-not-231* | Novel Transcript 231 | None | None | No | 240 | 240 | 240 | AA874734 | AA874735 | Tc-EST-231 |
| 42 | Tc-EST-243 | *Tc-not-243* | Novel Transcript 243 | None | None | No | 430 | 351 | 369 | AA874761 | AA874762 | Tc-EST-243 |
| 43 | Tc-EST-272 | *Tc-not-272* | Novel Transcript 272 | None | None | No | 550 | 399 | 463 | AA875779 | AA875780 | Tc-EST-272 |
| 44 | Tc-EST-276 | *Tc-not-276* | Novel Transcript 276 | None | None | No | 460 | 409 | 421 | AA87582 | AA875783 | Tc-EST-276 |
| 45 | Tc-EST-289 | *Tc-not-289* | Novel Transcript 289 | None | None | No | 700 | 462 | 620 | AA875795 | AA875796 | Tc-EST-289 |
| 46 | Tc-EST-302 | *Tc-not-302* | Novel Transcript 302 | None | None | No | 760 | 283 | 621 | AA879342 | AA879343 | Tc-EST-302 |
| 47 | Tc-EST-307 | *Tc-not-307* | Novel Transcript 307 | None | None | No | 258 | 258 | 258 | AA879344 | AA879345 | Tc-EST-307 |
| 48 | Tc-EST-310 | *Tc-not-310* | Novel Transcript 310 | None | None | No | 700 | 589 | 625 | AA879348 | AA879439 | Tc-EST-310 |
| 49 | Tc-EST-311 | *Tc-not-311* | Novel Transcript 311 | None | None | No | 394 | 394 | 368 | AA979717 | AA979718 | Tc-EST-311 |
| 50 | Tc-EST-313 | *Tc-not-313* | Novel Transcript 313 | None | None | No | 750 | 564 | 669 | AA879352 | AA879353 | Tc-EST-313 |
| 51 | Tc-EST-317 | *Tc-not-317* | Novel Transcript 317 | None | None | No | 123 | 123 | 119 | AA879356 | AA803052 | Tc-EST-317 |
| 52 | Tc-EST-327 | *Tc-not-327* | Novel Transcript 327 | None | None | No | 550 | 397 | 528 | AA879370 | AA879371 | Tc-EST-327 |
| 53 | Tc-EST-329 | *Tc-not-329* | Novel Transcript 329 | None | None | No | 250 | 237 | 249 | AA879372 | AA879373 | Tc-EST-329 |

### 3.3.8 *Homologues of Known Genes*

Table 3.7 shows a compilation of all the genes for which significant homologues of known function existed in the database. This includes 62 new genes together with 3 previously characterised sequences. Of the identified genes, there are 23 metabolic and respiratory enzymes, but remarkably few structural proteins (only actin, calponin and α-tubulin) and no products involved in DNA replication. This finding is consistent with the developmentally arrested state of the parasite. The identified genes have been categorised under distinct headings, and the findings are summarised in Figure 3.5 and Table 3.7 below. The pie-chart in Figure 3.5 shows a diagrammatic representation of the EST project and is also likewise categorised. The *ants* and *nots* have been omitted from this grouping.

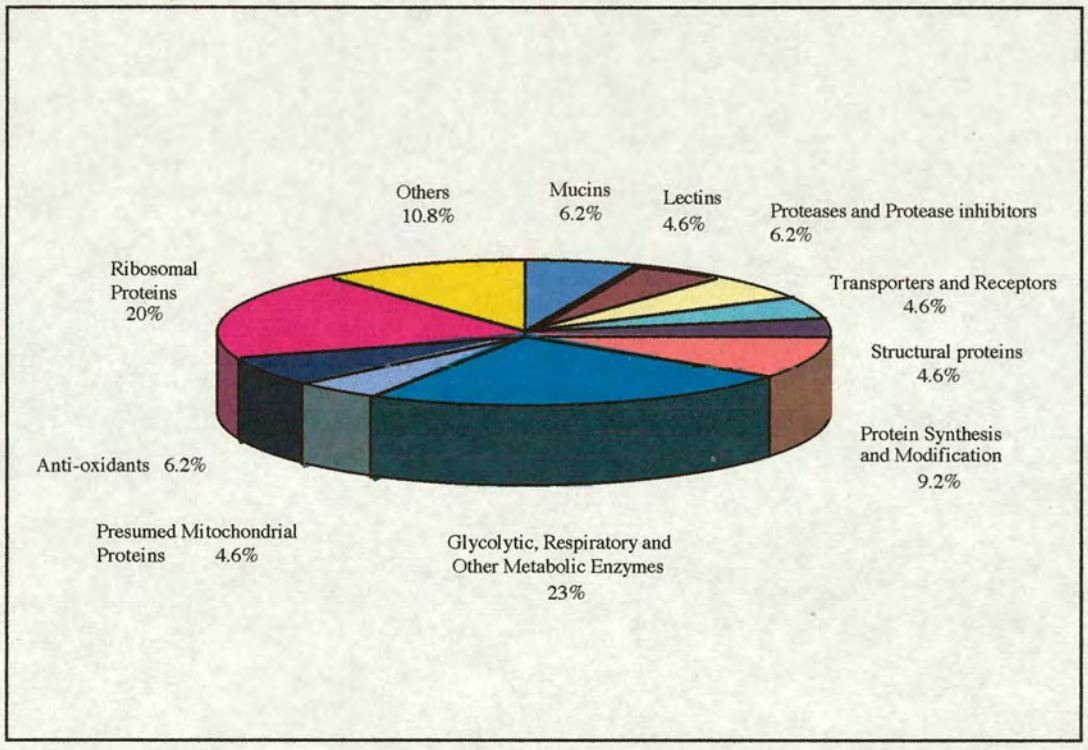



Figure 3.5. Pie-chart showing a summary of the EST project categorised by genes with homologues of known function only.

**Table 3.7.** Summarised table of the clones identified by homology searching

| | Gene Name | Identification | Closest Species | p (n) | Clone Number (• = Longest Clone) | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|---|---|
| MUCIN | | | | | | | | | | |
| 1 | *Tc-muc-1* | Mucin-1, Surface coat TES-120 | *Toxocara canis* | 6.3E-104 | • Tc-EST-087 | 810 | 557 | 539 | AA583105 | ‡ AA907963 |
| 2 | *Tc-muc-2* | Mucin-2 | *T. canis* | 3.8E-49 | • Tc-EST-130 | 990 | 612 | 694 | AA583106 | AA583107 |
| 3 | *Tc-muc-3* | Mucin-3 | *T. canis* | 2.6E-36 | • Tc-EST-162 | 1010 | 603 | 555 | AA583111 | AA583112 |
| 4 | *Tc-muc-4* | Mucin-4 | *T. canis* | 1.6E-45 | Tc-EST-186 | 560 | 553 | 586 | AA583108 | AA583109 |
| LECTINS | | | | | | | | | | |
| 5 | *Tc-ctl-1* | C-type lectin/TES-32 glycoprotein | | | •Tc-EST-209 | 830 | 615 | 393 | AA874713 | ‡ AA874714 |
| 6 | *Tc-ctl-2* | C-type lectin-2/Variant of TES-32 | | | Tc-EST-036 | 780 | 596 | 618 | AI 078880 | AI 078881 |
| 7 | *Tc-ctl-3* | C-type lectin-3/Variant of TES-32 | | | Tc-EST-155 | 610 | 590 | 737 | AI 080917 | AI 080918 |
| PROTEASES AND PROTEASE INHIBITORS | | | | | | | | | | |
| 8 | *Tc-aep-1* | Asparaginyl Endopeptidase | *Homo sapiens* | 3.5E-28 | Tc-EST-262 | 1300 | 520 | 648 | AA825116 | AA875769 |
| 9 | *Tc-api-1* | Aspartyl proteinase inhibitor | *Caenorhabditis elegans* | 1.5E-09 | Tc-EST-096 | 650 | 570 | 638 | AA820019 | AA820020 |
| 10 | *Tc-cpl-1* | Cathepsin L cysteine proteinase | *C. elegans* | 2.3E-40 | Tc-EST-013 | 1150 | 604 | 390 | AA563518 | ‡ AI 078878 |
| 11 | *Tc-cpz-1* | Cathepsin Z proteinase precursor | *Onchocerca volvulus* | 1.2E-73 | Tc-EST-205 | 1000 | 568 | 701 | AI 080921 | AI 080922 |
| TRANSPORTERS AND RECEPTORS | | | | | | | | | | |
| 12 | *Tc-aat-1* | ADP/ATP translocase | *C. elegans* | 1.9E-69 | •Tc-EST-041 | 750 | 515 | 663 | AA603955 | AA683466 |
| 13 | *Tc-acr-1* | Acetylcholine Receptor (tar-1) | *T. colubriformis* | 1.2E-35 | Tc-EST-121 | 1300 | 529 | 547 | AA835586 | AA835587 |
| 14 | *Tc-peb-1* | Phosphatidylethanolamine binding protein (TES-26) | *T.canis* | 6.1E-76 | •Tc-EST-164 | 1200 | 406 | 459 | AA864177 | ‡AA864178 |
| STRUCTURAL PROTEINS | | | | | | | | | | |
| 15 | *Tc-act-1* | Actin | *C. elegans* | 1.3E-81 | Tc-EST-097 | 1450 | 524 | | AA820021 | |
| 16 | *Tc-cap-1* | Calponin | *M.javanica* | 4.2E-21 | •Tc-EST- | 620 | 462 | 537 | AA875799 | AA907961 |
| 17 | *Tc-tua-1* | Tubulin alpha-3 chain | *C. griseus* | 1.5E-69 | •Tc-EST- | 1500 | 400 | 439 | AA979712 | AA907952 |

**Table 3.7** (continued)

| | Gene Name | Identification | Closest Species | p (n) | Clone Number (• = Longest Clone) | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|---|---|
| PROTEIN SYNTHESIS AND MODIFICATION | | | | | | | | | | |
| 18 | *Tc-efa-1* | Translation elongation factor 1a | *C. elegans* | 3.5E-84 | Tc-EST-300 | 1600 | 498 | 572 | AA879374 | AA879375 |
| 19 | *Tc-efb-1* | Elongation Factor Beta-1 | *C. elegans* | 3.1E-42 | Tc-EST-081 | 830 | 355 | 756 | AA888759 | AA888760 |
| 20 | *Tc-pam-1* | Peptidylglycine alphahydroxylating monooxygenase | *D. melanogaster* | 1.1E-07 | •Tc-EST-200 | 420 | 385 | 357 | AA874705 | AI 080786 |
| 21 | *Tc-pdi-1* | Protein disulphide isomerase | *Gallus gallus* | 5.9E-18 | Tc-EST-235 | 474 | 445 | 468 | AA874740 | AA874741 |
| 22 | *Tc-ppi-1* | Peptidyl-prolyl cis-trans isomerase | *M. musculus* | 3.2E-15 | Tc-EST-019 | 1200 | 319 | | AA563572 | |
| 23 | *Tc-ptp-1* | Protein Tyrosine Phosphatase | *M. musculus* | 5.9E-49 | •Tc-EST-118 | 1590 | 589 | 628 | AA820013 | AA820014 |
| GLYCOLYSIS, RESPIRATION AND MODIFICATION | | | | | | | | | | |
| 24 | *Tc-aca-1* | Acetyl-CoA Acetyltransferase | *C. elegans* | 9.8E-33 | Tc-EST-059 | 626 | 626 | 560 | AA610965 | AA907958 |
| 25 | *Tc-aki-1* | Adenylate kinase isoenzyme I | *Gallus gallus* | 9.2E-39 | Tc-EST-328 | 800 | 452 | 505 | AA873889 | AA873920 |
| 26 | *Tc-ald-1* | Aldehyde dehydrogenase | *Bos bovis* | 1.3E-45 | Tc-EST-299 | 1700 | 559 | 360 | AA979715 | AA979716 |
| 27 | *Tc-ata-1* | Alanine aminotransferase | *C.elegans* | 4.0E-42 | Tc-EST-294 | 1440 | 472 | 357 | AA875802 | AA875803 |
| 28 | *Tc-cgl-1* | Cystathionine gamma lyase | *C.elegans* | 6.4E-54 | Tc-EST-090 | 750 | 555 | 685 | AA583117 | AA583118 |
| 29 | *Tc-cyb-5* | Cytochrome b5 | *H. sapiens* | 4.5E-10 | Tc-EST-032 | 406 | 406 | | AA569419 | |
| 30 | *Tc-fba-1* | Fructose biphosphate aldolase | *O. volvulus* | 2.1E-98 | Tc-EST-183 | 1010 | 480 | 533 | AA915868 | AA915869 |
| 31 | *Tc-fdh-1* | Fumarate dehydrogenase precurs. | *H. sapiens* | 1.7E-26 | Tc-EST-237 | 317 | 317 | 314 | AA285389 | AA874746 |
| 32 | *Tc-frb-1* | Fumarate reductase, cytochrome b small subunit. | *Ascaris suum* | 2.9E-32 | Tc-EST-083 | 1560 | 452 | 480 | AA888752 | AA888753 |
| 33 | *Tc-gdh-1* | Glutamate dehydrogenase | *H. contortus* | 4.7E-74 | Tc-EST-026 | 1740 | 584 | 331 | AA568090 | AI 078879 |
| 34 | *Tc-pcc-1* | Proprionyl CoA carboxylase | *C.elegans* | 8.6E-64 | Tc-EST-074 | 1560 | 320 | 399 | AA888750 | AA888751 |
| 35 | *Tc-pck-1* | Phosphoenolpyruvate carboxykinase | *Ascaris suum* | 4.4E-69 | Tc-EST-004 | 2740 | 599 | | AA470327 | |
| 36 | *Tc-pgk-1* | Phospho-glycerate kinase | *S. mansoni* | 1.8E-41 | Tc-EST-109 | 1400 | 576 | 805 | AA825117 | AA825118 |
| 37 | *Tc-sdi-1* | Succinate dehydrogenase iron-sulphur protein | *Ascaris suum* | 3.0E-73 | Tc-EST-166 | 630 | 518 | | AA864181 | |
| 38 | *Tc-ubo-1* | NADH ubiquinone oxidoreductase 18 kDa subunit | *Bos taurus* | 1.9E-39 | Tc-EST-201 | 521 | 501 | 480 | AA874708 | AI 080786 |

**Table 3.7** (continued)

| | Gene Name | Identification | Closest Species | p (n) | Clone Number (• = Longest Clone) | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|---|---|
| PRESUMED MITOCHONDRIAL PROTEINS | | | | | | | | | | |
| 39 | *Tc-atp-6* | ATP synthase A chain (Protein 6) | *Ascaris suum* | 1.9E-68 | Tc-EST-064 | 610 | 576 | 557 | AA612565 | AA979708 |
| 40 | *Tc-cox-2* | Cytochrome c oxidase II | *Ascaris suum* | 7.1E-75 | •Tc-EST-038 | 850 | 496 | 669 | AA569422 | AA569423 |
| 41 | *Tc-ubo-4* | NADH-Ubiquinone oxidoreductase chain 4L | *Ascaris suum* | 2.6E-29 | Tc-EST-163 | 220 | 213 | 216 | AA864175 | AA864176 |
| ANTI-OXIDANTS | | | | | | | | | | |
| 42 | *Tc-sod-2* | Superoxide dismutase 2 | *C.albicans* | 1.5E-23 | Tc-EST-084 | 850 | 401 | 340 | AI 078885 | AI 078886 |
| 43 | *Tc-sod-3* | Superoxide dismutase 3 | *Zea mays* | 2.1E-18 | •Tc-EST-100 | 900 | 449 | 608 | AI 080774 | AI 080775 |
| 44 | *Tc-sod-4* | Superoxide dismutase 4 | *C.elegans* | 3.6E-40 | •Tc-EST-143 | 804 | 630 | 520 | A1 080778 | AI 080779 |
| 45 | *Tc-sod-5* | Superoxide dismutase 5 | *C.elegans* | 2.4E-39 | Tc-EST-266 | 900 | 731 | 730 | AI 080782 | AI 080783 |
| RIBOSOMAL PROTEINS | | | | | | | | | | |
| 46 | *Tc-rpl-10* | 60S Ribosomal Protein L10 | *C. elegans* | 8.6E-44 | Tc-EST-271 | 800 | 321 | 590 | AA875777 | AA875778 |
| 47 | *Tc-rpl-19* | 60S Ribosomal Protein L19 | *M.musculus* | 2.1E-60 | •Tc-EST-244 | 900 | 439 | 555 | AA874749 | AA874750 |
| 48 | *Tc-rpl-3* | 60S Ribosomal Protein L3 | *T. canis* | 5.5E-60 | Tc-EST-257 | 500 | 498 | 457 | AA875762 | ‡AI 080789 |
| 49 | *Tc-rpl-31* | 60S Ribosomal Protein L31 | *H. sapiens* | 4.8E-49 | Tc-EST-072 | 580 | 440 | 491 | AA728649 | AA728650 |
| 50 | *Tc-rpl-37a* | 60S Ribosomal Protein L37 | *S.mansoni* | 3.6E-34 | Tc-EST-267 | 400 | 372 | 381 | AA875772 | AA875773 |
| 51 | *Tc-rpl-7* | 60S Ribosomal Protein L7A | *H. sapiens* | 7.2E-29 | •Tc-EST-054 | 750 | 513 | 560 | AA610961 | AA907956 |
| 52 | *Tc-rpl-9* | 60S Ribosomal Protein L9 | *H.contortus* | 1.1E-54 | •Tc-EST-182 | 630 | 396 | 480 | AA873890 | AI 080784 |
| 53 | *Tc-rpp-0* | 60S Ribosomal Protein p0 | *C. elegans* | 4.4E-35 | •Tc-EST-122 | 1100 | 476 | 552 | AA835588 | AA835589 |
| 54 | *Tc-rpp-2* | 60S Ribosomal Protein P2 | *B. malayi* | 3.6E-08 | Tc-EST-088 | 288 | 288 | 275 | AA907964 | AA907965 |
| 55 | *Tc-rps-4* | 40S Ribosomal Protein S4 | *C.griseus* | 4.1E-61 | Tc-EST-006 | 750 | 582 | | AA563574 | |
| 56 | *Tc-rps-5* | 40S Ribosomal Protein S5 | *C. elegans* | 1.4E-70 | Tc-EST-215 | 610 | 571 | 675 | AA874719 | AA874720 |
| 57 | *Tc-rps-8* | 40S Ribosomal Protein S8 | *H. sapiens* | 7.4E-66 | Tc-EST-277 | 640 | 481 | 600 | AI 080929 | AI 080930 |
| 58 | *Tc-rps-9* | 40S Ribosomal Protein S9 | *R. norvegicus* | 8.6E-38 | Tc-EST-270 | 410 | 38 | | AA875776 | |

**Table 3.7** (continued)

| | Gene Name | Identification | Closest Species | p (n) | Clone Number (• = Longest Clone) | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|---|---|
| OTHERS | | | | | | | | | | |
| 59 | *Tc-gep-1* | Granulin/epithelin precursor | *M. musculus* | 8.4E-31 | Tc-EST-141 | 1000 | 654 | 625 | AA836692 | AA836693 |
| 60 | *Tc-hih-4* | Histone H4 | *C. elegans* | 3.5E-60 | Tc-EST-174 | 250 | 444 | 459 | AA557121 | AA873915 |
| 61 | *Tc-lah-1* | Lupus Autoantigen Homologue | *C. elegans* | 1.5E-22 | Tc-EST-177 | 450 | 607 | 520 | AI 083048 | AI 083049 |
| 62 | *Tc-mps-1* | Metallopanstimulin (=rps27) | *S. ratti* | 8.3E-41 | Tc-EST-178 | 1430 | 367 | 400 | AA873918 | AA979709 |
| 63 | *Tc-ofm-1* | Olfactomedin (F11C3.2) | *C.elegans* | 1.6E-54 | •Tc-EST-079 | 1870 | 479 | 628 | AA728643 | AA728644 |
| 64 | *Tc-pth-1* | PC4/TIS7/"Interferon-related" | *M. musculus* | 4.1E-20 | •Tc-EST-075 | 1440 | 616 | 299 | AA618627 | AI 078882 |
| 65 | *Tc-tlp-1* | Tubby-like protein | *H.sapiens* | 1.6E-42 | Tc-EST-242 | 1530 | 506 | 585 | AA874759 | AA874760 |

Notes : Numerous sequences for which "closest species" listed is not a nematode show a higher BLASTX score with a *C.elegans* sequence for which there is either no assignation, or which i described as similar to the non-nematode sequence given here. Full species names are as follows *: B.malayi = Brugia malayi; C.albicans = Candida albicans; C.griseus= Cricetulus griseus H.contortus = Haemonchus contortus; M.javanica= Meloidogyne javanica; O.volvulus= Onchocerca volvulus; S.mansoni= Schistosoma mansoni; T. colubriformis = Trichostrongylus colubriformis S. ratti* = Strongyloides ratti.

‡ Denotes that a full-length, verified sequence has been deposited in GenBank. Accession numbers : *Tc-cpl-1*, U53172; *Tc-ctl-1*,AF041023; *Tc-muc-1*, U39815; *Tc-peb-1*, U29761; *Tc-rpl-*: P49149.

Figure 3.5 and Table 3.7 show a breakdown of the EST project into categories and as such manage to highlight the relative absence of developmental and reproductive genes. Structural proteins make up 6.3% of the isolated clones and only 4.6% of the identified genes from the project. In stark contrast to this is the percentage of the library made up of enzymes classified as glycolytic, respiratory and other metabolic enzymes, which account for 23%, representing the largest segment of identified genes from the library.

### 3.3.9 *Mucins*

Three new genes were discovered which bore similarity to a known *T. canis* mucin, *Tc-muc-1*. This gene was reported to be abundant both in mRNA prepared by amplification with SL and oligo-dT primers (Gems and Maizels, 1996), and in the conventional cDNA library described here. Consistent with this, 7 clones of *Tc-muc-1* were recorded. Three new genes with similarity to *Tc-muc-1* were found, each of which contain similar mucin domains and flanking NC6/SXC domains (Blaxter, 1998; Gems et al., 1995; Gems and Maizels, 1996). The mucins differ in the composition of repeats within the mucin domains (serine/threonine content), and in the number of NC6/SXC domains. A separate analysis of these mucins with respect to the family of TES-120 glycoproteins associated with the parasite surface coat is in progress (Loukas *et al* ms in preparation). Chapter 5 deals with those clones isolated from the EST project that possessed NC6 domains.

### 3.3.10 *C-type Lectins*

One of the most abundant transcripts (16/263) was found to correspond to the peptide sequence obtained from TES-32, a prominent secreted glycoprotein, and the full-length sequence showed homology to C-type lectins. The clone designated ctl-1 represented 6.1% of the clones isolated as part of the EST project. The similarity exhibited by the clone through database searching to asialoglycoprotein, a mammalian C-type lectin, was only weakly apparent when initial EST sequences were analysed, and the significance threshold of $10^{-6}$ only achieved once full-length sequence was available. A detailed analysis of the functional lectin properties of TES-32/*Tc-ctl-1* is shown in (Loukas et al.,

1998). Two variants of this sequence have since been identified and assigned *Tc-ctl-2* and *Tc-ctl-3*.

Confirmation of the sequences' relation to TES 32 came from peptides obtained from tryptic digestion of the TES 32 native protein, which were found in the sequence. Due to the modest length of the clone, full-length sequence was obtained from the EST project using the universal primers M13 R/F and T3/T7. Sequence data was later confirmed by a postdoctoral investigator, Alex Loukas, who carried-out some elegant experiments to define this molecule (Loukas et al submitted). Essentially, the recombinant protein, derived from the pET 29-T vector cloned *Tc-ctl-1* was used to immunise Balb/c mice. The resultant antibody was then used to probe Western blots to confirm the presence of TES 32 in TES (Loukas et al., 1998). Later, experiments were carried out in which different monosaccharides were bound to beaded agarose in order to determine the ligand(s) concerned. Although the physiological ligand of TES 32 is yet to be determined, modelling predictions with its closest homologue mannose binding protein (MBP) have suggested that Tc-ctl-1 would be capable of ligating Gal-type of man-type sugars. Further work, in particular site directed mutagenesis, would be required to confirm this hypothesis (Loukas et al., 1998).

As can be seen from Figure 3.6, the three lectin clones, *Tc-ctl-1, -2,* and *-3,* share a high degree of similarity but with significant levels of variation. From the Figure it can be seen that *ctl-2* appears to be the most divergent of the two lectins at the aa level. Although the carbohydrate recognition domain, QPD, is identical in all three lectins, so too is the $Ca^{2+}$ ligation site. It is not yet known whether these differences are due to allelic polymorphisms within the parasite population, or represent distinct gene products.

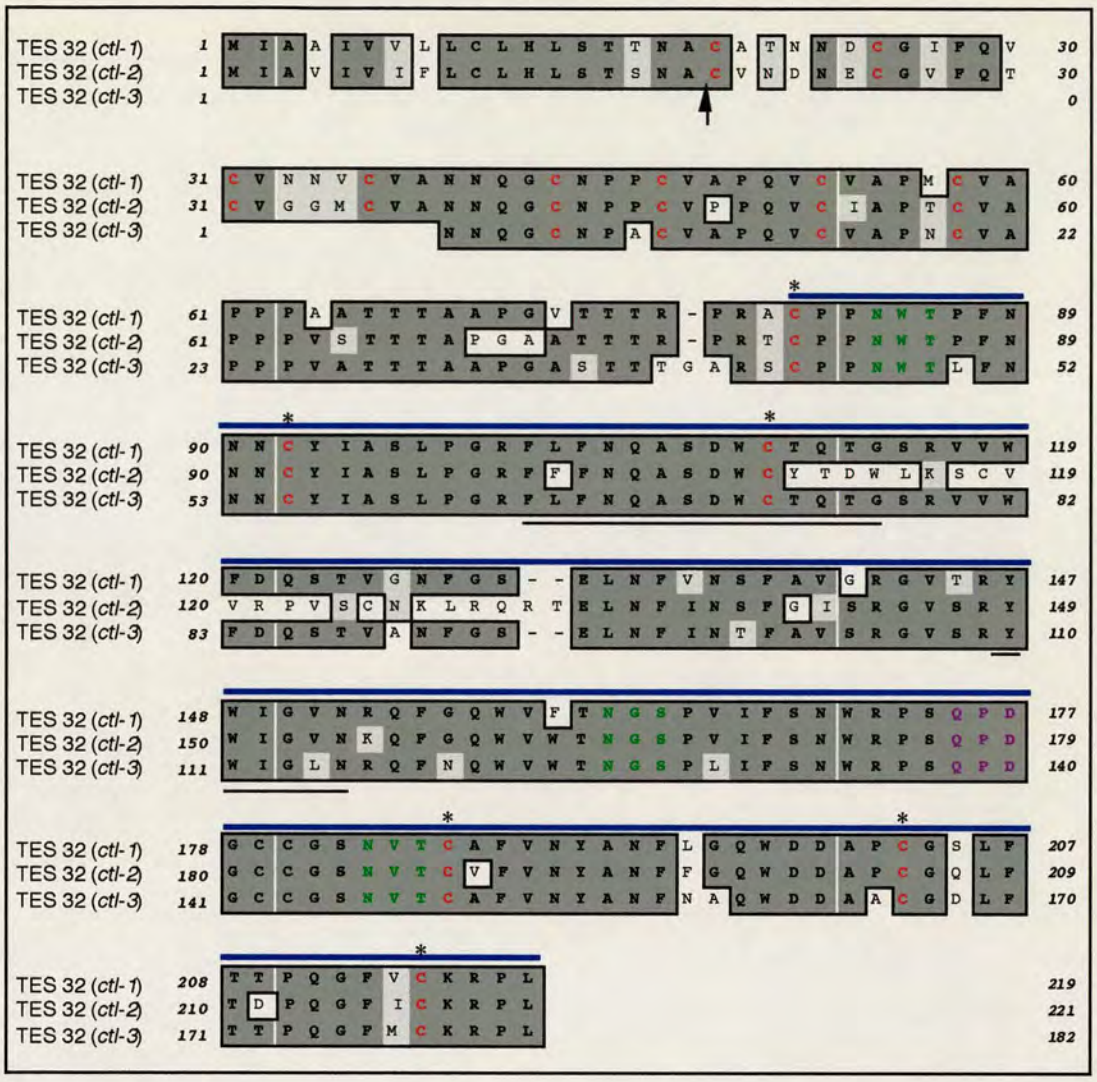



**Figure 3.6.** Amino acid pileup of the 3 C-type lectins isolated from the EST project. The arrow identifies the start methionine. N-linked glycosylation sites are identified in green, and conserved cysteine are in red; those cysteines bearing an asterix are those involved in forming the conserved $Ca^{2+}$ ligation site in mannose binding protein. The blue lines identify the C-type lectin domain, whereas the black lines show the tryptic peptides obtained via aa sequencing of *ctl-1* (Loukas *et al* ms. submitted). The carbohydrate binding motif, QPD is highlighted in purple.

### 3.3.11 *Proteases*

Proteolytic enzymes have been prominent in most studies of parasitic helminths at the biochemical (Healer et al., 1991) and molecular (Lustigman et al., 1996) levels. Three transcripts, with strong similarities to cathepsin L (Loukas et al., 1998) and cathepsin Z (Lustigman et al., 1996) and asparaginyl endopeptidase (Chen et al., 1997), were each present as single clones. Full-length sequences of the cathepsins L (Loukas et al., 1998) and Z (Falcone *et al* ms.in preparation) have been determined. A protease inhibitor, similar to the aspartyl protease inhibitor of *Ascaris* and *Brugia* (Bm33) (Dissanayake et al., 1993) was isolated as a single clone.

### 3.3.12 *Asparaginyl endopeptidase (Legumain)*

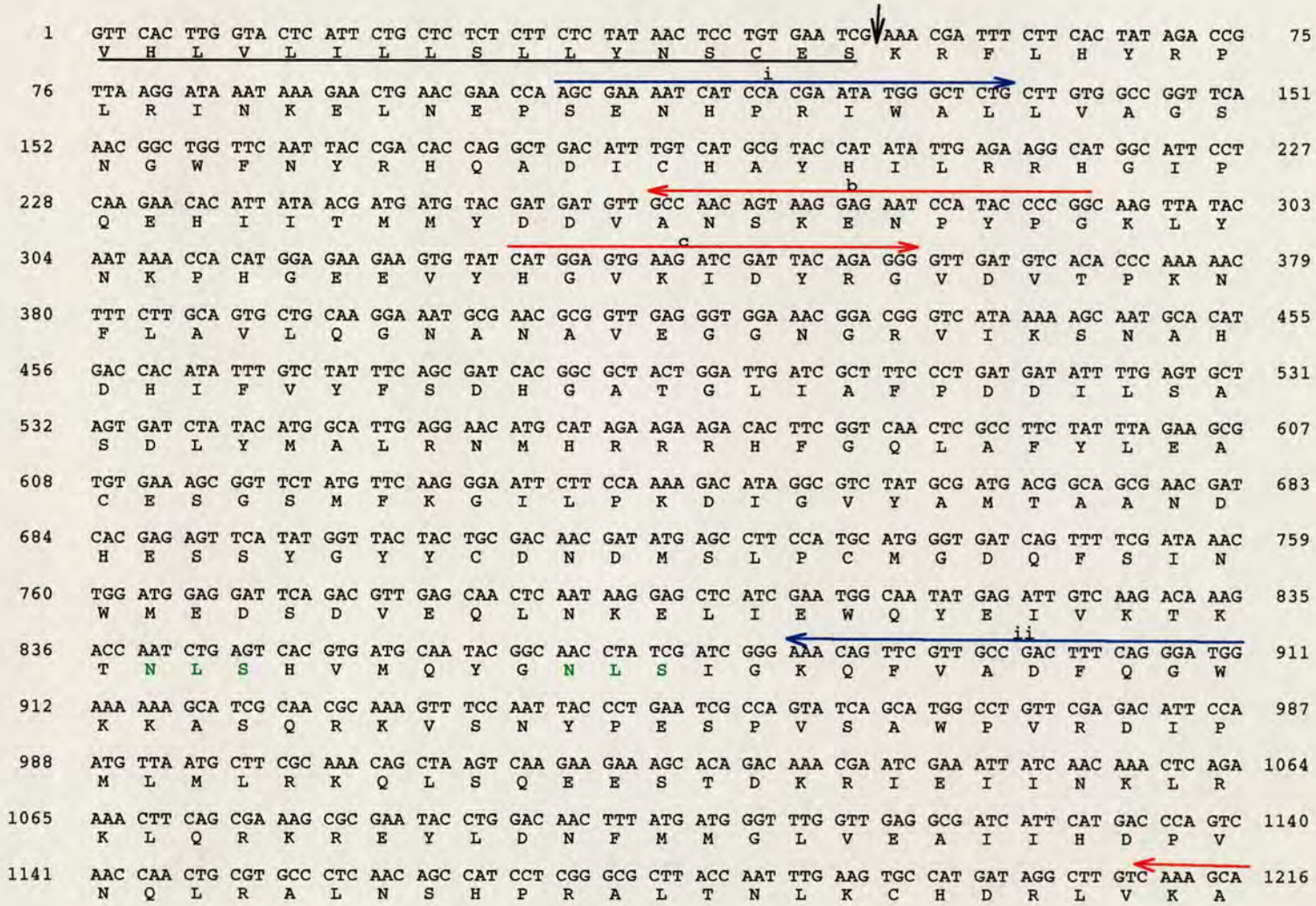
The identification of clone Tc-EST-262 from the *Toxocara* EST Project - designated *Tc-aep*-1 (GenBank accession nos : 5' EST : AA825116 & 3' EST : AA875769) - represents the first such peptidase isolated from a parasitic nematode and only the second identified in parasitic helminths to date; sequence homologous to asparaginyl endopeptidase has also been identified as part of the *C. elegans* genome project. An EST from the Filarial Genome Project (FGP) representing partial sequence with homology to AEP has also been identified, and was partially sequenced as part of the analysis of this clone (gift from Giuliano and Blaxter).

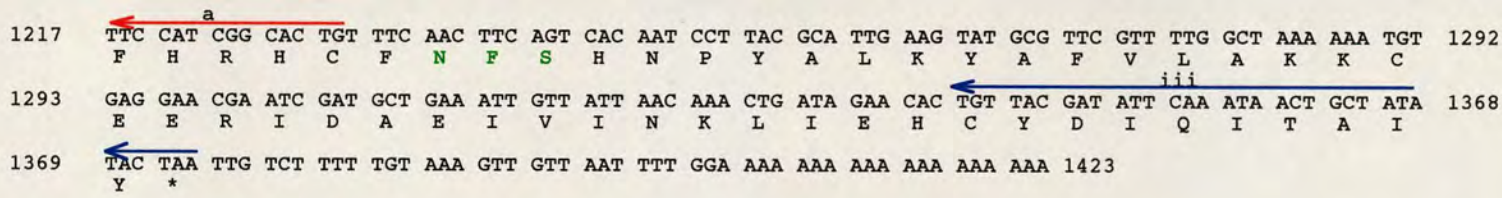
Asparaginyl endopeptidase, or legumain (EC 3.4.22.34), is a member of a novel family of cysteine peptidases first isolated from leguminous plants. Abe (1993) purified an asparaginyl endopeptidase from *Canavalia ensiformis* (Jack bean), and Takeda *et al* (1994) isolated and analysed cDNA encoding the legumain precursor. Legumain is classified as belonging to the C13 family of peptidases, a characteristic of which is the strict specificity that they show for the hydrolysis of aparaginyl bonds (Barrett and Rawlings, 1996); Abe et al (1993) found exceptions to this specificity exist when the asparagine is located at the $NH_2$-terminus, the second position from the $NH_2$-terminus or where it is glycosylated. Members of this group differ from papain in their ineffective inhibition by cystatin and E64 (Kembhavi et al., 1991) and also by their rapid inactivation by iodoacetamide in preference to iodoacetate (Csoma and Polgar, 1984; Kembhavi et al., 1991).

The best characterised of the animal asparaginyl endopeptidases is that derived from *Schistosoma mansoni*, Sm 32. Sm32 has had something of a labyrinthine route to discovery. Evidence for the presence of a proteolytic enzyme with a pH optimum of 3.9 in extracts of *S. mansoni* had previously been shown (Dalton et al., 1995a). Again in 1979 an important advance in the characterisation of this elusive peptidase came from Dresden and Deelder (1979). A purified form of the cysteine peptidase was isolated by Lindquist *et al* (1986). Later clones were obtained, through immunological screening using human and mouse infection sera, of two antigens, termed Sm31/32 (Klinkert et al., 1989; Klinkert et al., 1987). Davis et al first cloned the gene in 1987 (Davis et al., 1987) but it was not until 1994, when Takeda *et al* (1994) isolated their jack bean legumain and carried out database searches that they found their sequence to be homologous to the *S. mansoni* hemoglobinase.

The enzyme is thought to be located in the lysozomes of the animal and has been implicated in the digestion of host haemoglobin (Brindley et al., 1997; Dalton et al., 1995b; Davis et al., 1987; Wasilewski et al., 1996). Sm 32 was originally identified as a minor component from adult parasites capable of eliciting a potent immune response, but was later found to show similarity to an asparaginyl endopeptidase from Jack bean (*Canavalia ensiformis*) (Takeda et al., 1994). An asparaginyl endopeptidase has been cloned and sequenced from humans and one has been isolated and characterised from pig kidney (Chen et al., 1997); more recently an asparaginyl endopeptidase has been cloned and sequenced from mouse and the recombinant protein expressed and characterised (Chen et al., 1998).

**Figure 3.7** Nucleotide and deduced amino acid sequence of the *Toxocara canis* legumain (*Tc-aep-1*)






| | | |
|---|---|---|
| A | Tc-EST-262 R | 5′ CAG TGC CGA TGG AAT GCT TTG AC 3′ |
| B | Tc-EST-262 R | 5′ CCG GGG TAT GGA TTC TCC TTA CTG TTG GC 3′ |
| C | Tc-EST-262 F | 5′ CAT GGA GTG AAG ATC GAT TAC AGA GGG 3′ |
| i | Tc-262-pET 15b F | 5′ CGG CGC **GGA TTC** AGC GAA AAT CAT CCA CGA ATA TGG GCT CTG 3′ |
| ii | Tc-262-pET 15b R 1 | 5′ CGG CGC **GGA TCC** *TTA* CCA TCC CTG AAA GTC GGC AAC GAA CTG TTT 3′ |
| iii | Tc-262-pET 15b R 2 | 5′ CGG CGC **GGA TCC***TTA* AGT ATA TAG CAG TTA TTT GAA TAT CGT AAC A 3′ |

The translated sequence shows the primers designed to obtain the full-length clone from the cDNA library. Also indicated are the pET 15b (Novagen) primers designed for future expression of the AEP in the bacterial system. The arrow indicates the predicted cleavage sites and the underlined segment refers to the putative leader peptide. The potential *N*-glycosylation sites are shadow underlined.

Figure 3.7 shows the deduced amino acid sequence of *Tc-aep-1*. The full-length sequence as determined from the insert from the cDNA library was obtained as outlined in Chapter 2: materials and methods. Two 3' primers were designed (Tc-262 pET 15b R1 and R2) based on the cleavage patterns highlighted in (Takeda et al., 1994). The full-length cDNA sequence comprises 1423 bp (inclusive of poly A tail) with an open reading frame of 455 amino acid residues. Amino acid translation of Tc-aep-1 showed that the 5' end of the cDNA was missing albeit by one amino acid (Figure 3.8). PCR was performed using the gene specific primer Tc-EST-262 R with vector primers to amplify the remaining fragment from the cDNA library, but this proved unsuccessful.

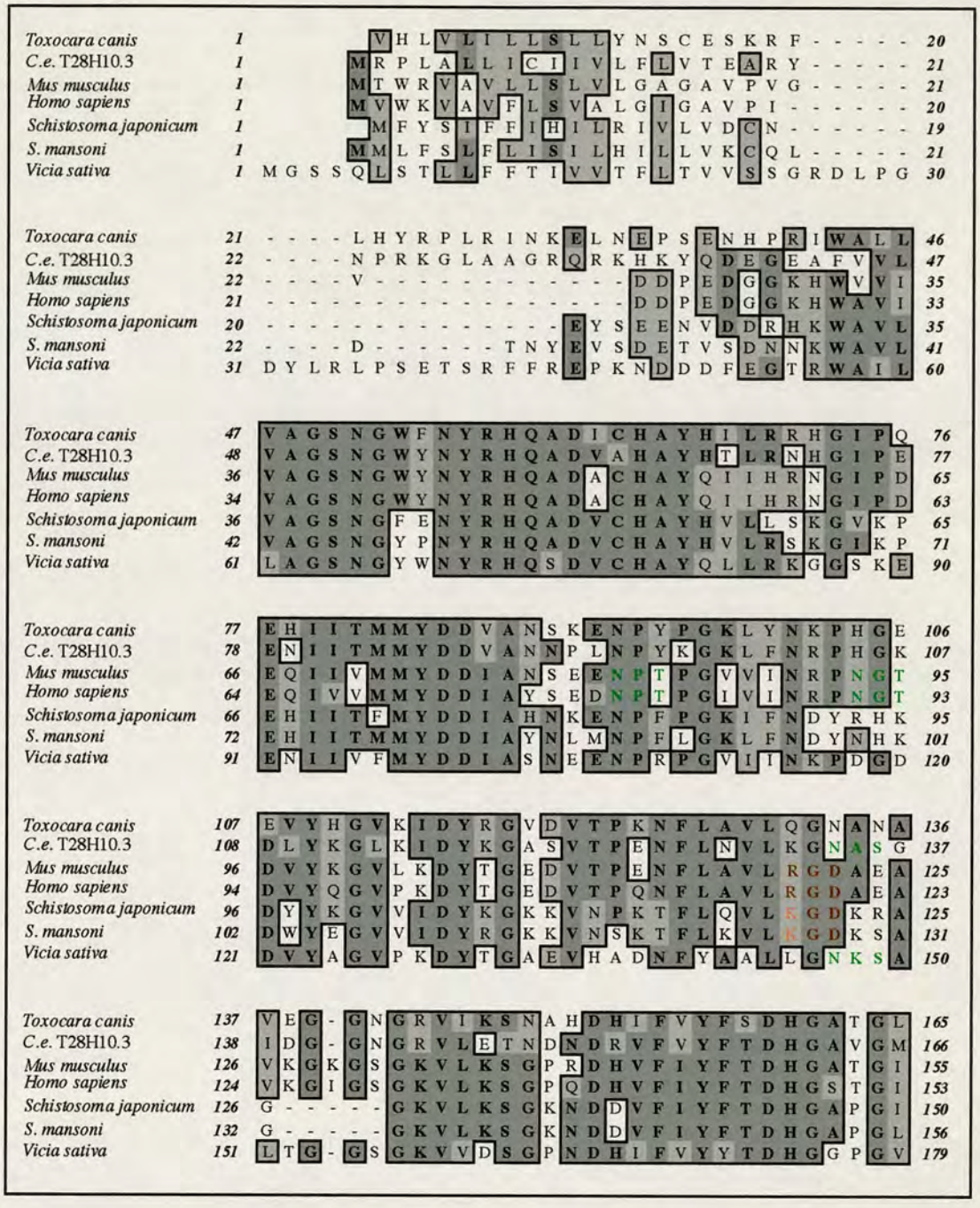
The pileup in Figure 3.8 shows Tc-AEP-1 with six of the highest ranking homologues as determined by database searching. As can be seen from this figure, the *Toxocara canis* AEP shares a high degree of homology with species as diverse as mammals and plants. At present no active site domains have been elucidated for this novel family of cysteine peptidases. Chen et al (1997; 1998) showed the presence of an RGD motif in both mouse and human AEP, a similar sequence appear in those of the flukes as a KGD motif. The amino acid sequence QGN appear in *Tc-aep-1* which may represent a divergent member of the RGD sequence motif. This motif may be involved in processing of the mature enzyme. RGD or Arg-Gly-Asp is a motif present in cell adhesive proteins and allows binding to their respective receptors. It has been shown that a mutation of this motif results in a loss of processing activity of intracellular proprotein convertase (PC1) (Lusson et al., 1997).

The clone *Tc-aep-1,* as shown in the pileup of Figure 3.8 bears strong homology to asparaginyl endopeptidases (AEP) from a number of different species, including those from *Schistosoma japonicum* and *S. mansoni.* In *schistosoma*, AEP is thought to be a factor involved in host blood digestion (Dalton et al., 1995a; Dalton et al., 1995b; Davis et al., 1987) . This clone was isolated only once out of the initial 332 clones picked, suggesting that it is not a major product of this lifecycle stage. Sequence analysis shows that significant differences exist between the species. In terms of *N*-linked glycosylation the *T. canis* clone has three potential glycosylation sites, whereas the *C. elegans*, *S. japonicum*, *S. mansoni* and *Vicia sativa* (vetch) all have two. In contrast both known mammalian AEPs have 5 *N*-linked glycosylation sites. The function of AEPs is still somewhat unknown. In flukes it is thought to function as a haemoglobinase in digesting host blood. In plants these novel enzymes are thought to play roles in the processing of

seed storage proteins (Bottari et al., 1996; Jung et al., 1998; Scott et al., 1992) and also in the processing of lectins (Min and Jones, 1994) (Sheldon et al., 1996).

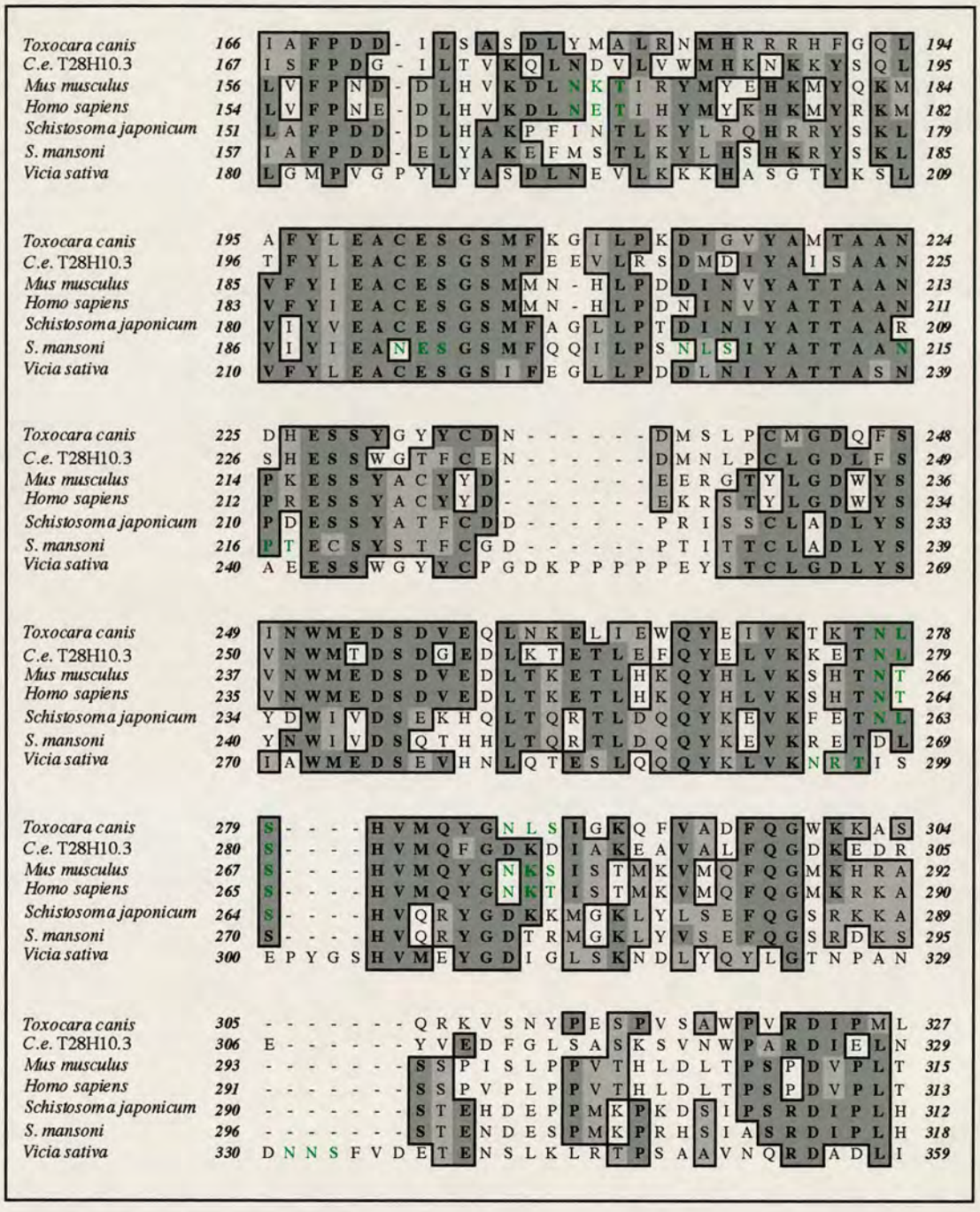
Through a collaboration with Dr. Andrew Dowd (Dublin City University) AEP activity has been confirmed in somatic extracts (TEX) of larval *T. canis*. Further work is presently being carried out to determine the presence of AEP activity within TES material.

