Functional analysis of a parasitic nematode GATA transcription factor using *C. elegans* as a heterologous expression system

Annabelle Couthier

Supervisor: John Gilleard

Department of Veterinary Parasitology

University of Glasgow

2004

ProQuest Number: 10390855

ProQuest 10390855

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

Declaration

I, Annabelle Couthier, declare that the following thesis embodies the results of my own special work, that it has been composed by myself and that it does not include work forming part of a thesis presented successfully for a degree in this or another university.

Publications

Couthier, A., J. Smith, P. McGarr, B. Craig and J. S. Gilleard (2004). "Ectopic expression of a *Haemonchus contortus* GATA transcription factor in *Caenorhabditis elegans* reveals conserved function in spite of extensive sequence divergence." <u>Molecular and Biochemical Parasitology.</u> **133**: 241-253.

Summary

In recent years there has been a growing interest in the use of *C. elegans* in parasitic nematode research both as a model and as a surrogate expression system. However, relatively little is know about the extent to which different aspects of *C. elegans* biology is conserved in particular parasitic nematode species. Also, there are as yet very few examples of the use *C. elegans* as an expression system to study parasite gene function. The aim of this work was to exploit the in depth understanding of *C. elegans* GATA transcription factor function in order to explore conservation with parasitic nematode species and to develop *C. elegans* as a heterologous expression system for the study of parasite gene function.

The endoderm is a relatively simple tissue which is present in all nematodes. In *C. elegans*, the endoderm has been extensively studied and its specification and development have been well described. Several GATA transcription factors, namely *med-1/2*, *end-1/3* and *elt-2/7*, are involved in the regulation of endodermal development and maintenance of endodermal function. The first objective of this project was to isolate homologues of *C. elegans* GATA transcription factors from parasitic nematodes by searching available DNA sequence databases and by using molecular biology approaches. The next objectives were then to examine the conservation of both the function and regulation of these using *C. elegans* as a surrogate expression system.

A search of the *Brugia malayi* genome sequence database revealed five potential GATA transcription factors. Interestingly, sequence analysis suggested that three of these were potential homologues of the three *C. elegans* GATA factors with essential functions (*elt-1*, *elt-2* and *elt-5*). A search of the parasitic nematode EST

databases identified a single GATA transcription factor from the species *Ancylostoma ceylanicum*, *Strongyloides ratti* and *Meloidogyne arenaria*. The relationship of these genes to individual members of the *C. elegans* GATA transcription factor family was not clear.

A GATA transcription factor was also isolated from the strongylid parasitic nematode *Haemonchus contortus*. This was achieved by PCR using degenerate primers corresponding to the consensus sequence encoding the conserved GATA zinc finger of several *C. elegans* family members. A full-length cDNA λZAPII clone corresponding to this gene was isolated and sequence analysis suggested this was homologue of the *C. elegans elt-2* gene. Consequently this gene was named *Hc-elt-2*. Fusion proteins corresponding to two different regions of the Hc-ELT-2 polypeptide were used to generate specific rabbit polyclonal antisera. Immunofluorescence experiments with the specific antisera demonstrated that the Hc-ELT-2 was expressed in the *H. contortus* endoderm and the overall expression pattern was very similar to that of *C. elegans* ELT-2. However one difference between the *C. elegans* and *H. contortus* ELT-2 expression patterns was that additional expression was seen in two (probably neuronal) cells either side of the *H. contortus* pharynx. These Hc-ELT-2 immunolocalisation experiments also revealed that development of the endoderm lineage is extremely similar in *C. elegans* and *H. contortus.*

The *H. contortus elt-2* gene was ectopically expressed in transgenic *C. elegans* using the *C. elegans hsp16-2* promoter. A series of experiments showed that expression of the Hc-ELT-2 polypeptide could be induced by heat shock in transgenic *C. elegans*. Expression of Hc-ELT-2 in early *C. elegans* embryos resulted in embryonic arrest with the absence of any signs of morphogenesis. Arrested embryos consisted of largely uniform balls of 200-300 endoderm-like cells. The ectopic experiments showed that the *H. contortus elt-2* gene was capable of

activating a program of endodermal differentiation in a very similar manner to the *C. elegans elt-2* gene itself. Hence the function and specificity of the *elt-2* gene appears to be highly conserved between *C. elegans* and *H. contortus*.