**Figure 3.8.**



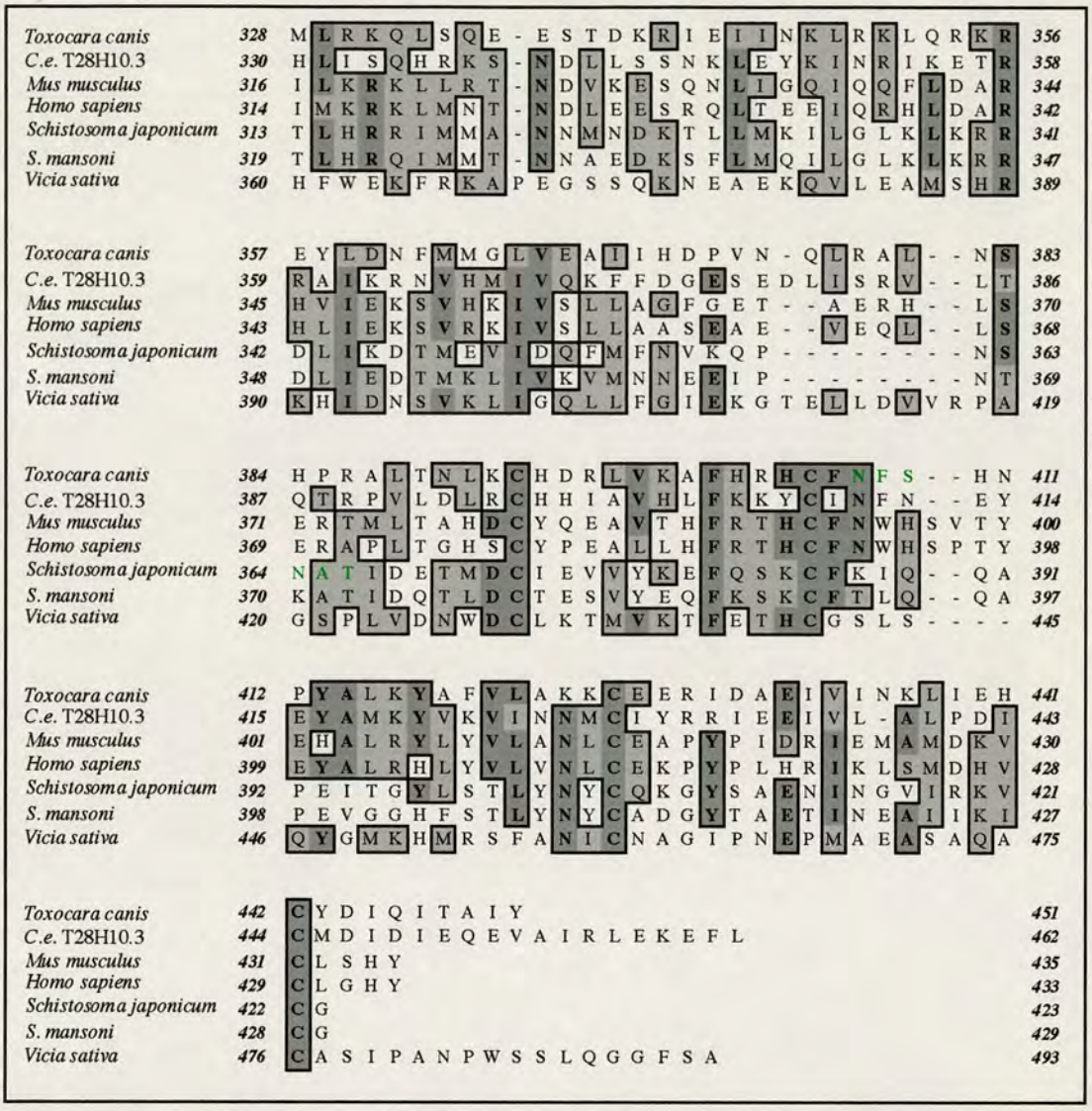

MacVector amino acid alignment of legumain sequences. Regions of identity are shown in bold capitals; regions of change are shown in capitals. The *N*-glycosylation sites are highlighted in green. The RGD sequence is indicated in brown with the conservative change in orange. The arrow indicates the putative *N*-terminal $Gly^{26}$ as hypothesised in ref (Chen et al., 1997).

**Figure 3.8 (continued)**




The GenBank accession numbers are: *Tc-aep-1* (5': AA825116 & AA875769; full-length sequence has not yet been deposited), *C. elegans* T28H10.3 (Z75551), *Mus musculus* (legumain) (AJ000990), *Homo sapiens* (Cysteine protease) (D55696)

**Figure 3.8 (continued)**




*Schistosoma japonicum*/SJ32 (Hemoglobinase) (X70967), *Schistosoma mansoni*/SM32 (Hemoglobinase) (M21308) and *Vicia sativa* (Cysteine proteinase) (Z34899)

### 3.3.13 *Transporters and Receptors*

One clone homologous to the acetylcholine receptors from other nematodes (*C. elegans*, *O. volvulus* and *Trichostrongylus colubriformis*) (Fleming et al., 1996) was isolated. A relatively frequent transcript (5/263) encodes *Tc*-PEB-*1*, a previously identified phosphatidylethanolomine-binding protein (Gems et al., 1995) which was present in parasite secretions as a 26-kDa protein (TES-26).

### 3.3.14 *Structural Proteins*

Few structural proteins were identified in the EST dataset, probably reflecting the arrested state of the parasite stage from which the cDNA library was constructed. Actin (Zeng and Donelson, 1992) and calponin (Irvine et al., 1994) from parasitic nematodes have previously been characterised. For nematode tubulins, most attention has focused on the ß-tubulins (Guenette et al., 1991), with which benzimidazole resistance is associated. *Tc-tua-1* has similarity to *C. elegans* and *S. japonicum* α-tubulins but no sequence from a parasitic nematode appears yet to have been reported.

### 3.3.15 *Protein Synthesis and Modification*_

Two genes essential for protein synthesis (elongation factors 1a and 1b) were identified, as well as peptidyl-glycine alpha-hydroxylating monooxygenase, which modifies the C-terminus of peptides. Protein disulphide isomerase (PDI) is a well-known requisite for protein folding and correct formation of disulphide bonds and a homologue from *O. vovulus* has been characterised (Wilson et al., 1994). Peptidyl-prolyl cis-trans-isomerase is similarly essential for correct protein conformation; the *Tc-ppi-1* however is not related to the multi-gene cyclophilin family of PPIs described from *C. elegans* (Page et al., 1996). Interestingly, although 4 clones were found to encode a protein tyrosine phosphatase, all correspond to the same transcript. No protein kinases were found by homology searches.

### 3.3.16 *Glycolysis, Respiration and other Metabolic Enzymes and Citric Acid Cycle_*

Some 18 distinct metabolic enzymes were found to have high levels of similarity ($p = 10^{-25}$ or higher) to *T. canis* ESTs, and to represent the major metabolic pathways of glycolysis and aerobic respiration, as well as essential processes such as amino acid synthesis and degradation. Particularly prominent among these is cytochrome c oxidase subunit II (25/263 = 9.6% of all clones). It is difficult to account for the extremely high level of this transcript, but it may be connected with the imperative for nematodes to minimise oxidative stress.

### 3.3.17 *Anti-oxidants:_Superoxide Dismutases (SODs)*

Oxidative stress is known to have highly detrimental effects on both parasitic (Ou et al., 1995) and free-living (Ishii et al., 1998) nematodes. In tissue-dwelling parasites, reactive oxygen intermediates are likely to be released by aggressive granulocytes, and expression of anti-oxidants such as glutathione peroxidase (GPX) and/or superoxide dismutase (SOD) is to be expected. In previous work, a SOD (*Tc-sod-1*) was obtained via a degenerate PCR from *T. canis* cDNA, although no GPX activity was detectable in parasite extracts (Matzilevich et al unpublished). The EST dataset contained 7 clones encoding SOD isoforms, all quite distinct from *Tc-sod-1* (approximately 66% divergence in amino acid sequence) and showing more similarity to *C. elegans* gene F55H2.1. Moreover, the 7 clones fall into 4 consensus sequences, each showing 10 - 20% divergence in nucleotide and deduced amino acid sequence, including two codon insertion/deletions (indels). Although sequence data is still incomplete for one of the clones (Tc-EST-084 = *Tc-sod-2*), this level of divergence led us to designate these 4 isoforms as separate genes, namely *Tc-sod-2*, *-3*, *-4* and *-5*. Figure 3.9 shows an amino acid pileup of the 5 SODs. From the pileup it can be seen that all the *T. canis* clones share the same catalytic and metal ion binding sites as those as those shown for a number of helminths in (Liddell and Knox, 1998). The cysteine involved in intrachain disulphide binding is also highlighted.

Both *Tc-sod-1* and *-sod-3* have 2 potential *N*-linked glycosylation sites, whereas *Tc-sod-2, -sod-4* and *-sod-5* have three such sites. The results of the BLASTX database survey are suggestive that all five SODs are probable extracellular proteins. These results also serves

to highlight the level of variation existing between each SOD. *Tc-sod-1* has 72% (121/168 amino acid) identity to the copper/zinc (Cu/Zn) extracellular SOD precursor from *Brugia pahangi*; the partial sequence for *Tc-sod-2* and *Tc-sod-5* show a 40% (54/133 amino acid) and 44% (80/182 amino acid) identity respectively, to an SOD-4 isoform from the free-living nematode *C. elegans*; *Tc-sod-3* shows a 42% (77/182 amino acid) identity to a Cu/Zn SOD from the bacterium *Candida albicans*; *Tc-sod-4* exhibits a 47% (86/182 amino acid) identity to the predicted Cu/Zn extracellular SOD from *C. elegans*. Both *Tc-sod-2* and *-sod-5* have been sequenced a sufficient number of times to significantly reduce the possibility of sequencing error. Also from the blast result the regions of homology are significantly different to suggest that although both homologous to the same protein they are in fact different genes.

Confirmation of this assignation will require full-length sequencing and hybridisation of isoform-specific probes to Southern blots of *T. canis* DNA.

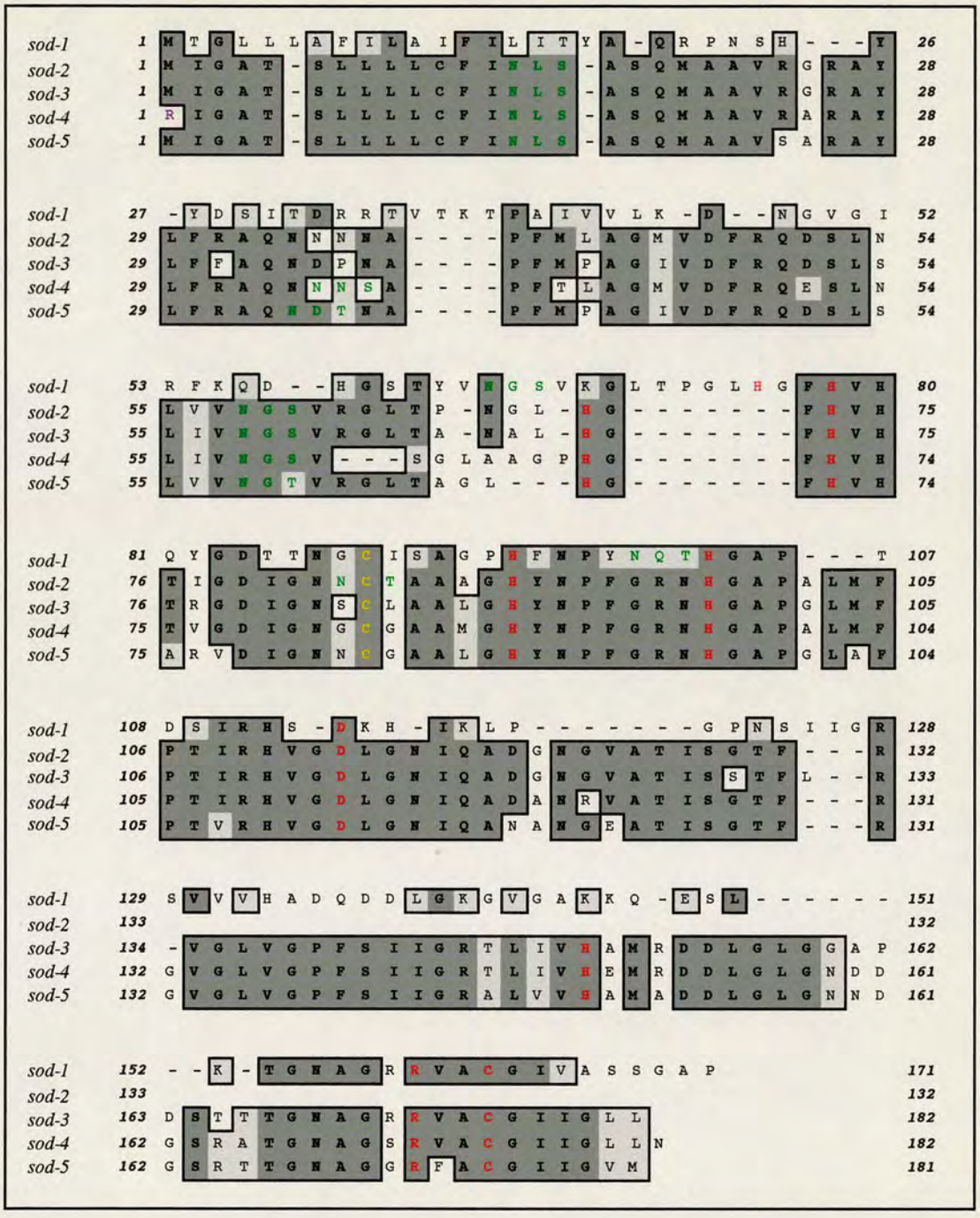



**Figure 3.9.** Amino acid sequence alignment showing the level of divergence between the 5 superoxide dismutases (SOD). The putative *N*-glycosylation sites are indicated in green. The arginine (R) highlighted in purple may represent an alternative start.

### 3.3.18 *Ribosomal Proteins*

Twenty ESTs (comprising 13 different genes; 8.4%) encoded ribosomal proteins. This is within the range found with EST projects of other nematodes; for example, 8.5% (1,339/15,811) *Brugia malayi* ESTs deposited are ribosomal (Blaxter, M. pers. comm.), as are 5.0% (11/218) of *Necator americanus* ESTs (Blaxter, M. pers. comm.).

### 3.3.19 *Other Proteins*

Some ESTs had homology to mammalian proteins for which a function in nematodes is not obvious. Five particularly interesting findings were noted:

### 3.3.20 *Granulin/epithelin precursor (Tc-gep-1)*

Granulins are synthesised as large (500 - 600 aa) precursors from which typically are derived 7 small (~ 60 aa) 12-cysteine peptides with growth-factor-like activity (Bhandari et al., 1992) (Belcourt et al., 1993) (Zhou et al., 1993). Figure 3.10(a) shows these domains outlined on the amino acid pileup based on Plowman *et al* (1992) Mammalian and fish kidney epithelial cells are all rich sources of these peptides (Belcourt et al., 1993) (Plowman et al., 1992), as are human and rodent leukocytes, suggesting that granulins may fulfil cytokine-like functions (Bateman et al., 1990). If so, the synthesis of a granulin-like homologue by *T. canis* larvae may be important in the interaction between parasite and the host immune system.

Mohammed Shoyab et al (Shoyab et al., 1990) isolated a pair of cysteine-rich growth regulators which they termed epithelins 1 and 2; epithelin 3 was later identified using the same purification protocol (Culouscou et al., 1993). These proteins were found to be both heat and acid stable (Plowman et al., 1992). The epithelins were first isolated from purified rat kidneys by their ability to inhibit the growth of a human epidermal carcinoma cell line (A431). Both epithelin 1 and 2 were structurally related but exerted antagonistic effects on each other (Shoyab et al., 1990). Epithelin 1 was shown to stimulate the proliferation of murine keratinocytes, whereas epithelin 2 inhibited the action of epithelin 1

on these cells. At around the same time Bateman et al (Bateman et al., 1990) isolated and characterised a family of novel cysteine rich peptides, of approximately 6 kDa, purified from leukocytes of the bone marrow of rats and humans. This family was composed of four peptides termed granulins or grn- A, B, C and D; partial sequence for four more granulin domains (grn-E, -F, -G and paragranulin) were also sequenced (Bateman et al., 1990). By virtue of being the most abundant, granulin A was selected for a detailed sequence analysis. It was later found that both the rat granulins and epithelins were in fact identical, a fact alluded to by Bateman et al (Bateman et al., 1990) and later by Bhandari et al (1992).

The pileup in Figure 3.10(a) shows the *T.canis* clone, designated *Tc-egh-1*, with its six closest homologues. The seven granulin domains as defined in (Plowman et al., 1992) for mouse, rat and human have been outlined.

Due to the high cysteine content (approximately 20%) of these molecules it is suggested that the secondary structure of these growth factors may be an essential component with regards to their biological function. Evidence which is borne out by the evolutionary conservation shown in their sequence structure (Bateman et al., 1990; Hrabal et al., 1996; Shoyab et al., 1990). The structure formed by this concentration of cysteines has been termed the hairpin stack fold. This is a novel protein architecture unique to the epithelin/granulins. The structure also contains a subdomain shared by small toxins, protease inhibitors and the epithelin growth factor-like protein modules (Hrabal et al., 1996).

To further characterise the epithelin/granulin family of growth molecules Culouscou et al (1993) utilised the human breast carcinoma cell line, MDA-MB-468 in order to determine whether the epithelins had specific cell surface receptors. Using $^{125}$I-epithelin in a competitive binding it was shown that the three epithelins were capable of binding the same receptor on the MDA-MB-468 carcinoma cell line. Despite the binding of all three recombinant epithelins to the same receptor this did not preclude the possibility of each naturally occurring epithelin having its own specific receptor (Culouscou et al., 1993) as has been demonstrated for insulin growth factor I and II (IGF-1 & -II) in (Casella et al., 1986). Apart from a unique subdomain shared with transforming growth factor-e (TGF-e) (Hrabal et al., 1996), the epithelin/granulins showed no strong sequence homology with

other molecules, growth factors or otherwise, in the database. Yet Shoyab *et al* (1990) showed that epithelin 1 elicited identical biological responses to epithelin growth factor (EGF), such as the stimulation of Balb/MK cells and the induction of anchorage independent growth of rat kidney cells in a dose dependent manner.

To date the epithelin/granulins have been identified from the rats, humans (Shoyab et al., 1990) (Bateman et al., 1990) (Bhandari et al., 1992) (Belcourt et al., 1993), carp (*Cyprinus carpio*) (Belcourt et al., 1993) and locust brain (PMP-D1) (Nakakura et al., 1992) and now in *Toxocara canis*. The precursor for the epithelin/granulin was isolated by Bhandari et al (Bhandari et al., 1992) and was shown to be a 593-residue prepropeptide containing seven tandem repeats of 12-cysteine epithelin/granulin domains (Bhandari et al., 1992). Figure 3.10(b) shows a schematic representation highlighting the presence of these domains in both *Toxocara canis* and *C. elegans* compared to those previously identified in rats (Plowman et al., 1992). A definitive function is still to be elucidated for these candidate growth factors. Xu et al (1998) demonstrated that the epithelin/granulin peptide was capable of stimulating growth in a R- 3T3 cell line. R- 3T3 are cells deficient in the insulin growth factor receptor, and therefore incapable of growth in serum-free media supplemented with growth factors that would ordinarily facilitate growth in 3T3 cells without the disruption (Xu et al., 1998). The results of these findings suggest that epithelin/granulins are able to bypass the insulin receptor substrate-1 (IRS-1) mitogenic pathway, rejoining at a point common to many growth factors and mitogenic stimuli (Sale et al., 1995; Schlaepfer and Hunter, 1998; Whitmarsh et al., 1995). The significance of this finding is as yet difficult to ascertain due to the complexities of the mitogenic signalling pathways. More so in light of the fact that as yet no receptor for the granulin precursor has been identified (Xu et al., 1998).

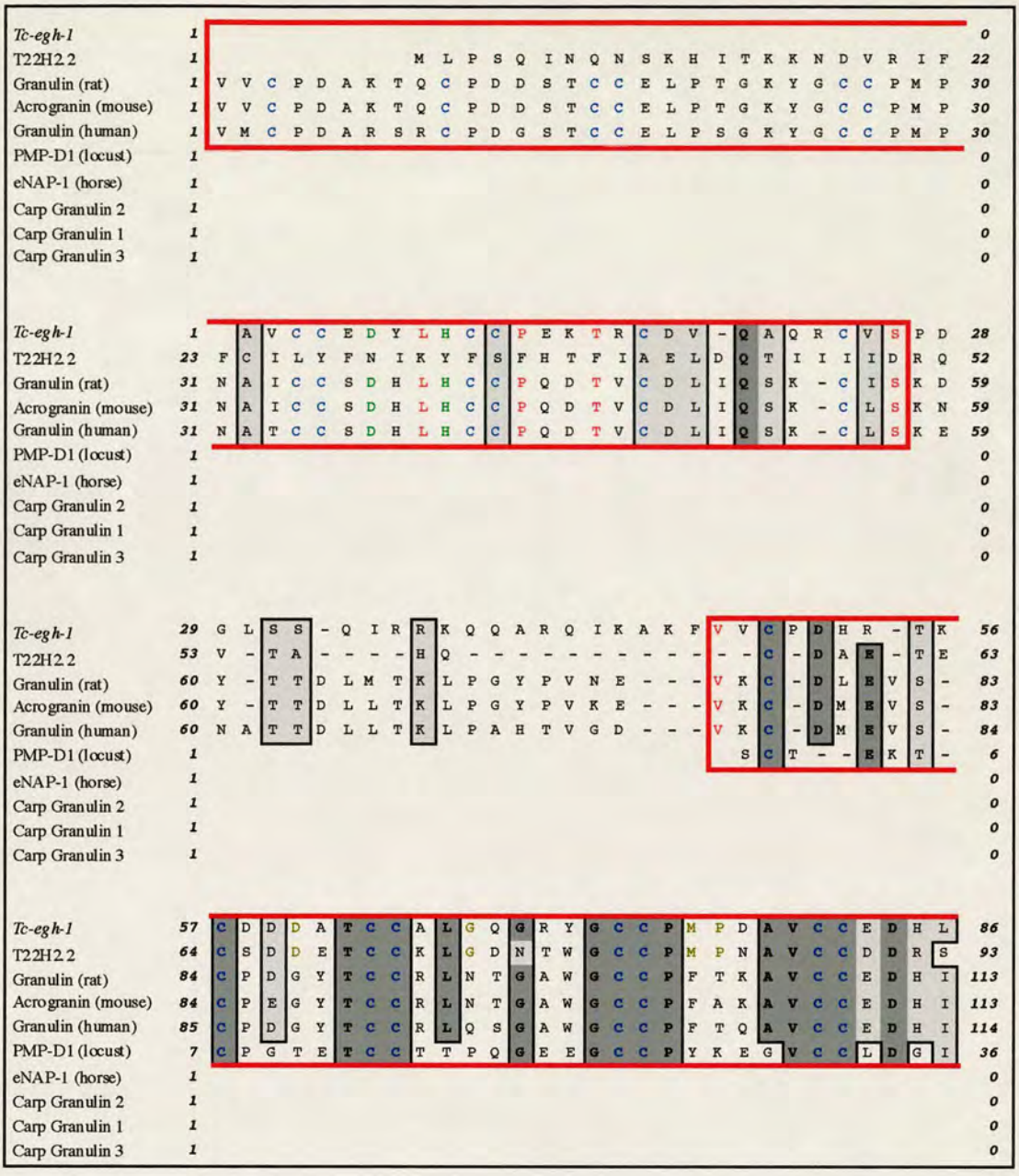



**Figure 3.10 (a).** Partial amino acid pileup of *Tc-egh-1* with 9 relevant homologues as determined by database searching. Conserved cysteine residues are highlighted in blue. The red boundaries mark the 7 epithelin domains as determined by sequence comparison from human, rat and mouse (Plowman et al., 1992). The aspartic acid and histidine residues in green are found in all epithelins/granulins. Sequences used in the pileup were obtained from GenBank and SwissProt.

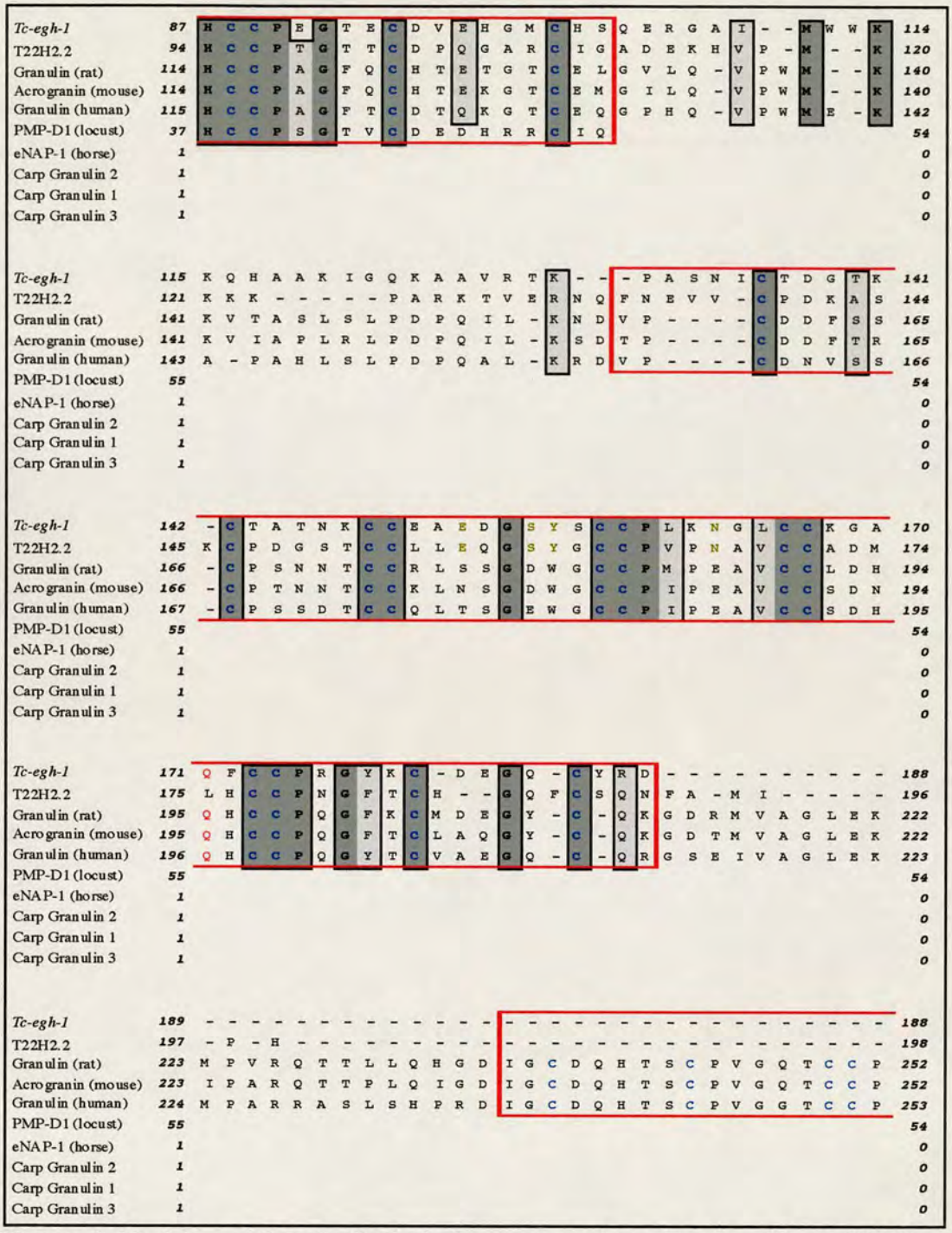



**Figure 3.10 (a). continued**. GenBank Accession numbers: *Tc-egh-1* (5': AA836692 & 3': AA836693), *C. elegans*/T22H2 (Z81595), *Rattus norvegicus* (Granulin) (M97750), *Mus musculus* (Acrogranin) (D16195)

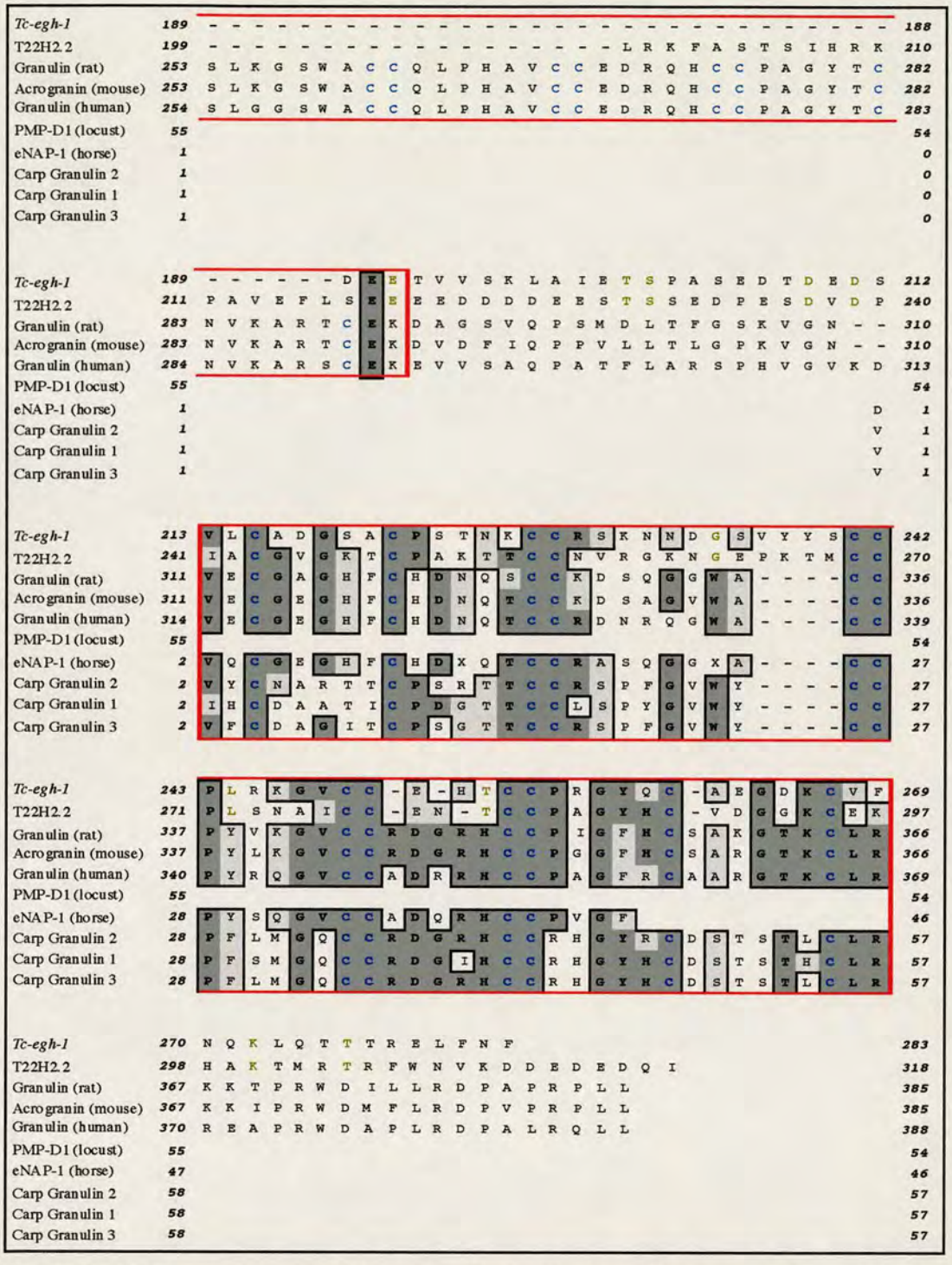



**Figure 3.10 (a). continued**. SwissProt Identification Numbers: *Homo sapiens* (Granulin) (sp|P28799), Locust (PMP-D1) (sp|P80059), Horse (eNAP-1) (sp|P80930), *Cyprinus carpio* (Granulin 2 - partial) (sp|P81014), (Granulin 1) (sp|P81013), (Granulin 3) (sp|P81015)

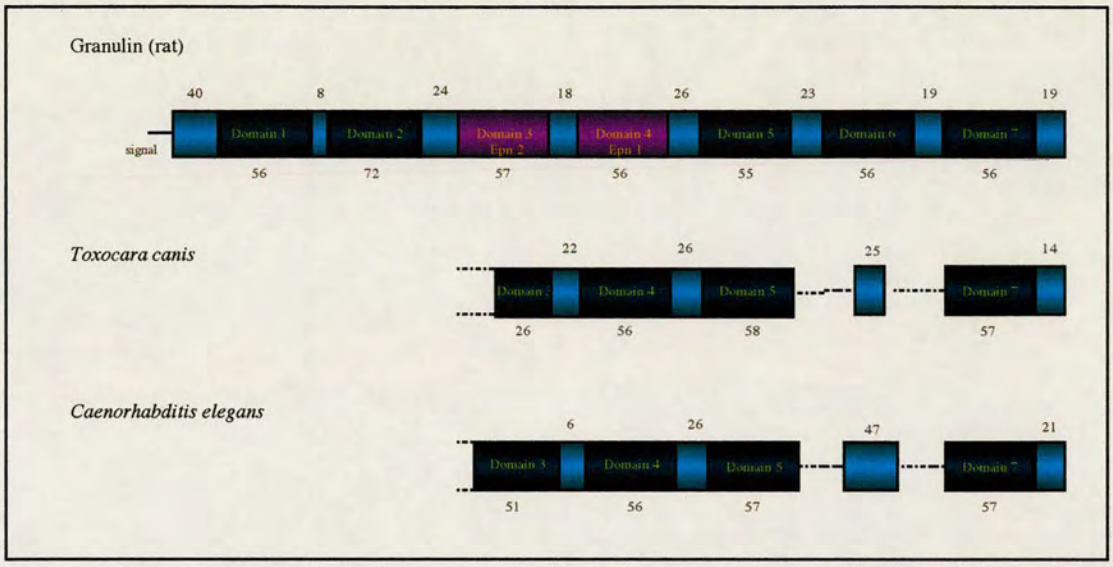



**Figure 3.10 (b)**. Schematic representation of the aligned granulin sequences obtained from rat (accession no. M97750), *T. canis*/ *Tc-egh-1* (accession nos. 5' AA836692 and 3' AA836693) and *C. elegans* /T22H2.2 (accession no. Z81595)

### 3.3.21 *Lupus autoantigen homologue* (Tc-lah-1)

The autoimmune disease systemic lupus erythrematosus (SLE) is characterised by high levels of autoantibodies to DNA and nucleic acid-associated proteins (Kotzin, 1996). The Lupus Autoantigen (also known as Sjogren Syndrome type B antigen, or SS-B) is a ribonucleoprotein which shows a high degree of conservation between animal phyla. Autoimmune responses can be initiated by infectious agents (Oldstone, 1987), and in a related example, the RAL-1 product of *Onchocerca volvulus*, which is homologous to the Sjogren Syndrome type A antigen, has been shown to stimulate high levels of antibody (McCauliffe et al., 1990).

### 3.3.22 *Olfactomedin* (Tc-ofm-1)

A *T. canis* homologue of the olfactomedin gene family was found. The prototype gene was found to encode a 57 kDa glycoprotein in the extracellular mucous matrix of the olfactory neuroepithelium of frogs (Yokoe and Anholt, 1993), but additional homologues are now known to be widely expressed in mammalian brain (Karavanich and Anholt, 1998). An olfactomedin homologue has been found in *C. elegans*, and *Tc-ofm-1* shows maximum similarity to this sequence (Figure 3.11). As *T. canis* larvae produce high levels of mucins, including those constituting the surface coat (Gems and Maizels, 1996; Page et al., 1992b), *Tc-ofm-1* protein may be involved with the mucous layers in this parasite.

Analysis of the primary structure of olfactomedin has linked the glycoprotein with extracellular matrix proteins of which include the neural cell adhesion molecule (N-CAM), laminin, fibronectin and proteoglycans. These extracellular proteins have all been implicated in adhesion, maintenance, migration, growth, or differentiation of neurones (Reichardt and Tomaselli, 1991), thus by analogy it could be suggested that olfactomedin may be able to function in a similar manner to the olfactory epithelium (Yokoe and Anholt, 1993). Recently, Nguyen *et al* (1998) showed the presence of a glycoprotein in the trabecular-meshwork-inducible glucocorticoid response gene (TIGR) of the eye that showed significant homology to olfactomedin. Mutations in these olfactomedin related domains have been linked closely with a large number of cases of primary open-glaucoma, a disease responsible for blindness in 1 out of every 100 individuals over the age of 40

(Stone et al., 1997). Kubota *et al* (1997) described the identification of a protein they termed "myocilin" as a result of its homology to myocin and its preferential localisation to the ciliary rootlet and basal body of the connecting cilium of photoreceptor cells. Myocilin was shown to have significant homology with myosin in the N-terminal region and with olfactomedin in the C-terminal region, a characteristic not seen in the frog olfactomedin (Yokoe and Anholt, 1993) or the rat related olfactomedin (Danielson et al., 1994); it was however found in the non-muscle myosin of *Dictyostelium discoideum.* The myocilin cDNA sequence differs from TIGR by only a 2 base pair GA insertion at nucleotides 104-105 (Kubota et al., 1997).

The olfactomedin homologue isolated as part of the EST project bears little similarity to myocillin or TIGR. The homology obtained is due to the fact that both the aforementioned proteins have an olfactomedin like domain within their sequence. From the pileup in Figure 3.11 it can be seen that there is significant similarity between the *C. elegans* clone at the 5 prime end of the gene. This similarity extends to encompass similarities with both the frog olfactomedin and the rat AMZ protein (see Chapter 1: Introduction). The pileup of the conserved sequences in Figure 3.12 show this level of homology more clearly. As can be seen from both Figures 3.11 and 3.12 there are 7 domains found in *C. elegans*, frog and rat. Of these 7, Domain 1 is absent from the *T. canis* sequence. The reason for this absence could be attributable to the fact that the *T. canis* sequence is 5' truncated. This suggestion seems somewhat unlikely as there is a putative start methionine present, although this is yet to be confirmed. The more likely explanation would be that the loss of Domain 1 was an evolutionary modification of the parasites protein. Perhaps this domain had no relevant function in terms of allowing the parasite to colonise a host or in parasite survival within the host. Identification of olfactomedin homologues from other parasitic nematodes bearing similar evolutionary changes would support the likelihood of this event having occurred.

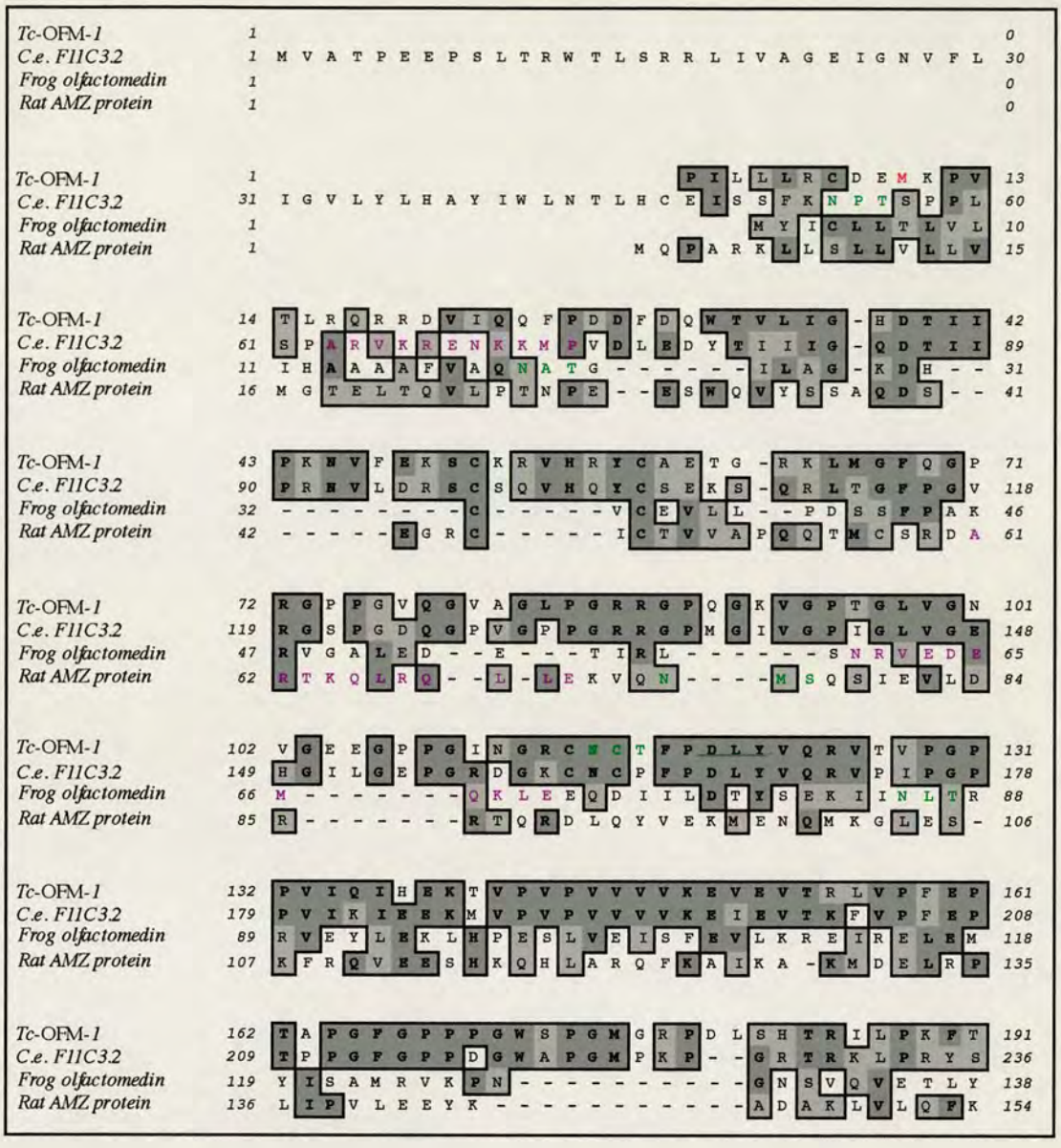



**Figure 3.11.** Amino acid pileup following that as outlined by Karavanich and Anholt in (1998). N-glycosylation sites are highlighted in green. The purple bars mark the conserved regions as defined from *C. elegans*, frog, rat and mouse (Karavanich and Anholt, 1998) that are shared by *T. canis*. Sequences were obtained from GenBank as referenced in (Karavanich and Anholt, 1998): *Tc-ofm-1* (Tc-EST-070: 5'AA728641 & AA728642; Tc-EST-079: AA728643 & AA728644), *C.e.* F11H3.2 (Z81499), Frog (*Rana catesbeiana*) (L13595), Rat (*Rattus norvegicus*) (U03416)

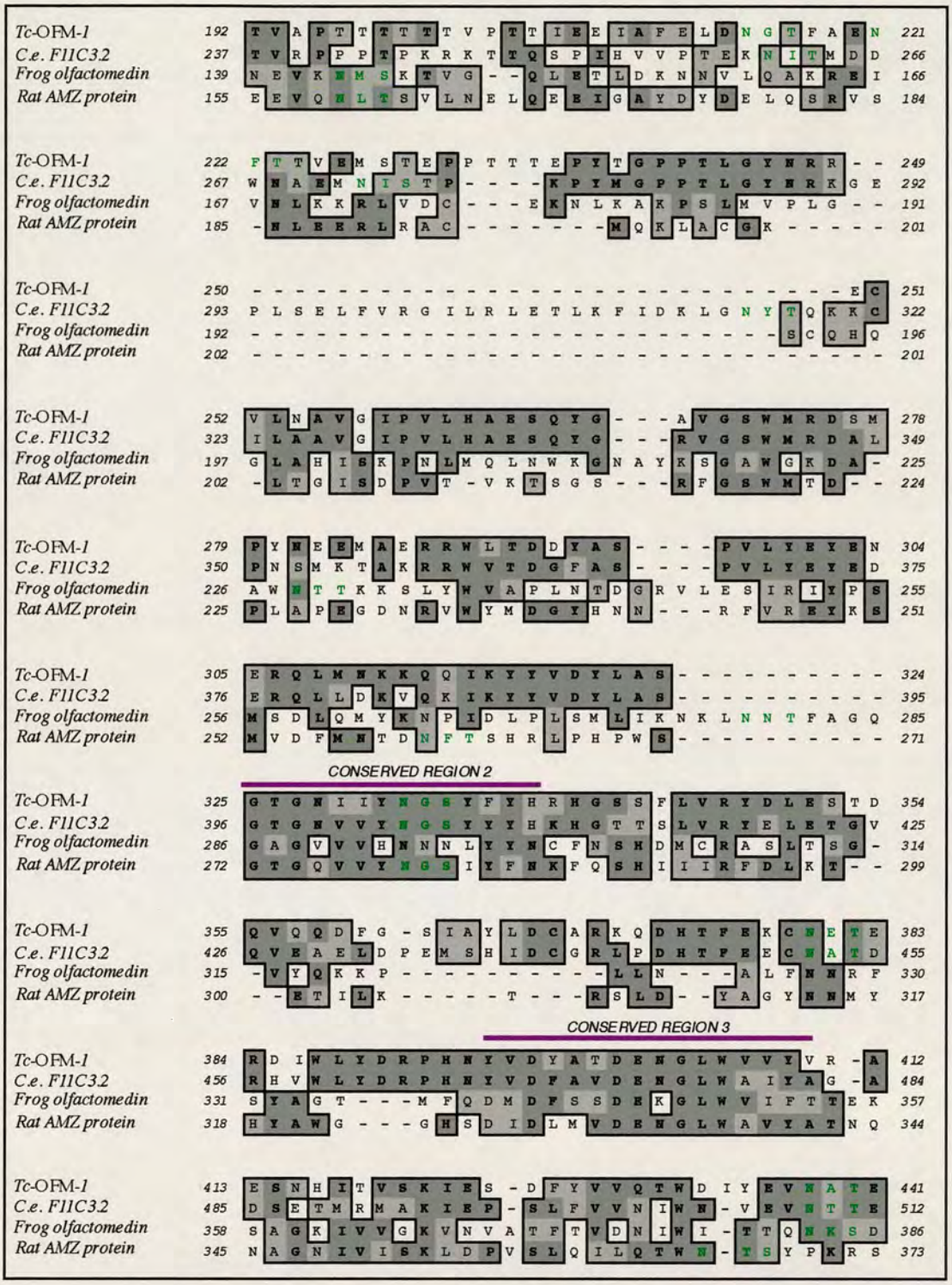



**Figure 3.11 continued**

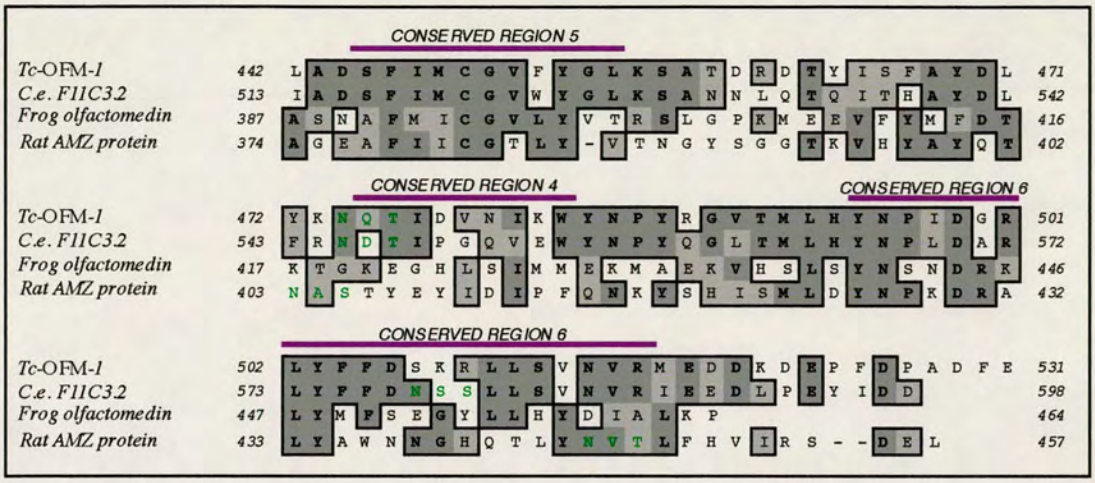



**Figure 3.11 continued**

**CONSERVED REGION 1**

| | |
|---|---|
| *Toxocara canis* | |
| *Caenorhabditis elegans* | A R V K R E N K K M P |
| Frog (*Rana catesbeiana*) | N R V E D E N Q K L E |
| Rat (*Rattus norvegicus*) | A R T K Q L R Q L L E |

**CONSERVED REGION 2**

| | |
|---|---|
| *Toxocara canis* | G T G N I I Y N G S Y F Y H |
| *Caenorhabditis elegans* | G T G N V V Y N G S Y Y Y H |
| Frog (*Rana catesbeiana*) | G A G V V V H N N N L Y Y N |
| Rat (*Rattus norvegicus*) | G T G Q V V Y N G S I Y F N |

**CONSERVED REGION 3**

| | |
|---|---|
| *Toxocara canis* | Y V D Y A T D E N G L Y V V Y V |
| *Caenorhabditis elegans* | Y V D F A V D E N G L W A I Y A |
| Frog (*Rana catesbeiana*) | D M D F S S D E K G L W V I F T |
| Rat (*Rattus norvegicus*) | D I D L M V D E N G L W A V Y A |

**CONSERVED REGION 4**

| | |
|---|---|
| *Toxocara canis* | Q T I D V N I K W |
| *Caenorhabditis elegans* | S L F V V N I W W |
| Frog (*Rana catesbeiana*) | T F T V V N I W I |
| Rat (*Rattus norvegicus*) | S L Q I L Q T W N |

**CONSERVED REGION 5**

| | |
|---|---|
| *Toxocara canis* | S F I K C G V F Y G L |
| *Caenorhabditis elegans* | S F I M C G V W Y G L |
| Frog (*Rana catesbeiana*) | A F M I C G V L Y V I |
| Rat (*Rattus norvegicus*) | A F I I C G T L Y V T |

**CONSERVED REGION 6**

| | |
|---|---|
| *Toxocara canis* | Y N P I D G R L Y F F D S K R L L S V N V R M |
| *Caenorhabditis elegans* | Y N P L D A R L Y F F D N S S L L S V N V R I |
| Frog (*Rana catesbeiana*) | Y N S N D R K L Y M F S E G Y L L H Y D I A L |
| Rat (*Rattus norvegicus*) | Y N P K D R A L Y A W N N G H Q T L Y N V T L |

**Figure 3.12.** The six conserved domains identified in olfactomedin and the respective homologues. Residues highlighted in red are those found in all the aligne members. Those highlighted in green refer to aa found in three of the four aligned sequences of which *T. canis* is one. Residues highlighted in blue are those amino acid shared by *T. canis* and *C. elegans* only

### 3.3.23 *PC4/TIS7 "Interferon-related protein"* (Tc-pth-1)

A *T. canis* gene has been isolated which is similar to a mammalian product variously described as PC4 or TIS7, induced in cell lines by activators such as nerve growth factor or phorbol esters (Tirone and Shooter, 1989). The same mammalian gene has also been designated interferon-related protein, due to an erroneous deposition of this sequence in 1985 as murine beta-interferon (accession no J00424). We have named the *T. canis* gene, which bears no similarity to mammalian beta- or gamma-interferon's, *Tc-pth-1* (PC4/TIS7 homologue).

### 3.3.24 *Tubby-like protein* (Tc-tlp-1)

Recently, a new multigene family related to the mouse gene *tubby* has been recognised (Kleyn et al., 1996). Mutations in *tubby* in the mouse result in obesity and degenerative changes in adult life; however the existence of conserved homologues in nematodes (including *C. elegans*) and in plants indicate that this gene encodes a protein which fulfils a fundamental function in all higher organisms.

## 3.4 Discussion

This chapter deals with the small scale analysis of expressed genes from a *Toxocara canis* cDNA library. The hypothesis behind this investigative approach was to identify cDNAs encoding abundant primarily surface and secreted proteins that were represented as abundant mRNAs from this immunologically important parasite lifestage. In this regard the project has realised its aim by identifying numerous clones encoding many genes of hypothetical immunological importance.

### 3.4.1 *ESTs generated from serendipitously selected clones*

The cDNA library used in this study was derived from the infective larval stage of *Toxocara canis* , which is reproductively and developmentally arrested, yet is very much metabolically active. Reflecting this, very few clones relating to developmental genes were identified.

Of the 332 selected clones, 266 were found to contain inserts and so were used in subsequent analysis. Detailed analysis on these clones showed that 5 of the 266 clones were in fact chimeric. That is to say, a number of the clones possessed internal *Eco RI* sites which ligated to each other prior to packaging within the phage.

To conclude the outcome of this small-scale EST strategy suggests that it may be applicable to any metazoan organisms which have relatively large genome sizes ($=10^8$ bp) and in which interest is focused on genes expressed at moderate to high levels. In addition, genes restricted to life-cycle stage can be identified in this manner. Further evidence for the success of this approach is seen in the Filarial Genome Project which in three years has deposited over 15,000 sequences in dbEST, which are estimated to represent some 5,000 separate gene transcripts, or around 33% of the total gene complement of *B. malayi*. The availability of the full genome sequence of *C. elegans* lends an exceptional opportunity to compare free-living and parasite gene sequences and structure, with the potential to identify adaptations requisite for parasitism at the molecular level.

Although this project is a modest effort compared to the wealth of genome projects presently being undertaken. Many of the genes identified by database homology searching proved to be novel not only to parasitic nematodes but also to parasitic systems as a whole. This work has also made the first step towards a comprehensive

gene catalogue of a biologically intriguing and clinically significant parasite, providing a resource and springboard for future studies.

## Chapter 4 *Abundant Novel Transcripts*

## 4.1 Introduction

A significant proportion of genes identified in genome projects turn out to be novel. That is to say, they have no homologues in any of the existing databases. Approximately 40% of the genes isolated as part of the *Toxocara* EST project were found to be novel genes. *Ascaris*, hookworm (*Necator*) and *Trichuris* ESTs have all been found to contain 16% (22.56/141), 42% (86.1/205) and 51% (75.99/149) novels genes (Daub J and Blaxter M. L. pers. comm.).

This chapter deals with the identification of four abundantly represented gene clusters termed *a*bundant *n*ovel *t*ranscripts (*ants*) 3, 5, 30 and 34 (see chapter 3). Tables 4.1-.4 shows the clones that comprised each gene cluster (also included (Table 4.5) are the clones representative of the *no*vel *t*ranscript (*not*) 018 which in the original classification was termed an abundant novel transcript). These clones were somewhat unexpected, because although hyper abundant within the library - occurring at a frequency of between 2.6 - 6.0% of all cDNA clones randomly selected - their sequences showed no homology with any known proteins in the database. It would appear that whatever proteins these clones encode are novel. The abundance of these clones, which together made up 18% of the library – *not*-018 accounts for 1.1% -, suggest an important role in parasite survival.

Described here is the characterisation of these antigens and subsequent attempts to assign putative functions to these families and thus to hypothesise putative roles with regards to survival at the parasite/host interface.

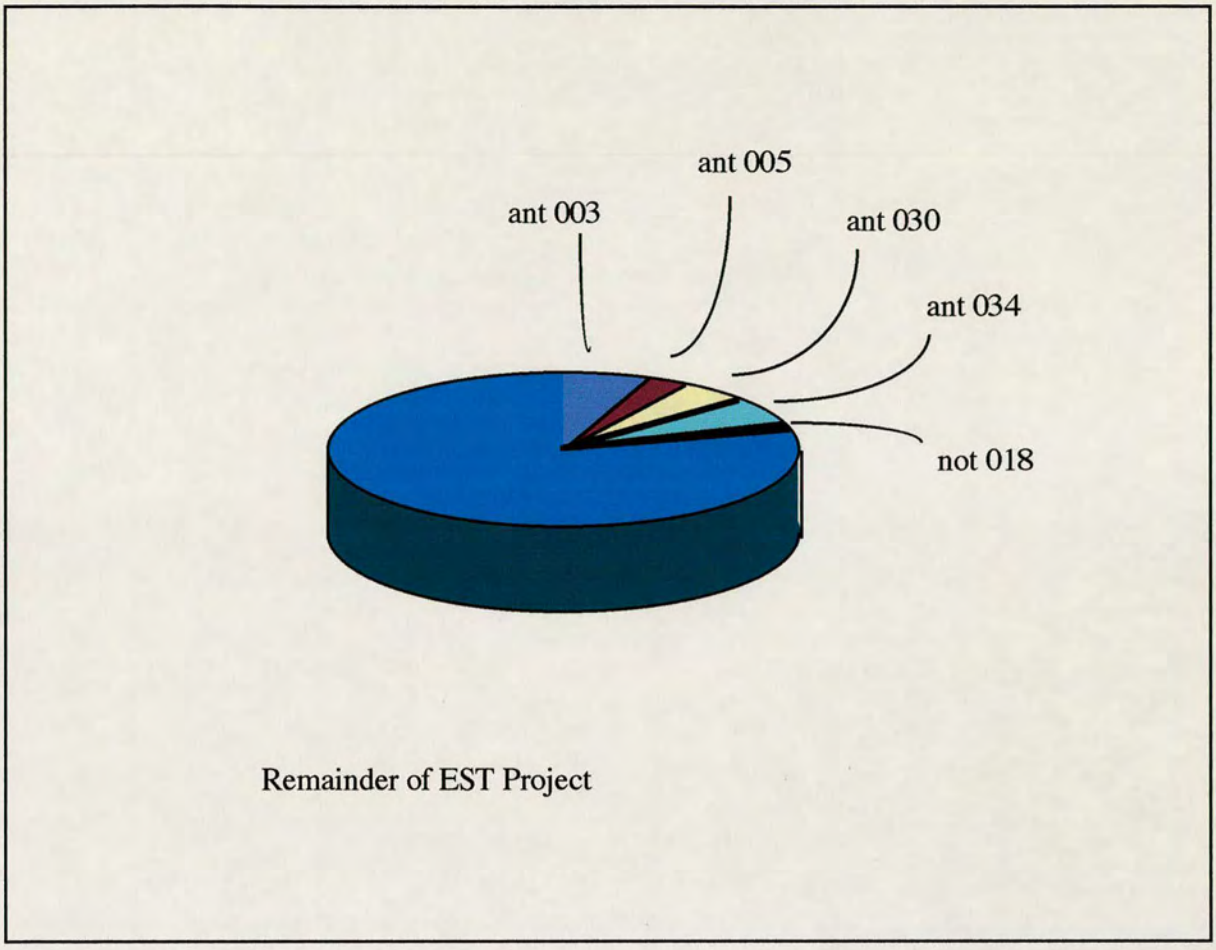



Figure 4.1. Pie-chart showing the abundance of the *ant* genes and the *not 018* gene with relation to the remainder of the EST project.

## 4.2 MATERIALS AND METHODS:

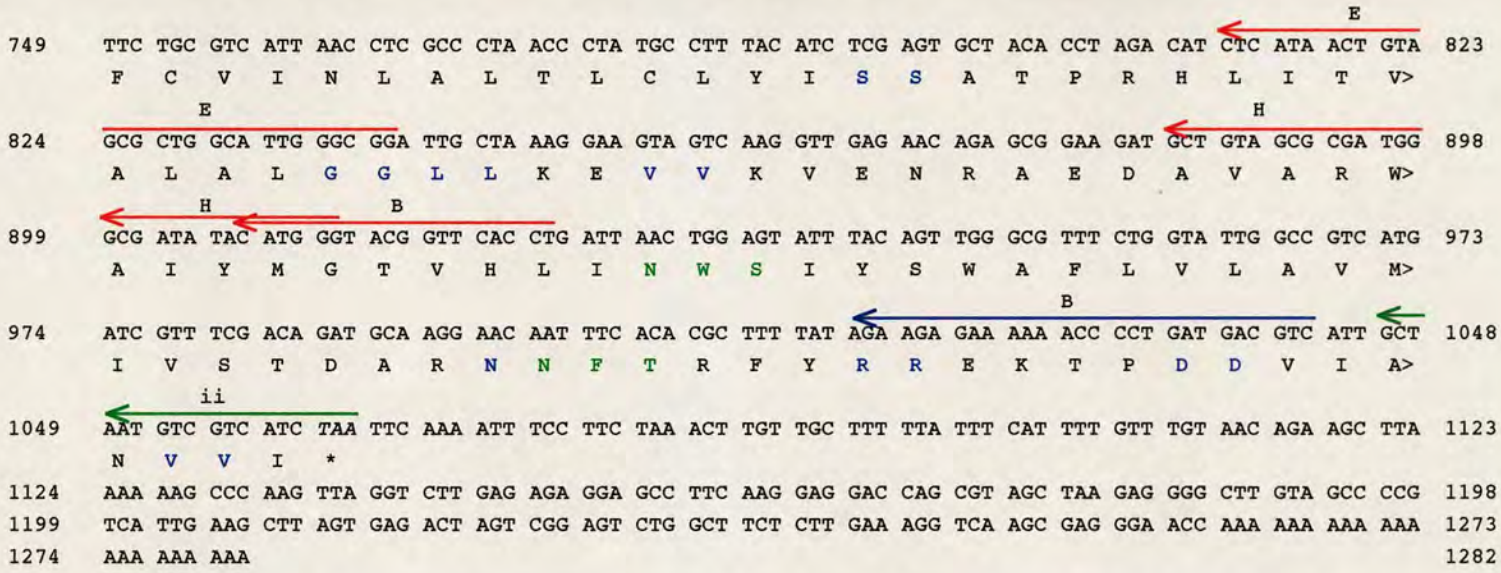
### 4.2.1 *Identification of clones.*

The clones relating to the *T. canis ant* genes were initially isolated as part of the *T. canis* EST project (chapter 3). The procedure employed is as outlined in Chapter 2: Materials and Methods.

### 4.2.2 *Sequencing of the Ants*

Double stranded DNA templates for the following members of the abundant novel transcripts (*ant 003*: *Tc*-003; *Ant 005*: *Tc* 228; *ant 030*: *Tc 128* and *ant 034*: *Tc 315*) were sequenced using the universal primers M13 forward and reverse and subsequent gene specific internal primers (see figure 4.2 to 4.6 for primer design). Clone were selected based on the initial analysis of the gene clusters. Later, more detailed analysis of the clusters identified longer clone members, but it was deemed prudent to complete analysis on clones selected (Tables 4.1 to 4.5). Sequencing was performed as outlined in the materials and methods of chapter 2. Initial analysis of the *ant* clusters, EST clones Tc-EST-003, 228, -128, and -315 were selected for analysis. Analysis of the nucleotide and deduced amino acid composition and database searches were performed as outlined in chapters 2 and 3. Sequence analyses of the members of each *ant* cluster were performed using the MacVector software (Oxford Molecular) in order to determine open reading frames (ORFs). Contigs for each cluster were created using the AssemblyLign software (Oxford molecular).


**Figure 4.2** Nucleotide and deduced amino acid sequence of *a*bundant *n*ovel *t*ranscript (*ant*) 003. The reverse and forward pCDNA3.1(+/-) (Invitrogen) (see later) primers are indicated by green arrows. pET 15b primers (see later) are shown by the blue arrows. Restriction sites are highlighted in bold and the stop codon is indicated by the asterix. Internal primers are shown by the red arrows. The sequence for *ant-003* was derived from clone Tc-EST-003. (Figure continued over page)




| | | |
|---|---|---|
| i | ANT 3 pcDNA3.1(-) F: | 5′ **GAA TTC** GCC ACC *ATG* GTG CTC GCC GTG GTT TAC GGT 3′ |
| ii | ANT 3 pcDNA3.1(-) R: | 5′ **GAA TTC** *TTA* GAT GAC GAC ATT AGC AAT 3′ |
| A | ANT 3 pET 15b F: | 5′ CGG CGC **CAT ATG** TAC GGT TTG AGC TAT ACC ATT GGG GAT ATT GAC 3′ |
| B | ANT 3 pET 15b R: | 5′ CGG CGC **CAT ATG** TTT *TCA* AGA AAC ACC TTT GAG AGG AAT TAT 3′ |
| A | Tc-EST-003/323 F | 5′ AAC TTA CTG GAA TGG TTT 3′ |
| B | Tc-EST-323 R | 5′ AGG TGA ACC GTA CCC ATG 3′ |
| C | Tc-EST-281 (gc) R | 5′ GCT TAG CTC TCG CGT CGC CCA ACA C 3′ |
| D | Tc-EST-281 (gc) F | 5′ GTT TGG CCT GGT TGT GAA CAC TGG G 3′ |
| E | Tc-EST-281 R | 5′ CCG CCC AAT GCC AGC GCT ACA GTT ATG AG 3′ |
| F | Tc-GC/003 380 R | 5′ GGC GTA GAT GTC ATC AAG ATG GAA GCC C 3′ |
| G | Tc-GC/003 479 F | 5′ GGT TTT GAG CCT GTT CAA TGG GTT CGG ACC 3′ |
| H | Tc-GC/003 109 R | 5′ CCC ATG TAT ATC GCC CAT CGC GCT ACA GC 3′ |

**Figure 4.2** (continued). *N*-linked glycosylation sites are indicated in green.

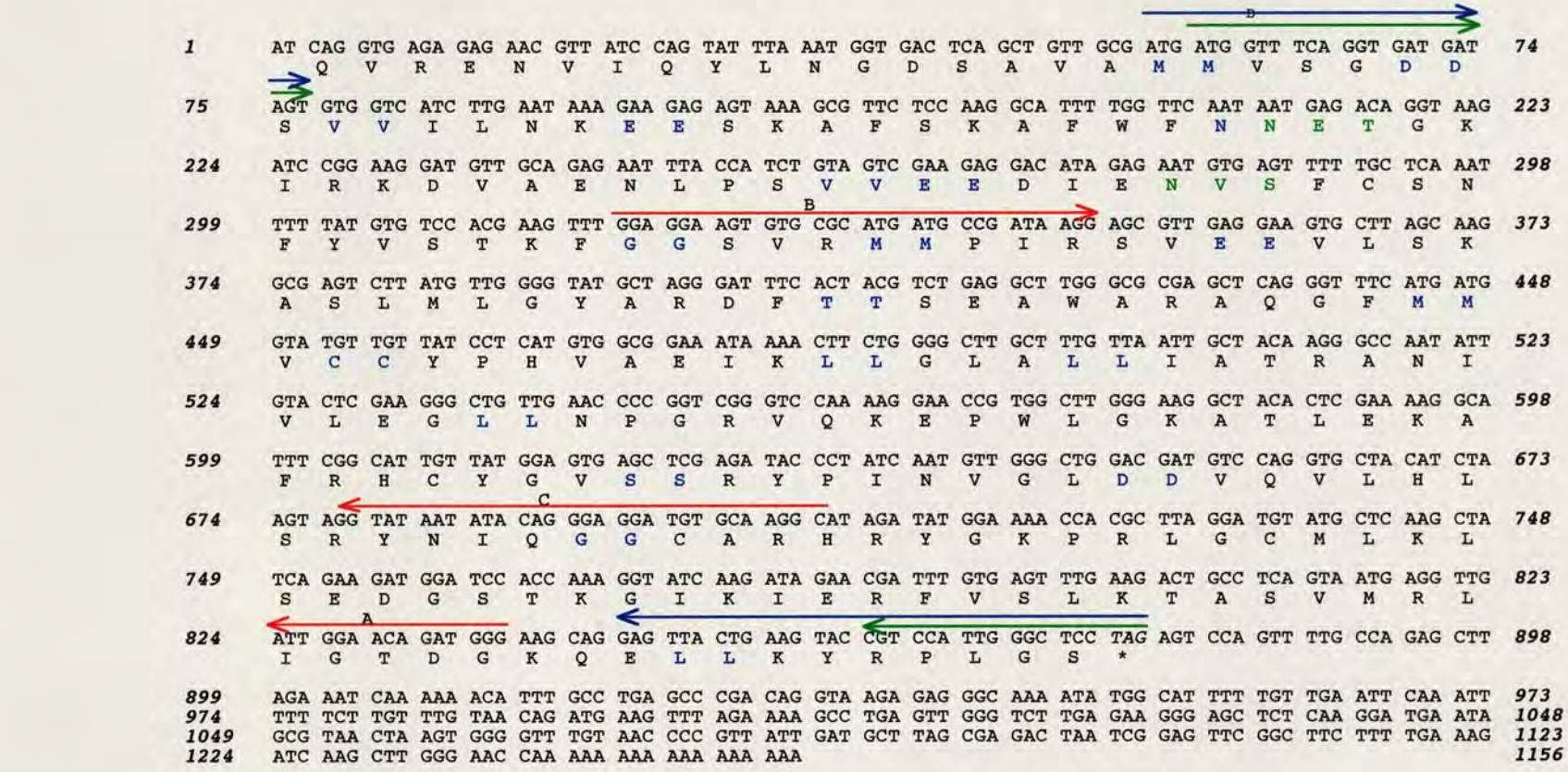



**Figure 4.3** Nucleotide and deduced amino acid sequence of *a*bundant *n*ovel *t*ranscript (*ant*)-*005*. pcDNA3.1(+/-) primers are shown by the green arrows for both the 5' and 3' ends. pET 15b primers are shown by the blue arrows. Internal primers are shown by red arrows with *N*-linked glycosylation sites in green.

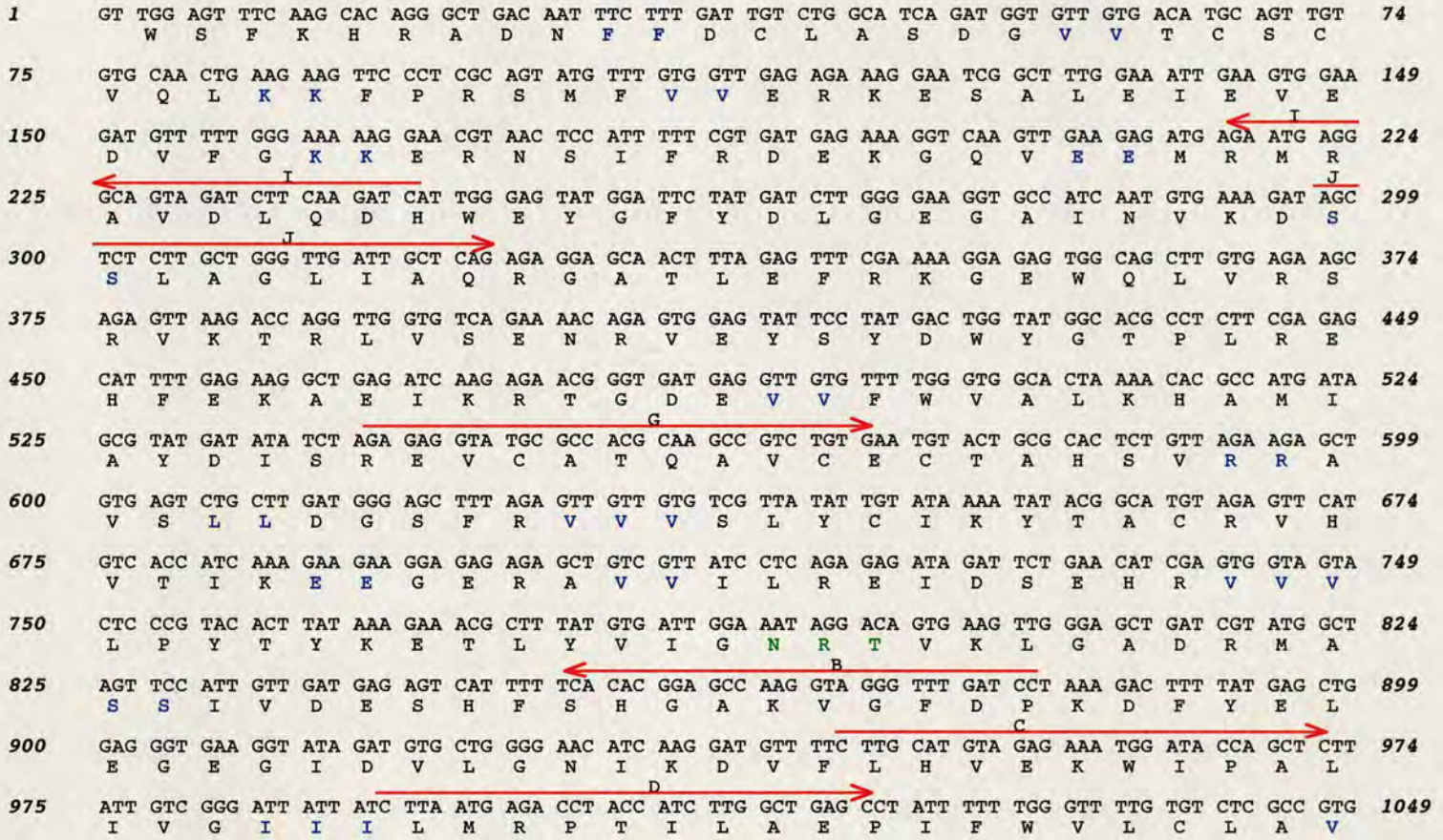



**Figure 4.4** Nucleotide and deduced amino acid sequence of *a*bundant *n*ovel *t*ranscript (*ant*) 030. Internal primer sites are highlighted in red. The stop codon is shown in italics. The *N*-linked glycosylation sites are in shown in green.

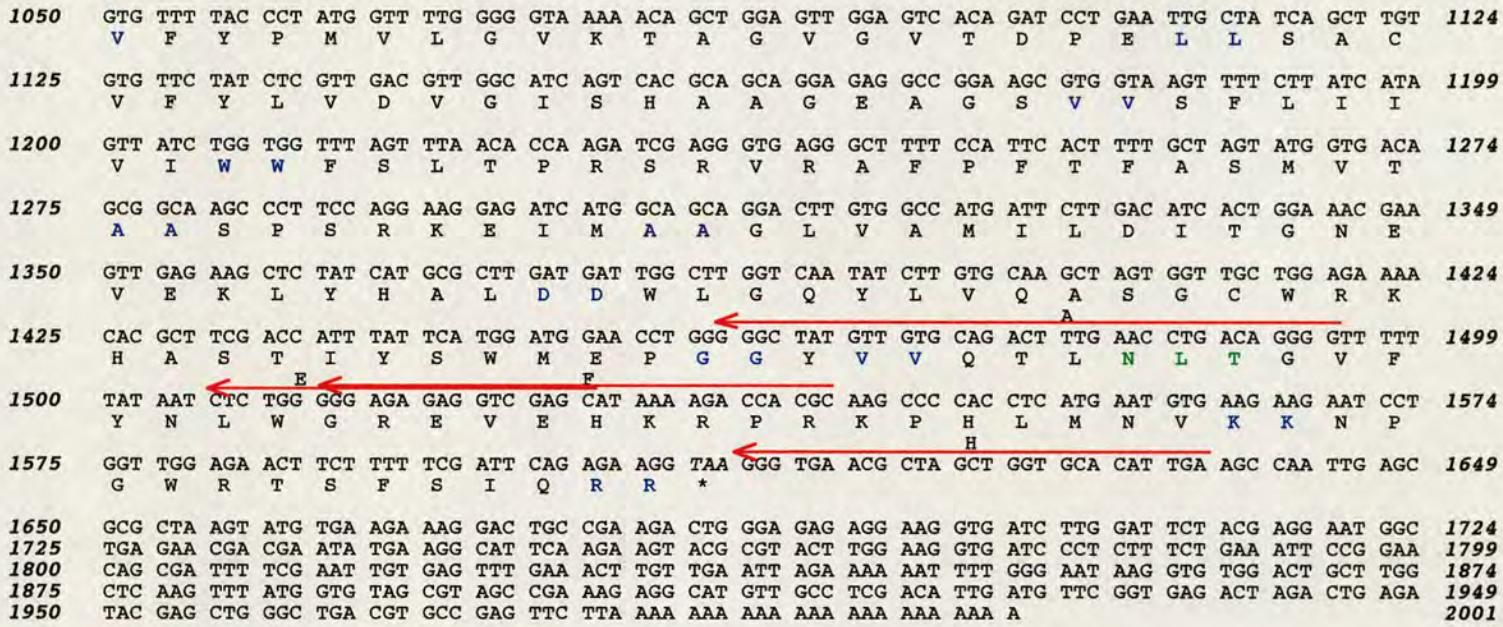
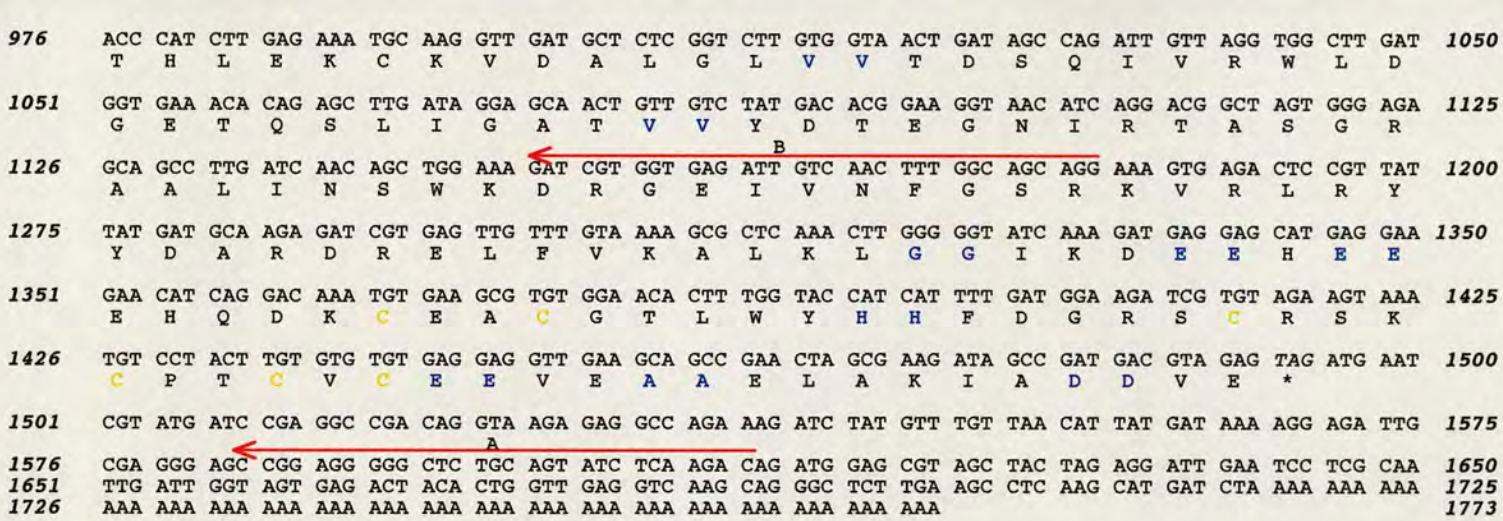


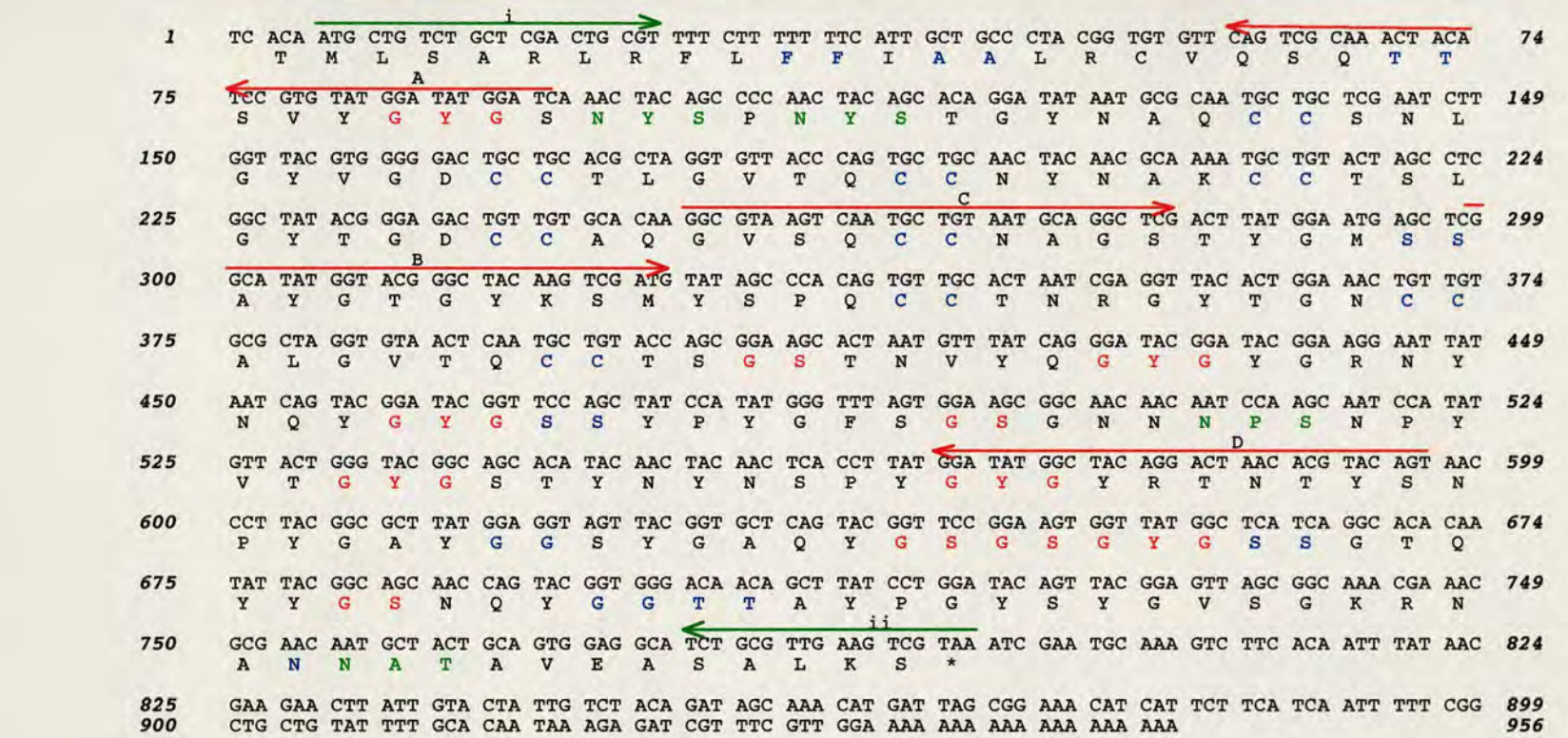

**Figure 4.4** continued


**Figure 4.5** Nucleotide and deduced amino acid sequence of *a*bundant *n*ovel *t*ranscript (*ant*)-*034*. The internal primer sites are shown by red arrows. The *N*-linked glycosylation sites of *ant-034* are shown by the green arrows.




| | | |
|---|---|---|
| A | Tc-GC/034 279 R: | 5′ GTC TTG AGA TAC TGC AGA GCC CCC TCC GG 3′ |
| B | Ant 034 AR 142-110: | 5′ CCT GCT GCC AAA GTT GAC AAT CTC ACC ACG ATC 3′ |
| C | Ant 034 BR 131-102: | 5′ GCA TTG ATG TTT ATG CCG CAT TCT GAC ATG 3′ |
| D | Ant 034 BR 291-321: | 5′ CCA CAG TAT ACC ATC AGC CGA CCT ATG GAG 3′ |
| E | Tc-315(Ant 034) 364 F: | 5′ GGC CCA TTT CTC AGA TAC GAA GAC GAT AGC 3′ |

**Figure 4.5** continued




| | | |
|---|---|---|
| i | ANT 18 pcDNA3.1 (-) F: | 5′ **GAA TTC** GCC ACC *ATG* CTG TCT GCT CGA CTG CGT 3′ |
| ii | ANT 18 pcDNA3.1 (-) R: | 5′ **GAA TTC** *TTA* CGA CTT CAA CGC AGA TGC 3 |
| A | ANT 18 R 136-102: | 5′ GAT CCA TAT CCA TGC ACG GAT GTA GTT GTC GAC TC 3′ |
| B | ANT 18 F 340-368: | 5′ CGG CAT ATG GTA CGG GCT ACA ACT CGA TG 3′ |
| C | Tc 31 F 252-280: | 5′ GGC GTA AGT CAA TGC TGT AAT GCA GGC CTC 3′ |
| D | Tc 31 R 100-71: | 5′ ACT GTA CGT GTT AGT CCT GTA GCC ATA TCC 3′ |

**Fugure 4.6** Nucleotide and deduced amino acid sequence of *no*vel *t*ranscript (*not*)-*018*. pcDNA primers are shown by green arrows and internal primers are shown in red. The *N*-linked glycosylation sites are shown in green

| | Clone Number | Gene Name | Identification | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|
| 1 | Tc-EST-003 | *Tc-ant-003* | Abundant Novel Transcript 003 | 1300 | 594 | 620 | AA470326 | AA683460 |
| 2 | Tc-EST-007 | *Tc-ant-003* | Abundant Novel Transcript 003 | 330 | 280 | | AA563575 | |
| 3 | Tc-EST-015 | *Tc-ant-003* | Abundant Novel Transcript 003 | 420 | 407 | 416 | AA563520 | |
| 4 | Tc-EST-060 | *Tc-ant-003* | Abundant Novel Transcript 003 | 2050 | 456 | 462 | AA888761 | AA888762 |
| 5 | Tc-EST-067 | *Tc-ant-003* | Abundant Novel Transcript 003 | 1280 | 534 | 695 | AA873909 | AA873910 |
| 6 | Tc-EST-106 | *Tc-ant-003* | Abundant Novel Transcript 003 | 534 | 519 | 492 | AA824224 | AA824225 |
| 7 | Tc-EST-137 | *Tc-ant-003* | Abundant Novel Transcript 003 | 544 | 491 | 537 | AA835623 | AA835624 |
| 8 | Tc-EST-172 | *Tc-ant-003* | Abundant Novel Transcript 003 | 1700 | 393 | 564 | AA873911 | AA873912 |
| 9 | Tc-EST-191 | *Tc-ant-003* | Abundant Novel Transcript 003 | 860 | 497 | 460 | AA873896 | AA873897 |
| **10** | **Tc-EST-239** | ***Tc-ant-003v*** | **Abundant Novel Transcript 003** | **640** | **326** | **477** | **AA874755** | **AA874756** |
| 11 | Tc-EST-258 | *Tc-ant-003* | Abundant Novel Transcript 003 | 300 | 276 | 276 | AA875763 | AA875764 |
| 12 | Tc-EST-268 | *Tc-ant-003* | Abundant Novel Transcript 003 | 400 | 372 | 395 | AA875774 | AA875775 |
| 13 | Tc-EST-281 | *Tc-ant-003* | Abundant Novel Transcript 003 | 1200 | 558 | 620 | AA875786 | AA875787 |
| 14 | Tc-EST-291 | *Tc-ant-003* | Abundant Novel Transcript 003 | 600 | 250 | 443 | AA875797 | AA875798 |
| 15 | Tc-EST-318 | *Tc-ant-003* | Abundant Novel Transcript 003 | 320 | 290 | 299 | AA879357 | AA879358 |
| 16 | Tc-EST-323 | *Tc-ant-003* | Abundant Novel Transcript 003 | 1000 | 380 | 726 | AA879366 | AA879367 |

**Table 4.1.** Table showing EST clones belonging to the *ant-003* cluster. Line 10 is highlighted in bold to indicate a variant member of the *ant-003* cluster. Line 4 is shown in blue text to indicate the longest member of the cluster

| | Clone Number | Gene Name | Identification | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST Genbank |
|---|---|---|---|---|---|---|---|---|
| 1 | Tc-EST-005 | *Tc-ant-005* | Abundant Novel Transcript 005 | 510 | 426 | 435 | AA470328 | AA683461 |
| 2 | Tc-EST-065 | *Tc-ant-005* | Abundant Novel Transcript 005 | 1950 | 509 | 543 | AA728637 | AA728636 |
| 3 | Tc-EST-111 | *Tc-ant-005* | Abundant Novel Transcript 005 | 890 | 329 | 562 | AA825121 | AA825122 |
| 4 | Tc-EST-134 | *Tc-ant-005* | Abundant Novel Transcript 005 | 1100 | 443 | 459 | AA835617 | AA835618 |
| 5 | Tc-EST-228 | *Tc-ant-005* | Abundant Novel Transcript 005 | 1150 | 570 | 390 | AA874729 | AA979710 |
| 6 | Tc-EST-234 | *Tc-ant-005* | Abundant Novel Transcript 005 | 1960 | 419 | 496 | AA874738 | AA874739 |
| 7 | Tc-EST-295 | *Tc-ant-005* | Abundant Novel Transcript 005 | 1150 | 183 | 333 | AA875804 | AA875805 |

**Table 4.2.** Table showing EST clones belonging to the *ant-005* cluster. Line 6 is highlighted in blue to indicate the longest member of the *ant-005* cluster.

| | Clone Number | Gene Name | Identification | Insert Size | 5' Read | 3' REad | 5' EST GenBank | 3' EST Genbank |
|---|---|---|---|---|---|---|---|---|
| 1 | Tc-EST-018 | *Tc-not-018* | Novel Transcript 018 | >1270 | 600 | 671 | AA683463 | AA683464 |
| 2 | Tc-EST-031 | *Tc-not-018* | Novel Transcript 018 | 1000 | 382 | 555 | AA568095 | AA568096 |
| 3 | Tc-EST-136 | *Tc-not-018* | Novel Transcript 018 | 1000 | 620 | 565 | AA835621 | AA835622 |

**Table 4.3.** EST clones belonging to the *not-018* cluster. Line 1 is highlighted in blue to indicate a the longest member of the cluster

| | Clone Number | Gene Name | Identification | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|
| 1 | Tc-EST-030 | *Tc-ant-030* | Abundant Novel Transcript 030 | 652 | 652 | 652 | AA568093 | AA568094 |
| 2 | Tc-EST-045 | *Tc-ant-030* | Abundant Novel Transcript 030 | 1300 | 519 | 652 | AA728634 | AA728635 |
| 3 | Tc-EST-050 | *Tc-ant-030* | Abundant Novel Transcript 030 | 1200 | 598 | 663 | AA683467 | AA683468 |
| 4 | Tc-EST-092 | *Tc-ant-030* | Abundant Novel Transcript 030 | 490 | 371 | 402 | AA810507 | AA825113 |
| 5 | Tc-EST-110 | *Tc-ant-030* | Abundant Novel Transcript 030 | 1800 | 357 | 693 | AA825119 | AA825120 |
| 6 | Tc-EST-128 | *Tc-ant-030* | Abundant Novel Transcript 030 | 2000 | 485 | 419 | AA907967 | AA907968 |
| 7 | Tc-EST-133 | *Tc-ant-030* | Abundant Novel Transcript 030 | 427 | 414 | 381 | AA907969 | AA907970 |
| 8 | Tc-EST-149 | *Tc-ant-030* | Abundant Novel Transcript 030 | 630 | 602 | | AA836698 | |
| 9 | Tc-EST-165 | *Tc-ant-030* | Abundant Novel Transcript 030 | 280 | 279 | 285 | AA864179 | AA864180 |
| 10 | Tc-EST-199 | *Tc-ant-030* | Abundant Novel Transcript 030 | 1950 | 335 | 483 | AA888755 | AA888756 |
| 11 | Tc-EST-219 | *Tc-ant-030* | Abundant Novel Transcript 030 | 800 | 576 | 663 | AA874723 | AA874724 |
| 12 | Tc-EST-280 | *Tc-ant-030* | Abundant Novel Transcript 030 | 1200 | 567 | 432 | AA875784 | AA875785 |
| 13 | Tc-EST-308 | *Tc-ant-030* | Abundant Novel Transcript 030 | 800 | 458 | 579 | AA879346 | AA897347 |

**Table 4.4.** EST clones belonging to the *ant-030* cluster. Line 10 is highlighted in blue to indicate the longest member of the *ant-030* cluster

| | Clone Number | Gene Name | Identification | Insert Size | 5' Read | 3' Read | 5' EST GenBank | 3' EST GenBank |
|---|---|---|---|---|---|---|---|---|
| 1 | Tc-EST-034 | *Tc-ant-034* | Abundant Novel Transcript 034 | 230 | 230 | 230 | AA728632 | AA728633 |
| 2 | Tc-EST-055 | *Tc-ant-034* | Abundant Novel Transcript 034 | 1100 | 543 | | AA603960 | |
| 3 | Tc-EST-080 | *Tc-ant-034* | Abundant Novel Transcript 034 | 400 | 350 | 360 | AA618629 | AA618630 |
| 4 | Tc-EST-086 | *Tc-ant-034* | Abundant Novel Transcript 034 | 630 | 632 | 435 | AA820017 | AI 078887 |
| 5 | Tc-EST-108 | *Tc-ant-034* | Abundant Novel Transcript 034 | 1050 | 543 | 689 | AA824228 | AA824229 |
| 6 | Tc-EST-148 | *Tc-ant-034* | Abundant Novel Transcript 034 | 432 | 432 | 431 | AA836696 | AA836697 |
| 7 | Tc-EST-151 | *Tc-ant-034* | Abundant Novel Transcript 034 | 730 | 517 | | AA836699 | |
| 8 | Tc-EST-175 | *Tc-ant-034* | Abundant Novel Transcript 034 | >1000 | 538 | 664 | AA563519 | AA873916 |
| 9 | Tc-EST-192 | *Tc-ant-034* | Abundant Novel Transcript 034 | 379 | 379 | 374 | AA873898 | 873899 |
| 10 | Tc-EST-223 | *Tc-ant-034* | Abundant Novel Transcript 034 | 414 | 401 | 414 | AI 080787 | AI 080788 |
| 11 | Tc-EST-265 | *Tc-ant-034* | Abundant Novel Transcript 034 | 810 | 535 | 657 | AA915870 | AA915871 |
| 12 | Tc-EST-283 | *Tc-ant-034* | Abundant Novel Transcript 034 | 900 | 537 | 618 | AA875788 | AA875789 |
| 13 | Tc-EST-298 | *Tc-ant-034* | Abundant Novel Transcript 034 | 400 | 320 | 395 | AA979713 | AA979714 |
| 14 | Tc-EST-315 | *Tc-ant-034* | Abundant Novel Transcript 034 | 1760 | 639 | 520 | AA879354 | AA879355 |
| 15 | Tc-EST-319 | *Tc-ant-034* | Abundant Novel Transcript 034 | 1050 | 704 | 529 | AA879350 | AA879360 |

**Table 4.5.** Table showing EST clones belonging to the *ant-034* cluster. Line 14 is highlighted in blue to highlight the longest member of the *ant-034* cluster.

Table 4.6 below summarises the amino acid composition of each gene in terms of individual amino acids and the nature of their side groups (i.e. hydrophobic, polar, basic or acidic). Immediately, one striking difference that can be seen is between the *ant* genes and *not-018*. The proportion of neutral hydrophobic amino acids is almost exactly half that of the *ant* genes whereas the situation is almost perfectly reversed in the case of the neutral polar amino acids. There are also significantly fewer basic and acidic amino acids present in the *not-018* sequence. Between the *ants* these differences are more subtle. There is a significantly higher number of valines present in both *ant-003* and *-030* compared to *ant-005* and *-034*. The number of isoleucines in *ant-030* is again significantly less than is found in the remaining three ants. Such differences can also be seen in the other amino acid groups. Whereas such differences could also be found in randomly picked genes, this level of analysis is required in order to give another descriptive parameter to these novel sequences.

A common feature of each of the *ants* is the presence of amino acid repeats which appeared to have no discernible pattern. Although the proportions of the amino acids vary, as would be expected, the presence of single pairs of amino acids, some which are repeated as many as 7 times, may be a salient feature of these genes. These amino acid pairs, and in the case of *ant 030* triplets, are highlighted in blue in Figures 4.2 - 4.6 above. Also of interest was the fact that *ant-034* had a cysteine rich stretch of amino acids at its 3' terminal end. These residues are shown in yellow in Figure 4.5. One of *not 018s* key feature was the presence of double amino acid repeats, a feature shared with the *ants* although in this case cysteine was the predominant amino acid doublet. Also of interest was the presence of GYG and GS/T motifs which are shown in red in Figure 4.6.

The results of the plaque screening assay enabled us to tentatively categorise the transcripts in terms of their possibility of being secreted products. The cloning of *ant-005* and *not-018* into the pcDNA3.1 (-) (Invitrogen)DNA vaccine vector (Figure 4.7) enabled the further characterisation of the two genes. Both *ant-005* and *not-018* have not only been confirmed, through ELISA, to be TES products but have also been shown to be poorly represented in somatic tissue (TEX) (Figures 4.14 – 4.17)).

**Table 4.6**

| Nature of side-groups | amino acids | *ant 003* | *ant 005* | *ant 030* | *ant 034* | *not 018* |
|---|---|---|---|---|---|---|
| | | Percentage Composition Of Amino Acids | | | | |
| N/H | A | 8.75 | 6.36 | 7.08 | 6.35 | 6.06 |
| | V | 11.01 | 8.61 | 11.56 | 9.11 | 3.78 |
| | L | 9.60 | 9.73 | 7.27 | 6.56 | 3.40 |
| | I | 7.62 | 4.86 | 0.37 | 5.93 | 0.37 |
| | P | 1.97 | 2.99 | 2.98 | 3.17 | 2.65 |
| | W | 1.41 | 1.12 | 2.23 | 1.48 | 0.0 |
| | F | 4.80 | 3.74 | 4.85 | 3.17 | 1.51 |
| | M | 2.25 | 3.37 | 2.23 | 2.33 | 1.13 |
| Percentage of amino acids per group | | 47.41 | 40.78 | 38.57 | 38.10 | 18.90 |
| N/P | G | 7.34 | 8.23 | 6.34 | 6.14 | 16.67 |
| | S | 5.36 | 8.23 | 6.52 | 4.02 | 1.13 |
| | T | 5.36 | 3.37 | 4.10 | 6.56 | 8.71 |
| | Y | 5.93 | 3.37 | 2.98 | 1.69 | 14.77 |
| | C | 1.69 | 2.24 | 2.05 | 2.96 | 1.19 |
| | Q | 0.84 | 2.62 | 1.49 | 3.60 | 4.92 |
| | N | 4.23 | 4.49 | 2.05 | 3.38 | 9.09 |
| Percentage of amino acids per group | | 30.75 | 32.55 | 25.53 | 28.37 | 56.48 |
| B | K | 3.38 | 6.36 | 5.22 | 5.93 | 1.51 |
| | R | 5.36 | 5.24 | 6.34 | 7.20 | 2.27 |
| | H | 1.69 | 1.49 | 2.79 | 2.75 | 0.0 |
| Percentage of amino acids per group | | 10.43 | 13.09 | 14.32 | 15.88 | 3.78 |
| A | D | 4.51 | 3.74 | 4.47 | 6.35 | 0.75 |
| | E | 5.08 | 6.74 | 5.78 | 6.56 | 0.37 |
| Percentage of amino acids per group | | 9.59 | 10.48 | 10.25 | 12.91 | 1.12 |

**Table 4.6** Summary table of the percentage amino acid composition of the *abundant novel transcripts* and *novel transcript 018.* N/H: neutral hydrophobic amino acids; N/P: neutral polar amino acids; B: basic amino acids and A: acidic amino acids.

Each of the genes obtained so far have lacked a start methionine and a signal peptide sequence. This together with the insert sizes, suggests that the full length transcripts have not yet been found. As such, there may be more salient features that may be present in the missing 5' sequences of the genes; so too may any potential regions of homology.