The genomic locus of the *Hc-elt-2* gene, including 3.3kb of 5'flanking sequence was isolated and the putative *Hc-elt-2* promoter region was able to direct the expression of a GFP reporter construct in transgenic *C. elegans*. The highest levels of GFP expression were in the endoderm and the spatial expression pattern directed by the *Hc-elt-2* regulatory elements was broadly similar to that of the endogenous *H. contortus* and *C. elegans elt-2* genes. Although in addition to the endodermal expression, the GFP reporter was expressed in some neuronal cells: possibly amphids and phasmids. Also the temporal pattern of GFP expression pattern directed by the *Hcelt-2* promoter in transgenic *C. elegans* was quite different to the endogenous expression of Hc-ELT-2 polypeptide in *H. contortus:* the transgene expression was only initiated during late embryogenesis compared to the early expression of the Hc-ELT-2 polypeptide in the 2E cells of the early *H. contortus* embryo.

A transgenic *C. elegans* line was established which carried the *Hc-elt-2* genomic locus on an extrachromosomal array. This line was used to show that the *H. contortus elt-2* gene could at least partially rescue the *C. elegans* RNAi phenotype. Finally, an attempt was made to "knock down" *H-celt-2* function during *H. contortus* L1 to L3 development using RNA-mediated interference applied by feeding. However these attempts were unsuccessful.

In summary, this work has shown that the function and regulation of the endodermal GATA factor *elt-2* is highly conserved between *C. elegans* and *H. contortus*. Function was sufficiently conserved for the *H. contortus elt-2* gene to function appropriately when expressed in transgenic *C. elegans*. Hence, for

strongylid nematodes at least, *C. elegans* should be a valuable tool for the direct in *vivo* study parasite gene regulation and function.

Table of contents

List of tables and illustrations

Tables

Figures

1 Introduction

Parasitic nematodes are difficult experimental subjects. One of the major problems is the lack of *in vitro* culture systems and the requirement to propagate life cycles *in vivo*. For some parasites, model laboratory animal host systems are available but for many species this is not the case. As a consequence, technology and genomic resources lag behind that available for most other pathogen groups. In spite of this, there has recently been major progress in the genomic resources available for parasitic nematodes and this situation will continue to improve. There are EST sequencing projects for over 30 different parasitic nematode species and there are now two parasitic nematode species with genome sequencing projects. A shotgun-sequencing project for the human filarial nematode *Brugia malayi* genome has reached over 5-fold genome coverage and annotation is underway (http://www.tigr.org/tdb/e2k1/bma1/). The construction of an integrated HAPPY /BAC clone map and shotgun sequencing have been initiated for the sheep nematode *Haemonchus contortus* (J. Gilleard, personal communication). Hence there is going to be an increasing requirement for methods of studying parasitic nematode gene function. One approach is to use *C. elegans* as a surrogate expression system for parasite genes. This thesis presents an investigation of the functional conservation of the *C. elegans* GATA transcription factor *elt-2* in the parasitic nematode *Haemonchus contortus*. The use of *C. elegans* as a surrogate expression system to investigate the function and regulation of the *H. contortus elt-2* homologue is explored.

This introductory chapter discusses the potential of *C. elegans* as a model system for parasitic nematodes and its phylogenetic relationship to other nematode species. Some of the recent literature concerning the evolutionary conservation and

variation of a number of different aspects of nematode biology are then discussed. This is presented to illustrate the principle that conservation of biological processes and gene function cannot be taken for granted, even in quite closely related nematode species. This introduction also presents background information on the parasitic nematode *Haemonchus contortus,* the process of endodermal development in nematodes and GATA factor function in *C. elegans*

1.1 *Caenorhabditis* elegans as a model nematode

1.1.1 What is *C. elegans*?

C. elegans is a free-living nematode with a worldwide distribution. The only requirements for growth and reproduction are a humid environment, appropriate ambient temperature, atmospheric oxygen and bacteria as a food source; consequently *C. elegans* is particularly cheap and easy to maintain in the laboratory. The adults are only 1 mm long and can be grown in the laboratory on agar plates seeded with a lawn of *E. coli* bacteria. The standard *E. coli* strain used for this purpose is OP50, which has a limited growth rate on standard agar plates due to it being a uracil auxotrophe. This prevents the bacteria overgrowing the nematodes on plates.