### 4.2.3 *Production of antisera to* ant *proteins using pcDNA 3.1(-)*

Inserts for each of the *ants* and *not-018* were generated from a PCR reaction using the pcDNA 3.1(-) primers (Figures 4.3 and 4.5) on plasmid template for *ant-005* (Tc 228) and *not-018* (Tc 031). The PCR products obtained were run on a 1% agarose, excised and cleaned of contaminants using the Qiaquick Gel Extraction kit (Qiagen). The cleaned products were then subcloned into the pGEM-T (Promega) vector and transformed into JM109 competant cells (Promega). The transformed cells were plated out onto ampicillin plates and grown overnight at $37^0$C. Positive colonies were screened for using the specific primers *ant 5* pcDNA3.1 (-) R and *not 18* pcDNA3.1 (-) with the T7 vector primer. Plasmid template was prepared from the pGEM-T construct and the insert excised from the vector using the restriction enzymes *Apa 1* and *Eco RI* (Promega) for *ant-005* and *not-018* respectively. The excised inserts were then ligated into the pcDNA3.1 (-) ( Invitrogen) vector under the control of the T7 promoter using the *Apa 1* and *EcoRI* restriction enzyme sites (see Figure 4.7) for *ant -005* and *not - 018*. Ligation reactions were carried out at $16^0$C overnight. The ligation reactions were then transformed into JM109 competent cells in the presence of super optimal catabolite (S.O.C.) (Gibco BRL) medium and plated out onto LB-amp plates and incubated at $37^0$C overnight. Positve colonies were screened for by PCR using the T7 vector primer and gene specific primers (see Figure 4.8 ). Positive colonies were grown o/n in 10mls LB-amp media at $37^0$C and plasmids subsequently prepared and digested with the *Apa I* and *Eco RI* restriction enzymes to confirm the inserts (Figure 4.9). Upon confirmation of inserts, 1L o/n cultures were prepared and plasmid subsequently prepared following a modification of the Qiagen Plasmid Mega kit (Qiagen) protocol. Briefly, the protocol was adhered to except that plasmid were resuspended in 0.9% sodium chloride solution (Sigma) as opposed to the resuspension buffer provided. This was done so as not to introduce any contaminants into the immunogen. The purity was ascertained by taking optical density readings between $OD_{260}$ and $OD_{280}$.

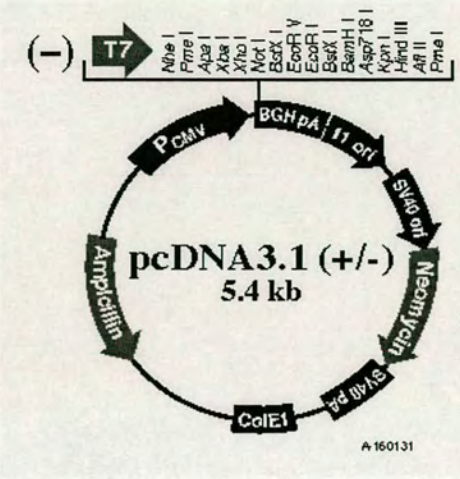



**Figure 4.7.** pcDNA (-) plasmid (Invitrogen) schematic showing the salient features of the vector. Of particular note is the multiple cloning site (mcs) the *Apa 1* and *EcoRI* sites of which were used to clone *ant-005* and *not-018*.

Note: The figure was reproduced from the Invitrogen web catalogue (www.invitrogen.com/vectors.html)

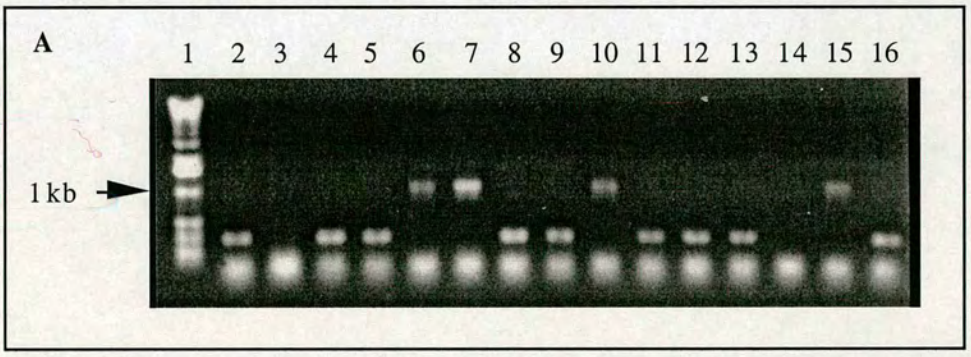



**Figure 4.8 (a).** 1% agarose gel showing results of a PCR colony screen of pcDNA constructs. Lane 1: 1kb ladder; Lanes 2-16 *ant-005* colony screen. Lanes 6, 7, 10 and 15 are positive for the *ant-005* pcDNA construct in the correct orientation.

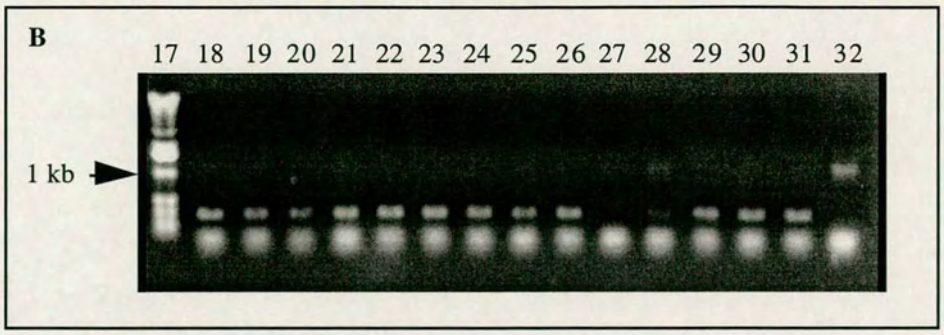



**Figure 4.8 (b)** 1% agarose gel showing results of a PCR colony screen of pcDNA constructs (continued). Lane 17: 1 kb ladder; lane 18: *ant-005*; lanes 19-32: *not-018*. Lanes 28 and 32 are positive for *not-018*. Arrow indicates the 1kb band.

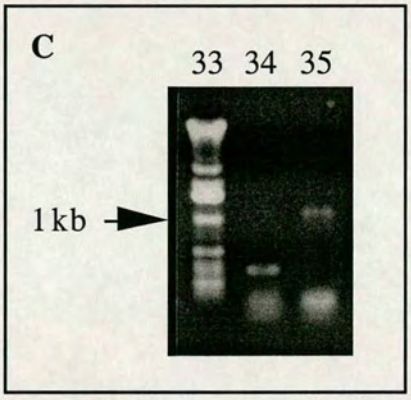



**Figure 4.8.(c)** 1% agarose gel showing results of a PCR colony screen of pcDNA constructs. Lane 33: molecular weight marker; Lanes 34 & 35: *not* 018. 35 is positive for the *not-018* pcDNA construct. Arrow indicates the 1kb band.

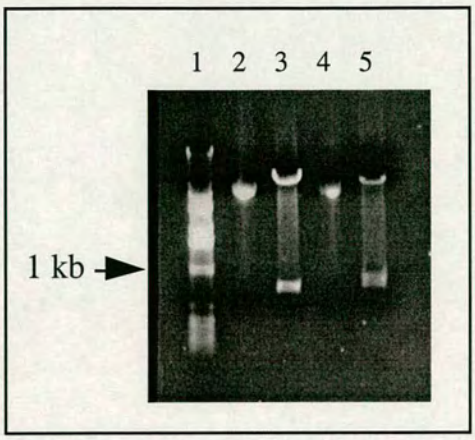



**Figure 4.9.** 1% agarose gel showing restriction enzyme digest of pcDNA constructs of *ant-005* and *not-018*. Lane 1: 1 kd ladder; lane 2: undigested pcDNA/*ant-005* plasmid construct; lane 3: pcDNA/*ant-005* plasmid digested with *Apa 1*; lane 4 undigested pcDNA/*not-018* plasmid construct; lane 5: pcDNA/*not-018* plasmid digested with *EcoRI*.

4.2.4 *Immunoscreening of transcripts*

The procedure was as outlined in the "Screening Protocols: Plaque Screening" of the picoBlue™ Immunoscreening kit instruction manual (Stratagene). Essentially, clones composing the *ants* and *nots* were dotted onto a gridded LB-amp plate on which a bacterial lawn composed of LB-top agar and XL1-Blue *E. coli* cells had previously been grown. It was determined that a phage dilution of $10^{-1}$ pfu/ml was optimal for this assay. The plates were then incubated at $42^0$C for approximately 4 hours, after which time plaques were clearly visible. During the incubation period 88 mm nitrocellulose filters were incubated in IPTG diluted to 10mM in $H_2O$, and then allowed to air dry. The nitrocellulose discs were marked with a soft pencil to aid orientation and identification of the discs. The discs were placed onto each plate and incubated at $37^0$C for 3.5 hours. Each plate was labelled appropriately in order to correspond with its respective disc. Following the incubation the membranes were removed and washed in Tris-Buffered-saline-Tween-20 (TBST), then blocked in 5% skimmed milk for 1 hour at room temperature. Membranes were then transferred to 5% skimmed milk blocking solution in which had been diluted the α-TES polyclonal antibody (1/100 dilution). The serum had been raised by immunising BALB/c mice with 15 μg via the subcutaneous route with TES. Incubation was for 1 hour at room temperature followed by washing in TBST. The membranes were incubated in the secondary antibody (Biorad: EIA affinity purified goat anti-mouse IgG horseradish peroxidase conjugate) diluted in 5% skimmed milk and incubated for 1 hour. Following this period of incubation, the blots were washed then developed using chemiluminescent Western blot detection system (ECL; Amersham Pharmacia Biotech)

4.2.5 *Immunisation of mice with pcDNA3.1 (-) constructs*

Immunisation of mice followed a modification of methods outlined in (Whalen, 1998). Essentially, mice were anaesthetised using 50μl of a Vetalar (Pharmacia & Upjohn) : Rompun (Bayer plc) mixture in the ratio of 2:1 respectively. Once anaethsised Immac (Reckitt and Colman), a hair removing agent, was used to remove the fur from both calf muscles of the animal. Briefly, using a 27GX$^{3/4}$" needle with a polyethylene tubing coat, cut such that the bevel of the needle enters the calf muscle and no further, 100μg of each pcDNA3.1(-) DNA vaccine construct was injected into each calf muscle. pcDNA3.1 (-) plasmid with no insert (100μg/calf muscle) and β-galactosidase

(50μg/calf muscle) were also used as immunogens to function as positive and negative controls. Both BALB/c and CBA/Ca mice strains were used as strain variations in antibody production are common (Maizels pers. comm.). Following immunisation the mice were left for 3 months without boosting prior to being bled.

### 4.2.6 *Enzyme Linked Immunosorbent Assay (ELISA)*

An ELISA was performed using microtitre plates coated with *T. c.* excretory/secretory (TES) antigen and somatic extract (TEX). The procedure used was as defined in [Maizels, 1991 #329]. Briefly, the TES and TEX were diluted in sodium carbonate buffer (0.06M, pH9.6) to 1μg/ml and added to Nunc-Immuno™ Plate MaxiSorp™ surface plates (Nunc) at 50μl/well. The plates were then incubated o/n at $4^0$C. Following incubation the plates were washed three times with PBS-T. Diluted pcDNA (in carbonate buffer) derived antibody (1/100 and 1/500) was aliquoted at 100μl/well and incubated at $37^0$C for 1 hour; the plates were then washed as before and 100μl/well of conjugate diluted in carbonate buffer was added and incubated at $37^0$C for 1 hour. Following this incubation period, the plates were extensively washed and before addition of 50μl of the substrate ABTS fluorogenic reagent (KPL). The absorbance values of the plates were read at 450nm.

## 4.3 RESULTS

### 4.3.1 *Sequence analysis of clones*

The identification of all the clones that made up each cluster was made possible by the sequencing strategy employed, which was to sequence in both the 5' and 3' directions using two pairs of vector primers (M13 R/F and T3/T7) (Chapter 2). Figure 4.10 shows a schematic alignment of the clones that comprised each *ant* cluster. As each of the clones were truncated, no start methionine was obtained, so the numbers indicated in the figure refer to the start of the cDNA and not the real start of the gene. The longest member of each cluster can be seen in Table 4.1 - 4.5. *Tc*-EST-239 is shown in bold as this clone appeared to be a variant member of the *ant-003* cluster in that sequence differences were apparent that were not attributable to sequencing error (Table 4.1).

The sequence analysis of each member of the cluster was performed as outlined in Chapter 2 in order to determine putative open reading frames.

As the clusters were novel no definitive sequence description could be obtained via database searching. The exception to this was *ant-034* (see Table 4.6). Both the BLASTX and BLASTP database search algorithms gave probability values of $1.9 \times 10^{-11}$ and $3.4 \times 10^{-12}$, which was above the threshold of $10^{-6}$ which was determined as significant for the EST project. Figure 4.11 shows an amino acid pileup of *ant-034* and the Japanese encephalitis virus protease. The level of identity between the two sequences was 23%. The *no*vel *t*ranscript-*018*, also had no significant homology to anything in the existing databases (see Table 4.6). The ProDom protein domain homology algorithm was also carried out. Using this method the sequence(s) submitted is used to search an automatically compiled database of protein motifs derived from SWISSPROT entries. Three of the transcripts, ant-003, -030 and not-018 gave no result using this method. The *ants 005* and *034* both yielded results, which are summarised by the ProDom alignments in Figure 4.12 (a) and (b). The fact that the results are not the same as those obtained for BLAST searching could be attributable to the fact that sequences are deposited into the SWISSPROT database at a slower rate than the NCBI databases. Entries into SWISSPROT are also compacted such that entries are non-redundant. The result obtained for *ant-005* suggested that the transcript had an $8 \times 10^{-11}$ probability of containing domain characteristics similar to a 30 member protein family of which the protein envelope genome protein is affiliated

suggested that the transcript had a $7 \times 10^{-15}$ probability of bearing homology to a 280 member protein family of which the protein helicase ATP-binding RNA-binding protein was a member (Figure 4.12 (b)).

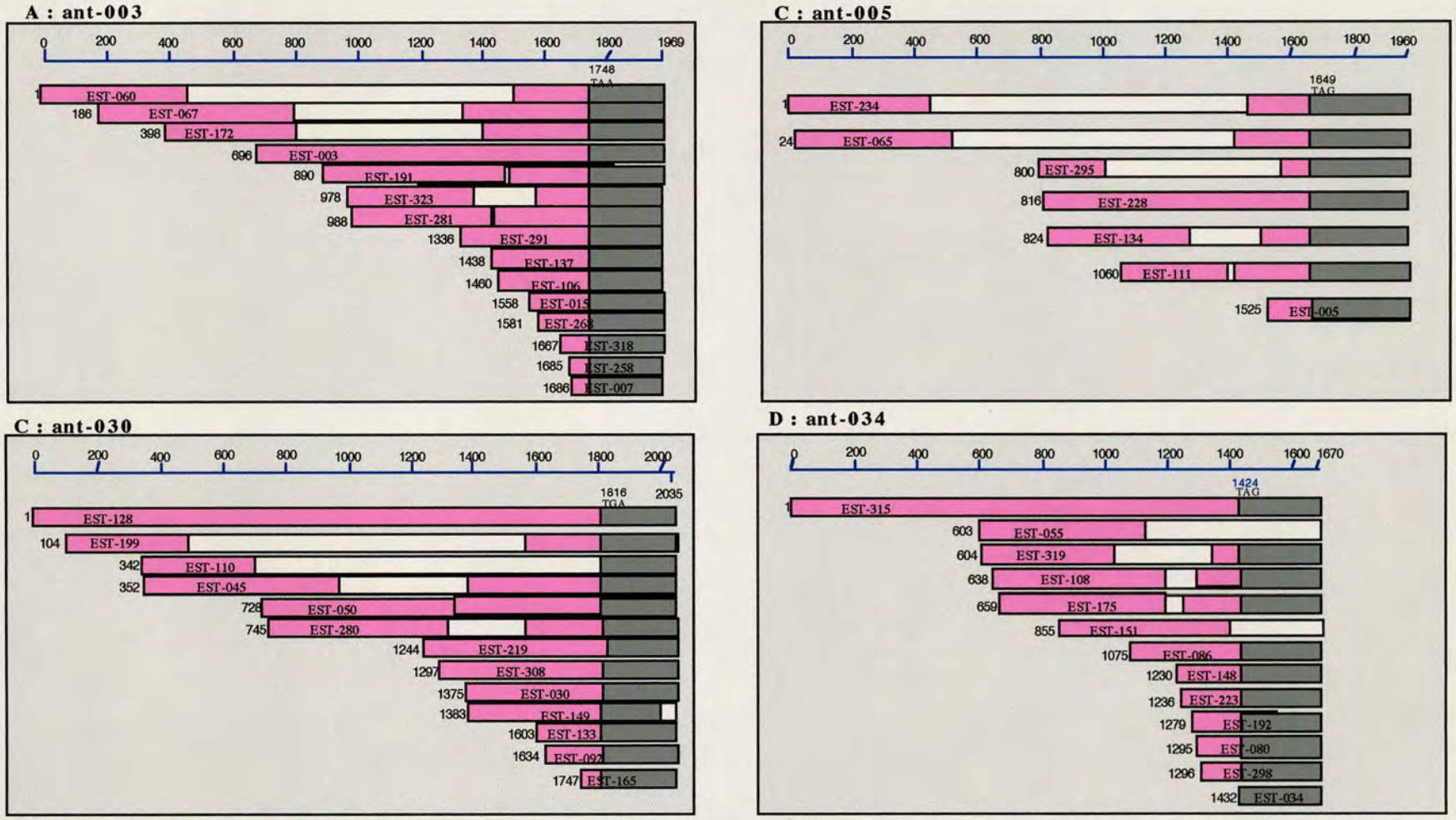



**Figure 4.19** Schematic of the *ant* clusters highlighting the importance of including 3' reads in the generation of ESTs. In particular the *ant-005* gene cluster would have been impossible to properly characterise without the inclusion of 3' reads. The number(s) in blue refers to the length of sequence to the stop codon. Unshaded segments refer to regions not yet sequenced; dark segments indicate 3'UTR. The pink sections refer to sequenced regions.

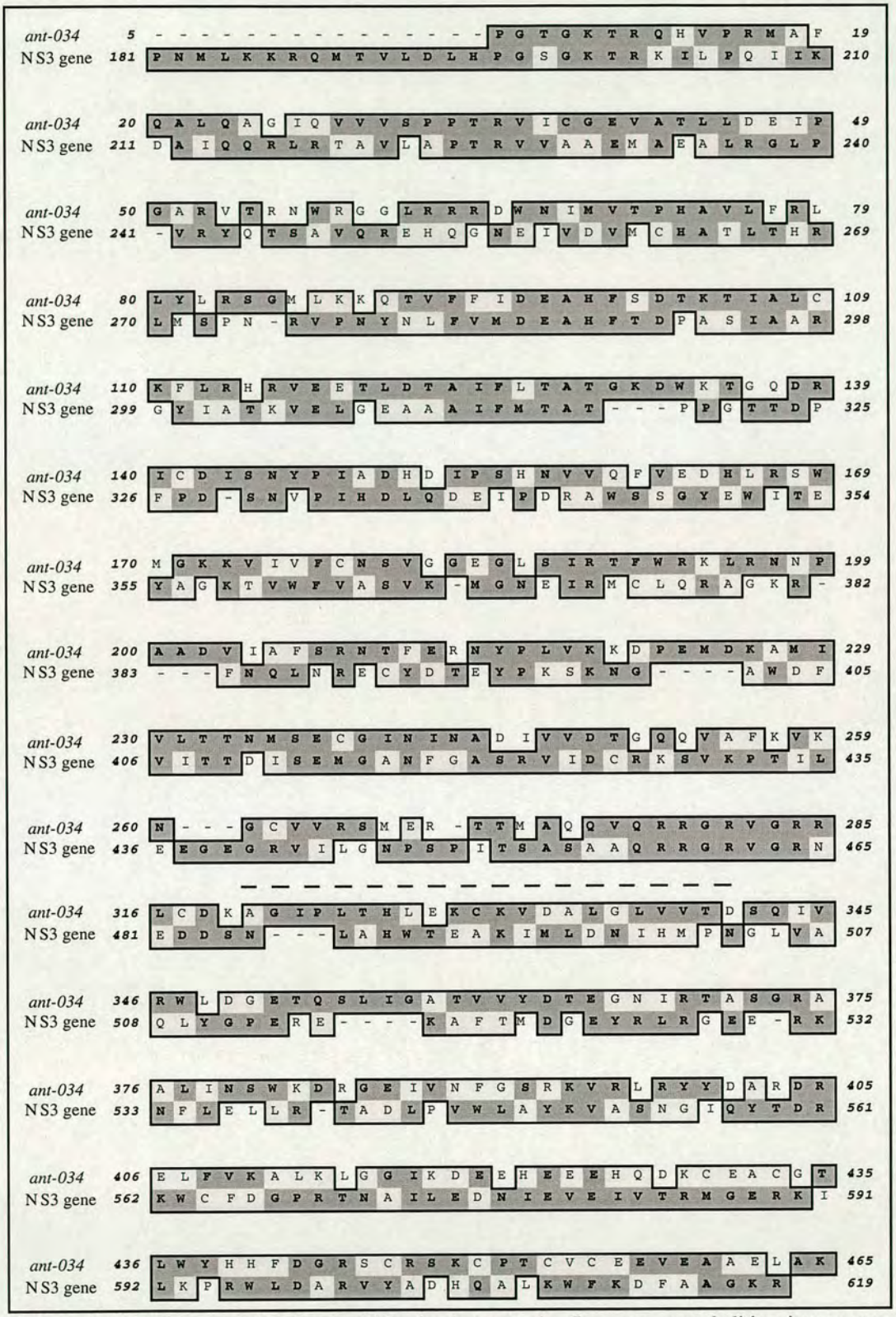



**Figure.4.11** Pileup showing level of identity between the Japanese encephalitis virus protease (Y16245) (NS3 gene) and *ant-034*. Identities are shown by dark shaded boxes, similarities are indicated in the lighter grey. The dashed line marks the break in the sequence alignment.

```
Number of sequences in family: 30
Most frequent protein names:    POLG(30)
Commentary (automatic):         PROTEIN ENVELOPE GENOME
              POLYPROTEIN CONTAINS: CAPSID C CORE MATRIX M
             Length = 1427
   Score = 156  (65.2 bits), Expect = 8e-11
  Identities = 45/137 (32%), Positives = 69/137 (49%), Gaps = 4/137 (2%)

 Query:     18 MMVSGDDSVVILNKEESKAFSKAFWFNNETGKIRKDVAENLPSVVEEDIENVSFCS-NFY 76
               M+VSGDD VV    K      F KA +F N+  K+RKDV E  PS+   + E V FCS +F+
 Sbjct:   3171 MLVSGDDCVV---KPIDDRFGKALYFLNDMAKVRKDVGEWEPSMGFTEWEEVPFCSHHFH 3227

 Query:     77 VSTKFGGSVRMMPIRSVEEVLSKASLMLGYARDFTTSEAWARAQGFMMVCCYPHVAEIKL 136
                    G   ++P R  +E++ +A +  G       +   ++A G M +  Y H  +++
 Sbjct:   3228 ELVMKDGRSLIVPCRDQDELVGRARVSPGCGWSVRETACLSKAYGQMWLLNYFHRRDLRT 3287

 Query:    137 LGLALLIATRANIVLEG 153
              LG A+  A   + V  G
 Sbjct:   3288 LGFAICSAVPVSWVPMG 3304
```

**Figure 4.12 (a).** ProDom alignment showing regions and levels of homology exhibited by *ant-005* with sequence for the protein envelope genome. A 30 member protein domain family.

```
Number of sequences in family: 280
Most frequent protein names:    POLG(44) SECA(20) UVRB(15)
Commentary (automatic):         PROTEIN
             HELICASE ATP-BINDING RNA-BINDING RNA ATP-DEPENDENT
             NUCLEAR FACTOR ENVELOPE PUTATIVE
            Length = 284
   Score = 193  (79.6 bits), Expect = 7e-15
  Identities = 79/256 (30%), Positives = 121/256 (46%), Gaps = 27/256 (10%)

 Query:      5 PGTGKTRQHVPRMAFQALQAGIQVVVSPPTRVICGEVATLLDEIPGARVTRNWRGGLRRR 64
             PG GKTR+ +P++  +A+   ++  V  PTRV+  E+A  L  +P   +        R
 Sbjct:   1701 PGAGKTRRILPQIIKEAINRRLRTAVLAPTRVVAAEMAEALRGLP---IRYQTSAVAREH 1757

 Query:     65 DWN--IMVTPHAVLFRLLYLRSGMLKKQTVFFIDEAHFSDTKTIALCKFLRHRVEETLDT 122
              + N  + V  HA L   L +    +    +F +DEAHF+D  +IA   ++  RVE
 Sbjct:   1758 NGNEIVDVMCHATLTHRL-MSPHRVPNYNLFVMDEAHFTDPASIAARGYISTRVELGEAA 1816

 Query:    123 AIFLTATGKDWKTGQDRICDISNYPIAD--HDIPSHNVVQFVEDHLRSWMGKKV-IVFCN 179
              AIF+TAT      G        SN PI+D   +IP         E  +  ++GK V  V
 Sbjct:   1817 AIFMTAT----PPGTSDPFPESNAPISDLQTEIPDRAWNSGYE-WITEYIGKTVWFVPSV 1871

 Query:    180 SVGGE-GLSIRTFWRKLRNNPAADVIAFSRNTFERNYPLVKKDPEMDKAMIVLTTNMSEC 238
               +G E  L ++   +K        VI  +R ++E  YP  K D + D    V+TT++SE
 Sbjct:   1872 KMGNEIALCLQRAGKK--------VIQLNRKSYETEYPKCKND-DWD---FVVTTDISEM 1919

 Query:    239 GININADIVVDTGQQV 254
              G N  A  V+D+ + V
 Sbjct:   1920 GANFKASRVIDSRKSV 1935
```

**Figure 4.12 (b).** ProDom alignment showing regions and levels of homology exhibited by *ant-034* with sequence for the protein helicase ATP-binding RNA-binding protein. A 280 member protein domain family.

Although conclusive database homology could not be obtained for these transcripts an interesting characteristic was shown in the 3' segment of the genes. Figure 4.13(a) shows the 3' coding sequence of the transcripts highlighting the low level of sequence homology existing between the clusters. This is in contrast to the homology shown between the clusters at the 3' UTR (Figure 4.13 (b)).

| Transcript | Top Hit | Species/source | BLASTX | | Top Hit | species/source | BLASTP | |
|---|---|---|---|---|---|---|---|---|
| | | | Probability | Score | | | Probability | Score |
| *ant-003* | Lactate permease | *Streptococcus iniae* | 0.42 | 74 | As for BLASTX | As for BLASTX | 0.23 | 74 |
| *ant-005* | NS5 protein | Kyanasur forest disease virus | 1.1 x 10-09 | 83 | As for BLASTX | As for BLASTX | 4.7 x 10-10 | 83 |
| *ant-030* | Serine/threonine protein kinase | *Homo sapiens* | 0.63 | 54 | Hypothetical ABC Transporter ATP-binding protein MG467 homolog | *Mycoplasma pneumoniae* | 0.48 | 73 |
| *ant-034* | NS3 protease | Japanese encephalitis virus | 1.9 x 10-11 | 92 | As for BLASTX | As for BLASTX | 3.4 x 10-12 | 92 |
| *not-018* | Keratin, glycine/tyrosine-rich of hair | *Oryctolagus cuniculus* | 2.5 x 10-06 | 67 | As for BLASTX | As for BLASTX | 2.9 x 10-06 | 67 |

**Table 4.7**. Summary of the "top hits" obtained through BLASTX and BLASTP database searching. The scores and probabilities of the each database hit together with the description and source of each search result.

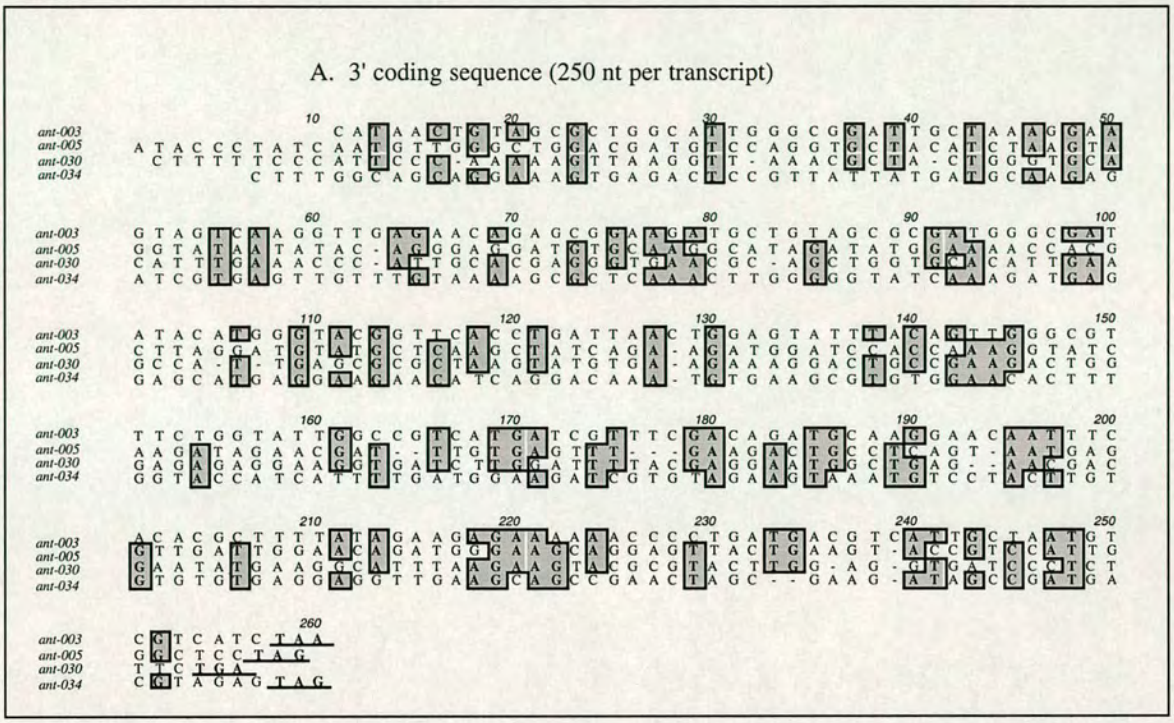



**Figure 4.13.(a).** Pileup of the 3' coding sequences of the four *ant* gene clusters. Identities are shown by the grey shaded boxes, with similarities shown in the lighter grey boxes.

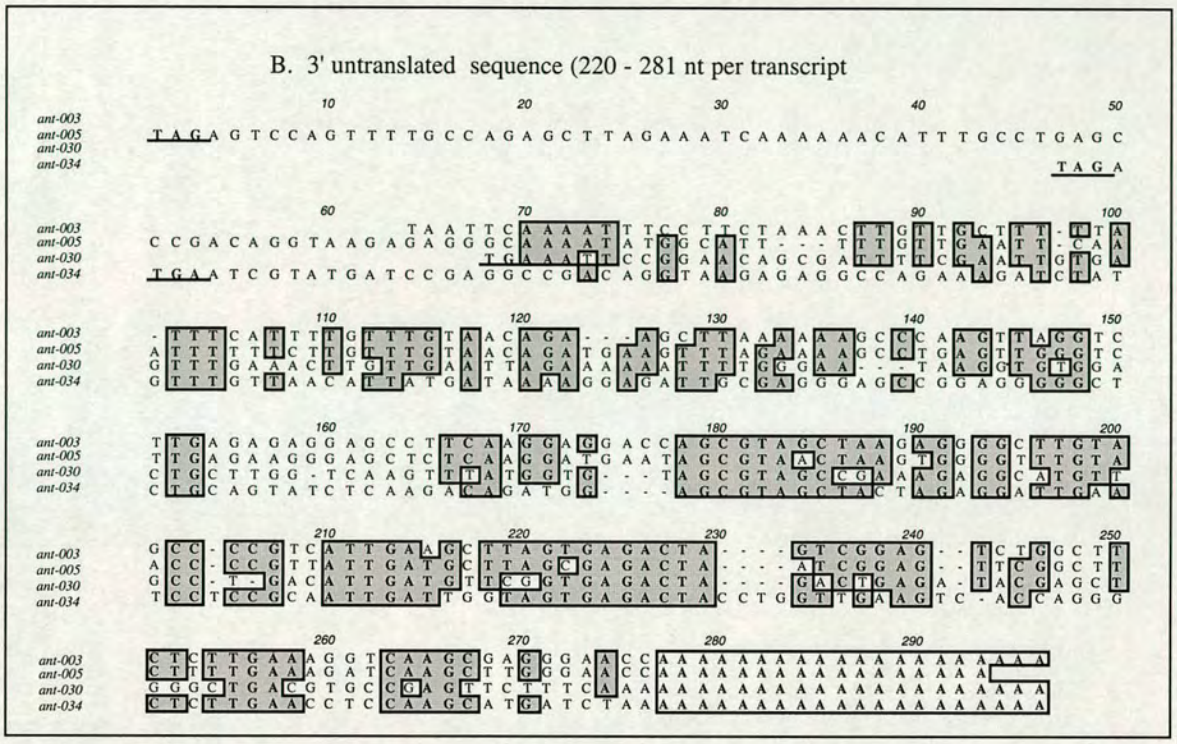



**Figure 4.13.b**. Pileup of the 3' untranslated region (UTR) of the *ant*s showing the high level of sequence conservation shared by the four genes in this region. Identities are shown by the grey shaded boxes, with similarities shown in the lighter grey boxes.

4.3.2 *Antibody Screening of Phage*

The results from this experiment are tabulated in Table 4.8 below. The positive controls used to ascertain positive antibody recognition were two variants of the abundant major TES product TES 32 (Clones 36 (variant: see Chapter 3) and 66). The negative controls in the experiment were those clones that were in the incorrect reading frame within the Bluescript vector and as a result incapable of protein production. These are indicated in the table shown overleaf.

| ANT Cluster | Clone Numbers | Correct Reading Frame? | α-TES |
|---|---|---|---|
| 003 | 106 | NO | - |
| | 137 | YES | ++ |
| | 239 | YES | ++ |
| | 258 | NO | - |
| | 268 | YES | ++ |
| | 281 | YES | + |
| | 318 | YES | + |
| | 323 | YES | ++ |
| *CONTROL* | 36 (TES 32v) | YES | +++ |
| | 66 (TES 32) | YES | +++ |
| 005 | 005 | YES | ++ |
| | 065 | NO | - |
| | 077 | NO | - |
| | 134 | YES | +++ |
| | 111 | YES | ++ |
| | 234 | YES | ++ |
| | 275 | YES | + |
| *CONTROL* | 36 (TES 32v) | YES | +++ |
| | 66 (TES 32) | YES | +++ |
| 030 | 030 | NO | - |
| | 045 | NO | - |
| | 050 | NO | - |
| | 092 | NO | - |
| | 128 | NO | - |
| | 133 | YES | ++ |
| | 149 | NO | - |
| | 165 | NO | - |
| | 219 | NO | - |
| | 231 | YES | ++ |
| | 280 | NO | - |
| | 308 | YES | ++ |
| *CONTROL* | 36 (TES 32v) | YES | ++ |
| | 66 (TES 32) | YES | ++ |
| 034 | 034 | NO | - |
| | 055 | NO | - |
| | 080 | NO | - |
| | 108 | NO | - |
| | 148 | NO | - |
| | 192 | NO | - |
| | 294 | NO | - |
| | 299 | NO | - |
| | 315 | NO | - |
| | 319 | YES | + |
| *CONTROL* | 36 (TES 32v) | YES | ++ |
| | 66 (TES 32) | YES | ++ |
| 018 | 018 | YES | + |
| 031 | 031 | YES | ++ |
| 136 | 136 | YES | + |
| *CONTROL* | 36 (TES 32v) | YES | + |
| | 66 (TES 32) | YES | + |

**Table 4.8.** Summarised data of plaque screening assay. The reactivity of the anti-TES antibody is indicated by + or - where the strength of signal is signified by + = weak signal, ++ = moderate signal, or +++ = strong signal. - denotes no signal.

*ELISA*

Figures 4.14 and 4.15 shows the responses obtained from TEX coated plates probed with the pooled sera from CBA/Ca or Balb/c mice respectively, each immunised with the two pcDNA constructs. The results show a poor antibody response for *ant-005* and *not-018* which are both barely above the background control (pcDNA alone). The results obtained from the analysis of the TES coated plates, shown in Figures 4.16 -.4.17, proved to be very different. Figure 4.16 shows that an antigen specific response was obtained from the CBA/Ca immunised with not-018 pcDNA. This was not the case for the α-*ant-005* antibody, the response of which was barely above background. The antibody response was almost completely reversed in the BALB/c derived antibody, in that it was the α-*ant-005* antibody that gave the antigen specific response to TES with the response from α-*not- 018* scarcely above background.

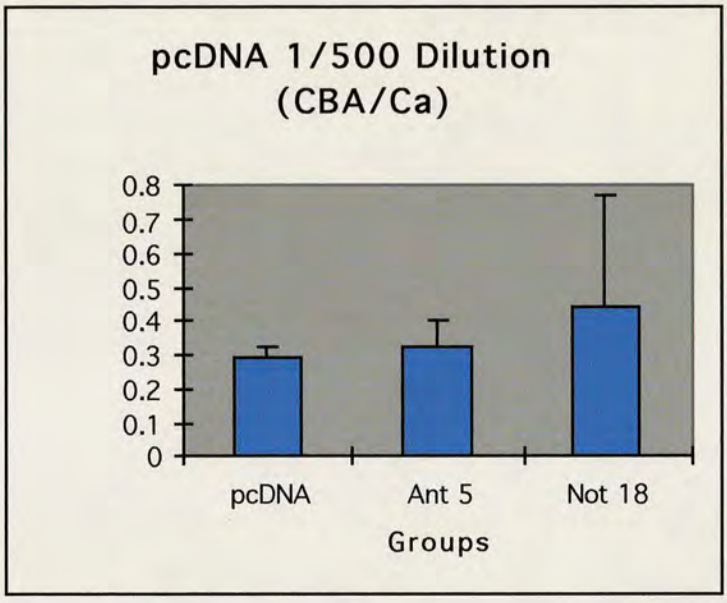



**Figure 4.14** Graph showing response to *Toxocara canis* somatic extract (TEX) coated ELISA plates probed by pcDNA3.1 (-) (Invitrogen) induced antibody to *ant 005* and *not 018*. The antibodies were raised in the CBA/Ca strain of mice. Sera were diluted 1:500

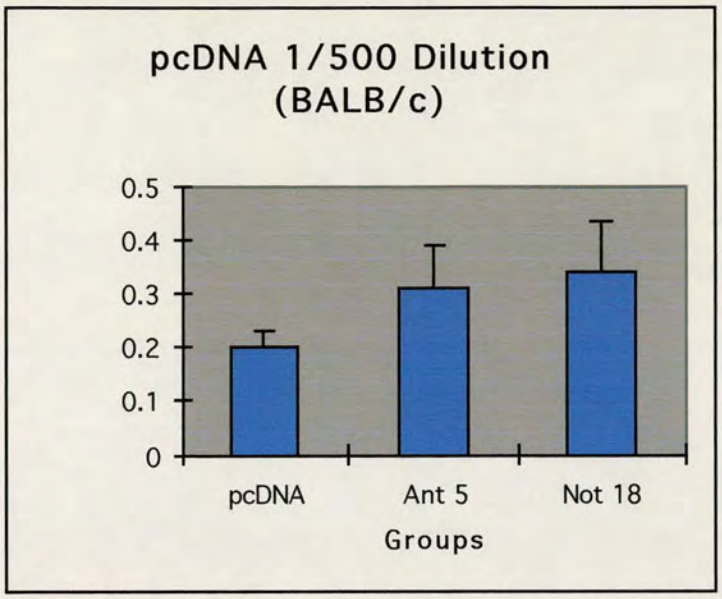



**Figure 4.15.** Graph showing response to *Toxocara canis* somatic extract (TEX) coated ELISA plates probed by pcDNA3.1 (-) (Invitrogen) induced antibody to *ant 005* and *not 018*. The antibodies were raised in the Balb/c strain of mice. Sera were diluted 1:500

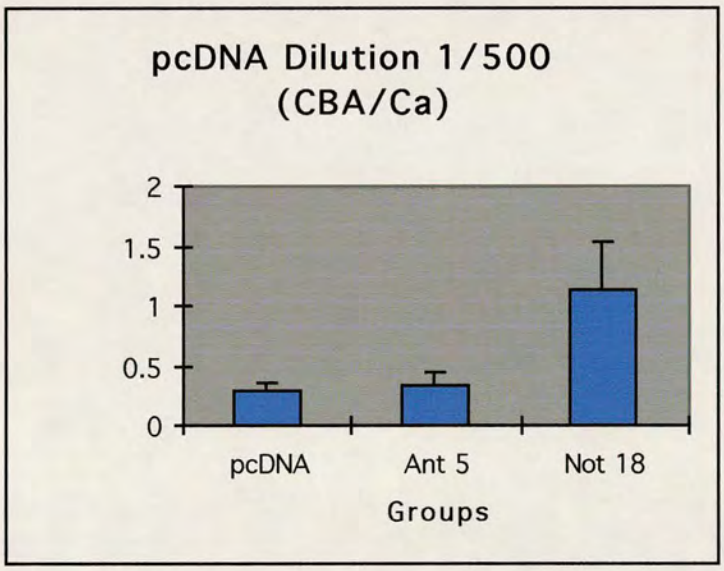



**Figure 4.16.** Graph showing response to *Toxocara canis* secretory/excretory antigen (TES) coated ELISA plates probed by pcDNA3.1 (-) (Invitrogen) induced antibody to *ant 005* and *not 018*. The antibodies were raised in the CBA/Ca strain of mice. Sera were diluted 1:500.

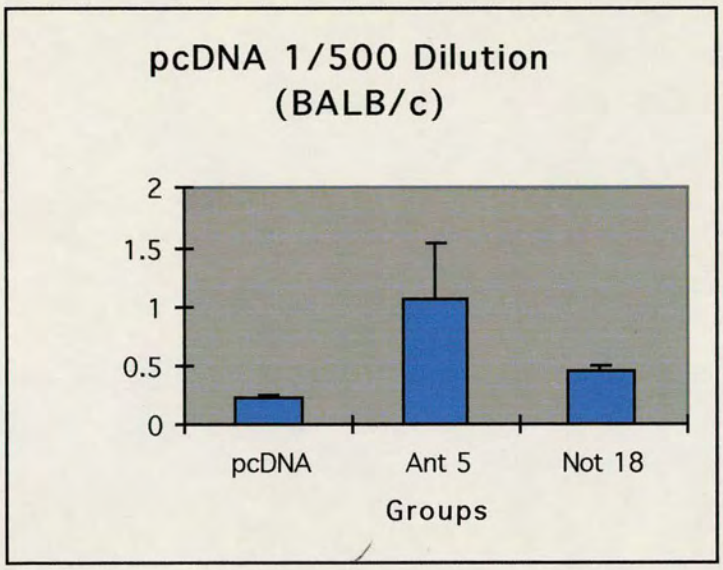



**Figure 4.17.** Graph showing response to *Toxocara canis* excretory/secretory antigen (TES) coated ELISA plates probed by pcDNA3.1 (-) (Invitrogen) induced antibody to *ant 005* and *not 018*. The antibodies were raised in the Balb/c strain of mice. Sera were diluted 1:500

## 4.4 DISCUSSION

The amino acid analysis of the transcripts yielded little definative information in terms of putative function or structure of the transcripts (see Table 4.8). From the table the composition of the transcripts indicated that on a comparative basis *ant*-003 was the most hydrophobic and *not*-018 the most basic.

Homology searching of the public databases again yielded little useful information. Table 4.7 summarises the results of these searches. Although significant search results were obtained for *ant-005* and *not-018* these proved to be false positives. This is best demonstrated by Figures 4.12 (a) and (b). The amino acid pileup in Figure 4.11 seems to suggest a sequence similarity to the Japanese encephalitis virus protease (accession number Y16245).

Interestingly a high degree of sequence conservation was shown at the 3' UTR for the four ants (Figure 4.13(b)). This level of homology was not exhibited at the 3' coding end of the genes (Figure 4.13 (a)). The high level of sequence conservation shown by the transcripts could be attributable to the fact that they may share a common ancestry. Perhaps, although of different putative function, the transcripts share similar gene regulatory elements. At present this is mearly speculation.

The results of the plaque screening assay provided preliminary evidence for the fact that the transcripts encoded putative TES products. Table 4.8 summarises the results of this assay. This data suggested the *ant-034* was not a TES product whereas the remaining *ants* and *not* 018 were. Support for this data came from the successful cloning of *ant-005* and *not-018* into the pcDNA (Invitrogen) DNA vaccine vector. Antibody obtained from immunisation of mice with the constructs gave strong specific antibody responses in ELISA's against TES coated plates (Figures 4.16 and 4.17). The data obtained from the assays involving TEX coated plates proved less conclusive (Figure 4.14 and 4.15).
To summarise, sequence data alone proved insufficient in providing definitive information about the nature of these genes or of their products. From the partial sequences that have been obtained to date it is known that these molecules have varying degrees of *N*-linked glycosylation. The genes of *ant 003, 005* and *030* all have 2 putative *N*-linked glycosylation sites, although *ant 034* has just 1 putative glycosylation site. The gene for

*not 018* has 4 putative *N*-linked glycosylation sites. But no real significance can be placed on the putative levels of glycosylation as full-length sequence is yet to be obtained.

Work is currenly in progress to obtain full-length sequence for all five transcripts. Following this the next phase is to clone the remaining transcripts into the pcDNA vaccine vectors and initiate characterisation of the respective gene products.

## Chapter 5 NC6/SXC rich proteins from *Toxocara canis*

## 5.1 Introduction

SXC (six cysteine), or NC6 (nematode cysteine 6) domains, as they were formerly known, refer to an array of six conservatively spaced cysteines ($CXD_{4-6}C_{4-8}C_{6-12}CX_3CX_2C$), the function of which has not yet been determined. The domain is a short structure of between 36-42 amino acids, which together with the conserved cysteine has a number of other conserved residues (Blaxter, 1998). The first definition of this structure came from a report by Gems *et al*, (1996) in which this novel domain was found in both a lipid binding protein (phosphatidylethanolamine binding protein) (Gems et al., 1995) and a mucin-like protein (Gems and Maizels, 1996) associated with the surface coat of *T. canis*. At present 75 genes have been identified from *C. elegans* that contain 184 SXC motifs (Blaxter, 1998). Of these are included tyrosinases (developmentally regulated enzymes implicated in the synthesis of nematode cuticles) (Gerrits, unpublished). SXC domains have also been found in a number of a other nematode genera such as *Ascaris, Trichuris muris* and *Necator americanus* (Blaxter, 1998).

SXC domains have also been identified in cnidarians *Anemonia sulcata* (Schweitz et al., 1995), *Stichodactyla helianthus* (Tudor et al., 1996), *Heteractis magnifica* (Gendeh et al., 1997) and *Bunodosoma granulifera* (Dauplais et al., 1997). In these species, these proteins function as toxins which specifically block potassium ion channels. These domains have also been found in the hydra, *Podocoryne carnea* (Pan et al., 1998), where they are thought to play a role in tissue remodelling or cell migration and also digestion in the case of the adults (Pan et al., 1998). A mammalian homologue was found in microfibril-associated glycoprotein factor (MAGP), a protein involved in the biology of fibrillin-containing proteins and associated with the aetiology of inherited connective tissue disease, has also been isolated from a human cDNA (Faraco et al., 1995). Bovine (Gibson et al., 1989) and murine (Chen et al., 1993) homologues of MAGP-1 have also been found.

This chapter deals with the identification of a number of proteins containing SXC domains isolated as part of the EST project described in Chapter 3. The common factor shared by these genes is the presence of a cysteine rich domain of unknown function termed NC6/SXC described in the text above.

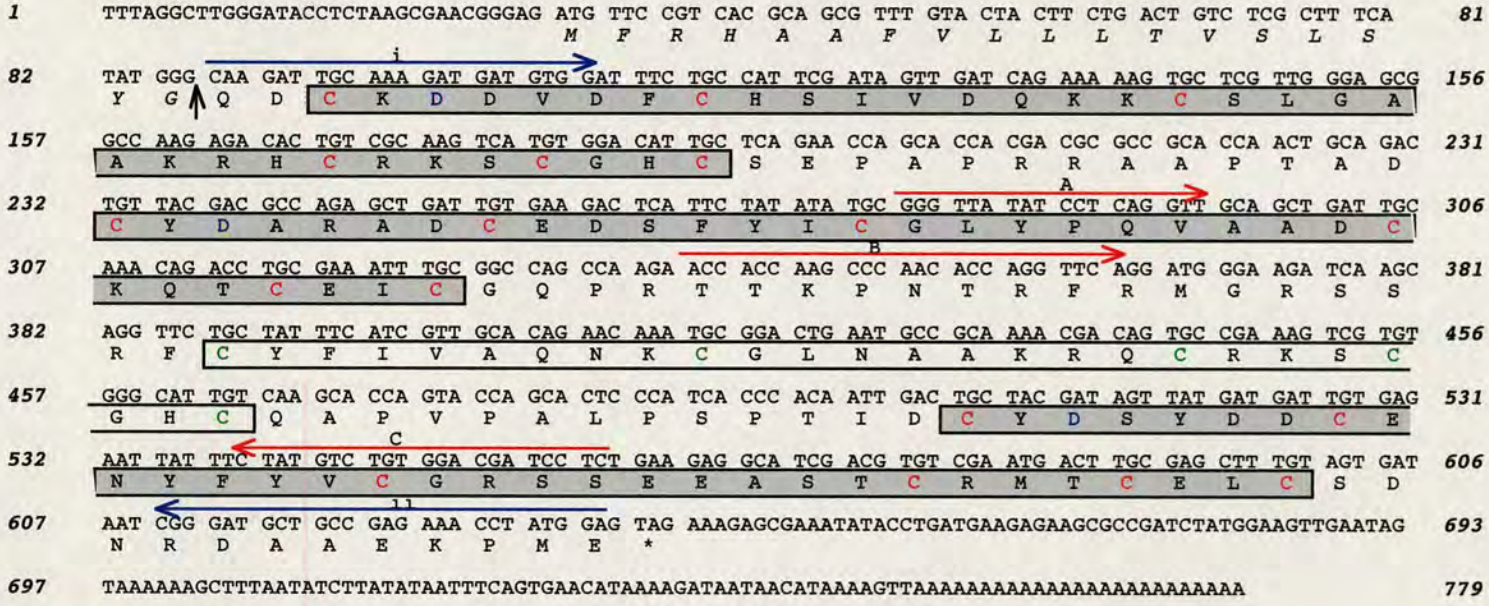
## 5.2 MATERIALS AND METHODS

### 5.2.1 *Identification of clones*

*Tc-muc-1, -2, -3, -4* and *Tc-huf-1* were identified following procedures as outlined in Chapter 2. *Tc-muc-5* was identified by expression screening of the *T. canis* cDNA library with a polyclonal antiserum raised to TES products (Loukas *et al* ms. in preparation).

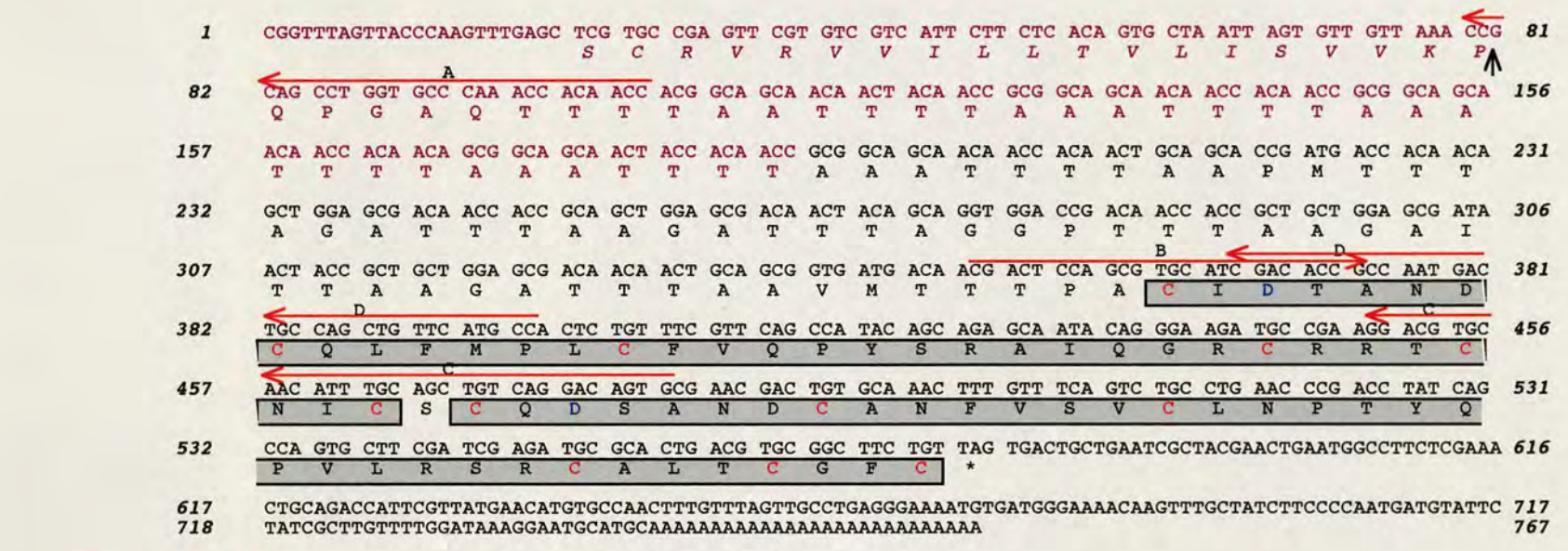
### 5.2.2 *Sequencing and sequence analysis*

Sequencing and sequence analysis were performed as outlined in Chapter 2 and as discussed in Chapter 3. Figure 5.1 shows the primer map for the SXC-rich protein *Tc-huf-1*. Figures 5.2 and 5.3 show the primer maps for *Tc-muc-2* and *-muc-4* respectively; sequence information for *Tc-muc-1* was previously obtained by Dr. D. Gems and Prof. R. M. Maizels (Gems and Maizels, 1996). Further sequencing of *Tc-muc-3* and sequence information for *Tc-muc-5* were obtained by Dr. A. Loukas et al (ms. in preparation).



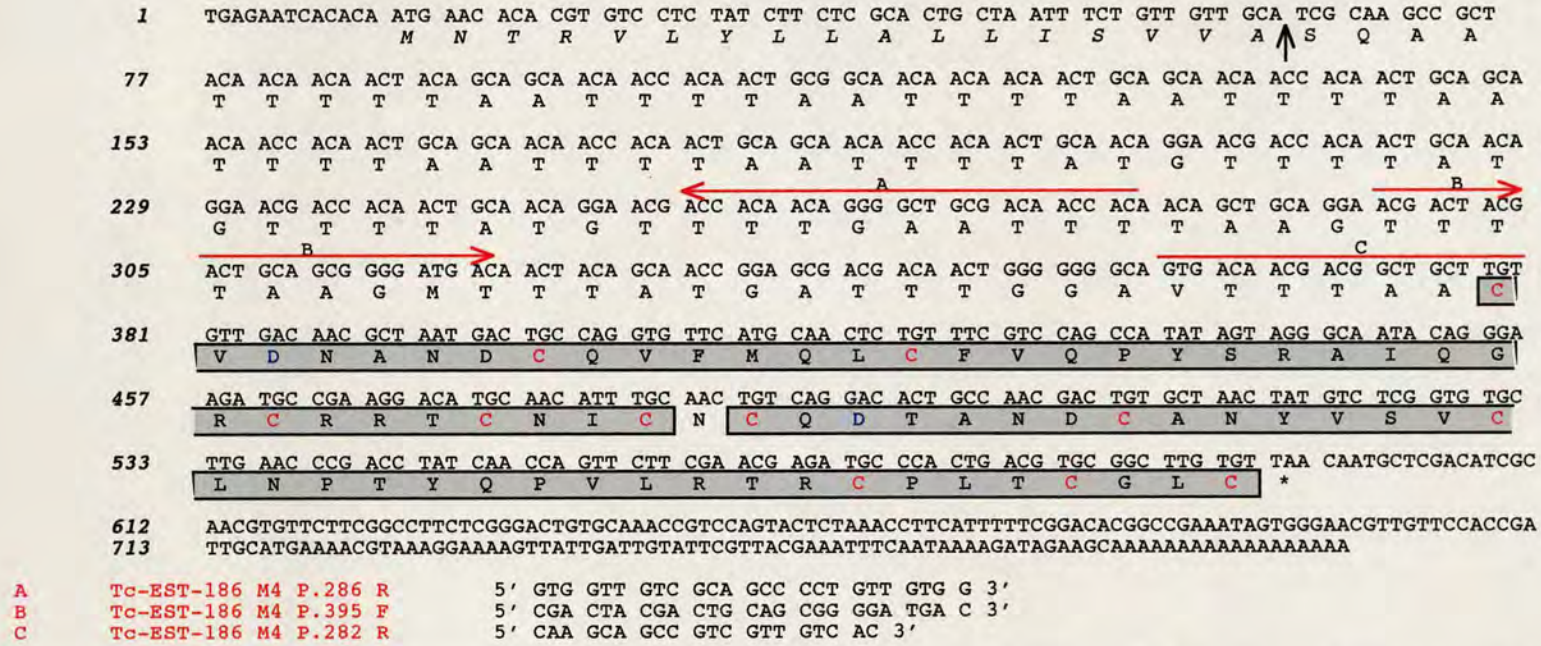

| | | |
|---|---|---|
| i | TclpETF1 | 5′ CCA AGA TTG CAA AGA TGA TGT GG 3′ |
| ii | TclpETR1 | 5′ CCT CCA TAG GTT TCT CGG CAG CAT CCC G 3′ |
| A | TC-EST-001 F | 5′ GGG TTA TAT CCT CAG GTT 3′ |
| B | TC 1 F | 5′ CCA AGC CCA ACA CCA GGT TCA 3′ |
| C | Tc-HUF-1R | 5′ GAG GAT CGT CCA CAG ACA TAG 3′ |

**Figure 5.1** Nucleotide and deduced amino acid sequence of an NC6/SXC rich protein of unknown function termed *Tc-huf-1*. Primer design and direction are indicated by the horizontal arrows. pET primers (see section entitled " Bacterial expression of Tc-huf-1") are shown in blue and internal primers in red. The putative signal peptide is indicated in italics with the cleavage point shown by the black arrow. The SXC domains are shown by the grey shaded boxes with the conserved cysteines in red and conserved aspartic acid residues in blue. The cysteines forming the partial SXC domain or FC 1 (Five cysteine 1) domain are highlighted in green.



| | | |
|---|---|---|
| A | Tcmuc2R | 5′ GGT TGT GGT TTG GGC ACC AGG CTG CGG 3′ |
| B | Tcmuc2F | 5′ CGA CTC CAG GGT GCA TCG ACA CCG 3′ |
| C | Tc-EST-076 P.263 R | 5′ CAC TGT CCT GAC AGC TGC AAA TGT TGC ACG TCC 3′ |
| D | Tc-EST-241 M2 239 R | 5′ GGC ATG AAC AGC TGG CAG TCA TTG GCG GTG TCG 3′ |

**Figure 5.2.** Nucleotide and deduced amino acid sequence of *Tc-muc-2* showing primer design/direction. The signal peptide is shown in italics with the putative cleavage point indicated by the black arrow. The two SXC domains are shown by the grey shaded boxes The conserved cysteines are highlighted in red and the conserved aspartic acids are shown in blue. The original EST is shown in black text with sequence generated from 5' PCR shown in dark magenta




| | | |
|---|---|---|
| A | Tc-EST-186 M4 P.286 R | 5′ GTG GTT GTC GCA GCC CCT GTT GTG G 3′ |
| B | Tc-EST-186 M4 P.395 F | 5′ CGA CTA CGA CTG CAG CGG GGA TGA C 3′ |
| C | Tc-EST-186 M4 P.282 R | 5′ CAA GCA GCC GTC GTT GTC AC 3′ |

**Figure 5.3.** Nucleotide and deduced amino acid sequence of *Tc-muc-4* showing direction and position of primers used in obtaining full-length sequence of the transcript. The signal peptide is hown in italics with the putative cleavage site indicated by the black arrow. The two SXC domains of *Tc-muc-4* are indicated by the grey shaded boxes. The conserved cysteine residues are shown in red and the conserved aspartic acid residues are shown in blue.

5.2.3 *Bacterial expression of* Tc-huf-1

The *Tc-huf*-1 insert was amplified from phagemid template using the pET primers Tc1 F1 and Tc1 pET R1. The resulting amplicon was ligated into the pET 29-T (Novagen) vector according to the manufacturer's instructions. The ligated product was transformed into competent XL1-Blue *E. coli* cells which lack the T7 RNA polymerase gene, and grown overnight on LB-kanamycin (LB-kan) plates at 37°C. A single colony of cells carrying the plasmid insert was used to inoculate 10mls of LB-kan broth which was then grown in a 37°C shaking incubator for 16-18 hours. Plasmid was purified from the bacterial culture using the Qiagen Spin miniprep (Qiagen) kit as outlined in the manual. The purified plasmid was diluted 1: 50 and transformed into *E. coli* strain BL21 (Novagen) carrying the DE3 episome, which express the T7 RNA polymerase gene. Expression of the protein from the BL21 DE3 *E. coli* was as outlined in the Novagen manual. Briefly, a single colony of the transformed BL21 DE3 expression cells was used to inoculate 1 litre of LB-kanamycin broth and allowed to grow at 37°C for approximately 3.5 hrs or until an OD of 0.5 was reached. Following this period, a 1ml aliquot was removed and centrifuged, before addition of isopropyl β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.4 mM. Further samples were collected 1 and 2 hours post induction and the bacteria pelleted at 13,000 rpm. The resultant pellets were resuspended in SDS PAGE sample buffer and analysed on a 15% SDS PAGE minigel.

Cells grown in a 1L culture fluid were pelleted in a Sorval centrifuge at 9000 rpm, 4°C for 10 mins. The pellet was subsequently resuspended in 5 ml of binding buffer, and sonicated on ice using an MSE Soniprep 150 sonicator (Sanyo Gallenkamp) at an amplitude of 7 microns. Recombinant Tc-HUF-1 (rHUF-1) protein was only soluble in the presence of 6M urea, so subsequent purifications all included urea as the denaturant. Purification of the recombinant protein was achieved by using a metal chelating resin (His Bind Resin; Novagen) under denaturing conditions according to manufacturers instructions. A Bradford protein determination assay was performed on 4 chelating column-purified fractions of 0.5 ml using the Coomassie Plus Protein Assay Reagent (Pierce) on Nunc-immuno plates (Nunc). The concentration of each fraction was read using an Anthos Reader 2001 with the Delta Soft II ELISA analysis program. Protein concentrations were estimated at 2.0 mg/ml, 1.7 mg/ml, 2.2 mg/ml and 1.9 mg/ml for

fractions 1-4 respectively. The concentration of the recombinant protein post dialysis was 1.3 mg/ml.

Fractions containing protein were pooled and dialysed over three days in a stepwise manner using decreasing molarities of urea in PBS. This was done to reduce the likelihood of the recombinant protein precipitating; this only proved partially successful.

### 5.2.4 *Immunisation of mice with* Tc-HUF-1

Four BALB/c and four CBA/Ca mice were immunised with 20$\mu$g in 100$\mu$l with 100 $\mu$l of Freund's complete adjuvant rHUF-1. The mixture was emulsified by sonicating on ice for a total of 3 mins. The primary immunisation was followed by 2 boosts at days 28 and 42 with 20 $\mu$g of rHUF-1 in incomplete Freund's adjuvant, delivered via the subcutaneous route. Sera were collected at day 54. Both strains of mice were used as strain related differences in the level of antibody production with different recombinant proteins are common (R. Maizels pers. comm.). Blood collected from the BALB/c and CBA/Ca mice were left at 4$^{0}$C overnight. The samples were centrifuged at 13,000 rpm for 20 mins, the sera were then transferred to fresh tubes and stored at -20$^{0}$C.

### 5.2.5 *Western blots with* $\alpha$-Tc-HUF-1 antibody

rHUF-1 (1.0 $\mu$g), rMUC-1 (1.0 $\mu$g), TES (2.5 $\mu$g) and TEX (5.0 $\mu$g) were electrophoresed in quadruplicate lanes on a 5 - 25 % SDS PAGE gradient gel and proteins transferred onto nitrocellulose membrane at 45 mA for 1 hour using a Novablot electrophoretic transfer apparatus (LKB Bromma). The nitrocellulose membrane was then treated as for a Western blot (see Chapter 2: Materials and Methods). Once transfer was complete the nitrocellulose membrane was cut into four panels, each panel containing all four antigen preparations, and probed with normal mouse sera (NMS) or sera from mice immunised with either rHUF-1, rMUC-1 (Chapter 4; negative control) or Tcn 2 (control monoclonal antibody that recognises carbohydrate epitopes on the majority of TES molecules; (Maizels et al., 1987) - see Chapter 1); all sera were diluted 1 : 5,000 in PBS Tween (PBS-T), followed by incubation with horseradish peroxidase-conjugated rabbit

anti-mouse IgG (diluted 1:2,000; BioRad). A second experiment was carried out with extracts prepared by sequentially sonicating *T. canis* larvae in PBS then 0.5% Empigen (non-ionic detergent) and finally 1% SDS. Extracts were electrophoresed, transferred and probed with α-rHUF-1 and the monoclonal antibody Tcn 1 (which shows strong recognition for a 70 kDa protein in TES (Maizels et al., 1987) - see Chapter 1). Blots were developed using the ECL chemiluminescent substrate (Amersham).

### 5.2.6 *5' PCR of* Tc-muc-2

The original EST clones relating to *Tc-muc-2* (Tc-EST-076, -130, and -241) were all truncated at the 5' end. In order to obtain the 5' end of the gene, a PCR reaction was carried out on the library using an internal primer. Tc-EST-076M2-P.263 R (Figure 5.4) a 5' specific primer was used in conjunction with the universal primers M13 R & F and also T3 & T7; and also the spliced leader 1(SL1) and oligo dT primers. As *Tc-muc-1* and *-muc-3* both had Spliced Leader 1 the SL1 primer was used in this instance because it was thought likely that the other mucin members would also be SL1 transpliced. Bands obtained from Tc-EST-076M2-.P.263 R/SL and Tc-EST-076M2-.P.263 R /T3 were excised from the gel and cleaned of agarose and ethidium bromide using the Qiaquick Gel Extraction Kit (Qiagen). The cleaned PCR products were then ligated into the pGEM-T vector (Promega) as outlined in the pGEM-T manual. The ligated products were transformed into JM109 competent cells (Promega) and plated out onto LB-amp with Xgal and IPTG to screen for blue white colonies. Incubation of the plates was carried out overnight at $37^0$C. The following day the white colonies obtained (white denoting recombinant) were screened by PCR for the presence of inserts using Tc-EST-076M2-P.263/T7. Positive clones were grown in 10 mls of LB-amp broth and plasmids prepared using the Qiagen Spin Miniprep Kit (Qiagen), which were subsequently used in dye terminator sequencing reactions (Perkin Elmer) (see Chapter 2 and 3).

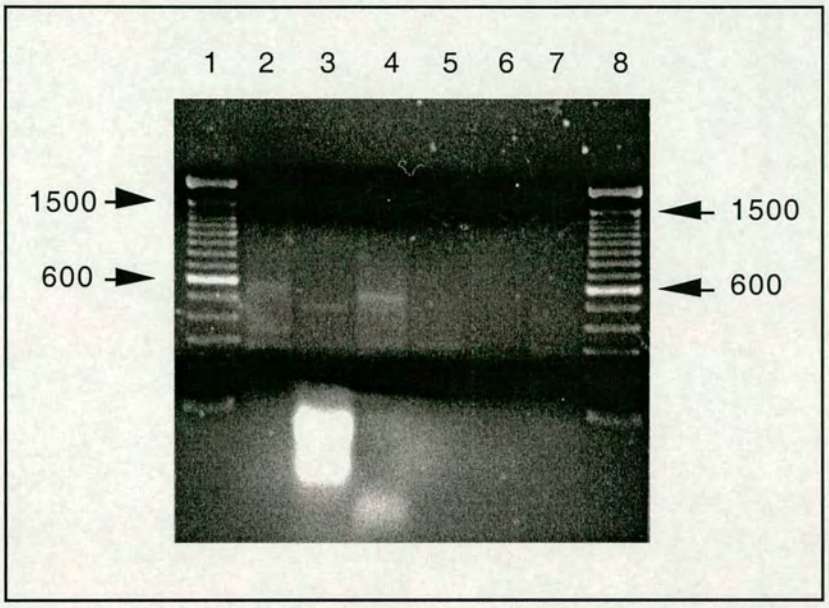



**Figure 5.4.** Results of 5' library PCR using the specific primer Tc-EST-076M2-P.236 R in combination with /M13 R (vector primer), /Spliced leader 1 (SL1), /T3 (vector primer), /M13F (vector primer), oligo dT and /T7 (vector primer). The markers used were the 100 Kb DNA ladder.

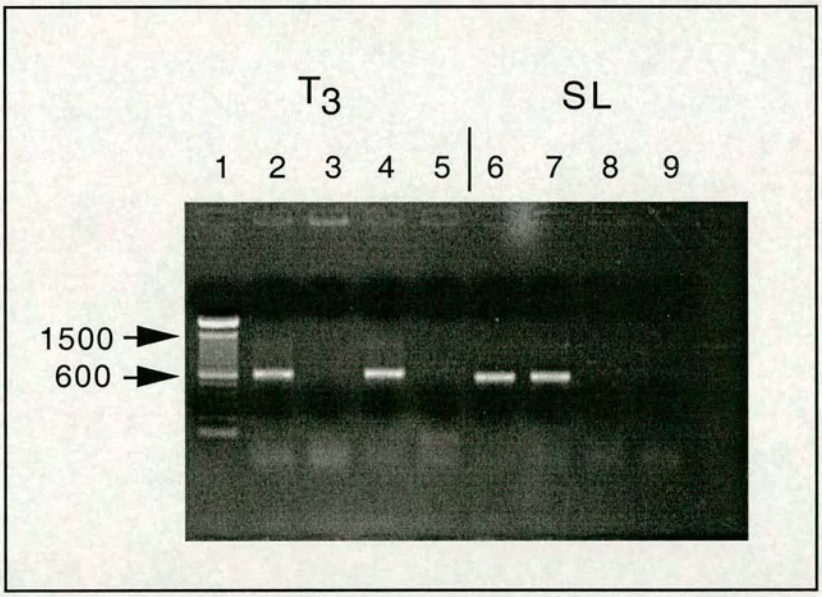



**Figure 5.5.** Results of PCR colony screen using the specific primer Tc-EST-076M2-P.263 in combination with the T3 (vector primer) and SL primers. The markers used were the 100 Kb DNA ladder.

## 5.3 Results and Discussion

### 5.3.1 Tc-huf-1*: an SXC rich protein of unknown function*

The clone *Tc-huf-1* was represented in the EST project (Chapter 3; Tetteh *et al.*, 1999) by two clones, Tc-EST-001 and Tc-EST-058. *Tc-huf*-1 contains a cDNA insert of 779 base pairs including a poly A+ tail. The deduced full-length protein contains 202 amino acids including an 18-aa potential signal peptide with a putative cleavage site after residue 18 (figure 5.1). The gene encodes a protein containing 3 SXC domains and 1 partial SXC (FC1) domain with a gap of 20 aa separating SXC domain 2 from FC1 (Figure 5.1 and 5.6). A 13 amino acid gap is also present between SXC domain 1 and domain 2, and also between FC domain 1 and SXC domain 3 (Figure 5.1 and 5.6). The partial SXC domain termed here FC1 (Five cysteine 1) domain bears the consensus sequence $CX_8CX_8CX_3CX_2C$ and was present in the sequences of both Tc-EST-001 and -058. The protein does not possess any putative *N*-linked glycosylation sites, although the possibility remains that *O*-linked glycosylation might occur. The predicted size of the protein was 22, 437 Da, although the recombinant protein migrated at 43,000 Da on SDS PAGE including the 5 kDa added on due to the pET tag. Figure 5.7 shows an SDS PAGE gel with a duplicate analysis of the bacterial expressed *Tc*-HUF-1 (rHUF-1) recombinant protein. Figure 5.8 shows the purified fractions of the His column purified rHUF-1 as shown by SDS PAGE analysis.

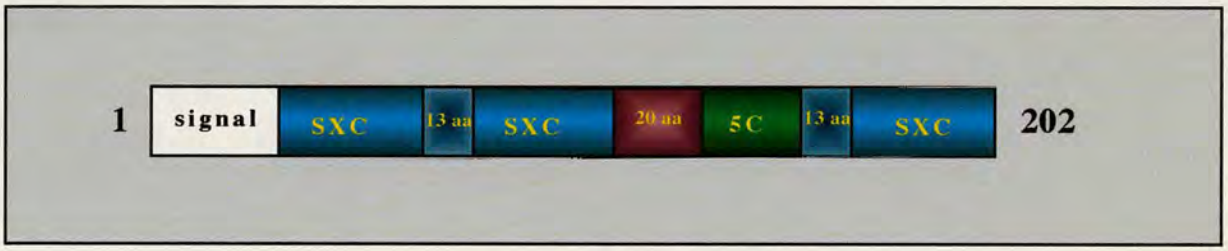



**Figure 5.6.** Schematic representation of *Tc*-HUF-1 showing the SXC domain arrangements. The putative signal sequence is shown by the open box. The partial SXC domain is designated 5C and is shown in green, while the SXC domains are in blue.

The closest homologues of *Tc-huf-1* are a predicted protein from *C. elegans* (accession no. Z81529) and *Tc-peb-1* (phophatidylethanolamine binding protein-TES 26) from *Toxocara canis* (Gems et al 1995).

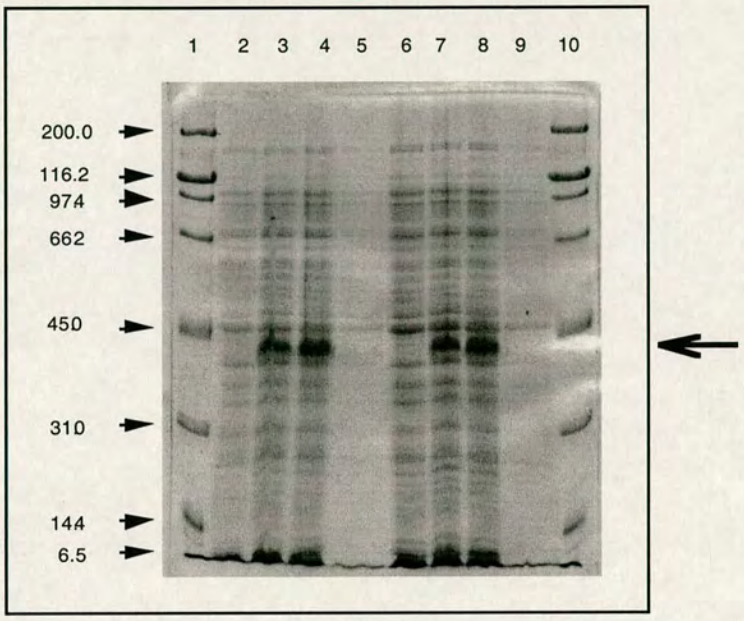



**Figure 5.7**. Fifteen percent polyacrylamide gel showing expression of recombinant *Tc*-HUF-1 (rHUF-1) by *E. coli* BL21 DE3 expression cells over a 3 hour time course (in duplicate). Lanes 1 & 10: molecular weight markers; Lanes 2 & 6: uninduced BL21 DE3 cells, pre-induction; Lanes 3 & 7: induced BL21 DE3 cells, 1hr; Lanes 4 & 8: induced BL21 DE3 cells, 2hrs; Lanes 5 & 9: supernatant from sonicated pellet. The arrow indicates the position of rHUF-1

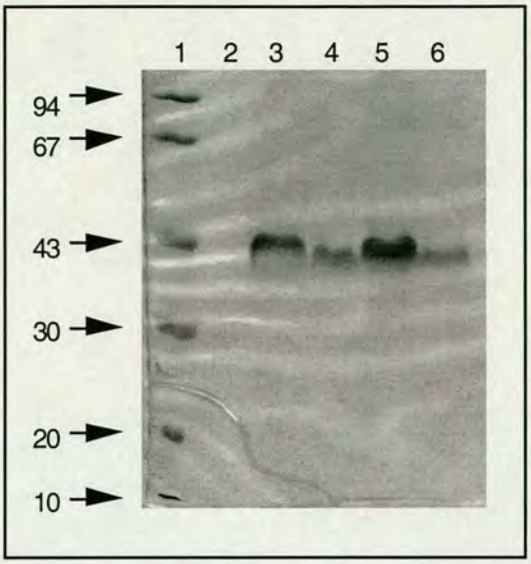



**Figure 5.8.** Fifteen percent polyacrylamide gel showing rHUF-1 purified using nickel-chelate resin. Lane 1: molecular weight markers; Lane 2: blank; Lanes 3-6: His column purified fractions 1-4
Note: Fraction numbers refer to the 0.5ml aliquots obtained during the His column purification.

Antibodies raised to the purified recombinant, rHUF-1, were used to TES and TEX products of *T. canis* by Western blot. The results of the blots showed the antibody recognised two bands (Figure 5.9), a band of approximately 66 kDa in both TES and TEX, but also a higher band at 70 kDa in TES alone. Neither of these bands were recognised by the anti-carbohydrate monoclonal antibody Tcn 2, anti-MUC-3 antibody, or by the normal mouse serum (NMS) control (Figure 5.9). Additional reaction was seen by the rHUF-1 antibody at approximately 32 kDa in TES alone (Figure 5.9). Full-length sequence of TES 32, a C-type lectin and the most abundant protein in TES product, has been obtained (Loukas *et al* ms. submitted) and is shown not to contain SXC domains. There is no obvious explanation for the binding seen, excepting perhaps that recognition is a result of the abundance of this protein product. A second blot was run in which sequential detergent extracted parasite antigen was probed with rHUF-1 and the monoclonal antibody Tcn 1 (shown to be specific for TES 70; (Page et al., 1992a)) (Figure 5.9). The monoclonal antibody Tcn 1 showed strong recognition of a band at approximately 70 kDa in both the PBS and SDS extracted TEX and to a lesser extent the Empigen® extracted TEX; banding patterns could also be seen at 66 kDa in the PBS and

SDS fractionated antigen only. In contrast, the α-HUF-1 antibody showed recognition of a 70 kDa band in the PBS and SDS extracted antigen only, and a second band at 66 kDa in all three lanes. Thus it would appear that the MabTcn 1 and anti-HUF-1 are reactive to different antigens.

Although the signal of recognition appeared specific for 66-70 kDa targets, the predicted size of the protein is around 22 kDa. The size discrepancy between the predicted protein and the reactive bands is not due to *N*-linked glycosylation but could be due to extensive *O*-glycosylation. It is impossible to predict from the primary sequence exactly where *O*-linked glycosylation sites might reside as consensus sequences have yet to be established (Gooley et al., 1994). However, very heavily glycosylated proteins tend to migrate very diffusely on SDS PAGE gels, whereas the proteins recognised by anti-HUF-1 resolve sharply (Fig. 5.9 and 5.10). A number of other reasons may explain the size discrepancy. Firstly, there is the possibility of oligomerisation such as through tyrosine cross-linking. As no definitive function(s) have been assigned to SXC domains, the possibility exists that they allow the formation of stable non-reversible structures (dimers, multimers etc). In this case such interactions would have to resist boiling in a reducing agent (mercaptoethanol, dithiothreitol) in the preparation of samples for SDS PAGE. The second possibility is that cross-reactivity might be occurring between the SXC domains of *Tc-huf-1* and other SXC containing proteins. So far the only TES proteins found to contain SXC domains are the mucins (*Tc-muc-1* to *muc-5*), and TES 26 (PE binding protein) (Gems et al., 1995). It could be that the SXC domain(s) of a hitherto unidentified SXC containing protein with an electrophoretic mobility of approximately 70kDa (perhaps the as yet uncharacterised TES 70) is proving to be immunologically cross-reactive with the α-rHUF-1 serum.

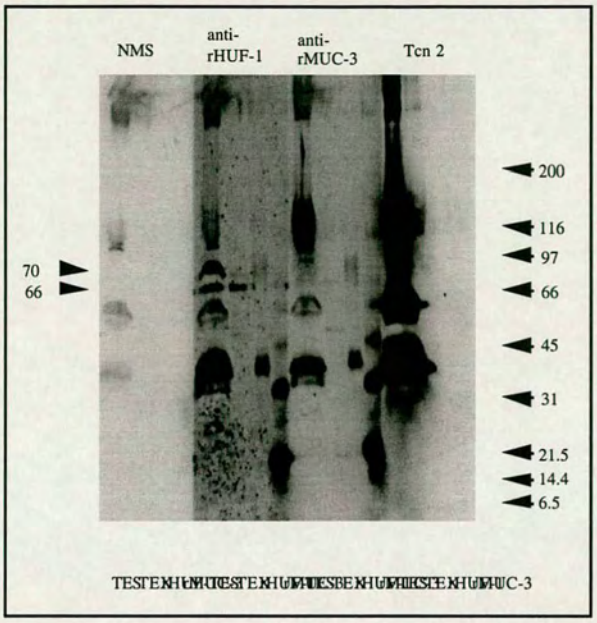



**Figure 5.9.** Western blot showing four panels, each with TES, TEX, rHUF-1 and rMUC-3 antigen*. The panels were probed from left to right with NMS (normal mouse serum), α-rHUF-1, α-rMUC-3 and Tcn 2 (monoclonal antibody, see Chapter 1). The bands recognised by α-rHUF-1 are indicated by the arrows on the left side of the figure at 70 and 66 kDa. Molecular weight markers are indicated on the right

Note: * Recombinant protein and antiserum provided by Dr. A. Loukas.

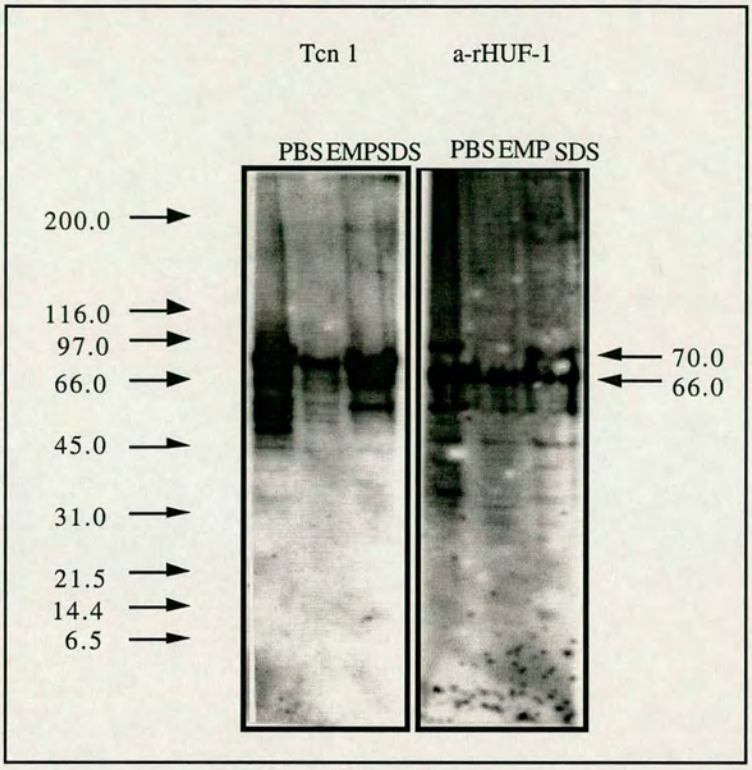



**Figure 5.10.** Western blot showing probing of sequentially extracted TEX [using PBS (phosphate buffered saline), EMP (empigen) and SDS (sodium dodecyl sulphate)] with anti-rHUF-1 and Tcn1 monoclonal antibody. The two arrows on the right side of the blots indicates the positions of the, α-rHUF-1 identified bands at 70 and 66 kDa.

### 5.3.2 *A putative family of mucins from* T. canis

Earlier work by Page *et al* (1992c) showed that while preparing parasites for electron microscopy using a graded alcohol series, the surface coat of the parasite was removed from the underlying cuticle in 35% alcohol. Analysis of the alcohol-derived fraction showed that the predominant protein in this fraction was TES-120 (see Chapter 1) (Page et al., 1992b). This protein is one of the most abundant antigens secreted by the parasite and has also been shown to be the major component of the surface coat of the parasite (Page et al., 1992b). It has now recently been shown that the TES-120 proteins consist of a putative family of mucins, only one of which had previously been reported (Gems and Maizels, 1996).

To date 5 mucins have been isolated from *Toxocara canis* (Figure 5.11). Of these *Tc-muc-1* was first isolated by spliced leader (SL1) PCR (Gems and Maizels, 1996) and was also isolated from this EST project (see Chapter 3) as 7 clones which represent 2.6% of the 266 clones serendipitously sequenced as part of the project. This indicates the high level of expression of *Tc-muc-1* within the parasite. In contrast *Tc-muc-2*, *-muc-3* and *-muc-4* were not as abundantly represented, appearing 3, 2 and 1 times respectively, which accounted for 1.1%, 0.75% and 0.37% respectively of the 266 clones analysed as part of the EST project (Tetteh *et al* ms submitted). The fifth mucin was recently identified, not as a clone from the EST project, but expression screening with a mouse polyclonal antiserum raised to TES products (Loukas A. *et al* ms. in preparation). Full-length sequence was obtained for *Tc-muc-2 -muc-4* (Figure 5.2 and 5.3), and work is currently in progress by A Loukas to confirm the full-length sequence for *muc-5* (previously obtained by A. Doedens). Figure 5.11 shows a schematic representation of all five mucins. Each of the cDNAs is 5' transpliced with Spliced leader 1 (SL1).

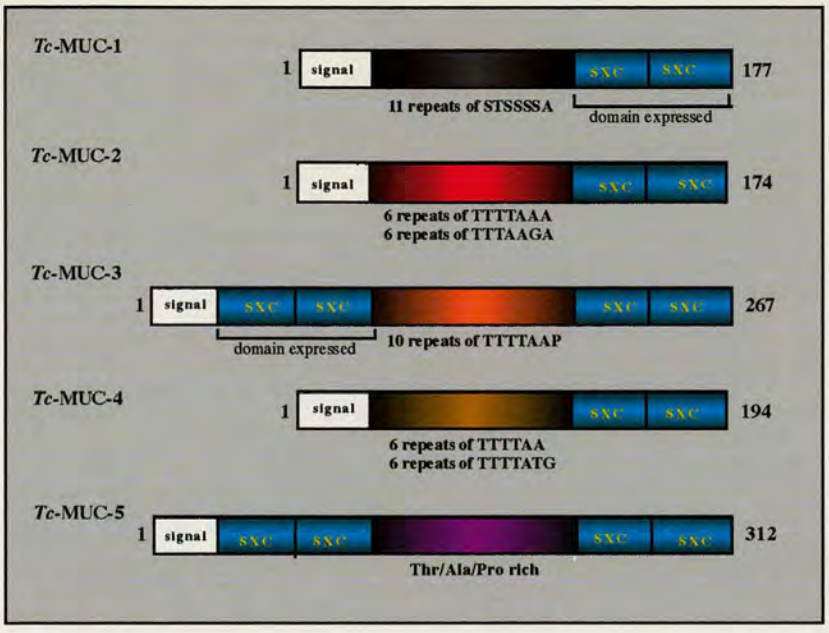



**Figure 5.11**. Schematic showing a diagrammatic alignment of the 5 *T. canis* mucins identified to date. The relative positions of the SXC domains is clearly indicated

Note: *Tc-muc-5* was identified as part of an antibody screen of the cDNA library (Doedens and Loukas pers. comm.) and was not isolated as part of the EST project (Tetteh et al submitted)

As is shown in Figure 5.11 each of the translated mucin proteins have a putative signal peptide. All the proteins possess C-terminal paired SXC domains and *Tc-muc-3* and-*5* each contain paired N-terminal and C-terminal domains. The SXC domains of the mucins show a remarkable level of sequence conservation (Figure 5.12). The SXC domains of MUC-1, -2 and -4 share between 86-97% identity to each other. The N-terminal SXC domains of MUC-3 proved to be less similar to both the C-terminal domains of itself (58%) and that of the remaining four mucins (65-70%) (Loukas A. per. comm.) (Figure 5.12). Each single SXC domain of the mucins was compared in a pairwise manner to each SXC domain of the remaining mucins. The data for each pairing was represented as a percentile value as is demonstrated in Table 5.1. The compiled data shows that the highest levels of similarity existed between equivalent mucin SXC domains. When comparing SXC 1 of MUC-1 to the analogous domains (where SXC 3 and 4 of MUC-3 and MUC-5 correspond to SXC 1 and 2 of the remaining mucins) this value ranges from 50-97%. This appears to be the rule for MUC 1 to 4, but does not apply as stringently to MUC-5 and highlights the level of divergence existing between MUC-5 and MUC 1 to 4 (Table 5.1).

(Table 5.1). The SXC domains of the mucins were then compared, for percentage identity, to the other SXC

| muc-1 sxc1 | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 40 | muc-1 SXC 2 | | | | | | | | | | | |
| 97 | 40 | MUC-2 SXC 1 | | | | | | | | | | |
| 40 | 97 | 40 | MUC 2 SXC2 | | | | | | | | | |
| 68 | 42 | 65 | 42 | MUC-3 SXC1 | | | | | | | | |
| 37 | 71 | 37 | 71 | 34 | MUC-3 SXC 2 | | | | | | | |
| 74 | 34 | 74 | 34 | 62 | 37 | MUC 3 SXC3 | | | | | | |
| 25 | 62 | 25 | 60 | 31 | 51 | 22 | MUC 3 SXC 4 | | | | | |
| 85 | 40 | 88 | 40 | 68 | 37 | 74 | 25 | MUC 4 SXC 1 | | | | |
| 40 | 88 | 40 | 85 | 37 | 71 | 37 | 54 | 37 | MUC 4 SXC 2 | | | |
| 37 | 57 | 37 | 57 | 45 | 54 | 37 | 42 | 37 | 54 | MUC 5 SXC1 | | |
| 42 | 54 | 42 | 54 | 40 | 54 | 37 | 42 | 40 | 54 | 51 | MUC 5 SXC 2 | |
| 37 | 57 | 37 | 57 | 40 | 60 | 34 | 48 | 37 | 54 | 68 | 60 | MUC 5 SXC 3 |
| 37 | 48 | 37 | 48 | 37 | 45 | 37 | 42 | 37 | 54 | 60 | 54 | 65 | MUC 5 SXC 4 |

**Table 5.1.** Table showing percentage similarity between SXC domains from the *T. canis* mucins. Percentile values were obtained by pairwise comparison of each domain using the MacVector software (Oxford molecular).

domains isolated from *T. canis* (*Tc-peb-1, Tc-huf-1* and *Tc-vah-1*), but also to the SXC domains isolated from two *C. elegans* predicted proteins, *P. carnea* metalloproteinase (PMP1), potassium channel blocking toxins and the human MAGP1 protein (Table 5.2). Interestingly, the highest values were obtained in a comparison of the mucin SXC domains with SXC 1 and 2 of *Tc-peb-1*. For SXC domain 1 of *Tc-peb-1*, the highest values were again obtained for the equivalent SXC domains in the mucins. This pattern was repeated for SXC 2 of *Tc-peb-1* for MUC-1 to 4, but for MUC-5 identical scores were obtained for SXC 1 to 3 with a marginal decrease with SXC 4. Table 5.3 shows the compiled data of the comparative analysis between the *T. canis* SXC domains of non-mucin origin compared with the SXC domains from two *C.elegans* predicted proteins, PMP1, the cnidarian toxin homologues and the human derived MAGP1. The same patterns of SXC domain identity as demonstrated in Tables 5.1 and 5.2 also apply.

Figure 5.12 shows an amino acid alignment of the SXC domains of five mucins, the PE binding protein, *Tc*-HUF-1, a number of potassium channel blocking toxins from sea anemones and hydra, and a microfibril associated glycoprotein (MAGP 1) from humans. The SXC domains across these diverse species show a high level of conservation. Structural modelling of the SXC domains from *Tc*-MUC-1 based on the known structure of *Bunodosoma granulifera* (BgK) toxin (Dauplais et al., 1997) reveals that they share similar folding patterns (Loukas *et al* manuscript in preparation). Figure 5.13 shows the manner in which the cysteine residues are thought to interact with each other in folding and is based on the NMR structure ofBgK (Dauplais et al., 1997).

Previous analysis of the serine/threonine content of TES products using radiolabelled amino acids(Page et al., 1991) showed TES-120 to have high levels of serine and threonine. Subsequent cloning of *Tc-muc-1* (Gems and Maizels, 1996) suggested that the protein product of the MUC-1 gene was responsible for the high serine content and that MUC-1 was TES-120. The identification of 4 new mucins - 3 from the EST project (Tetteh *et al* ms. submitted) and 1 from antibody screening (Loukas *et al* ms. in preparation) has shown that TES-120 consists of a family of mucins and not alternatively processed forms of one mRNA. Indicated in Figure 5.11 and in Figure 5.14 is the presence of serine/threonine rich regions which act as sites of potential *O*-linked glycosylation within the mucins thus accounting for their high molecular weights. MUC-1 is the only one of the mucins identified to date that possesses serine in its repeat region.

The other mucins each have a predominance of threonine with varying amounts of glycine, alanine and proline (Figure 5.11).

| | *Toxocara canis* mucin SXC domains | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | MUC1/1 | MUC1/2 | MUC2/1 | MUC2/2 | MUC3/1 | MUC3/2 | MUC3/3 | MUC3/4 | MUC4/1 | MUC4/2 | MUC5/1 | MUC5/2 | MUC5/3 | MUC5/4 |
| PE/1 | 48 | 42 | 48 | 42 | 57 | 40 | 57 | 31 | 48 | 40 | 45 | 34 | 40 | 34 |
| PE/2 | 34 | 51 | 34 | 51 | 34 | 51 | 28 | 42 | 34 | 54 | 48 | 48 | 48 | 45 |
| HUF/1 | 25 | 25 | 25 | 25 | 22 | 28 | 25 | 28 | 25 | 25 | 25 | 34 | 31 | 25 |
| HUF/2 | 34 | 25 | 34 | 25 | 31 | 28 | 37 | 20 | 34 | 25 | 28 | 28 | 31 | 31 |
| HUF/F1 | 20 | 20 | 20 | 20 | 17 | 22 | 20 | 20 | 20 | 20 | 17 | 25 | 20 | 17 |
| HUF/3 | 28 | 34 | 28 | 34 | 31 | 28 | 31 | 22 | 28 | 37 | 28 | 34 | 31 | 37 |
| VAH/1 | 25 | 25 | 25 | 25 | 22 | 28 | 25 | 28 | 25 | 25 | 25 | 31 | 28 | 25 |
| VAH/2 | 34 | 28 | 34 | 28 | 31 | 31 | 34 | 20 | 34 | 20 | 31 | 31 | 37 | 34 |
| VAH/3 | 25 | 25 | 25 | 25 | 22 | 28 | 25 | 28 | 25 | 25 | 22 | 31 | 25 | 25 |
| VAH/4 | 34 | 28 | 34 | 28 | 34 | 31 | 34 | 20 | 34 | 31 | 31 | 34 | 37 | 37 |
| *Ce*T20G8.5/1 | 28 | 34 | 28 | 37 | 25 | 34 | 25 | 25 | 31 | 31 | 37 | 31 | 34 | 28 |
| *Ce*K09C8.3/1 | 25 | 34 | 25 | 34 | 28 | 31 | 28 | 37 | 25 | 28 | 31 | 37 | 28 | 25 |
| PMP1/1 | 22 | 17 | 22 | 17 | 25 | 17 | 25 | 20 | 25 | 17 | 28 | 20 | 25 | 22 |
| ShK/1 | 37 | 28 | 37 | 28 | 25 | 28 | 28 | 22 | 31 | 28 | 25 | 37 | 28 | 28 |
| HmK/1 | 28 | 25 | 28 | 25 | 25 | 28 | 31 | 22 | 28 | 25 | 25 | 40 | 31 | 25 |
| BgK/1 | 25 | 28 | 25 | 31 | 28 | 25 | 22 | 25 | 25 | 31 | 31 | 31 | 28 | 34 |
| AsKs/1 | 25 | 25 | 25 | 28 | 25 | 28 | 28 | 22 | 25 | 28 | 28 | 37 | 34 | 28 |
| MAGP1/1 | 22 | 22 | 22 | 25 | 25 | 22 | 25 | 25 | 25 | 22 | 28 | 20 | 20 | 17 |

**Table 5.2.** Tabulated data showing the comparison between the mucin SXC domains with the other *T.c.* derived SXC domains, two predicted *C.elegans* proteins (*Ce*T20G8.5 and *Ce*K09C8.3), Podocyrne metalloproteinase (PMP1), four cnidarian potassium channel blocking toxins (ShK, HmK, BgK, AsKs) and microfibril associated glycoprotein 1 (MAGP1). The figures are shown as percentages based on thelevel of similarity between each domain pairs.

| PE/1 | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 37 | PE/2 | | | | | | | | | | | | | | | |
| 20 | 25 | HUF/1 | | | | | | | | | | | | | | |
| 28 | 25 | 24 | HUF/2 | | | | | | | | | | | | | |
| 14 | 20 | 57 | 21 | HUF/FC | | | | | | | | | | | | |
| 28 | 28 | 27 | 50 | 26 | HUF/3 | | | | | | | | | | | |
| 20 | 25 | 90 | 25 | 65 | 28 | VAH/1 | | | | | | | | | | |
| 28 | 22 | 30 | 75 | 23 | 53 | 27 | VAH/2 | | | | | | | | | |
| 20 | 31 | 69 | 28 | 69 | 28 | 69 | 25 | VAH/3 | | | | | | | | |
| 31 | 25 | 30 | 59 | 23 | 59 | 27 | 75 | 24 | VAH/4 | | | | | | | |
| 28 | 40 | 33 | 25 | 26 | 31 | 36 | 31 | 27 | 31 | CeT20/1 | | | | | | |
| 28 | 34 | 33 | 21 | 34 | 25 | 30 | 21 | 30 | 25 | 28 | CeK09/1 | | | | | |
| 25 | 25 | 27 | 25 | 30 | 18 | 27 | 25 | 36 | 25 | 22 | 27 | PMP1/1 | | | | |
| 28 | 34 | 33 | 25 | 34 | 31 | 33 | 25 | 36 | 25 | 31 | 27 | 25 | ShK/1 | | | |
| 25 | 31 | 39 | 28 | 38 | 31 | 36 | 28 | 36 | 28 | 28 | 24 | 25 | 60 | HmK/1 | | |
| 28 | 31 | 24 | 15 | 19 | 15 | 27 | 18 | 21 | 18 | 31 | 21 | 25 | 39 | 48 | BgK/1 | |
| 25 | 28 | 33 | 21 | 26 | 21 | 33 | 21 | 30 | 21 | 40 | 21 | 25 | 39 | 45 | 52 | AsK/1 |
| 28 | 22 | 27 | 25 | 26 | 18 | 27 | 18 | 27 | 21 | 28 | 29 | 22 | 27 | 24 | 30 | 34 | MAGP1/1 |

**Table 5.3.** Table showing percentage similarity between *T. canis* SXC domains of non-mucin origin, two *C.elegans* predicted proteins, Podocyrne metalloproteinase, potassium channel blocking toxins and the human MAGP1.

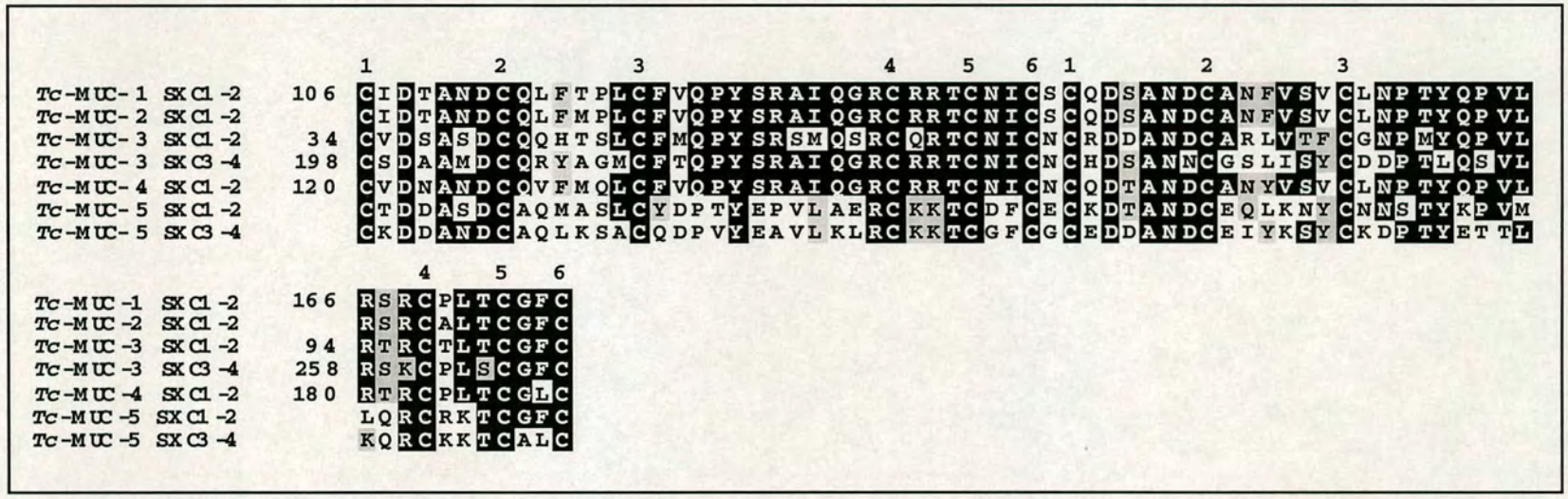



**Figure 5.12**. Amino acid pileup showing the paired SXC domains of the *T. canis* mucins.

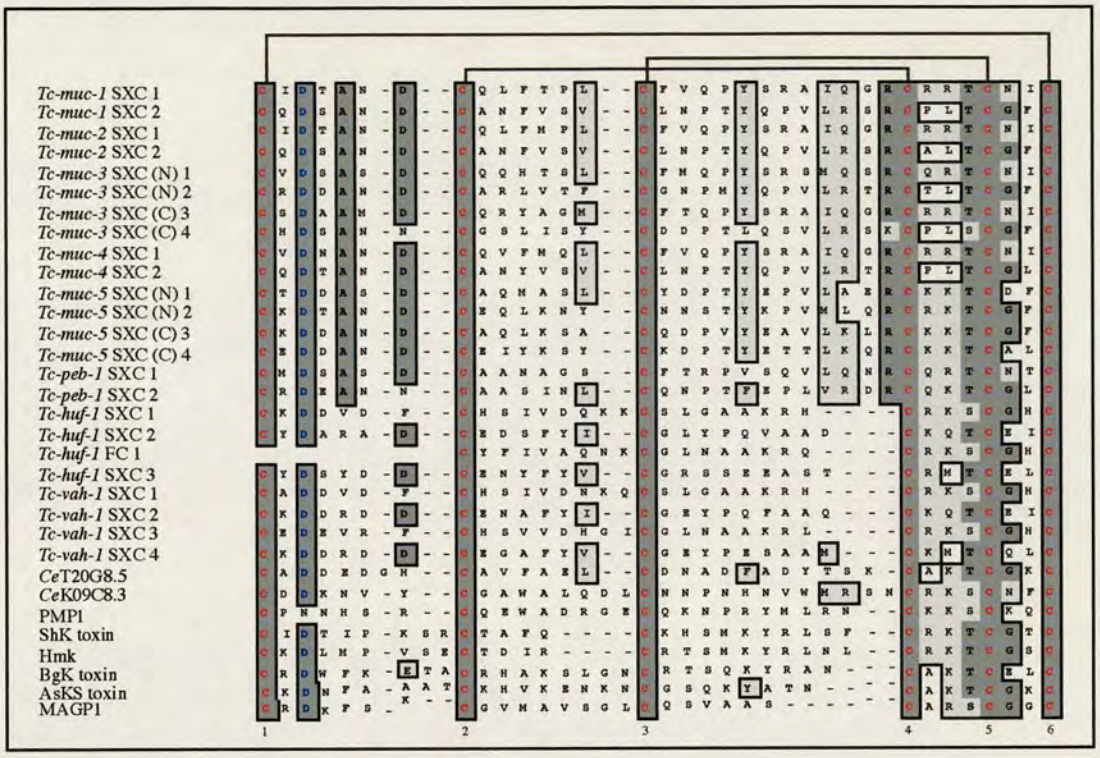



**Figure 5.13**. Amino acid pileup showing the SXC domains from predicted proteins from *Toxocara canis* (*Tc-muc-1, -muc-2, muc-3, muc-4 & -muc-5*, Tc-peb-1* (TES 26), *Tc-huf-1*, two *C. elegans* predicted proteins (*Ce T20G8.5* [Acc. No. Z304230] and *Ce K09C8.3* [Acc. No. Z68006], a metalloproteinase from *Podocoryne carnea*, PMP1 (AJ005052), and potasium channel inhibitors from *Stichodactyla helianthus*, ShK toxin (P29187), *Heteractis magnifica,* HmK toxin (AF020047), *Bunodosoma granulifera*, BgK toxin (P29186) and *Anemonia sulcata*, AsKS toxin (1181911) and the C-terminus of the microfibril associated glycoprotein (MAGP 1) from humans (P55002)

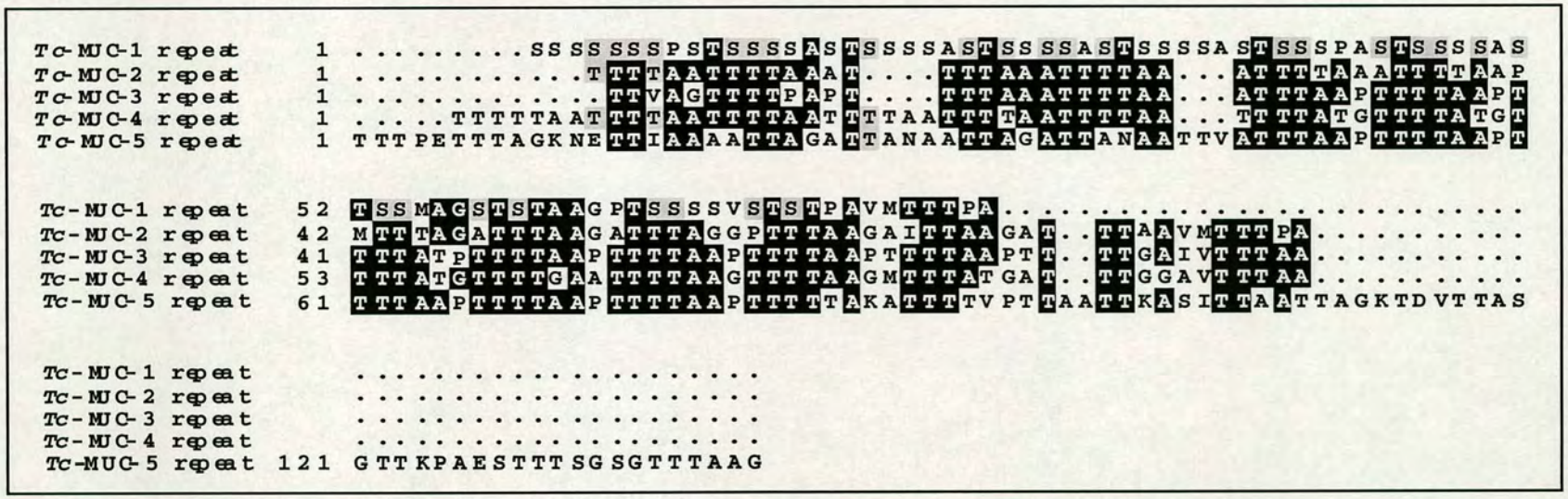



**Figure 5.14.** Pileup of the serine/threonine/alanine/glycine/proline repeat regions from the five *Toxocara mucins*

## 5.4 *Concluding remarks*

First identified as a secreted component from *T. canis* (Gems et al., 1995) SXC domains have been identified in a number of diverse species (Blaxter, 1998; Pan et al., 1998). The *C. elegans* project alone has identified 75 genes that contain 184 SXC domains (Blaxter, 1998). A number of these *C. elegans* open reading frames are composed solely of SXC domains (up to four), others are mucin-like with the SXC domains separated by serine or threonine rich regions, and yet others still possess SXC domains coupled to known enzymatic domains, such as tyrosinases, myeloperoxidases, and astacin-like zinc metalloproteinases (Blaxter, 1998). Yet the function(s) of these domains remains elusive. It has been hypothesised (Blaxter, 1998) that these domains are involved in protein-protein interactions with a probable specificity to the nematode cuticle. It has also been suggested that perhaps these domains are analogous in function to that of the epidermal growth factor (EGF) domains (Blaxter, 1998). While these suggestions may be valid, there is no supporting data as yet to lend credence to such hypotheses. It seems unlikely that these domains would function in a manner analogous to EGF as the arrangement of cysteines ($CX_7CX_5CX_{10}CXCX_8C$) and therefore the folding of each molecule would be different (figure 5.14).

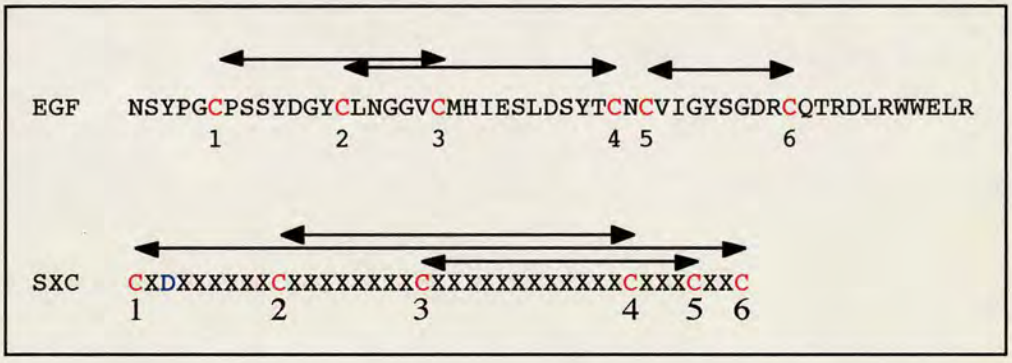



**Figure 5.15.** Figure showing the characteristic disulphide linkages of EGF compared to the deduced disulphide linkages of the SXC consensus The conserved cysteines are highlighted in red and the conserved aspartic acid in the SCX domain is shown in blue.

The SXC rich protein, *Tc*-HUF-*1* was first assumed to be a member of the TES-120 family of mucins. But analysis of the completed sequence showed it to be too dissimilar from the other members of this putative mucin family.

It may be of note to point out that the domains of *Tc-huf-1* are separated from each other by a 13 aa gap, followed by a 20 aa gap and then by another 13 amino acid gap (Figure 5.1 and 5.6). The amino acids of these gaps are predominantly composed of serine, threonine, proline, alanine and glycine, amino acids indicative of the core proteins of mammalian mucins (Devine and McKenzie, 1992).

Mucin-like structures have been found in a number of different species, such as the novel *Drosophila* surface mucin, hemomucin (Theoplod et al., 1996); dSR-C1, a class C macrophage specific scavenger receptor also from *Drosophila* (Pearson et al., 1995). Mucins have been reported from parasitic protozoans and are thought to be involved in immune evasion. Examples are the mucin-like proteophosphoglycans of *Leishmania* species (Ilg et al., 1996), a filamentous macromolecule thought to be implicated in cell aggregation (Ilg et al., 1996). The membrane mucin Ag C10 from *Trypanosoma cruzi* modulates the host immune response by binding to the surfaces of macrophages and specifically inhibiting monoclonal antibody binding against CD62L/L-selectin (de Diego et al., 1997). Glycosylphosphatidylinositol-anchored mucin-like glycoproteins *from T. cruzi* were shown to be potent inducers of nitric oxide synthesis in addition to induction of microbicidal activity in murine macrophage (Camargo et al., 1997).

Although the exact function of the mucin rich surface coat of the parasite is unknown, it appears that one of its main roles is to serve in a protective capacity. Badley *et al* (,1987) showed that this structure was frequently shed from the larval epicuticular surface under immune attack when incubated in infection serum and immune cells. It seems safe to assume that this structure plays a major role in the parasite's ability to survive for extended periods of time in the hostile environment of the host's immune system. The mucinous nature of the surface coat may also provide lubrication for the parasite during its tissue migratory phase. Extrapolating from de Diego *et al* (1997), perhaps the surface coat of *T. canis* might be effecting similar immune modulatory roles that aid its survival within the host.

# Chapter 6 Final Discussion

## 6.1 Final Discussion

### 6.0.1 *Relevance of study*

It is an intuitive assumption that the mechanisms employed by parasitic organisms have been developed over evolutionary time to be able to serve the parasite in terms of immune modulation and evasion. As such the establishment of an infectious organism within a host depends on the abilities of both the parasite and the host. In order to understand the subtle mechanisms involved at the host parasite interface both parasite derived and host derived components must be investigated. The release of functional products from parasites has long been thought to significantly confer a major physiological adaptation to survival within the host.

For these reasons, a number of excreted components have long been of interest to investigators. Proteases, in particular cysteine proteases have attracted a lot of interest due to the vital biological roles that they play both in the extracellular and intracellular environments. Of all the cysteine proteases the papain family is probably the best characterised (Barret and Rawlings, 1996).

Members of the C1 family of the CA clan have been shown to play vital roles at the host parasite interface (Barret and Rawlings, 1996). Cysteine proteases isolated and characterised from the protozoan parasite *Entamoeaba histolytica* were demonstrated to be released during infection. These proteins were shown to be responsible for the formation of liver abscesses in SCID mice (Stanley et al., 1995) and also in modulation of the host immune response by degradation of the anaphylatoxic complement components C3a and C5a (Reed et al., 1995). The effect of degrading the complement components initiates the activation of the alternate complement pathway which promotes lysis of the nonpathogenic, but not pathogenic strains of the parasite (Reed et al., 1995).

The ability of parasitic nematodes to penetrate and/or migrate within the host has been linked with parasite derived proteases. Such proteases have been identified and partially characterised in a number of systems (Haffner et al., 1998) including *Toxocara canis* (Robertson et al., 1989). An excellent example of host penetration and tissue migration is demonstrated in the parasitic nematode *Strongyloides stercoralis*, in which the parasite has been shown to travel at speeds in excess of 10 cm per hour through the dermal

extracellular matrix (McKerrow et al., 1990). The proteases involved in this process are thought to be neutral metalloproteinases, of which a homologous metaloprotease has been found in *Onchocera volvulus*, a leading cause of blindness and chronic dermatitis in both the African and South American continents (Haffner et al., 1998).

The components that comprise parasite excretory/secretory antigens are not restricted to proteases. A number of novel products have also been identified. The parasitic nematode *Strongyloides venezuelensis* secretes a glycoprotein mucin-like substance with strong adhesive properties (Maruyama and Nawa, 1997). Further characterisation of this mucosubstance is required although investigators hypothesis that the secretion has roles in the invasion and establishment of the parasite in the host epithelia (Maruyama and Nawa, 1997).

### 6.0.2 *Future work*

Just over 53% of the clones identified from the EST project had known homologues in the database. These included a number of metabolic and respiratory enzymes, remarkably few structural proteins and no DNA replication proteins. This finding is consistent with the developmentally arrested state of the parasite. There were also a small number of transcripts that were heavily represented within the library. Of these the novel C-type lectin, TES 32 (Loukas et al manuscript submitted) was the second most abundant (16/261= 6.1%). Work has been planned to determine the host ligand or ligands for which TES 32 is the receptor, and thereby determine a putative role for this novel lectin. A detailed description of these genes can be found in Chapter 3.

Included as part of the 53% of the clones with identified homologues were a group of clones for which functions within the nematode system were not intuitively obvious. It is, however, assumed that at least some of these genes will prove to be valuable subjects for immunological and biochemical studies. These included three venome allergen homologues; lupus autoantigen homologues; epithelin/granulin precurser homologues; olfactomedin; and the tubby-like protein, which bears homology to the mouse obesity gene. A number of these clones have been partially characterised in terms of gene structure (Chapter 3). In order to investigate the potential function of thses clones, the

most likely stratergy will involve a two pronged approach. Firstly, full-length transcripts will be cloned into appropriate vectors and antibodies raised to recombinant proteins or against DNA vaccine vectors. At around the same time proteins purified from the cloned constructs would be submitted to an appropriate institute (e.g. ICMB, Edinburgh University, BBSRC centre for structural biology, UCL), for the construction of 3 dimensional structures. It is theorised that from the 3D structures created using nuclear magnetic resonance (NMR), crystalography, or electron microscopy putative functions of these molecules can be better elucidated. Antibodies raised against these molecules can be used to further characterise these genes in terms of distribution within the parasite as well as to determine putative functions.

Five clones for which homologues were identified but for which functions within the nematode system were not obvious proved to be of interest. Of these clones, granulin/epithelin precursor, lupus autoantigen homologue, olfactomedin, interferon-related protein and the tubby-like protein, only olfactomedin and the epithelin/granulin homologues have been partially characterised in terms of gene structure (see Chapter 3). The first step is to obtain full-length sequence for these clones. Following this they will be expressed and antibodies raised in order to further characterise these genes in terms of location and function within the parasite.

The abundant novel transcripts (see Chapter 4) represented by four clusters (*ant-003, -005, -030, -034*) represented 2.7 - 5.7% respectively. Of these *ant-005* has been cloned into the DNA vaccine vector pcDNA3.1(-) (Invitrogen) and shown to be a component of TES. Work is presently ongoing to clone the remaining members of this novel clad and raise antibodies in order to be able to localise and deduce putative functions for them. This work is presently ongoing.

6.0.3 *Conclusion*

The scope of substances secreted by nematode systems represent a very complex antigenic mixture designed to enhance the parasites ability to survive the inhospitable environment that is the host defence machinery. In order to be able to apply suitable strategies for the effective treatment and/or control these ingenious organisms investigators must first understand them. The generation of the *Toxocara canis* EST database provides a major step forward towards the beginning of a comprehensive database of biological information.

In depth genomic analysis of nematodes is still in its infancy. Although with the completion of the *C. elegans* genome the significance of such ventures have been well advertised (Blaxter, 1998). Analytical tools are available and are also constantly being improved that will improve the rate at which genome projects are being completed. It will also allow questions to be asked of the *T. canis* genome in terms of the presence or absence of operons and the nature of the promoters. Thus shedding light on the gene organisation of the parasite. This level of further characterisation of the genome will enable the identification of gene loci through which possible variations across *T. canis* isolates can be mapped in terms of geography and between canid species.

It should be noted that the EST approach could enable researchers to identify 'new' gene sequences more rapidly than whole genome sequencing. Care should be taken that the relevance of abundant clones is not over-emphasised. The lack of complete genome information means important experiments such as transgenics cannot be performed. Although an excellent nematode transgenic model exist in *C. elegans*, the potential benefits of having a parasite model in which transgenics experiments could be performed should not be ignored.

Clones isolated from the EST project have been submitted to the dbEST public database and are available to investigators on request. The main aim of this project was to enhance existing knowledge about *Toxocara canis* and other zoonotic parasites in global sense. And in this regard the project was success.

## BIBLIOGRAPHY

Abe Y., Shirane K., Yokosawa H., Matsushita H., Mitta M., Kata I. and Ishii S.-i. (1993). Asparaginyl endopeptidase of jack bean seeds Purification, characterisation and high utility in protein sequence analysis. Journal of Biological Chemistry *268*, 3525 - 3529.

Adams M. D., Dubnick M., Kerlavage A. R., Moreno R., Kelley J. M., Utterback T. R., Nagle J. W., Fields C. and Venter J. C. (1992). Sequence identification of 2,375 human brain genes. Nature *355*, 632-634.

Adams M. D., Kelley J. M., Gocayne J. D., Dubnick M., Polymeropoulos M. H., Xiao H., Merril C. R., Wu A., Olde B., and Moreno R. F., et al (1991). Complementary DNA sequencing: expressed sequence tags and human genome project. Science *252*, 1651-1656.

Badley J. E., Grieve R. B., Rocke J. H., and Glickman L. T. (1987). Immune-mediated adherence of eosinophils to *Toxocara canis* infective larvae: the role of excretory-secretory antigens. Parasite Immunology *9*, 133-143.

Banfi S., Guffanti A. and Borsani G. (1998). How to get the best of dbEST. Trends in Genetics *14*, 80-81.

Barrett A. J. and Rawlings N. D. (1996). Families and clans of cysteine peptidases. Perspectives in Drug Discovery and Design 6, 1-11

Bateman A., Belcourt D., Bennet H., Lazure C. and Solomon S. (1990). Granulins, a novel class of peptide from leukocytes. Biochemical and Biophysical Research Communications *173*, 1161 - 1168.

Beaver P. C. (1966). Zoonoses, with particular reference to parasites of veterinary importance. In Biology of Parasites, E. J. L. Soulsby, ed. (A.P., N.Y., pp. 215-227.

Belcourt D. R., Lazure C. and Bennet H. P. J. (1993). Isolation and Primary Structure of the Three Major Forms of Granulin-like Peptides from Hematopoietic Tissues of a Telost Fish (*Cyprinus carpio*). The Journal of Biological Science *268 (13)*, 9230

Bhandari V., Palfree R. G. E. and Bateman A. (1992). Isolation and sequence of the granulin precursor cDNA from human bone marrow reveals tandem cysteine-rich granulin domains. Proc. Natl. Acad. Sci. *89*, 1715 - 1719.

Bird D. M. and Riddle D. L. (1994). A genetic nomenclature for parasitic nematodes. The Journal of Nematology *26*, 138-143.

Blattner F. R., Plunkett G. r., Bloch C. A., Perna N. T., Burland V., Riley M., Collado-Vides J., Glasner J. D., Rode C. K., Mayhew G. F., Gregor J., Davis N. W., Kirkpatrick H. A., Goeden M. A., Rose D. J., Mau B. and Shao Y. (1997). The complete genome sequence of *Escherichia coli* K-12. Science *277*, 1453-1474.

Blaxter M. (1998). *Caenorhabditis elegans* is a nematode. Science *282*, 2041-2046.

Blaxter M. L. and Robertson W. M. (1998). 2. The Cuticle. In The Physiology and Biochemistry of Free-living and Plant-parasitic Nematodes, R. N. Perry and D. J. Wright, eds.: CAB INTERNATIONAL), pp. 25 - 48.

Boguski M. S., Lowe T. M. J. and Tolstoshev C. M. (1993). dbEST - database for "expressed sequence tags". Nature Genetics *4*, 332-333.

Bottari A., Capocchi A., Galleschi L., Jopova A. and Saviozzi F. (1996). Asparaginyl endopeptidase during maturation and germination of durum wheat. Physiologia Plantarum *97*, 475 - 480.

Brindley P. J., Kalinna B. H., Dalton J. P., Day S. R., Wong J. Y., Smythe M. L. and McManus D., P. (1997). Proteolytic degradation of host hemoglobin by schistosomes. Molecular and Biochemical Parasitology *89*, 1-9.

Bult C. J., White O., Olsen G. J., Zhou L., Fleischmann R. D., Sutton G. G., Blake J. A., FitzGerald L. M., Clayton R. A., Gocayne J. D., Kerlavage A. R., Dougherty B. A., Tomb J. F., Adams M. D., Reich C. I. Overbeek R., Kirkness E. F., Weinstock K. G., Merrick J. M., Glodek A., Scott J. L., Geoghagen N. S. M. and Venter J. C. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii.* Science *273*, 1058-1073.

Camargo M. M., Andrader A. C., Almeida I. C., Travassos L. R. and Gazzinelli R. T. (1997). Glycoconjugates isolated from *Trypanosoma cruzi* but not *Leishmania* sp membrane trigger nitric oxide synthesis as well as microbicidal activity in IFN-gamma-primed macrophages. Journal of Immunology 159(12), 6131-9

Casella S. J., Han, V. K., D'Ercole A. J., Svoboda M. E. and Van Wyk J. J. (1986). Insulin-like growth factor II binding to the type I somatomedin receptor. Evidence for two high affinity binding sites. The Journal of Biological Chemistry *261*, 9268-9273.

Chen J.-M., Dando P. M., Stevens R. A. E., Fortunato M. and Barrett A. J. (1998). Cloning and expression of mouse legumain, a lysosomal endopeptidase. Biochemical Journal *335*, 111-117.

Chen J.-M., Dando P. M., Rawlings N. D., Brown M. A., Young N. E., Stevens R. A., Hewitt E., Watts C. and Barrett A. J. (1997). Cloning, Isolation,and Characterisation of Mammalian Legumain, an Aspariginyl Endopeptidase. The Journal of Biological Chemistry *272 (12)*, 8090-8098.

Chen Y., Faraco J., Yin W., Germiller J., Francke U. and Bonadio J. (1993). Structure, chromosomal localization, and expression pattern of the murine Magp gene. The Journal Chemistry *268*, 27381-27389.

Csoma C. and Polgar L. (1984). Proteinase from germinating bean cotyledons. Evidence for involvement of a thiol group in catalysis. Biochemical Journal *222*, 769-776.

Culouscou J.-M., Carlton G. W. and Shoyab M. (1993). Biochemical Analysis of the Epithelin Receptor. The Journal of Biological Chemistry *268*, 10458 - 10462.

Cypess R. H., Karol M. H., Zidian J. L., Glickman L. T. and Gitlin D. (1977). Larva-specific Antibodies in Patients with Visceral Larva Migrans. The Journal of Infectious Diseases *135*, 633-640.

Dalton J., P., Smith A., M., Clough K., A. and Brindley P., J. (1995a). Digestion of haemoglobin by Schistosomes: 35 years on. Parasitology Today *11*, 299-303.

Dalton J. P., Hola-jamriska L. and Brindley P. J. (1995b). Aspariginyl endopeptidase activity in adult *Schistosoma mansoni*. Parasitology *111*, 575-580.

Danielson P. E., Forss-Petter S., Battenberg E. L., deLecea L. Bloom F. E., and Sutcliffe, J. G. (1994). Four structurally distinct neuron-specific olfactomedin-related glycoproteins produced by differential promoter utilization and alternative mRNA splicing from a single gene. Journal of Neuroscience Research *38*, 468-478.

Dauplais M., Lecoq A., Song J., Cotton J., Jamin N., Gilquin B., Roumestand C., Vita C., de Medeiros C. L. C., Rowan E., G., Harvey A. L. and Menez A. (1997). On the convergent evolution of animal toxins. Conservation of a diad of functional residues in potassium channel-blocking toxins with unrelated structures. The Journal of Biological Chemistry *272*, 4302-4309.

Davis A. H., Nanduri J. and Watson D. C. (1987). Cloning and gene expression of *Schistosoma mansoni* protease. Journal of Biological Chemistry *262*, 12851-12855.

de Diego J., Punzon C., Duarte M. and Fresno M. (1997). Alteration of macrophage function by *Trypanosoma cruzi* membrane mucins. Journal of Immunology 159(10), 4983-9

de Savigny D. H. (1975). In vitro maintenance of *Toxocara canis* larvae and a simple method for the production of *Toxocara* ES antigen for use in serodiagnosis test for visceral larva migrans. Journal of Parasitology *61*, 781-782.

Dent J. H., Nichols R. L., Beaver P. C. Carrera G. M., and Staggers R. J. (1956). Visceral Larva Migrans With A Case Report. American Journal of Pathology *32*, 777-803.

Devine P. L. and McKenzie F. C. (1992). Mucins: Structure, Function, and Associations with Malignancy. Bioassays *14*, 619-625.

Dissanayake S., Xu M., Nkenfou C. and Piessens W. F. (1993). Molecular cloning and serological characterisation of a Brugia Malayi pepsin inhibitor homolog. Molecular and Biochemical Parasitology *62*, 143 - 146.

Dowd A. J., Dalton J. P., Loukas A. C., Provic P. and Brindley P. J. (1994). Secretion of cysteine proteinase activity by the zoonotic hookworm Ancylostoma caninum. American Journal of Tropical Medicine and Hygiene *51*, 341-347.

Dresden M. H. and Deelder A., M. (1979). Schistosoma mansoni: thiol proteinase properties of adult worm "hemoglobinase". Experimental Journal of Parasitology *48*, 190-197.

El-Sayed A. M. N. and Donelson J. E. (1997). A survey of the *Trypanosoma brucei* rhodesiense genome using shotgun sequencing. Molecular and Biochemical Parasitology *84*, 167-178.

Faraco J., Bashir M., Rosenbloom J. and Francke U. (1995). Characterization of the human gene for microfibril-associated glycoprotein (MFAP2), assignment to chromosome 1p36.1-p35, and linkage to D1S170. Genomics *25*, 630-637.

Finkelman F. D., Pearce E. J., Urban Jr J. F. and Sher A. (1991). Regulation and biological function of helminth-induced cytokine responses. Immunology Today 12, A62

Fleming J. T., Baylis H. A., Sattelle D., B. and Lewis J., A. (1996). Molecular cloning and in vitro expression of *C. elegans* and parasitic nematode ionotropic receptors. Parasitology *113*, S175-190.

Gamble H. R., Fetterer R. H. and Mansfield L. S. (1996). Developmentally regulated zinc metaloproteinases from third- and fourth-stage larvae of the ovine nematode *Haemonchus contortus*. Journal of Parasitology *82*, 197-202.

Gems D., Ferguson C. J., Robertson B. D., Nieves R., Page A. P., Blaxter M. L. and Maizels R. M. (1995). An Abundant, *trans*-spliced mRNA from *Toxocara canis* Infective Larvae Encodes a 26-kDa Protein with Homology to Phosphatidylethanolamine-binding Proteins. The Journal Of Biological Chemistry *270 No. 31*, 18517-18522.

Gems D. and Maizels R. M. (1996). An abundantly expressed mucin-like potein from Toxocara canis infective larvae: The precursor of the larval surface coat glucoproteins. Proceedings of the National Academy of Science *93*, 1665-1670.

Gendeh G., S.,Young L., C., de Medeiros C. L., Jeyaseelan K., Harvey A. L. and Chung, M. C. (1997). A new potassium channel toxin from the sea anemone *Heteractis magnifica*: isolation, cDNA cloning, and functional expression. Biochemistry *36*, 11461-11471.

Ghendler Y., Parizade M., Arnon R., McKerrow J. H. and Fishelson Z. (1996). *Schistoma mansoni*: Evidence for a 28-kDa Membrane-Anchored Protease on Schistosomula. Experimental Parasitology *83*, 73-82.

Gibson M., A., Kumaratilake J. S. and Cleary E., G. (1989). The protein components of the12-nanometer microfibrils of elastic and nonelastic tissues. The Journal of Biological Chemistry *264*, 4590-4598.

Gillespie S. H. (1988). The epidemiology of *Toxocara canis*. Parasitology Today *4*, 180-182.

Gillespie S. H. (1987). Human Toxocariasis. Journal of Applied Bacteriology *63*, 473-479.

Gillespie S. H., Bidwell D., Voller A., Robertson B. D. and Maizels, R. M. (1993). Diagnosis of human toxocariasis by antigen capture enzyme linked immunosorbent assay. Journal of Clinical Pathology *46*, 551-554.

Glickman L. T. (1993). The epidemiology of human toxocariasis. In Toxocara and Toxocariasis: Clinical, epidemiological and molecular perspectives, J. W. Lewis and R. M. Maizels, eds. (London: The Institute of Biology and the British Society for Parasitology), pp. 3-10.

Gooley A. A., Pisano A., Packer N. H., Ball M., Jones A., Alewood P. F., Redmand J. W. and Williams K. L. (1994). Characterisation of a single glycosylated asparagine site on a glycopeptide using solid-phase Edman degredation. Glycoconjugate Journal *11(3)*, 180-6.

Guenette S., Prichard R. K., Klein R. D. and Matlashewski G. (1991). Characterization of a beta-tubulin gene and a beta-tubulin gene products of Brugia pahangi. Molecular and Biochemical Parasitology *44*, 153-164.

Haffner A., Guilavogui A. Z., Tischendorf F. W. and Brattig N., W. (1998). *Onchocerca volvulus*: microfilariae secrete elastinolytic and males nonelastinolytic matrix-degrading serine and metalloproteases. Experimental Parasitology *90*, 26-33.

Hammond M. P. and Bianco A. E. (1992). Genes and Genomes of Parasitic Nematodes. Parasitology Today *8*, 299 - 305.

Hawdon J. M., Jones B. F., Hoffman D. R. and Hotez P. J. (1996). Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. The Journal Biological Chemistry *271*, 6672-6678.

Healer J., Ashall F. and Maizels R., M. (1991). Characterization of proteolytic enzymes from larval and adult *Nippostrongylus brasiliensis*. Parasitology *103*, 305-314.

Himmelhoch S., Kisiel M. J. and Zuckerman B. M. (1977). *Caenorhabditis briggsae*: Electron Microscopic Analysis of Changes in Negative Surface Charge Density of the Outer Cuticular Membrane. Experimental Parasitology *41*, 118-123.

Himmelhoch S. and Zuckerman B. M. (1983). Caenorhabditis elegans: Characters of Negatively Charged Groups on the Cuticle and Intestine. Experimental Parasitology *55*, 299-305.

Hong X., Bouvier J., Wong M. M., Yamagata G. Y. L. and McKerrow, J. H. (1993). Brugia pahangi: Identification and characterisation of an Aminopeptidase Associated with Larval Molting. Experimental Parasitology *76*, 127-133.

Hoog C. (1991). Isolation of a large number of novel mammalian genes by a differential cDNA library screening strategy. Nucleic Acid Research *19*, 6123-6127.

Hotez P. J. and Cerami A. (1983). Secretion of a proteolytic anticoagulant by *Ancylostoma* hookworms. Journal of Experimental Medicine *157*, 1594-1603.

Hrabal R., Chen Z., James S., Bennett H. P. J. and Ni F. (1996). The hairpin stack fold, a novel protein architecture for a new family of protein growth factors. Nature Structural Biology *3*, 747 - 752.

Ilg T., Stierhof Y.-D., Craik D., Simpson R., Handman E. and Bacic A. (1996). Purification and Structural Characterisation of a Filamentous, Mucin-like Proteophosphoglycan Secreted by *Leishmania* Parasites. The Journal Of Biological Chemistry *271*, 21583-21596.

Irvine M., Huima T., Prince A. M. and Lustigman S. (1994). Identification and characterization of an *Onchocerca volvulus* cDNA clone encoding a highly immunogenic calponin-like protein. Molecular and Biochemical Parasitology *65*, 135-146.

Ishii N., Fujii M., Hartman P., S., Tsuda M., Yasuda K., Senoo-Matsuda N., Yanase S., Ayusawa D. and Suzuki K. (1998). A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. Nature *394*, 694-697.

Jacquier P., Gottstein B., Stingelin Y. and Eckert J. (1991). Immunodiagnosis of Toxocarosis in Humans: Evaluation of a New Enzyme-Linked Immunosorbent Assay Kit. Journal of Clinical Microbiology *29*, 1831-1835.

Jacquiet P., Cabaret J., Dia M. L., Cheikh D. and Thiam E. (1996). Adaptation to arid environment: *Haemonchus logistipes* in dromedaries of Saharo-Sahelian areas of Mauritania. Veterinary Parasitology *66*, 193-204.

Jansson H.-B., Jeyaprakash A., Coles G. C., Marban-Mendoza N. and Zuckerman B. M. (1986). Fluorescent and Ferritn Labelling of Cuticle Surface Carbohydrates of Caenorhabditis elegans and Panagrellus redivivus. Journal of Nematology *18*, 570-574.

Jung R., Scott M. P., Nam Y. W., Beaman T. W., Bassuner R., Saalbach I., Muntz K. and Nielsen N. C. (1998). The role of proteolysis in the processing and assembly of 11S seed globulins. Plant Cell *10*, 343-357.

Karanu F. N., Rurangirwa F. R., McGuire T. C. and Jasmer D. P. (1993). *Haemonchus contortus*: Identification of proteases with diverse characteristics in adult worm exctretory-secretory products. Experimental Parasitology *77*, 362-371.

Karavanich C. A. and Anholt R. R. H. (1998). Molecular Evolution of Olfactomedin. Molecular Biology and Evolution *15*, 718 - 726.

Kembhavi A., A., Buttle D. J., Rauber P. and Barrett A. J. (1991). Clostripain: characterization of the active site. FEBS letters *283*, 277-280.

Kim Y. S., Gum J. J. and Brockhausen, I. (1996). Mucin glycoproteins in neoplasia. Glycoconjugate Journal *13*, 693-707.

Kleyn P. W., Fan W., Kovats S. G., Lee J. J., Pulido J. C., Wu Y., Berkemeier L. R., Misumi D. J., Holmgren L., Charlat O., Woolf E. A., Tayber O., Brody T., Shu P., Hawkins F., Kennedy B., Baldini L., Ebeling C., Alperin G., D., Deeds J., Lakey N. D., Culpepper J., Chen H., Glucksmann-Kuis M. A., Moore K. J., et. al. (1996). Identification and characterization of the mouse obesity gene tubby: a member of a novel gene family. Cell *85*, 281-290.

Klinkert M. Q., Felleisen R., Link G., Ruppel A. and Beck E. (1989). Primary structures of Sm31/32 diagnostic proteins of Schistosoma mansoni and their identification as proteases. Molecular and Biochemical Parasitology *33*, 113-122.

Klinkert M. Q., Ruppel A. and Beck E. (1987). Cloning of diagnostic 31/32 kilodalton antigens of Schistosoma mansoni. Molecular and Biochemical Parasitology *25*, 247-255.

Kotzin B. L. (1996). Systemic lupus erythrematosus. Cell *85*, 303-306.

Kubota R., Noda S., Wang Y., Minoshima S., Asakawa S., Kudoh J., Mashima Y., Oguchi Y. and Shimizu N. (1997). A novel myosin-like protein (myocilin) expressed in the connecting cilium of the photoreceptor: molecular cloning, tissue expression, and chromosomal mapping. Genomics *41*, 360-369.

Liddell S. and Knox D. P. (1998). Exracellular and cytoplasmic Cu/Zn superoxide dismutase from *Haemonchus contortus*. Parasitology *116*, 383-394.

Lindquist R. N., Senft A. W., Petitt M. and McKerrow J. (1986). *Scistosoma mansonin:* Purification and characterisation of the major acidic proteinase from abdult worms. Experimental Parasitology 61(3), 398-404

Loukas A., Selzer P. M. and Maizels R. M. (1998). Characterisation of *Tc-cpl-1*, a cathepsin L-like cysteine protease from *Toxocara canis* infective larvae. Molecular and Biochemical Parasitology *92*, 275-289.

Loukas A., Mullin N. P., Tetteh K. K. A., Moens L. and Maizels R. M. (1999). A Novel C-type Lectin is the Major Secretory Product of a Tissue-Dwelling Parasitc Nematode. Current Biology 9(12), 825-828.

Lusson J., Benjannet S., Hamelin J., Savaria D., Chretien M. and Seidah N. G. (1997). The integrity of the RRGDL sequence of the proprotein convertase PC1 is critical for its zymogen and C-terminal processing and for its cellular trafficking. Biochemical Journal *326*, 737-744.

Lustigman S., McKerrow J. H., Shah K., Lui J., Huima T., Hough M. and Brotman B. (1996). Cloning of a Cysteine Protease Required for the Molting of *Onchocerca volvulus* Third Stage Larvae. The Journal of Biological Chemistry *271*, 30181 - 30189.

Lynch N. R., Eddy K., Hodgen A. N., Lopez R. I. and Turner K. J. (1988). Seroprevalence of *Toxocara canis* infection in tropical Venezuela. Transactions of the Royal Society of Tropica Medicine and Hygiene *82*, 275-281.

Maizels R. M., Bundy D. A., Selkirk M., E, Smith D. F. and Anderson R. M. (1993). Immunological modulation and evasion by helminth parasites in human populations. Nature *365*, 797-805.

Maizels R. M., Blaxter M. L., Robertson B. D. and Selkirk M. E. (1991). 6. Iodination of Soluble Proteins. In Parasite Antigens, Parasite Genes A Laboratory Manual for Molecular Parasitolgy, C. U. Press, ed.

Maizels R. M., Kennedy M. W., Meghji M., Robertson B. D. and Smith H. V. (1987). Shared carbohydrate epitopes on distinct surface and secreted antigens of the parasitic nematode *Toxocara canis*. The Journal of Immunology *139*, 207-214.

Maizels R. M. and Meghji M. (1984). Repeated patent infection of adult dogs with *Toxocara canis*. Journal of Helminthology 58, 327-333

Maizels R. M. and Robertson B. D. (1991). Toxocara canis : secreted glycoconjugate antigens in immunology and immunodiagnosis. In Parasitic nematodes - antigens, membranes and genes, M. W. Kennedy, ed. (London. New York. Philedelphia: Taylor and Francis), pp. 95-115.

Maruyama H. and Nawa Y. (1997). *Strongyloides venezuelensis*: Adhesion of Adult Worms to Culture Vessels by Orally Secreted Mucosubstances. Experimental Parasitology *85*, 10-15.

McCauliffe D. P., Zappi E., Lieu T. S., Michalak M., Sontheimer R. D., and Capra, J., D. (1990). A human Ro/SS-A autoantigen is the homologue of calreticulin and is highly homologous with onchocercal RAL-1 antigen and an aplysia "memory molecule". Journal of Clinical Investigation *86*, 332-335.

McCombie W. R., Adams M. D., Kelley J. M., FitzGerald M. G., Utterback T. R., Khan M., Dubnick M., Kerlavage A. R., Venter J. C. and Fields C. (1992). *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. Nature Genetics *1*, 124-131.

McConville M. J. and Bacic A. (1990). The glycoinositolphospholipid profiles of two *Leishmania major* strains that differ in lipophosphoglycan expression. Molecular and Biochemical Parasitology *38*, 57-68.

McKerrow J. H., Brindley P., Brown M., Gam A. A., Staunton C. and Neva F. A. (1990). *Strongyloides stercoralis*: identification of a protease that facilitates penetration of skin by the infective larvae. Experimental Parasitology *70*, 134-143.

Meghji M. and Maizels R. M. (1986). Biochemical properties of larval excretory-secretory glycoprteins of the parasite nematode *Toxocara canis*. Molecular and Biochemical Parasitology *18*, 155-170.

Miklos G. L. and Rubin G. M. (1996). The role of the genome project in determining gene function: insights from model organisms. Cell *86*, 521-529.

Min W. and Jones D. H. (1994). In vitro splicing of concanavalin A is catalyzed by asparaginyl endopeptidase. Nature Structural Biology *1*, 502-504.

Moore J., Todorova V. and Kennedy M. W. (1995). A cDNA encoding ribosomal protein L3 from the parasitic nematode Toxocara canis. Gene *165*, 239-242.

Morgan B. B. and Hawkins, P. A. (1953). Helminths of the dog and cat. In Veterinary Helminthology, B. B. Morgan,. Hawkins, eds. (Minneapolis: Burgess Publishing Company), pp. 234-236.

Nakakura N., Hietter H., Van Dorsselaer A., and Luu B. (1992). Isolation and structural determination of three peptides from the insect *Locusta migratoria.* Identification of a deoxyhexose-linked peptide. European Journal of Biochemistry *204*, 147 - 153.

Nguyen T. D., Chen P., Huang W. D., Chen H., Johnson D. and Polansky J. R. (1998). Gene structure and properties of TIGR, an olfactomedin-related glycoprotein cloned from glucocorticoid-induced trabecular meshwork cells. The Journal of Biological Chemistry *273*, 6341-6350.

Nichols R. L. (1956). The etiology of visceral larval migrans. The Journal of Parasitology *42*, 349-390.

Oaks J. A. and Kayes S. G. (1979). Artificial Hatching and Culture of *Toxocara canis* Second Stage Larvae. Journal of Parasitology *65*, 969-970.

Obwaller A., Jensen-Jarolim E., Auer H., Huber A., Kraft D. and Aspock H. (1998). *Toxocara* infestations in humans: symptomatic course of toxocariasis correlates significantly with levels of IgE/anti-IgE immune complexes. Parasite Immunology 20, 311-317

Oldstone, M. (1987). Molecular mimicry and autoimmune disease. Cell *50*, 819-820.

Ou X., Thomas R., Chacon M. R., Tang L. and Selkirk M. E. (1995). Brugia malayi: Differential Susceptibility to and Metabolism of Hydrogen Peroxide in Adults and Microfilariae. Experimantal Parasitology *80*, 530-540.

Page A. P. (1991). Biosynthesis, Localisation and Secretion of the Surface-Associated Antigens of *Toxocara canis*. PhD Thesis

Page A. P., Richards D. T., Lewis J. W., Omar H. M. and Maizels R. M. (1991). Comparison of isolates and species of *Toxocara* and *Toxascaris* by biosynthetic labelling of somatic and ES proteins from infective larvae. Parasitology *103*, 451-464.

Page A. P., MacNiven K. and Hengartner M., O. (1996). Cloning and biochemical characterization of the cyclophilin homologues from the free-living nematode *Caenorhabditis elegans*. Biochemical Journal *317*, 179-185.

Page A. P., Hamilton A. J. and Maizels R. M. (1992a). *Toxocara canis* : Monoclonal Antibodies to Carbohydrate Epitopes of Secreted (TES) Antigens Localize to Different Secretion-Related Structures in Infective Larvae. Experimental Parasitology, 56-71.

Page A. P. and Maizels R. M. (1992b). Biosynthesis and glycosylation of serine/threonine-rich secreted proteins from *Toxocara canis* larvae. Parasitology *105*, 297-308.

Page A. P., Rudin W., Fluri E., Blaxter M. L. and Maizels R. M. (1992c). *Toxocara canis* : A Labile Antigenic Surface Coat Overlying the Epicuticle of Infective Larvae. Experimental Parasitology *75*, 72-86.

Page A. P., Rudin W. and Maizels R. M. (1992d). Lectin binding to secretory structures, the cuticle and the surface coat of Toxocara canis infective larvae. Parasitology *105*, 285-296.

Pan T., Groger H., Schmid V. and Spring J. (1998). A toxin homology domain in an astacin-like metalloproteinase of the jellyfish Podocoryne carnea with a dual role in digestion and development. Development Genes and Evolution *208*, 259-266.

Pearson A., Lux A. and Krieger M. (1995). Expression cloning of dSR-C1, a class C macrophage specific scavenger receptor from *Drosophila melanogaster*. Proceedings of the National Academy of Science 29(9), 4056-60

Plowman G. D., Green J. M., Neubauer M. G., Buckley S. D., McDonald V. L., Todaro G. J. and Shoyab M. (1992). The Epithelin Precursor Encodes Two Proteins with Opposing Activities on Epithelial Cell Growth. The Journal of Biological Chemistry *267, No. 18*, 13073 - 13078.

Pritchard D. I., Leggett K. V. Rogan M. T., McKean P. G., and Brown A. (1991). *Necator americanus* secretory acetycholinesterase and its purification from excretory-secretory products by affinity chromatography. Parasite Immunology *13*, 187-199.

Reed S. L., Ember J. A., Herdman D. S., DiScipio R. G., Hugli T. E. and Gigli I. (1995). The extracellular neutral cysteine proteinase of *Entamoeba histolytica* degrades anaphylatoxins C3a and C5a. Journal of Immunology *55*, 266-274.

Reichardt L. F. and Tomaselli K., J. (1991). Extracellular matrix molecules and their receptors: functions in neural development. Annual Reviews in Neuroscience *14*, 531-570.

Rhoads M. L. and Fetterer R. H. (1995). Developmentally regulated secretion of cathepsin L-like cysteine proteases by *Haemonchus contortus*. Journal of Parasitology *81*, 505-512.

Robertson B. D., Bianco T. A., McKerrow J. H. and Maizels R. M. (1989). *Toxocara canis*: Proteolytic enzymes secreted by the infective larvae in vitro. Experimental Parasitology *69*, 30-36.

Sale E., M., Atkinson P. G. and Sale G. J. (1995). Requirement of MAP kinase for differentiation of fibroblasts to adipocytes, for insulin activation of p90 S6 kinase and for insulin or serum stimulation of DNA synthesis. The EMBO Journal *14*, 674-684.

Schallig H. D. F. H. and van Leeuwen M. A. W. (1997). Protective immunity to the blood-feeding nematode *Haemonchus contortus* induced by vaccination with parasite low molecular weight antigens. Parasitology *114*, 293-299.

Schallig H. D. F. H., van Leeuwen M. A. W., Verstrepen B. E. and Cornelissen A. W. C. A. (1997). Molecular characterisation and expression of two putative protective excretory secretory proteins of *Haemonchus contortus*. Molecular and Biochemical Parasitology *88*, 203-213.

Schlaepfer D. D. and Hunter T. (1998). Integrin signalling and tyrosine phosphorylation: just the FAKs? Trends in Cell Biology *8*, 151-157.

Schweitz H., Bruhn T., Guillemare E., Moinier D., Lancelin J., M., Beress L. and Lazdunski M. (1995). Kalicludines and kaliseptine. Two different classes of sea anemone toxins for voltage sensitive K+ channels. The Journal of Biological Chemistry *270*, 25121-25126.

Scott M. P., Jung R., Muntz K. and Nielsen N. C. (1992). A protease responsible for post-translational cleavage of a conserved Asn-Gly linkage in glycinin, the major seed storage protein of soybean. Proceedings of the National Academy of Sciences *89*, 658 - 662.

Sheldon P., S., Keen J. N. and Bowles D. J (1996). Post-translational peptide bond formation during concanavalin A processing in vitro. Biochmical Journal *320*, 865-870.

Shoyab M., McDonald V. L., Byles C., Todaro G. J. and Plowman G. D. (1990). Epithelins 1 and 2: Isolation and characteriastion of two cysteine-rich growth-modulating proteins. Proceeding of the National Academy of Science *87*, 7912 - 7916.

Smith H. V. (1989). A rapid method for hatching infective eggs of *Toxocara canis*. Transactions Of the Royal Society Of Tropical Medicine And Hygiene *83*, 215.

Smith H. V. (1991). Antibody reactivity in human toxocariasis. In *Toxocara* and Toxocariasis: Clinical, epidemiological and molecular perspectives, J. W. Lewis and R. M. Maizels, eds. (London: Institute of Biology and the British Society for Parasitology), pp 91-109.

Smyth J. D. (1994). Introduction to Parasitology (Third Edition). J. D. Smyth, ed. (Cambridge: Cambridge University Press).

Sonnhammer E. L. and Durbin R. (1997). Analysis of protein domain families in Caenorhabditis elegans. Genomics *46*, 200-216.

Soulsby L. (1993). Introduction. In *Toxocara* and Toxocariasis: Clinical, epidemiological and molecular perspectives, J. W. Lewis and R. M. Maizels, eds. (London: Institute of Biology and the British Society for Parasitology), pp. 1-2.

Stanley S. L. Jr, Zhang T., Rubin D. and Li E. (1995). Role of the Entamoeba histolytica cysteine proteinase in amebic liver abscess formation in severe combined immunodeficient mice. Infection and Immunity *63*, 1587-1590.

Stone E. M., Fingert J. H., Alward W. L. M., Nguyen T. D., Polansky J. R., Sunden S. L. F., Nishimura D., Clark A. F., Nystuen A., Nichols B. E., Mackey D. A., Ritch R., Kalenak J. W., Craven E. R. and Sheffield V. C. (1997). Identification of a gene that causes primary open angle glaucoma. Science *275*, 668-670.

Takeda O., Miura M., Matsushita H., Kato I., Abe Y., Yokosawa H. and Ishii S.-i. (1994). Isolation and Analysis of cDNA Encoding a Precursor of *Canavalia ensiformis* Asparaginyl Endopeptidase (Legumain). The Journal of Biochemistry *116*, 541 - 546.

Taylor M. R. H. (1993). Toxocariasis in Ireland. In *Toxocara* and Toxocariasis Clinical, epidemiological and molecular perspectives, J. W. Lewis and R. M. Maizels, eds. (London: The Institute of Biology and the British Society for Parasitology), pp. 71 - 80.

Tetteh K. K. A., Loukas A., Tripp C. and Maizels R. M. (1999). Identification of Abundantly Expressed Novel and Conserved Genes from the Infective Larval Stage of *Toxocara canis* by an Expressed Sequence Tag Strategy. Infection and Immunity 67(9), 4771-9

Theopold U., Samakovlis C., Erdjument-Bromage H., Dillon N., Axelsson B., Schimdt O., Tempst P. and Hulfmark D. (1996). *Helix pomatia* lectin, an inducer of *Drosophilla* immune response, binds to hemomucin, a novelsurface mucin. The Journal of Biological Chemistry 271(22), 12708-15

Tirone F., and Shooter E., M. (1989). Early gene regulation by nerve growth factor in PC12 cells: induction of an interferon-related gene. Proceedings of the National Academy of Science *86*, 2088-2092.

Tudor J., E., Pallaghy P. K., Pennington M. W., and Norton R. S. (1996). Solution structure of ShK toxin, a novel potassium channel inhibitor from a sea anemone. Nature Structure Biology *3*, 317-320.

Uga S., Minami T., and Nagata K. (1996). Defecation habits of cats and dogs and contamination by *Toxocara canis* in public park sandpits. American Journal of Tropical Medicine and Hygiene *54*, 122 - 126.

van Knapen F., and Buijs J. (1993). Diagnosis of *Toxocara* infection. In *Toxocara* and *Toxocariasis*: Clinical, epidemiological and molecular perspectives, J. W. Lewis and R. M. Maizels, eds. (London: The Institute of Biology and the British Society for Parasitology), pp. 49 - 53.

Wasilewski M. M., Lim K. C., Phillips J., and McKerrow J. H. (1996). Cysteine protease inhibitors block *schistosome* hemoglobin degredation in vitro and decrease worm burden and egg production in vivo. Molecular and Biochemical Parasitology *81*, 179-189.

Whalen R. (1998). The DNA vaccine web.
http://www.genweb.com/DNAvax/dnavax.html

Whitmarsh A. J., Shore P., Sharrocks A. D., and Davis R. J. (1995). Integration of MAP kinase signal transduction pathways at the serum response element. Science *269*, 403-407.

Wilson W., R., Tuan R. S., Shepley K. J., Freedman D. O., Greene B. M., Awadzi K. and Unnasch T. R. (1994). The *Onchocerca volvulus* homologue of the multifunctional polypeptide protein disulfide isomerase. Molecular and Biochemical Parasitology *68*, 103-117.

Xu S.-Q., Tang D., Chamderlain S., Pronk G., Masiarz F. R., Kaur S., Prisco M., Zanocco-Marani T. and Baserga R. (1998). The Granulin/Epithelin Precursor Abrogates the Requirement for the Insulin-like Growth Factor 1 Receptor for Growth *in vitro*. The Journal of Biological Chemistry *273*, 20078 - 20083.

Yokoe H. and Anholt R. H. (1993). Molecular cloning of olfactomedin, an extracellular matrix protein specific to olfactory neuroepithelium. Proceedings of the National Academy of Science *90*, 4655 - 4659.

Yorke W. and Maplestone P. A., (1926). The Nematode Parasites Of Vertebrates, 1 Edition Edition (Philidelphia: P. Blakiston's & CO.).

Zeng, W., and Donelson, J., E. (1992). The actin genes of *Onchocerca volvulus*. Molecular and Biochemical Parasitlogy *55*, 207-216.

Zhong-Jin L., Guang-xi W., Chong-li Y., Chuan-hua L., Sheng-wen C. and Ling L. (1999). Detection of circulating antigens and antibodies in *Toxocara canis* infection among children in Chengdu, China. Journal of Parasitology 85(2), 252-256

Zhou J., Gao G., Crabb J. W. and Serrero G. (1993). Purification of an Autocrine Growth Factor Homologous with Mouse Epithelin Precursor from a Highly Tumourigenic Cell Line. The Journal of Biological Chemistry *268, No. 15*, 10863 - 10869.

Zuckerman B. M., Kahane I. and Himmelhoch S. (1979). *Caenorhabditis briggsae* and *C. elegans*: Partial Characterisation of Cuticle Surface Carbohydrates. Experimental Parasitology *47*, 419-424.

# Appendix

Chapter 2: Materials and Methods (pp.39 )

Solution M (Main Gel)

| | | |
|---|---|---|
| 480 | ml | 1M HCL |
| 363.3 | g | Tris base |
| 8 | g | Sodium dodecyl sulphate (SDS) |

Adjust pH to 8.9, make up volume to 1000 ml with water.

Solution S (Stack Gel)

| | | |
|---|---|---|
| 121.1 | g | Tris base |
| | | pH to 6.8 with HCL |
| 8 | g | SDS |

# Publications

Title: Identification of Abundantly Expressed Novel and Conserved Genes from the Infective larval Stage of *Toxocara canis* by an Expressed Sequence Tag Strategy.
Authors: **Kevin K. A. Tetteh**, Alex Loukas, Cindy Tripp and Rick Maizels.
Journal: Infection and Immunity
Publication status: September 1999

Title: A novel C-type lectin secreted by a tissue-dwelling parasitic nematode.
Authors: Alex Loukas, Nicholas Mullin, **Kevin K. A. Tetteh**, Luc Moens, and Rick Maizels.
Journal: Current Biology
Publication Status: August 1999

INFECTION AND IMMUNITY, Sept. 1999, p. 4771–4779 Vol. 67, No. 9
0019-9567/99/$04.00+0


# Identification of Abundantly Expressed Novel and Conserved Genes from the Infective Larval Stage of *Toxocara canis* by an Expressed Sequence Tag Strategy

KEVIN K. A. TETTEH,[1†] ALEX LOUKAS,[1‡] CINDY TRIPP,[2] AND RICK M. MAIZELS[1]*

*Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, Scotland, United Kingdom,[1] and Heska Corporation, Fort Collins, Colorado 80525[2]*

Received 2 March 1999/Returned for modification 13 May 1999/Accepted 1 June 1999

**Larvae of *Toxocara canis*, a nematode parasite of dogs, infect humans, causing visceral and ocular larva migrans. In noncanid hosts, larvae neither grow nor differentiate but endure in a state of arrested development. Reasoning that parasite protein production is orientated to immune evasion, we undertook a random sequencing project from a larval cDNA library to characterize the most highly expressed transcripts. In all, 266 clones were sequenced, most from both 3′ and 5′ ends, and similarity searches against GenBank protein and dbEST nucleotide databases were conducted. Cluster analyses showed that 128 distinct gene products had been found, all but 3 of which represented newly identified genes. Ninety-five genes were represented by a single clone, but seven transcripts were present at high frequencies, each composing >2% of all clones sequenced. These high-abundance transcripts include a mucin and a C-type lectin, which are both major excretory-secretory antigens released by parasites. Four highly expressed novel gene transcripts, termed *ant* (abundant novel transcript) genes, were found. Together, these four genes comprised 18% of all cDNA clones isolated, but no similar sequences occur in the *Caenorhabditis elegans* genome. While the coding regions of the four genes are dissimilar, their 3′ untranslated tracts have significant homology in nucleotide sequence. The discovery of these abundant, parasite-specific genes of newly identified lectins and mucins, as well as a range of conserved and novel proteins, provides defined candidates for future analysis of the molecular basis of immune evasion by *T. canis*.**

*Toxocara canis* is a common nematode parasite of dogs and related species. Adult worms live in the gastrointestinal tract, releasing eggs which enter the environment by the fecal route. Within the eggs, larval *T. canis* develop over a 2- to 3-week period (27). Embryonated eggs are then infective if ingested by a new host, as larvae hatch in the stomach and penetrate the epithelial layer. *T. canis* larvae show a remarkable lack of host specificity, infecting a wide range of taxa, including humans (27, 28). In the definitive (canid) host, larvae may mature by migrating to the intestine and developing to the adult stage; such maturation is favored in pups and in reproducing females (40). In most other dogs and in all paratenic hosts such as humans, development is arrested at the larval stage.

The arrested stage is remarkable for surviving for as long as 9 years in vivo (7), without reproduction or differentiation and without succumbing to attack by the host immune system. This diapausal state is mirrored in vitro, where larvae survive for many months in serum-free medium, secreting quantities of excretory-secretory antigens which have been characterized in biochemical (5, 50, 58) and immunological (44, 45, 56) terms.

We hypothesized that in the absence of cell division, tissue growth, or gametogenesis, a significant proportion of larval protein production, and therefore mRNA, is likely to be directed at immune evasion. To characterize abundant mRNAs, we conducted a small-scale expressed sequence tag (EST) project. EST sequencing was pioneered for mammalian cells (1, 2) and *Caenorhabditis elegans* (49) and is an important component of major parasite sequencing projects (15, 18, 22, 23, 46, 64). By this means, we have identified a series of abundantly expressed genes from larval *T. canis*, among which are likely to be important mediators of parasite immune evasion.

* Corresponding author. Mailing address: ICAPB, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, Scotland, United Kingdom. Phone: (44) 131 650 5511. Fax: (44) 131 650 5450. E-mail: r.maizels@ed.ac.uk.

† Present address: London School of Hygiene and Tropical Medicine, London WC1E 7HT, United Kingdom.

‡ Present address: Molecular Parasitology Unit, Queensland Institute of Medical Research, Queensland 4029, Australia.

## MATERIALS AND METHODS

**cDNA library.** *T. canis* larvae were hatched and maintained in vitro as previously described (20, 50) for a period of 6 months, with weekly changes of serum-free medium. From 200,000 cultured *T. canis* larvae, using a single-step guanidine-phenol-chloroform extraction, 265 μg of total RNA was recovered, from which 6 μg of poly(A)$^+$ RNA was isolated by oligo(dT) chromatography. cDNA synthesized from this mRNA was unidirectionally cloned into the Uni-Zap XR phage vector, using packaging extracts from Stratagene. The amplified library contained $1.9 \times 10^9$ PFU/ml with 91% recombinants. The possibility of host contamination was essentially eliminated because eggs were first incubated in vitro in formalin, and once hatched, larvae were cultured in serum-free RPMI 1640 medium.

**Isolation of cDNA clones and in vivo excision of phagemids.** Single clones were randomly picked phage from the plated out cDNA library. Phagemids were rescued in vivo by using ExAssist helper phage and nonsuppressing *Escherichia coli* SOLR (Stratagene).

**Selection and maintenance of clones.** Plasmids were prepared from overnight cultures by using a Qiaprep Spin Miniprep kit (Qiagen) and stored at −20°C. Insert sizes were determined by PCR using vector primers (M13 reverse and M13 forward primers or T3 and T7 primers). In the few cases where insert sizes could not be determined by PCR, restriction digestion of purified recombinant plasmids was performed with *Xba*I and *Xho*I (Promega). All clones are available to the research community on request.

**Sequencing.** Plasmid templates were sequenced by using dye terminator cycle sequencing chemistry with Amplitaq DNA polymerase (FS enzyme) on an ABI 377 automated sequencer (Applied Biosystems, Foster City, Calif.). Both 5′ and 3′ ends were sequenced in some cases where 5′ sequences gave high-quality sequence through the poly(A) tail or where the 5′ sequence showed unequivocal identity with a previously characterized *T. canis* clone.

**Analysis.** SeqEd version 1.0.3 (Applied Biosystems) was used to edit out vector sequence and flanking sequences. Sequences were aligned by using AssemblyLign and MacVector 6.0 software (Oxford Molecular). Nonredundant database searches used ungapped BLAST (basic local alignment search tool) (3) on

the National Center for Biotechnology Information server. Nucleotide sequences were subjected to BLASTX searches against the GenBank nr (nonredundant) protein database. Nucleotide sequence searches used BLASTN (on both nr and dbEST databases), and deduced protein sequence from any unambiguous open reading frame was used to search with TBLASTN (against the nr nucleotide database).

**Nomenclature.** Gene naming follows the convention for nematodes (11, 14) of a three-letter name and a number, and a prefix indicating the source organism, in this case *Tc*. Genes are denoted in italics; proteins are capitalized. Three sets of interim gene names were used where no functional homology exists: *Tc-not* (novel transcript), *Tc-ant* (abundant novel transcript, where ≥5 clones of 263 are identical), and *Tc-huf* (homolog of unknown function, where similar sequences have been found in other nematode species). For interim gene designations, the number of the EST clone first sequenced is retained; for cDNAs assigned functional names, clones are generally numbered sequentially (e.g., *Tc-ctl-1* and *-2* for C-type lectins), except where numbering is significant in other organisms, principally for ribosomal proteins (e.g., *Tc-rps-4*, *-5*, and *-9* to conform with established nomenclature).

**Database deposition.** All sequences have been deposited in NCBI dbEST (17) with separate entries for 5′ and 3′ reads.

## RESULTS AND DISCUSSION

**Sequencing and similarity searches.** A total of 261 clones, containing inserts ranging from 128 to 2,050 bp, were taken for analysis once nonrecombinant and chimeric clones were discarded. All were sequenced from the 5′ end, and 218 were also sequenced from the 3′ end. Most similarities were found, or were stronger, with the 5′ sequences, but a significant minority of similarity relationships were found only with 3′ sequence reads. In general, a probability value of $10^{-6}$ was sought as a minimum degree of similarity.

**Clustering analysis.** EST sequences were clustered on the basis of nucleotide identity. It was noted that in some of the larger clusters, identity of the 3′ sequences was critical to correctly group differentially truncated clones with nonoverlapping 5′ sequences.

As a result of these analyses, a total of 128 distinct gene products were identified. Of these, only three, mucin 1 (*Tc-muc-1* [25]), phosphatidylethanolamine-binding protein (*Tc-peb-1* [26]), and the 60S ribosomal protein L3 (*Tc-rpl-3* [51]), have been previously characterized. Approximately 50% (65/128) of the total number of genes have informative similarity to genes of known function from other species, a further 17 clones have database homologs of unknown function, and 47 genes (37%) show no similarity to known genes; among this last class, designated novel genes, 4 were classified as abundant transcripts (see below).

**Abundant transcripts.** A small number of transcripts are heavily represented in the library. Just 8 transcripts (6.2% of genes) account for 102 clones (39.1%) sequenced, while the 20 most abundant (all those isolated three or more times) accounted for 143 clones (54.8%). Table 1 presents the transcripts characterized in order of frequency, with the most highly sampled clone being cytochrome *c* oxidase subunit II (25/261 = 9.6% of clones), which is a mitochondrial DNA-encoded gene in other nematodes (36, 52). The second most common clone is a C-type lectin (16/261 = 6.1%) which in work to be published elsewhere we demonstrate encodes the major surface and secreted glycoprotein of *T. canis* larvae, TES-32 (42).

**Abundant novel transcripts.** Unexpectedly, four more abundant clones are all novel, with no similarities in the database or, in the coding regions, to each other. These transcripts each represent between 2.7 and 5.7% of ESTs and together account for more than 18% of the library. They have consequently been designated *ant* genes and retain the number of the EST clone from which they were first isolated (*ant-003*, *-005*, *-030*, and *-034*).

The four *ant* genes differ in length and composition, but all are 1.6 kb or more in length. The characterization of their full-length sequences, and of the protein products encoded by these genes, is currently under way, as none of the ESTs isolated include 5′ methionine start codons. Figure 1 presents a map of the EST clones isolated for each *ant* gene. From this it can be seen that 3′ sequencing proved essential in identifying all members of the cluster, because 5′ sequences do not in all cases overlap. Because the multiple clones all have different 5′ termini, the abundance of these transcripts appears not to be an artifactual amplification of single clones in the construction of the library.

TABLE 1. Frequency of transcripts

| Transcript (gene) | No. of clones sequenced | Frequency of transcripts (%) |
|---|---|---|
| Cytochrome *c* oxidase subunit II (*cox-2*) | 25 | 9.6 |
| C-type lectin, TES-32 (*ctl-1*) | 16 | 6.1 |
| Abundant novel transcript 003 (*ant-003*) | 15 | 5.7 |
| Abundant novel transcript 034 (*ant-034*) | 13 | 5.0 |
| Abundant novel transcript 030 (*ant-030*) | 13 | 5.0 |
| Abundant novel transcript 005 (*ant-005*) | 7 | 2.7 |
| Mucin 1, TES-120 (*muc-1*) | 7 | 2.7 |
| Phosphatidylethanolamine-binding protein, TES-26 (*peb-1*) | 5 | 1.9 |
| Novel transcript 095 (*not-095*) | 4 | 1.5 |
| Protein tyrosine phosphatase (*ptp-1*) | 4 | 1.5 |
| 60S ribosomal protein L19 (*rpl-19*) | 4 | 1.5 |
| ADP/ATP translocase (*aat-1*) | 3 | 1.1 |
| ASP homolog 1/TC-CRISP (*vah-1*) | 3 | 1.1 |
| Mucin 2 (*muc-2*) | 3 | 1.1 |
| Novel transcript 018 (*not-018*) | 3 | 1.1 |
| Novel transcript 120 (*not-120*) | 3 | 1.1 |
| 60S ribosomal protein P0 (*rpp-0*) | 3 | 1.1 |
| Superoxide dismutase 3 (*sod-3*) | 3 | 1.1 |
| Tubulin alpha-3 chain (*tua-3*) | 3 | 1.1 |
| 14 transcripts present as two clones[a] | 28 | 10.7 |
| 95 transcripts present as single clones | 96 | 36.7 |
| Total no. of clones sequenced | 261 | |

[a] The 14 genes present as two clones are *aki-1*, *ash-2*, *cap-1*, *huf-001*, *irp-1*, *muc-3*, *not-069*, *not-216*, *ofm-1*, *pam-1*, *pcc-1*, *rpl-7a*, *rpl-9*, and *sod-4*. Details for these, apart from *not* genes, are given in Table 3; details for *ash-2* and *huf-001* are in Table 2.

**Homology in 3′ UTRs of the four *ant* genes.** Although *ant-003*, *-005*, *-030*, and *-034* showed no similarities between coding regions, the 3′ untranslated regions (UTRs) of all four genes bore significant levels of identity. It is notable also that none of these 3′ ends contain consensus polyadenylation motifs such as AATAAA or similar sequence. This is not unprecedented among nematode genes: a recent survey of *C. elegans* 3′ UTRs found that 7% of mRNAs bore no identifiable polyadenylation signal (16).

An alignment of the 3′ coding regions and UTRs of these four transcripts shows little identity in coding region nucleotides (Fig. 2A) or amino acids (not shown) but similar sequences in all four genes immediately after the stop codon (Fig. 2B). No similar sequences were found in any other *T. canis* genes or in genes from other organisms such as *C. elegans*. In *C. elegans*, there are examples of 3′ UTR motifs associated with repression of translation in genes such as *tra-2* and *fem-3* in germ line cell differentiation (4, 68). This is unlikely to be a useful parallel for the *T. canis ant* genes, as there is no similarity in 3′ UTR sequence between the two species and there is no gonadal development in the larval stage of *T. canis*. Translational suppression of the *ant* genes would be a surprising event in light of their unusually high level of transcription.

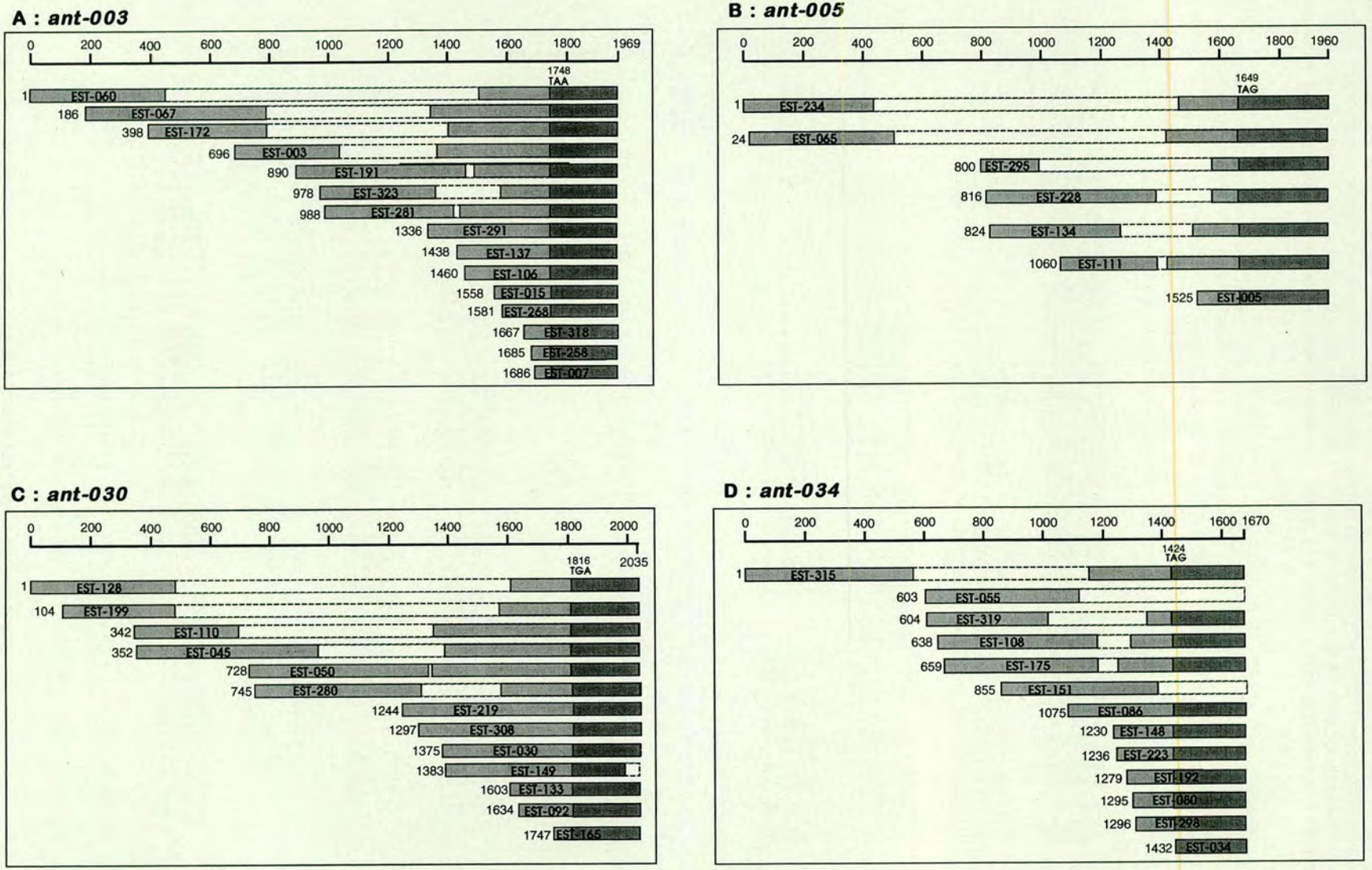



FIG. 1. Alignments of clones corresponding to *ant-003*, *-005*, *-030*, and *-034*. Segments sequenced in single-pass reactions are lightly shaded; unsequenced portions are shown unshaded in broken lines. Darker shading corresponds to 3′ UTRs. The position and codon of the stop signal are indicated for each transcript. Numbers to the left of the bars indicate the approximate start of the cDNA relative to the longest clone. No start codons have yet been identified for any of these genes, and hence numbering is provisional. Note that without 3′ sequence data, the longer EST clones belonging to *ant-005* and *ant-034* would not have been recognized.

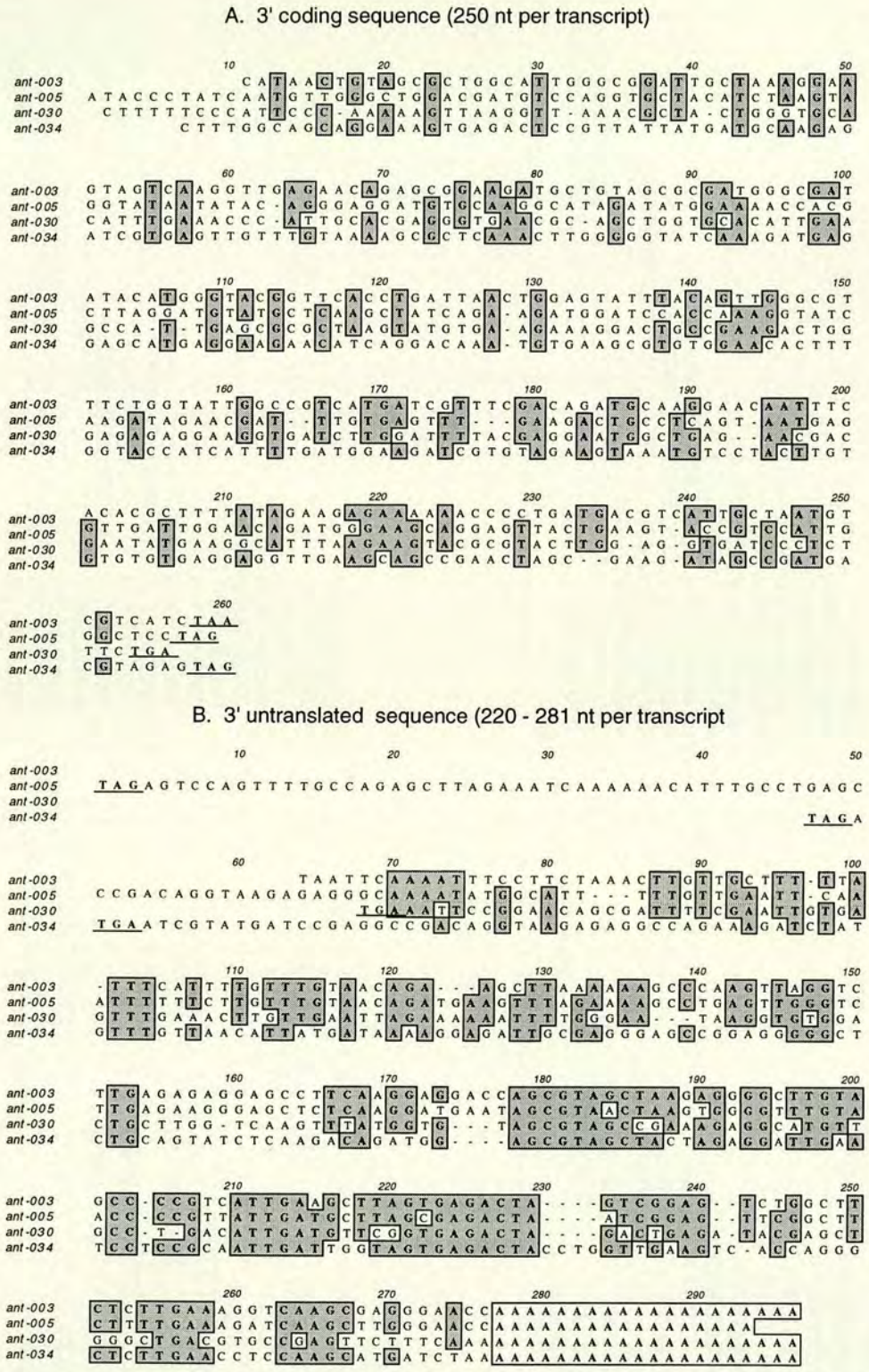



FIG. 2. Alignment of 250-nucleotide (nt) 3′-terminal coding sequences (A) and complete 3′ UTRs (B) of the four *ant* genes; identical settings were used for both alignments in MacVector ClustalW alignment (gap penalty = 10.0, extend gap penalty = 5.0). The stop codons at the end of the coding sequence and the beginning of the UTR are underlined. Bases identical in three or four of the sequences are boxed and shaded. The polyadenylated tail is shown boxed without shading.

**Homologs of genes of unknown function.** Sixteen clones showed significant similarities to known nematode sequences with no assigned function (Table 2). These were all designated *huf* genes, retaining the number of the corresponding EST clone. One clone, *Tc-huf-001*, contains a tandem array of four blocks of 36 amino acids each containing six cysteines in identical alignment. This motif, termed the NC6 (26) or SXC (12) domain, is found in *T. canis* and *C. elegans* proteins, particu-

TABLE 2. Homologs to proteins of unknown function in other nematodes[a]

| Gene | Similarity to: | *P* | EST name | Insert size | Read | | GenBank accession no. | |
|---|---|---|---|---|---|---|---|---|
| | | | | | 5′ | 3′ | 5′ | 3′ |
| *Tc-huf-001* | NC6/SXC domain | | Tc-EST-001 | 800 | 764 | 666 | AA683457 | AA683458 |
| *Tc-huf-012* | *C. elegans* R53.5 gene | 3.8E-06 | Tc-EST-012 | 850 | 599 | 342 | AA563573 | AI374481 |
| *Tc-huf-053* | *C. elegans* C13A10 gene | 2.8E-38 | Tc-EST-053 | 1,000 | 589 | 602 | AA610971 | AA610972 |
| *Tc-huf-062* | *C. elegans* F41D9.1 gene | 4.3E-42 | Tc-EST-062 | 760 | 482 | 521 | AA612566 | AA907959 |
| *Tc-huf-159* | *C. elegans* C56E6.5 gene | 7.6E-15 | Tc-EST-159 | 1,850 | 387 | 288 | AA836705 | AA836706 |
| *Tc-huf-169* | *C. elegans* K02E10.6 gene | 6.2E-17 | Tc-EST-169 | 1,050 | 466 | 393 | AA874757 | AA874758 |
| *Tc-huf-252* | *C. elegans* Y57G11C.12 gene | 7.7E-30 | Tc-EST-252 | 1,150 | 547 | 558 | AA684533 | AA684534 |
| *Tc-huf-264* | *C. elegans* B0034.3 gene | 1.4E-25 | Tc-EST-264 | 1,420 | 589 | 444 | AA684531 | AA684532 |
| *Tc-huf-287* | *C. elegans* K12H4.5 gene | 5.3E-23 | Tc-EST-287 | 452 | 429 | 416 | AA875794 | AA979711 |
| *Tc-huf-296* | *C. elegans* T24F1.6 gene | 1.9E-11 | Tc-EST-296 | 2,300 | 259 | 491 | AA875806 | AA875807 |
| *Tc-huf-302* | *B. malayi* MBAFCE2B09T3 | 2.3E-29 | Tc-EST-302 | 760 | 283 | 621 | AA879342 | AA879343 |
| *Tc-huf-316* | *C. elegans* T24D8.5 gene | 3.7E-21 | Tc-EST-316 | 720 | 542 | 578 | AA874711 | AA874712 |
| *Tc-huf-325* | *O. volvulus* L3CAN05C08SK | 7.4E-47 | Tc-EST-325 | 1,250 | 370 | 351 | AA879368 | AA879369 |
| *Tc-vah-1* | *C. elegans* T05A10.5 venom allergen/ASP homolog | 4.6E-15 | Tc-EST-269 | 1,870 | 459 | 590 | AI 083044 | AI 083045 |
| *Tc-vah-2* | *B. malayi* venom allergen/ASP homolog | 1.0E-13 | Tc-EST-303 | 1,600 | 550 | 699 | AI083048 | AI083049 |
| *Tc-vah-3* | *B. malayi* venom allergen/ASP homolog | 4.7E-16 | Tc-EST-249 | 650 | 507 | 512 | AI083050 | AI083051 |

[a] All cDNAs were isolated as single transcripts except *vah-1* (three clones), *vah-2* (two clones), and *huf-001* (two clones).

larly those associated with nematode surfaces, but the function of the array is not known.

In addition, we found three genes which show similarity to *Ancylostoma* secreted protein (ASP), which is associated with larval activation and development in other nematode species (10, 30, 31, 61). However, the *T. canis* stage from which the cDNA library was made is developmentally arrested and is not analogous to activated hookworm larvae. This gene family shows similarity to hymenopteran venom allergens and, in common with members reported for *Brugia malayi* and *Onchocerca volvulus*, has been designated *vah* (venom allergen homolog). *Tc-vah-1*, previously reported as Tc-CRISP (cysteine-rich secreted protein) by Liu (39a), is remarkable for containing two NC6/SXC domains in tandem with a VAH homology unit. Further characterization of these clones, *Tc-vah-1*, *-2*, and *-3* is in progress (51a).

**Other novel genes.** A total of 43 additional transcripts were found to have no significant similarities to any other sequences deposited in GenBank nr protein and dbEST databases. These were present at between one and four copies in the 261-member data set from *T. canis* and are designated *not* genes. Further studies on selected clones, such as *not-018*, for which antibodies to the protein product strongly recognize *T. canis* larval excretory-secretory antigens (61a) are ongoing.

**Homologs of known genes.** Sixty-five genes with database homologs were found; these are listed in Table 3. There are 23 metabolic and respiratory enzymes but remarkably few structural proteins (only actin, calponin, and α-tubulin) and no DNA replication proteins, consistent with the arrested state of the larval parasite. The various categories of genes are described below.

**Mucins.** A mucin gene, *Tc-muc-1*, has previously been reported to be abundantly expressed by *T. canis* larvae (25). Consistent with this, seven clones of *Tc-muc-1* were recorded (2.7% of the library). Three new mucin genes were found among the EST clones, each of which contains similar mucin domains and flanking NC6/SXC domains (26). The mucins differ in the composition of repeat motifs, and in the number and positions of NC6/SXC domains, and work in progress indicates that all are members of the TES-120 glycoprotein family associated with the parasite surface coat (40a).

**C-type lectins.** One of the most abundant transcripts (16/261) was found to correspond to peptide sequence obtained from TES-32, a prominent secreted glycoprotein. These ESTs showed weak similarity to C-type lectins, but the homology was firmly established from full-length sequence. A detailed analysis of the functional lectin properties of TES-32/*Tc-ctl-1* has been submitted for publication (42). Two variants of this sequence, designated *Tc-ctl-2* and *Tc-ctl-3*, were also noted.

**Proteases.** Proteolytic enzymes have been prominent in most studies of parasitic helminths at the biochemical (32) and molecular (43) levels. Three transcripts, with strong similarities to cathepsin L (41), cathepsin Z (43), and asparaginyl endopeptidase (19), were each present as single copies. Full-length sequences of the cathepsins L (41) and Z (22a) have been determined. A protease inhibitor similar to the aspartyl protease inhibitor of *Ascaris* and *Brugia* (Bm33) (21) was also isolated as a single clone.

**Transporters and receptors.** One clone homologous to the acetylcholine receptors reported from other nematodes (24) has been isolated. A relatively frequent transcript (5/263) encodes *Tc*-PEB-1, a previously identified phosphatidylethanolamine-binding protein (26) which is present in parasite secretions as a 26-kDa protein (TES-26).

**Structural proteins.** Few structural proteins were identified in the EST data set, probably reflecting the arrested state of this parasite stage. Actin (67) and calponin (33) have been sequenced from other parasitic nematodes. For tubulins, most attention has focused on the β-tubulins (29), with which benzimidazole resistance is associated. *Tc-tua-1* has strong similarity to α-tubulins from *Haemonchus contortus* (37) and *C. elegans*.

**Protein synthesis and modification.** Two genes essential for protein synthesis (elongation factors 1a and 1b) were identified, as well as peptidyl-glycine alpha-hydroxylating monooxygenase, which modifies the C termini of peptides. Protein disulfide isomerase is a well-known requisite for protein folding and correct formation of disulfide bonds. In *C. elegans*, the *pdi* gene is found in an operon with one cyclophilin gene (55); a homolog from *O. volvulus* has also been characterized (65). Peptidyl-prolyl *cis-trans*-isomerase is similarly essential for correct protein conformation; the *Tc-ppi-1* gene product is related to the FK506-binding proteins of mammals and not to the multigene cyclophilin family of peptidyl-prolyl *cis-trans* isomerases described for *C. elegans* (55, 57).

**Glycolysis, respiration, and other metabolic and citric acid cycle enzymes.** Some 18 distinct metabolic enzymes had very high levels of similarity to *T. canis* ESTs and represent the

TABLE 3. Homology of *T. canis* ESTs with known genes[a]

| Gene | Gene product | Closest species | *P* | Clone no. | Insert size | Read | | GenBank accession no. | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 5′ | 3′ | 5′ | 3′ | Full-length verified sequence |
| Mucins | | | | | | | | | | |
| *Tc-muc-1* | Mucin 1, surface coat TES-120 | *T. canis* | 6.3E-104 | Tc-EST-087[b] | 810 | 557 | 539 | AA583105 | AA907963 | U39815 |
| *Tc-muc-2* | Mucin 2 | *T. canis* | 3.8E-49 | Tc-EST-130[b] | 990 | 612 | 694 | AA583106 | AA583107 | AF167707 |
| *Tc-muc-3* | Mucin 3 | *T. canis* | 2.6E-36 | Tc-EST-162[b] | 1,010 | 603 | 555 | AA583111 | AA583112 | AF167708 |
| *Tc-muc-4* | Mucin 4 | *T. canis* | 1.6E-45 | Tc-EST-186 | 560 | 553 | 586 | AA583108 | AA583109 | AF167709 |
| Lectins | | | | | | | | | | |
| *Tc-ctl-1* | C-type lectin/TES-32 glycoprotein | | | Tc-EST-209 | 830 | 615 | 393 | AA874713 | AA874714 | AF041023 |
| *Tc-ctl-2* | C-type lectin-2/variant of TES-32 | | | Tc-EST-036 | 780 | 596 | 618 | AI078880 | AI078881 | |
| *Tc-ctl-3* | C-type lectin-3/variant of TES-32 | | | Tc-EST-155 | 610 | 590 | 737 | AI080917 | AI080918 | |
| Proteases and protease inhibitors | | | | | | | | | | |
| *Tc-aep-1* | Asparaginyl endopeptidase | *Homo sapiens* | 3.5E-28 | Tc-EST-262 | 1,300 | 520 | 648 | AA825116 | AA875769 | |
| *Tc-api-1* | Aspartyl proteinase inhibitor | *C. elegans* | 1.5E-09 | Tc-EST-096 | 650 | 570 | 638 | AA820019 | AA820020 | |
| *Tc-cpl-1* | Cathepsin L cysteine proteinase | *C. elegans* | 2.3E-40 | Tc-EST-013 | 1,150 | 604 | 390 | AA563518 | AI078878 | U53172 |
| *Tc-cpz-1* | Cathepsin Z proteinase precursor | *O. volvulus* | 1.2E-73 | Tc-EST-205 | 1,000 | 568 | 701 | AI080921 | AI080922 | AF143817 |
| Transporters and receptors | | | | | | | | | | |
| *Tc-aat-1* | ADP/ATP translocase | *C. elegans* | 1.9E-69 | Tc-EST-041[b] | 750 | 515 | 663 | AA603955 | AA683466 | |
| *Tc-acr-1* | Acetylcholine receptor (*tar-1*) | *Trichostrongylus colubriformis* | 1.2E-35 | Tc-EST-121 | 1,300 | 529 | 547 | AA835586 | AA835587 | |
| *Tc-peb-1* | Phosphatidylethanolamine-binding protein (TES-26) | *T. canis* | 6.1E-76 | Tc-EST-164[b] | 1,200 | 406 | 459 | AA864177 | AA864178 | U29761 |
| Structural proteins | | | | | | | | | | |
| *Tc-act-1* | Actin | *C. elegans* | 1.3E-81 | Tc-EST-097 | 1,450 | 524 | | AA820021 | | |
| *Tc-cap-1* | Calponin | *Meloidogyne javanica* | 4.2E-21 | Tc-EST-292[b] | 620 | 462 | 537 | AA875799 | AA907961 | |
| *Tc-tua-1* | Tubulin alpha-3 chain | *Cricetulus griseus* | 1.5E-69 | Tc-EST-297[b] | 1,500 | 400 | 439 | AA979712 | AA907952 | |
| Protein synthesis and modification factors | | | | | | | | | | |
| *Tc-efa-1* | Elongation factor 1a | *C. elegans* | 3.5E-84 | Tc-EST-300 | 1,600 | 498 | 572 | AA879374 | AA879375 | |
| *Tc-efb-1* | Elongation factor 1b | *C. elegans* | 3.1E-42 | Tc-EST-081 | 830 | 355 | 756 | AA888759 | AA888760 | |
| *Tc-pam-1* | Peptidyl-glycine alpha-hydroxylating monooxygenase | *Drosophila melanogaster* | 1.1E-07 | Tc-EST-200[b] | 420 | 385 | 357 | AA874705 | AI080786 | |
| *Tc-pdi-1* | Protein disulfide isomerase | *Gallus gallus* | 5.9E-18 | Tc-EST-235 | 474 | 445 | 468 | AA874740 | AA874741 | |
| *Tc-ppi-1* | Peptidyl-prolyl *cis-trans* isomerase | *Mus musculus* | 3.2E-15 | Tc-EST-019 | 1,200 | 319 | | AA563572 | | |
| *Tc-ptp-1* | Protein tyrosine phosphatase | *M. musculus* | 5.9E-49 | Tc-EST-118 | 1,590 | 589 | 628 | AA820013 | AA820014 | |
| Glycolysis, respiration, and other metabolic enzymes | | | | | | | | | | |
| *Tc-aca-1* | Acetyl coenzyme A acetyltransferase | *C. elegans* | 9.8E-33 | Tc-EST-059 | 626 | 626 | 560 | AA610965 | AA907958 | |
| *Tc-aki-1* | Adenylate kinase isoenzyme I | *G. gallus* | 9.2E-39 | Tc-EST-328 | 800 | 452 | 505 | AA873889 | AA873920 | |
| *Tc-ald-1* | Aldehyde dehydrogenase | *Bos bovis* | 1.3E-45 | Tc-EST-299 | 1,700 | 559 | 360 | AA979715 | AA979716 | |
| *Tc-ata-1* | Alanine aminotransferase | *C. elegans* | 4.0E-42 | Tc-EST-294 | 1,440 | 472 | 357 | AA875802 | AA875803 | |
| *Tc-cgl-1* | Cystathionine gamma lyase | *C. elegans* | 6.4E-54 | Tc-EST-090 | 750 | 555 | 685 | AA583117 | AA583118 | |
| *Tc-cyb-5* | Cytochrome $b_5$ | *H. sapiens* | 4.5E-10 | Tc-EST-032 | 406 | 406 | | AA569419 | | |
| *Tc-fba-1* | Fructose biphosphate aldolase | *O. volvulus* | 2.1E-98 | Tc-EST-183 | 1,010 | 480 | 533 | AA915868 | AA915869 | |
| *Tc-fdh-1* | Fumarate dehydrogenase precursor | *H. sapiens* | 1.7E-26 | Tc-EST-237 | 317 | 317 | 314 | AA285389 | AA874746 | |
| *Tc-frb-1* | Fumarate reductase, cytochrome *b* small subunit | *Ascaris suum* | 2.9E-32 | Tc-EST-083 | 1,560 | 452 | 480 | AA888752 | AA888753 | |
| *Tc-gdh-1* | Glutamate dehydrogenase | *H. contortus* | 4.7E-74 | Tc-EST-026 | 1,740 | 584 | 331 | AA568090 | AI078879 | |
| *Tc-pcc-1* | Propionyl coenzyme A carboxylase | *C. elegans* | 8.6E-64 | Tc-EST-074 | 1,560 | 320 | 399 | AA888750 | AA888751 | |
| *Tc-pck-1* | Phosphoenolpyruvate carboxykinase | *A. suum* | 4.4E-69 | Tc-EST-004 | 2,740 | 599 | 425 | AA470327 | AI374480 | |
| *Tc-pgk-1* | Phosphoglycerate kinase | *Schistosoma mansoni* | 1.8E-41 | Tc-EST-109 | 1,400 | 576 | 805 | AA825117 | AA825118 | |
| *Tc-sdi-1* | Succinate dehydrogenase iron-sulfur protein | *A. suum* | 3.0E-73 | Tc-EST-166 | 630 | 518 | 423 | AA864181 | AI374479 | |
| *Tc-ubo-1* | NADH ubiquinone oxidoreductase 18-kDa subunit | *Bos taurus* | 1.9E-39 | Tc-EST-201 | 521 | 501 | 480 | AA874708 | AI080786 | |

*Continued on following page*

TABLE 3—*Continued*

| Gene | Gene product | Closest species | *P* | Clone no. | Insert size | Read 5′ | Read 3′ | GenBank accession no. 5′ | GenBank accession no. 3′ | GenBank accession no. Full-length verified sequence |
|---|---|---|---|---|---|---|---|---|---|---|
| Presumed mitochondrial proteins | | | | | | | | | | |
| *Tc-atp-6* | ATP synthase A chain (protein 6) | *A. suum* | 1.9E-68 | Tc-EST-064 | 610 | 576 | 557 | AA612565 | AA979708 | |
| *Tc-cox-2* | Cytochrome *c* oxidase II | *A. suum* | 7.1E-75 | Tc-EST-038 | 850 | 496 | 669 | AA569422 | AA569423 | |
| *Tc-ubo-4* | NADH-ubiquinone oxidoreductase chain 4L | *A. suum* | 2.6E-29 | Tc-EST-163 | 220 | 213 | 216 | AA864175 | AA864176 | |
| Antioxidants | | | | | | | | | | |
| *Tc-sod-2* | Superoxide dismutase 2 | *C. albicans* | 1.5E-23 | Tc-EST-084 | 850 | 401 | 340 | AI078885 | AI078886 | |
| *Tc-sod-3* | Superoxide dismutase 3 | *Zea mays* | 2.1E-18 | Tc-EST-100 | 900 | 449 | 608 | AI080774 | AI080775 | |
| *Tc-sod-4* | Superoxide dismutase 4 | *C. elegans* | 3.6E-40 | Tc-EST-143 | 804 | 630 | 520 | AI080778 | AI080779 | |
| *Tc-sod-5* | Superoxide dismutase 5 | *C. elegans* | 2.4E-39 | Tc-EST-266 | 900 | 731 | 730 | AI080782 | AI080783 | |
| Ribosomal proteins | | | | | | | | | | |
| *Tc-rpl-10* | 60S ribosomal protein L10 | *C. elegans* | 8.6E-44 | Tc-EST-271 | 800 | 321 | 590 | AA875777 | AA875778 | |
| *Tc-rpl-19* | 60S ribosomal protein L19 | *M. musculus* | 2.1E-60 | Tc-EST-244[b] | 900 | 439 | 555 | AA874749 | AA874750 | |
| *Tc-rpl-3* | 60S ribosomal protein L3 | *T. canis* | 5.5E-60 | Tc-EST-257 | 500 | 498 | 457 | AA875762 | AI080789 | P49149 |
| *Tc-rpl-31* | 60S ribosomal protein L31 | *H. sapiens* | 4.8E-49 | Tc-EST-072 | 580 | 440 | 491 | AA728649 | AA728650 | |
| *Tc-rpl-37a* | 60S ribosomal protein L37 | *S. mansoni* | 3.6E-34 | Tc-EST-267 | 400 | 372 | 381 | AA875772 | AA875773 | |
| *Tc-rpl-7* | 60S ribosomal protein L7A | *H. sapiens* | 7.2E-29 | Tc-EST-054[b] | 750 | 513 | 560 | AA610961 | AA907956 | |
| *Tc-rpl-9* | 60S ribosomal protein L9 | *H. contortus* | 1.1E-54 | Tc-EST-182[b] | 630 | 396 | 480 | AA873890 | AI080784 | |
| *Tc-rpp-0* | 60S ribosomal protein p0 | *C. elegans* | 4.4E-35 | Tc-EST-122[b] | 1,100 | 476 | 552 | AA835588 | AA835589 | |
| *Tc-rpp-2* | 60S ribosomal protein P2 | *B. malayi* | 3.6E-08 | Tc-EST-088 | 288 | 288 | 275 | AA907964 | AA907965 | |
| *Tc-rps-4* | 40S ribosomal protein S4 | *C. griseus* | 4.1E-61 | Tc-EST-006 | 750 | 582 | | AA563574 | | |
| *Tc-rps-5* | 40S ribosomal protein S5 | *C. elegans* | 1.4E-70 | Tc-EST-215 | 610 | 571 | 675 | AA874719 | AA874720 | |
| *Tc-rps-8* | 40S ribosomal protein S8 | *H. sapiens* | 7.4E-66 | Tc-EST-277 | 640 | 481 | 600 | AI080929 | AI080930 | |
| *Tc-rps-9* | 40S ribosomal protein S9 | *Rattus norvegicus* | 8.6E-38 | Tc-EST-270 | 410 | 381 | | AA875776 | | |
| Others | | | | | | | | | | |
| *Tc-gep-1* | Granulin/epithelin precursor | *M. musculus* | 8.4E-31 | Tc-EST-141 | 1,000 | 654 | 625 | AA836692 | AA836693 | |
| *Tc-hih-4* | Histone H4 | *C. elegans* | 3.5E-60 | Tc-EST-174 | 250 | 444 | 459 | AA557121 | AA873915 | |
| *Tc-lah-1* | Lupus autoantigen homolog | *C. elegans* | 1.5E-22 | Tc-EST-177 | 450 | 607 | 520 | AI083048 | AI083049 | |
| *Tc-mps-1* | Metallopanstimulin (=rps27) | *Strongyloides ratti* | 8.3E-41 | Tc-EST-178 | 1,430 | 367 | 400 | AA873918 | AA979709 | |
| *Tc-ofm-1* | Olfactomedin (F11C3.2) | *C. elegans* | 1.6E-54 | Tc-EST-079[b] | 1,870 | 479 | 628 | AA728643 | AA728644 | |
| *Tc-pth-1* | PC4/TIS7/"interferon-related protein" | *M. musculus* | 4.1E-20 | Tc-EST-075[b] | 1,440 | 616 | 299 | AA618627 | AI078882 | |
| *Tc-tlp-1* | Tubby-like protein | *H. sapiens* | 1.6E-42 | Tc-EST-242 | 1,530 | 506 | 585 | AA874759 | AA874760 | |

[a] Numerous sequences for which closest species listed is not a nematode show a higher BLASTX score with a *C. elegans* sequence for which there is either no assignation or which is described as similar to the nonnematode sequence given here. One ribosomal RNA gene (Tc-EST-052, 5′ AA610968, 3′ AA610969) was also identified.

[b] Longest clone.

major metabolic pathways of glycolysis and aerobic respiration, as well as essential processes such as amino acid synthesis and degradation. Particularly prominent among these is the mitochondrial cytochrome *c* oxidase subunit II (25/261 = 9.7% of all clones). As mitochondrial mRNAs are not polyadenylated, the presence of mitochondrial DNA-encoded sequences in the cDNA library is difficult to interpret quantitatively.

**Antioxidants.** Oxidative stress is highly detrimental to both parasitic (54) and free-living (34) nematodes. In tissue-dwelling parasites, reactive oxygen intermediates from aggressive granulocytes may be countered by expression of antioxidants such as glutathione peroxidase and/or superoxide dismutase (SOD). Previous work characterized a SOD gene (*Tc-sod-1*) from *T. canis* larvae and showed that no glutathione peroxidase activity or gene sequence was detectable in this parasite (47). The EST data set contained seven clones encoding SOD isoforms, all quite distinct from *Tc-sod-1* (approximately 66% divergence in protein sequence) and more similar to *C. elegans* gene F55H2.1. The seven clones represent four distinct isoforms each showing 10 to 20% divergence in nucleotide and deduced amino acid sequence, including two-codon insertions/deletions. This level of divergence and the presence of triplet insertions/deletions led us to designate these four isoforms as separate genes, *Tc-sod-2*, *-3*, *-4*, and *-5*. Confirmation of this assignation is under way.

**Ribosomal proteins.** Twenty ESTs (comprising 13 different genes; 8.4%) encode ribosomal proteins. This is close to the range found with EST projects for other nematodes; for example, 8.5% (1,339/15,811) of *B. malayi* ESTs deposited are ribosomal (12a), as are 5.0% (11/218) of *Necator americanus* ESTs (12a).

**Other proteins.** Some ESTs had homology to mammalian proteins for which a function in nematodes is not obvious. Five particularly interesting findings were noted.

**(i) Granulin/epithelin precursor (*Tc-gep-1*).** Granulins are synthesized as large (500- to 600-amino-acid) precursors from which are derived seven small (~60-amino-acid) 12-cysteine peptides with growth factor-like activity (8, 9, 69). Mammalian and fish kidney epithelial cells are rich sources of these peptides (8, 60), as are human and rodent leukocytes, suggesting that granulins may fulfill cytokine-like functions (6). If this is so, the synthesis of a granulin homolog by *T. canis* larvae may be important in the interaction between parasite and the host immune system.

**(ii) Lupus autoantigen homolog (*Tc-lah-1*).** The lupus autoantigen (also known as Sjögren syndrome type B antigen) is

a highly conserved ribonucleoprotein which is a target of autoantibodies in systemic lupus erythrematosus (39). As autoimmune responses can be initiated by infectious agents (53), the expression of this homolog by *T. canis* may be significant. Similarly, *O. volvulus* expresses the RAL-1 product, which is homologous to the Sjögren syndrome type A antigen (48).

**(iii) Olfactomedin (*Tc-ofm-1*).** A *T. canis* member of the olfactomedin gene family was found. The prototype gene encodes a 57-kDa glycoprotein in the extracellular mucus matrix of the olfactory neuroepithelium of frogs (66), but additional homologs are widely expressed in mammalian brain (35). *Tc-ofm-1* shows maximum similarity to a *C. elegans* homolog. As *T. canis* larvae produce high levels of mucins, including those constituting the surface coat (25, 59), *Tc*-OFM-1 protein may be involved with mucus layers in this parasite.

**(iv) PC4/TIS7/"interferon-related protein" (*Tc-pth-1*).** A *T. canis* gene which is similar to a mammalian product described as PC4 or TIS7, induced in cell lines by activators such as nerve growth factor or phorbol esters (63), has been isolated. The same mammalian gene has also been designated interferon-related protein, due to an erroneous deposition of this sequence in 1985 as murine beta interferon (accession no. J00424). We have named the *T. canis* gene, which bears no similarity to mammalian interferons, *pth-1* (PC4/TIS7 homolog).

**(v) Tubby-like protein (*Tc-tlp-1*).** Recently, a new multigene family related to the mouse gene *tubby* has been recognized (38). Mutations in *tubby* in the mouse result in obesity and degenerative changes in adult life; however, the existence of conserved homologs in nematodes (including *C. elegans*) and in plants indicates that this gene encodes a protein which fulfills a fundamental—and as yet unrecognized—function in all higher organisms.

**Conclusion.** The analysis of expressed genes that we present here has achieved its aims of identifying a number of abundant transcripts, one of which (*ctl-1*) corresponds to a major secreted product and four of which (*ant-003*, *-005*, *-030*, and *-034*) represent novel gene sequences. We have also made the first steps toward a comprehensive gene catalogue of a biologically intriguing and clinically significant parasite, providing a resource and springboard for future studies.

The outcome of this study supports the proposition that the EST strategy is highly applicable to many metazoan organisms which have relatively large genome sizes ($\geq 10^8$ bp) (13, 14), especially where interest is focused on genes expressed at moderate to high levels. In addition, genes restricted to a life cycle stage can be identified in this manner. Further evidence for the success of this approach is seen in the Filarial Genome Project, which in 3 years has deposited in dbEST over 15,000 sequences, which are estimated to represent some 5,000 separate gene transcripts, or around 33% of the total gene complement of *B. malayi* (15, 62). The availability of the full genome sequence of *C. elegans* lends an exceptional opportunity to compare free-living and parasite gene sequences and structure, with the potential to identify adaptations requisite for parasitism at the molecular level. Along their evolutionary path, parasitic species must have developed a myriad of immune evasion mechanisms, and we anticipate that our study and others like it will be instrumental in identifying the novel immune evasion genes upon which parasite survival depends.

## ACKNOWLEDGMENTS

We thank the Medical Research Council for project grant support.

We thank Mark Blaxter for detailed critical comments on the manuscript.

## REFERENCES

1. **Adams, M. D., M. Dubnick, A. R. Kerlavage, R. Moreno, J. M. Kelley, T. R. Utterback, J. W. Nagle, C. Fields, and J. C. Ventner.** 1992. Sequence identification of 2,375 human brain genes. Nature **355:**632–634.
2. **Adams, M. D., A. R. Kerlavage, R. D. Fleischmeann, et al.** 1995. Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. Nature **377**(Suppl.)**:**3–17.
3. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
4. **Anderson, P., and J. Kimble.** 1997. mRNA and translation, p. 185–208. *In* D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess (ed.), *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
5. **Badley, J. E., R. B. Grieve, D. D. Bowman, L. T. Glickman, and J. H. Rockey.** 1987. Analysis of *Toxocara canis* larval excretory-secretory antigens: physicochemical characterization and antibody recognition. J. Parasitol. **73:**593–600.
6. **Bateman, A., D. Belcourt, H. Bennet, C. Lazure, and S. Solomon.** 1990. Granulins, a novel class of peptide from leukocytes. Biochem. Biophys. Res. Commun. **173:**1161–1168.
7. **Beaver, P. C.** 1966. Biology of parasites, p. 215–227. Academic Press, New York, N.Y.
8. **Belcourt, D. R., C. Lazure, and H. P. J. Bennet.** 1993. Isolation and primary structure of the three major forms of granulin-like peptides from hematopoietic tissues of a telost fish (*Cyprinus carpio*). J. Biol. Chem. **268:**9230–9237.
9. **Bhandari, V., R. G. E. Palfree, and A. Bateman.** 1992. Isolation and sequence of the granulin precursor cDNA from human bone marrow reveals tandem cysteine-rich granulin domains. Proc. Natl. Acad. Sci. USA **89:**1715–1719.
10. **Bin, Z., J. Hawdon, S. Qiang, R. Hainan, Q. Huiqing, H. Wei, X. Shu-Hua, L. Tiehua, G. Xing, F. Zheng, and P. Hotez.** 1999. *Ancylostoma* secreted protein 1 (ASP-1) homologues in human hookworms. Mol. Biochem. Parasitol. **98:**143–149.
11. **Bird, D. M., and D. L. Riddle.** 1994. A genetic nomenclature for parasitic nematodes. J. Nematol. **26:**138–143.
12. **Blaxter, M. L.** 1998. *Caenorhabditis elegans* is a nematode. Science **282:**2041–2046.

12a. **Blaxter, M. L.** Personal communication.

13. **Blaxter, M. L., M. Aslett, J. Daub, D. Guiliano, and The Filarial Genome Project.** Parasitic helminth genomics. Parasitology, in press.
14. **Blaxter, M. L., D. B. Guiliano, A. L. Scott, and S. A. Williams.** 1997. A unified nomenclature for filarial genes. Parasitol. Today **13:**416–417.
15. **Blaxter, M. L., N. Raghavan, I. Ghosh, D. Guiliano, W. Lu, S. A. Williams, B. Slatko, and A. L. Scott.** 1996. Genes expressed in *Brugia malayi* infective third stage larvae. Mol. Biochem. Parasitol. **77:**77–93.
16. **Blumenthal, T., and K. Steward.** 1997. RNA processing and gene structure, p. 117–145. *In* D. L. Riddle, T. Blumenthal, B. J. Meyer, and J. R. Priess (ed.), *C. elegans* II. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
17. **Boguski, M. S., T. M. J. Lowe, and C. M. Tolstoshev.** 1993. dbEST—database for "expressed sequence tags." Nat. Genet. **4:**332–333.
18. **Chakrabati, D., G. R. Reddy, J. B. Dame, E. C. Almira, P. J. Laipis, R. J. Ferl, T. P. Yang, T. C. Rowe, and S. M. Schuster.** 1994. Analysis of expressed sequence tags from *Plasmodium falciparum*. Mol. Biochem. Parasitol. **66:**97–104.
19. **Chen, J.-M., P. M. Dando, N. D. Rawlings, M. A. Brown, N. E. Young, R. A. Stevens, E. Hewitt, C. Watts, and A. J. Barrett.** 1997. Cloning, isolation and characterization of mammalian legumain, an asparaginyl endopeptidase. J. Biol. Chem. **272:**8090–8098.
20. **de Savigny, D. H.** 1975. In vitro maintenance of *Toxocara canis* larvae and a simple method for the production of *Toxocara* ES antigen for use in serodiagnosis test for visceral larva migrans. J. Parasitol. **61:**781–782.
21. **Dissanayake, S., M. Xu, C. Nkenfou, and W. F. Piessens.** 1993. Molecular cloning and serological characterization of a *Brugia malayi* pepsin inhibitor homolog. Mol. Biochem. Parasitol. **62:**143–146.
22. **El-Sayed, N. M. A., C. M. Alarcon, J. C. Beck, V. C. Sheffield, and J. E. Donelson.** 1995. cDNA expressed sequence tags of *Trypanosoma brucei rhodesiense* provide new insights into the biology of the parasite. Mol. Biochem. Parasitol. **73:**75–90.

22a. **Falcone, F. H., et al.** Unpublished data.

23. **Fan, J., D. J. Minchella, S. R. Day, D. P. McManus, W. U. Tiu, and P. J. Brindley.** 1998. Generation, identification, and evaluation of expressed sequence tags from different developmental stages of the Asian blood fluke *Schistosoma japonicum*. Biochem. Biophys. Res. Commun. **252:**348–356.
24. **Fleming, J. T., H. A. Baylis, D. B. Sattelle, and J. A. Lewis.** 1996. Molecular cloning and in vitro expression of *C. elegans* and parasitic nematode ionotropic receptors. Parasitology **113:**S175–S190.
25. **Gems, D., and R. M. Maizels.** 1996. An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins. Proc. Natl. Acad. Sci. USA **93:**1665–1670.
26. **Gems, D. H., C. J. Ferguson, B. D. Robertson, A. P. Page, M. L. Blaxter, and R. M. Maizels.** 1995. An abundant, *trans*-spliced mRNA from *Toxocara canis* infective larvae encodes a 26 kDa protein with homology to phosphatidyleth-

anolamine binding proteins. J. Biol. Chem. **270:**18517–18522.
27. **Gillespie, S. H.** 1987. Human toxocariasis. J. Appl. Bacteriol. **63:**473–479.
28. **Glickman, L. T., and P. M. Schantz.** 1981. Epidemiology and pathogenesis of zoonotic toxocariasis. Epidemiol. Rev. **3:**230–250.
29. **Guénette, S., R. K. Prichard, R. D. Klein, and G. Matlashewski.** 1991. Characterization of a β-tubulin gene and β-tubulin gene products of *Brugia pahangi*. Mol. Biochem. Parasitol. **44:**153–164.
30. **Hawdon, J. M., B. F. Jones, D. R. Hoffman, and P. J. Hotez.** 1996. Cloning and characterization of *Ancylostoma*-secreted protein. A novel protein associated with the transition to parasitism by infective hookworm larvae. J. Biol. Chem. **271:**6672–6678.
31. **Hawdon, J. M., S. Narasimhan, and P. J. Hotez.** 1999. *Ancylostoma* secreted protein 2: cloning and characterization of a second member of a family of nematode secreted proteins from *Ancylostoma caninum*. Mol. Biochem. Parasitol. **99:**149–165.
32. **Healer, J., F. Ashall, and R. M. Maizels.** 1991. Characterization of proteolytic enzymes from larval and adult *Nippostrongylus brasiliensis*. Parasitology **103:**305–314.
33. **Irvine, M., T. Huima, A. M. Prince, and S. Lustigman.** 1994. Identification and characterization of an *Onchocera volvulus* cDNA clone encoding a highly immunogenic calponin-like protein. Mol. Biochem. Parasitol. **65:**135–146.
34. **Ishii, N., M. Fujii, P. S. Hartman, M. Tsuda, K. Yasuda, N. Senoo-Matsuda, S. Yanase, D. Ayusawa, and K. Suzuki.** 1998. A mutation in succinate dehydrogenase cytochrome b causes oxidative stress and ageing in nematodes. Nature **394:**694–697.
35. **Karavanich, C. A., and R. R. H. Anholt.** 1998. Molecular evolution of olfactomedin. Mol. Biol. Evol. **15:**718–726.
36. **Keddie, E. M., T. Higazi, and T. R. Unnasch.** 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol. Biochem. Parasitol. **95:**111–127.
37. **Klein, R. D., S. C. Nulf, S. J. Alexander-Bowman, C. B. Mainone, C. A. Winterrowd, and T. G. Geary.** 1992. Cloning of a cDNA encoding alpha-tubulin from *Haemonchus contortus*. Mol. Biochem. Parasitol. **56:**345–348.
38. **Kleyn, P. W., W. Fan, S. G. Kovats, J. J. Lee, J. C. Pulido, Y. Wu, L. R. Berkemeier, D. J. Misumi, L. Holmgren, O. Charlat, E. A. Woolf, O. Tayber, T. Brody, P. Shu, F. Hawkins, B. Kennedy, L. Baldini, C. Ebeling, G. D. Alperin, J. Deeds, N. D. Lakey, J. Culpepper, H. Chen, M. A. Glücksmann-Kuis, G. A. Carlson, G. M. Duyk, and K. J. Moore.** 1996. Identification and characterization of the mouse obesity gene *tubby*: a member of a member of a novel gene family. Cell **85:**281–290.
39. **Kotzin, B. L.** 1996. Systemic lupus erythematosus. Cell **85:**303–306.
39a. **Liu, L.** Personal communication.
40. **Lloyd, S.** 1993. *Toxocara canis*: the dog, p. 11–24. *In* J. W. Lewis and R. M. Maizels (ed.), Toxocara and toxocariasis: epidemiological, clinical and molecular perspectives. Institute of Biology, London, England.
40a. **Loukas, A., M. Hintz, K. Tetteh, and R. M. Maizels.** Unpublished data.
41. **Loukas, A., P. M. Selzer, and R. M. Maizels.** 1998. Characterisation of *Tc-cpl-1*, a cathepsin L-like cysteine protease from *Toxocara canis* infective larvae. Mol. Biochem. Parasitol. **92:**275–289.
42. **Loukas, A., N. P. Mullin, K. K. A. Tetteh, L. Moens, and R. M. Maizels.** A novel C-type secreted by a tissue-dwelling parasitic nematode. Curr. Biol., in press.
43. **Lustigman, S., J. H. McKerrow, K. Sha, J. Lui, T. Huima, M. Hough, and B. Brotman.** 1996. Cloning of a cysteine protease required for the molting of *Onchocerca volvulus* third stage larvae. J. Biol. Chem. **271:**30181–30189.
44. **Maizels, R. M., D. H. Gems, and A. P. Page.** 1993. Synthesis and secretion of TES antigens from *Toxocara canis* infective larvae, p. 141–150. *In* J. W. Lewis and R. M. Maizels (ed.), Toxocara and toxocariasis: epidemiological, clinical and molecular perspectives. Institute of Biology, London, England.
45. **Maizels, R. M., M. W. Kennedy, M. Meghji, B. D. Robertson, and H. V. Smith.** 1987. Shared carbohydrate epitopes on distinct surface and secreted antigens of the parasitic nematode *Toxocara canis*. J. Immunol. **139:**207–214.
46. **Manger, I. D., A. Hehl, S. Parmley, L. D. Sibley, M. Marra, L. Hillier, R. Waterston, and J. C. Boothroyd.** 1998. Expressed sequence tag analysis of the bradyzoite stage of *Toxoplasma gondii*: identification of developmentally regulated genes. Infect. Immun. **66:**1632–1637.
47. **Matzilevich, D. A., C. Tripp, L. Tang, X. Ou, L. Matzilevich, M. K. Shaw, and R. M. Maizels.** Cloning, expression and localization of *Tc-sod-1*, an extracellular superoxide dismutase from infective stage larvae of *Toxocara canis*. Submitted for publication.
48. **McCauliffe, D. P., E. Zappi, T.-S. Lieu, M. Michalak, R. D. Sontheimer, and J. D. Capra.** 1990. A human Ro/SS-A autoantigen is the homologue of calreticulin and is highly homologous with onchocercal RAL-1 antigen and an aplysia "memory molecule." J. Clin. Investig. **86:**332–335.
49. **McCombie, W. R., M. D. Adams, J. M. Kelley, M. G. FitzGerald, T. R. Utterback, M. Khan, M. Dubnick, A. P. Kerlavage, J. C. Ventner, and C. Fields.** 1992. *Caenorhabditis elegans* expressed sequence tags identify gene families and potential disease gene homologues. Nat. Genet. **1:**124–131.
50. **Meghji, M., and R. M. Maizels.** 1986. Biochemical properties of larval excretory-secretory glycoproteins of the parasitic nematode *Toxocara canis*. Mol. Biochem. Parasitol. **18:**155–170.
51. **Moore, J., V. Todorova, and M. W. Kennedy.** 1995. A cDNA encoding ribosomal protein L3 from the parasitic nematode *Toxocara canis*. Gene **165:**239–242.
51a. **Murray, J., et al.** Unpublished data.
52. **Okimoto, R., J. L. Macfarlane, D. O. Clary, and D. R. Wolstenholme.** 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130:**471–498.
53. **Oldstone, M. B. A.** 1987. Molecular mimicry and autoimmune disease. Cell **50:**819–820.
54. **Ou, X., R. Thomas, M. R. Chacón, L. Tang, and M. E. Selkirk.** 1995. *Brugia malayi*: differential susceptibility to and metabolism of hydrogen peroxide in adults and microfilariae. Exp. Parasitol. **80:**530–540.
55. **Page, A. P.** 1997. Cyclophilin and protein disulfide isomerase genes are co-transcribed in a functionally related manner in *Caenorhabditis elegans*. DNA Cell Biol. **16:**1335–1343.
56. **Page, A. P., A. J. Hamilton, and R. M. Maizels.** 1992. *Toxocara canis*: monoclonal antibodies to carbohydrate epitopes of secreted (TES) antigens localize to different secretion-related structures in infective larvae. Exp. Parasitol. **75:**56–71.
57. **Page, A. P., K. MacNiven, and M. O. Hengartner.** 1996. Cloning and biochemical characterization of the cyclophilin homologues from the free-living nematode *Caenorhabditis elegans*. Biochem. J. **317:**179–185.
58. **Page, A. P., and R. M. Maizels.** 1992. Biosynthesis and glycosylation of serine/threonine-rich secreted proteins from *Toxocara canis* larvae. Parasitology **105:**297–308.
59. **Page, A. P., W. Rudin, E. Fluri, M. L. Blaxter, and R. M. Maizels.** 1992. *Toxocara canis*: a labile antigenic coat overlying the epicuticle of infective larvae. Exp. Parasitol. **75:**72–86.
60. **Plowman, G. D., J. M. Green, M. G. Neubauer, S. D. Buckley, V. L. McDonald, G. J. Todaro, and M. Shoyab.** 1992. The epithelin precursor encodes two proteins with opposing activities on epithelial cell growth. J. Biol. Chem. **267:**13073–13078.
61. **Schallig, H. D. F. H., M. A. W. van Leeuwen, B. E. Verstrepen, and A. W. C. A. Cornelissen.** 1997. Molecular characterization and expression of two putative protective excretory secretory proteins of *Haemonchus contortus*. Mol. Biochem. Parasitol. **88:**203–213.
61a. **Tetteh, K.** Unpublished data.
62. **The Filarial Genome Project.** Deep within the filarial genome: an update on progress in the Filarial Genome Project. Parasitol. Today, in press.
63. **Tirone, F., and E. M. Shooter.** 1989. Early gene regulation by nerve growth factor in PC12 cells: induction of an interferon-related gene. Proc. Natl. Acad. Sci. USA **86:**2088–2092.
64. **Verdun, R. E., N. Di Paolo, T. P. Urmenyi, E. Rondinelli, A. C. C. Frasch, and D. O. Sanchez.** 1998. Gene discovery through expressed sequence tag sequencing in *Trypanosoma cruzi*. Infect. Immun. **66:**5393–5398.
65. **Wilson, W. R., R. S. Tuan, K. J. Shepley, D. O. Freedman, B. M. Greene, K. Awadzi, and T. R. Unnasch.** 1994. The *Onchocerca volvulus* homologue of the multifunctional polypeptide protein disulfide isomerase. Mol. Biochem. Parasitol. **68:**103–117.
66. **Yokoe, H., and R. R. H. Anholt.** 1993. Molecular cloning of olfactomedin, an extracellular matrix protein specific to olfactory neuroepithelium. Proc. Natl. Acad. Sci. USA **90:**4655–4659.
67. **Zeng, W., and J. E. Donelson.** 1992. The actin genes of *Onchocerca volvulus*. Mol. Biochem. Parasitol. **55:**207–216.
68. **Zhang, B., M. Gallegos, A. Puoti, E. Durkin, S. Fields, J. Kimble, and M. P. Wickens.** 1997. A conserved RNA-binding protein that regulates sexual fates in the C. elegans hermaphrodite germ line. Nature **390:**477–484.
69. **Zhou, J., G. Gao, J. W. Crabb, and G. Serrero.** 1993. Purification of an autocrine growth factor homologous with mouse epithelin precursor from a highly tumorigenic cell line. J. Biol. Chem. **268:**10863–10869.

*Editor:* J. M. Mansfield

# A novel C-type lectin secreted by a tissue-dwelling parasitic nematode

Alex Loukas*, Nicholas P. Mullin†, Kevin K.A. Tetteh*, Luc Moens‡ and Rick M. Maizels*

**Many parasitic nematodes live for surprisingly long periods in the tissues of their hosts, implying sophisticated mechanisms for evading the host immune system. The nematode *Toxocara canis* survives for years in mammalian tissues, and when cultivated *in vitro*, secretes antigens such as TES-32. From the peptide sequence, we cloned TES-32 cDNA, which encodes a 219 amino-acid protein that has a domain characteristic of host calcium-dependent (C-type) lectins, a family of proteins associated with immune defence. Homology modelling predicted that TES-32 bears remarkable structural similarity to mammalian immune-system lectins. Native TES-32 acted as a functional lectin in affinity chromatography. Unusually, it bound both mannose- and galactose-type monosaccharides, a pattern precluded in mammalian lectins by a constraining loop adjacent to the carbohydrate-binding site. In TES-32, this loop appeared to be less obtrusive, permitting a broader range of ligand binding. The similarity of TES-32 to host immune cell receptors suggests a hitherto unsuspected strategy for parasite immune evasion.**

Addresses: *Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. †Department of Chemistry, University of Edinburgh, Edinburgh EH9 3JJ, UK. ‡Department of Biochemistry, University of Antwerp (UIA), Wilrijk B-2610, Belgium.

Correspondence: Rick M. Maizels
E-mail: Rick.Maizels@ed.ac.uk

Received: **30 March 1999**
Revised: **27 May 1999**
Accepted: **16 June 1999**

Published: **26 July 1999**

**Current Biology** 1999, 9:825–828
http://biomednet.com/elecref/0960982200900825



## Results and discussion

### Identification of TES-32 as a C-type lectin

TES-32 is a 32 kDa *N*-glycosylated protein secreted by tissue-stage *T. canis* parasites [1]. Secreted *Toxocara* excretory–secretory (TES) products collected from larval *T. canis* in culture [2] were separated by SDS–PAGE, electrotransferred to PVDF membrane and the TES-32 band excised. Peptides from trypsinised TES-32 were then separated by reverse-phase HPLC and sequenced. Peptide sequences were matched against a database of 250 expressed sequence tags (ESTs) from a larval *T. canis* cDNA library [3]. Clone *Tc*-EST-070 (*Tc-ctl-1*) encoded one of the TES-32 peptides. On complete sequencing, *Tc-ctl-1* proved to contain a full-length 762 bp cDNA insert, and to encode a second peptide derived from gel-purified TES-32.

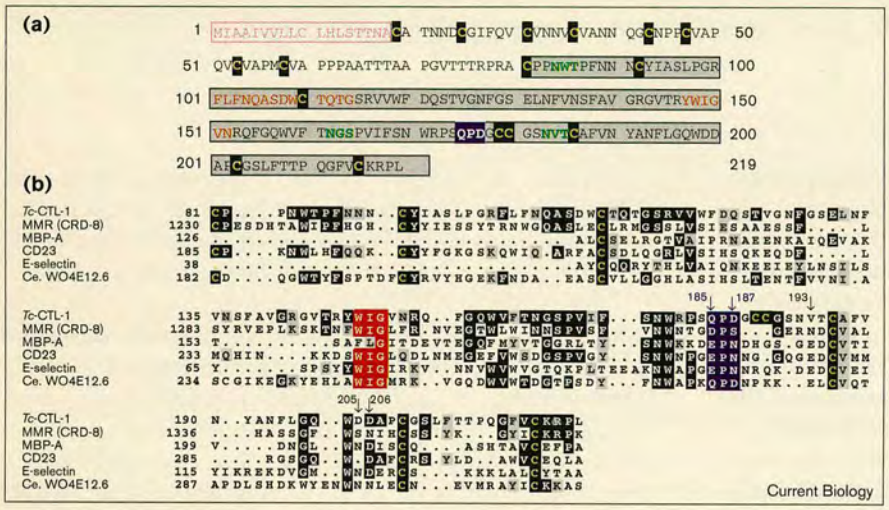
The deduced CTL-1 protein sequence comprises 219 amino acids, including a putative 18 amino-acid signal peptide and three potential *N*-glycosylation sites (Figure 1a). The amino-terminal portion (amino acids 19 to 80) of the mature protein includes eight cysteine residues, four of which are in (V)CVAP-like motifs (using the single-letter amino-acid code), and a short threonine-rich tract (Figure 1a). This amino-terminal portion shows no significant homology to other known proteins. The carboxy-terminal domain (amino acids 81 to 219), however, was found to have strong similarities to mammalian and invertebrate C-type lectins (CTLs), including the macrophage mannose receptor (MMR), CD23 and E-selectin (Figure 1b). Maximal amino-acid identity was 31% (43/139) between amino acids 81 to 219 of TES-32 and the carbohydrate-recognition domain (CRD) 8 of human MMR (amino acids 1230 to 1359). The gene encoding TES-32 was therefore designated *Tc-ctl-1*.

Antibodies to recombinant *Tc*-CTL-1 (rCTL-1) reacted with native TES-32 in Western blots (Figure 2), and with the 27 kDa recombinant protein (a dimeric form is also apparent). Furthermore, rCTL-1 was also recognised by monoclonal antibody Tcn-3, which specifically binds TES-32 [2] (Figure 2). The difference in size between the predicted mature *Tc*-CTL-1 protein (22 kDa) and the actual protein found in TES products can be attributed to *N*-glycosylation [4,5]. Monoclonal antibody Tcn-3 showed TES-32 to be localised to the parasite epicuticle [1]; an anti-rCTL-1 antiserum showed a similar surface-binding pattern (data not shown). Additional bands of 45 kDa and > 200 kDa were recognised in TES products by the anti-rCTL-1 antiserum; peptide sequence data indicated that the native 45 kDa component is a second lectin derived from a distinct gene (unpublished observations).

### TES-32 is a bispecific lectin

The lectin-like properties of native TES products were then tested on affinity columns of different monosaccharides in a $Ca^{2+}$-containing buffer. A 32 kDa protein was selectively retarded on a GalNAc column (Figure 3). Dissociation from the ligand occurred slowly, with protein evident throughout ten column-volume washes, and was completed in the presence of EDTA. Binding was greatly

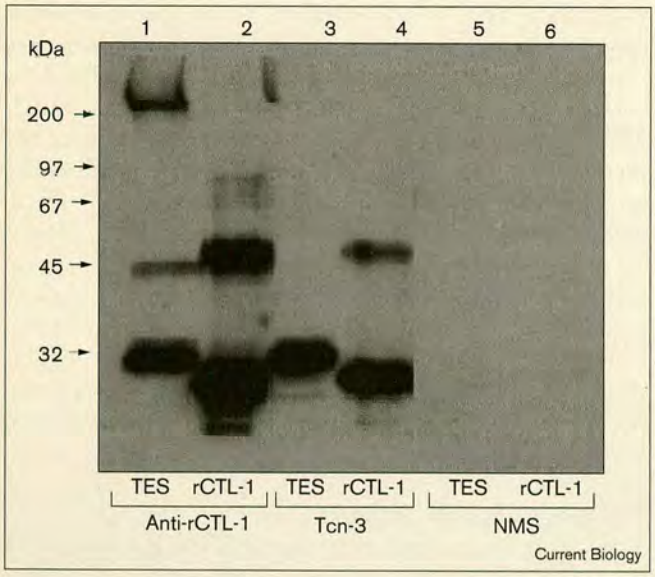
**Figure 1**




Sequence of *Tc*-CTL-1 and its similarity to other CTLs. **(a)** Amino-acid sequence of *Tc*-CTL-1, showing the signal peptide (pink), the CTL CRD (grey shading) and *N*-glycosylation sites (green lettering). Cysteine residues are yellow, the carbohydrate-binding motif is in white lettering on a blue background, and tryptic peptide sequences are red. **(b)** Alignment of CTLs. Cysteine residues are yellow; the five conserved $Ca^{2+}$-ligating residues (site 2) in MBP-A are indicated by arrows (MBP-A numbering). The sugar-binding motifs QPD and EPN are in white lettering on a blue background. Residues that are identical between *Tc*-CTL-1 and the other CTLs are in white lettering on a black background; similar amino acids are shaded grey. The WIG motif (red shading) in *Tc*-CTL-1 is conserved with minor substitutions [10]. GenBank accession numbers: human MMR, M93221; rat MBP-A, M14105; mouse CD23, M99371; human E-selectin, M24736. Ce WO4E12.6 is a predicted protein from *Caenorhabditis elegans*.

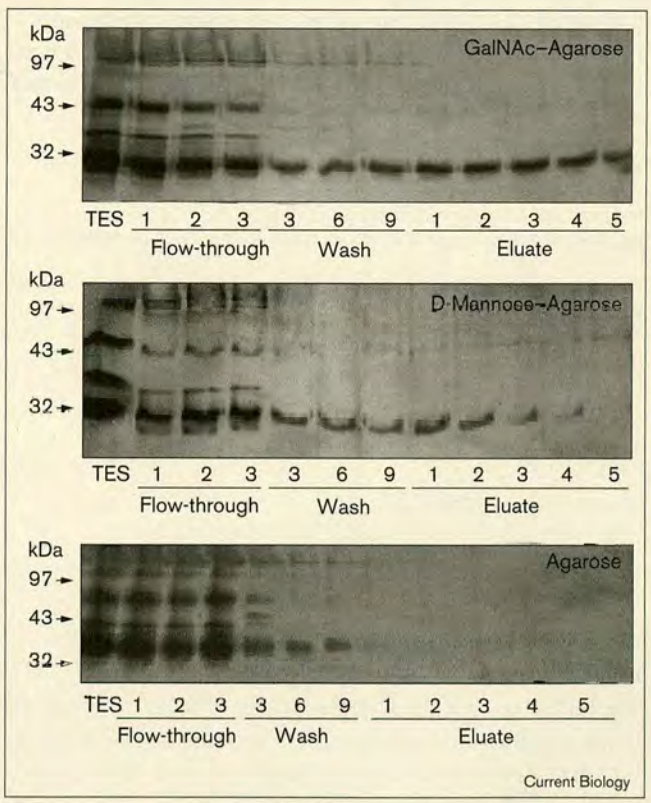
reduced when free GalNAc was added to the TES products before loading (data not shown). Surprisingly, the same 32 kDa molecule bound almost as well to D-mannose–agarose but not to unconjugated agarose alone (Figure 3). Although different affinity matrices may not bear comparable ligand densities, distinctly weaker

**Figure 2**




Identity of *Tc*-CTL-1 and TES-32. A western blot of native TES products (lanes 1,3 and 5) and of rCTL-1 (lanes 2,4 and 6) showing that both proteins are recognised by anti-rCTL-1 serum (lanes 1,2) and by monoclonal antibody Tcn-3 (lanes 3,4). Normal mouse serum (NMS; lanes 5,6) was used as a control.

**Figure 3**




Carbohydrate binding by *Tc*-CTL-1. SDS–PAGE profiles of TES products after affinity chromatography on GalNAc–agarose, D-mannose–agarose or unconjugated agarose. Fractions of 1 ml were collected from a 1 ml column loaded successively with 4 ml loading buffer (flow-through), 10 ml wash buffer and 5 ml elution buffer.

binding of the 32 kDa protein was observed to GlcNAc and D-galactose compared with GalNAc or D-mannose (data not shown). Both the anti-rCTL-1 serum and monoclonal antibody Tcn-3 bound the 32 kDa protein eluted from both GalNAc and D-mannose columns, confirming that the retarded native protein in each case was *Tc*-CTL-1. In addition, a second 45 kDa protein, similar to the one detected in western blotting, exhibited weak binding to D-mannose (Figure 3), consistent with data indicating that it is encoded by a distinct lectin gene (unpublished observations).

### Structural conservation between mammalian and nematode lectins

Co-crystallisation of rat mannose-binding protein-A (MBP-A) and an oligomannose ligand has identified the residues at which CTLs ligate $Ca^{2+}$ and saccharides [6]. We aligned the CRDs of *Tc*-CTL-1 and CTLs with known structure, and modelled the CRD of *Tc*-CTL-1 (Figure 4). In this prediction, *Tc*-CTL-1 shows a remarkable structural similarity to MBP-A, with the major beta sheets and alpha helices being in corresponding positions.

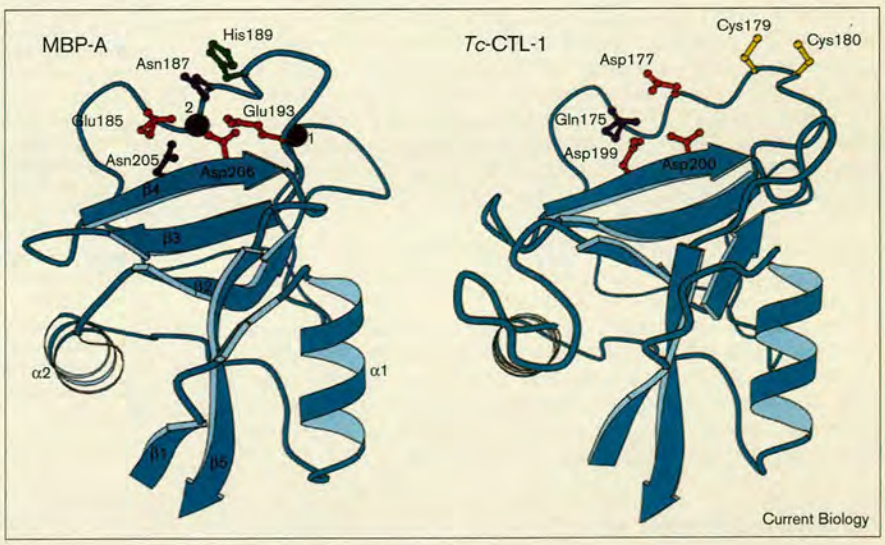
All CTLs possess a conserved $Ca^{2+}$-binding site (designated $Ca^{2+}$ number 2) [7], and *Tc*-CTL-1 contains four of the five residues necessary to ligate $Ca^{2+}$ in MBP-A (Figure 4); the fifth $Ca^{2+}$-binding residue cannot be determined from the alignment. But the similarity of this site to that of MBP-A indicates that sugar ligation by *Tc*-CTL-1 is likely to resemble that observed in the crystal structure of MBP-A [6]. Both MBP-A and tetranectin, but not E-selectin, contain a second (auxiliary) $Ca^{2+}$ binding site ($Ca^{2+}$ number 1), which does not bind directly to sugar. This site appears to be absent from *Tc*-CTL-1 which in this regard is more similar to E-selectin than to MBP-A or tetranectin.

CTLs can be loosely grouped into those that bind D-galactose and its derivatives (Gal-type), and those that bind D-mannose, D-glucose and their derivatives (Man-type). The specificity of MBP-A, a Man-type CTL, is determined by calcium-binding residues at positions 185 and 187 (MBP-A numbering) [6] and by residue 189 and the adjacent loop [8]. Thus, Man-type CRDs typically have Glu185 and Asn187, whereas Gal-type CRDs have Gln185 and Asp187, with Trp- or Phe189 [8,9]. The mannose-specific MBP-A can be converted to bind galactose by replacement of residues at positions 185 and 187 with Gal-type residues, and rendered galactose-specific by gaining Trp189 and the adjacent five-residue loop from a natural galactose-binding lectin [8]. In *Tc*-CTL-1, this loop appears significantly modified. Thus, in place of a fold which brings His/Trp189 close to the ligand, the protein backbone in *Tc*-CTL-1 is more extended and the bulky aromatic residue replaced by an unusual pair of cysteines at positions 179 and 180 (equivalent residues are 189 and 190 in MBP-A). The predicted position of these cysteine residues is too distant for them to have a direct role in forming the binding site for $Ca^{2+}$ and saccharides. As a consequence of these changes, the ligation of *Tc*-CTL-1 to either Gal-type or Man-type saccharides may be unimpeded. Site-directed mutagenesis to convert this region of *Tc*-CTL-1 would test this hypothesis.

### Biological significance of parasite lectins

This is the first report of a gene encoding a CTL from a pathogenic organism; in contrast, most other CTLs are involved in host defence against infection [10]. The latter include the selectins, the low affinity IgE receptor CD23, the NK cell receptors Ly-49 and CD94, and antigen-presenting cell lectins such as the MMR and the dendritic cell receptor DEC-205. Genes encoding CTLs have now

**Figure 4**




Similarity of mammalian and nematode lectins. Structural model of the CRD of *Tc*-CTL-1 based on the crystal structure of MBP-A. The His189 residue of MBP-A is close to the sugar-binding site, but no equivalent large residue exists in *Tc*-CTL-1. The conserved and auxiliary $Ca^{2+}$-binding sites in MBP-A are depicted as black spheres. It is not yet known whether the two newly identified cysteine residues (Cys179 and Cys180) in *Tc*-CTL-1 retain free SH groups, as shown. CRD sequences of *Tc*-CTL-1 (amino acids 81–219), rat MBP-A (M14105), human E-selectin (M24736) and human tetranectin (X64559) were aligned for the MODELLER program [19], and figures were produced using MOLSCRIPT [20]. Five separate models of *Tc*-CTL-1 with near-identical structures were produced, differing only in the carboxy-terminal loop (residues 201 onwards).

been identified in invertebrates, including arthropods, molluscs, echinoderms and tunicates, and some of the proteins encoded by these genes are thought to be involved in agglutinating pathogens [11]. The free-living nematode *C. elegans* contains 102 genes encoding CRD-like domains of CTLs [12]. Interestingly, *Tc*-CTL-1 has only weak similarity with these putative gene products, sharing greater identity with mammalian CTLs than with the closest *C. elegans* homologue (Figure 2b).

*T. canis* secretions contain a substantial glycan content — ~40% by weight [4] — and the predominant *O*-linked glycan contains a terminal fucose which is *O*-methylated at position 2 [13]. These carbohydrates are abundant on other secreted proteins such as mucins [2,14]. As adjacent hydroxyls at carbons 2 and 3 of fucose (or 3 and 4 of other hexoses) are required for binding to mammalian CTLs[6], *O*-methylation may prevent the parasite lectin from binding to carbohydrate structures being secreted at the same time.

It is intriguing that helminth parasites secrete proteins with significant homology to host molecules that are instrumental in the immune response against the very same pathogens. This finding suggests an exciting new parasite immune-evasion strategy [15]. Although there is much information on the role of host lectins in the innate and acquired immune responses to infectious organisms, little is known about the role of pathogen-derived CTLs. We are now investigating the potential host ligands of *Tc*-CTL-1, at both a molecular and cellular level, in an attempt to elucidate the biological function of this class of CTL proteins in host–parasite interactions.

## Materials and methods

*Parasites, protein sequencing and expression*

Larval *T. canis* were maintained *in vitro* in serum-free medium, from which secreted TES products were collected [16]. TES proteins were separated by SDS–PAGE and transferred to PVDF membrane as published [17]. TES-32 was excised and digested with trypsin [17], and HPLC-purified peptides sequenced on an ABI-471-B sequencer (Applied Biosystems). The larval cDNA library made in Uni-ZAP XR (Stratagene) was a generous gift from C. Tripp and R. Grieve (Heska). The mature region of *Tc*-CTL-1 (Cys19–Leu219) was expressed by cloning into the pET-29T expression vector (Novagen), and purified on metal chelation columns. BALB/c mice were immunised with three doses of 10 μg rCTL-1.

*Detection of lectin activity using immobilised monosaccharides*

Ligand-binding assays [18] employed 100 μg TES product diluted in 500 μl of low-calcium loading buffer (1.25 M NaCl, 25 mM Tris pH 7.4, 2.5 mM $CaCl_2$) and applied to 1 ml columns of 4% beaded agarose cross-linked to GalNAc (Sigma A2787, 6 hydroxy-linked), GlcNAc, β-D-glucose or D-mannose (Sigma A2278, G2019 and M6400, respectively; all hydroxy-linked). Columns were rinsed with 4 ml 2.5 mM $CaCl_2$ buffer and washed with 10 ml 25 mM $CaCl_2$ buffer. Bound proteins were eluted with 2.5 mM EDTA in calcium-free buffer.

## Acknowledgements

We thank M-L. Van Hauwaert for expert protein purification and sequencing, M. Hintz for additional protein sequence information, and K. Drickamer, J. Allen, M. Blaxter, I. Dransfield, F. Falcone, N. Gomez-Escobar, B. Gregory, A. MacDonald and X. Zang for helpful discussion and suggestions. This work was supported by the Medical Research Council.

## References

1. Page AP, Hamilton AJ, Maizels RM: ***Toxocara canis*: monoclonal antibodies to carbohydrate epitopes of secreted (TES) antigens localize to different secretion-related structures in infective larvae.** *Exp Parasitol* 1992, **75**:56-71.
2. Maizels RM, Kennedy MW, Meghji M, Robertson BD, Smith HV: **Shared carbohydrate epitopes on distinct surface and secreted antigens of the parasitic nematode *Toxocara canis*.** *J Immunol* 1987, **139**:207-214.
3. Tetteh KKA, Loukas A, Tripp C, Maizels RM: **Identification of abundantly-expressed novel and conserved genes from the infective larval stage of *Toxocara canis* by an expressed-sequence tag strategy.** *Infect Immun* 1999, in press.
4. Meghji M, Maizels RM: **Biochemical properties of larval excretory–secretory (ES) glycoproteins of the parasitic nematode *Toxocara canis*.** *Mol Biochem Parasitol* 1986, **18**:155-170.
5. Page AP, Maizels RM: **Biosynthesis and glycosylation of serine/threonine-rich secreted proteins from *Toxocara canis* larvae.** *Parasitology* 1992, **105**:297-308.
6. Weis WI, Drickamer K, Hendrickson WA: **Structure of a C-type mannose-binding protein complexed with an oligosaccharide.** *Nature* 1992, **360**:127-134.
7. Drickamer K: **Engineering galactose-binding activity into a C-type mannose-binding protein.** *Nature* 1992, **360**:183-186.
8. Iobst ST, Drickamer K: **Binding of sugar ligands to Ca(2+)-dependent animal lectins. II. Generation of high-affinity galactose binding by site-directed mutagenesis.** *J Biol Chem* 1994, **269**:15512-15519.
9. Kolatkar AR, Weis WI: **Structural basis of galactose recognition by C-type animal lectins.** *J Biol Chem* 1996, **271**:6679-6685.
10. Weis WI, Taylor ME, Drickamer K: **The C-type lectin superfamily in the immune system.** *Immunol Rev* 1998, **163**:19-34.
11. Wilson R, Chen C, Ratcliffe NA: **Innate immunity in insects: the role of multiple, endogenous serum lectins in the recognition of foreign invaders in the cockroach, *Blaberus discoidalis*.** *J Immunol* 1999, **162**:1590-1596.
12. The *C.elegans* Genome Sequence Consortium: **Genome sequence of *Caenorhabditis elegans*: a platform for investigating biology.** *Science* 1998, **282**:2012-2018.
13. Khoo K-H, Maizels RM, Page AP, Taylor GW, Rendell N, Dell A: **Characterisation of nematode glycoproteins: the major *O*-glycans of *Toxocara* excretory secretory antigens are methylated trisaccharides.** *Glycobiology* 1991, **1**:163-171.
14. Gems DH, Maizels RM: **An abundantly expressed mucin-like protein from *Toxocara canis* infective larvae: the precursor of the larval surface coat glycoproteins.** *Proc Natl Acad Sci USA* 1996, **93**:1665-1670.
15. Maizels RM, Bundy DAP, Selkirk ME, Smith DF, Anderson RM: **Immunological modulation and evasion by helminth parasites in human populations.** *Nature* 1993, **365**:797-805.
16. Page AP, Richards DT, Lewis JW, Omar HM, Maizels RM: **Comparison of isolates and species of *Toxocara* and *Toxascaris* by biosynthetic labelling of somatic and ES proteins from infective larvae.** *Parasitology* 1991, **103**:451-464.
17. Coligan JE, Kruisbeek AM, Margulies DH, Shevach EM, Strober W: *Current Protocols in Immunology.* New York: Greene Publishing Associates and Wiley-Interscience; 1997.
18. Burrows L, Lobst ST, Drickamer K: **Selective binding of N-acetylglucosamine to the chicken hepatic lectin.** *Biochem J* 1997, **324**:673-680.
19. Sali A, Blundell TL: **Comparative protein modelling by satisfaction of spatial restraints.** *J Mol Biol* 1993, **234**:779-815.
20. Kraulis PJ: **MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures.** *J Appl Crystallogr* 1991, **24**:946-950.