C. elegans has two sexes: male and hermaphrodite. Sex determination depends on the X chromosome dosage: XX individuals are hermaphrodite and XO individuals are male. The majority of individuals are hermaphrodites, which produce both sperm and oocytes and can reproduce by self-fertilization. The male produces

only sperm and so must mate with a hermaphrodite in order to reproduce. After mating, male sperm out-competes the hermaphrodites' own sperm in fertilizing the oocytes and most progeny are then the result of cross-fertilization. Cross-fertilization produces males and hermaphrodites in equal proportions, whereas self-fertilization produces only hermaphrodites. This mode of reproduction has particular value for genetic analysis. The use of hermaphrodites results in clonal populations, ideal to maintain strains, whereas the ability to mate hermaphrodites with males enables genetic crosses between strains. The *C. elegans* embryonic and post-embryonic cell lineage has been established and is essentially invariant between individuals (Sulston, J. E. *et al.* 1977; Sulston, J. E. *et al.* 1983). Hermaphrodites all have 959 cells and males have 1031 cells. The determination of the *C. elegans* cell lineage has been facilitated by the fact that both the embryonic and post-embryonic stages are transparent and this has made *C. elegans* a powerful model in which to study fundamental developmental processes such as cell fate specification, differentiation and apoptosis (Hekimi, S. *et al.* 1998; Marin, I. *et al.* 1998; Metzstein, M. M. *et al.* 1998).

Under laboratory conditions, the growth of *C. elegans* is rapid. The entire life cycle takes just 3.5 days at 20°C (or 3 days at 25°C and 6 days at 15°C). Population growth is fastest at 20°C, with brood sizes of more than 300 produced over a 4-day period. Development from fertilization to hatching is referred to as embryogenesis which takes 14 hours at 20°C. Post-embryonic development involves growth through four larval stages (L1 to L4) before the final moult to produce the adult worm. In the absence of food or at high population densities an alternative larval stage, the dauer, is formed instead of the normal L3. The dauer is a metabolically quiescent stage that is specialised for surviving adverse conditions and remains viable for several

months. When food becomes available it moults to a normal L4 and develops to an adult worm (Brenner, S. 1974). The rapid life cycle, large brood sizes and ability to form dauer larvae all make *C. elegans* a particularly attractive model genetic organism.

In addition to these basic biological features, a powerful arsenal of molecular techniques and bioinformatic resources has been developed by the *C. elegans* research community. Core technologies such as the creation of gene knock-outs and transgenic strains have been more recently supplemented by the completion of the genome sequence (Bird, D. M. *et al.* 1999), the development of RNA-mediated interference (RNAi) (Fire, A. *et al.* 1991; Fire, A. 1999; Timmons, L. *et al.* 2001) and the advent of DNA micro-arrays. For these reasons, *C. elegans* has become a widely used experimental model system in developmental biology, neurobiology and genetics. More recently there has been a great deal of interest in using *C. elegans* as a model for parasitic nematodes. Its relatively close evolutionary relationship to parasitic nematodes makes *C. elegans* likely to be a particularly powerful system to study gene function in this group of organisms which are often poorly suited to direct experimental manipulation (Blaxter, M. L. 1998; Blaxter, M. L. *et al.* 1998; Burglin, T. R. *et al.* 1998; Geary, T. G. *et al.* 2001; Hashmi, S. *et al.* 2001).

1.1.2 Nematode Phylogeny

In the last decade interest in the nematode phylum has increased dramatically. Nematodes are present in all ecosystems and it is thought they represent the largest phylum of the animal kingdom (Sommer, R. J. 2000). Although the large majority of nematodes are free-living organisms, many are parasites of

4

animals, plants and humans and consequently represent an important health and economic burden (Anderson, R. C. 1992). Traditionally, the phylogeny of nematodes was based on morphological and ecological criteria, resulting in several substantially different classifications between authors because the relative simple body plan of nematodes masks the underlying complexity of the phyla. The traditionally accepted taxonomy based on morphological and ecological criteria is that of Anderson (Anderson, R. C. 1992). More recently, molecular phylogenetic approaches have been applied to nematode classification. A phylogenetic tree has been produced, based on the sequences of the 18S rRNA subunit (figure 1.1) (Blaxter, M. L. *et al.* 1998). This classification results in a dramatically different picture from the traditional view. One fact of particular relevance to the relationship of *C. elegans* to parasitic nematodes is that the order *Rhabitidida* has been split into several groups, separating organisms were formerly considered to be closely related. For example, *Strongyloides* (*Rhabditidae*) which was traditionally thought to be a close relative of *C. elegans* is now placed in a separate clade. Also, the order *Strongylida*, which contains many important parasite species, is now the closest clade of parasitic nematodes to *C. elegans* in the molecular-based classification. This hypothesis is supported by current parasitic nematode EST sequencing projects which show that the strongylid nematodes appear to have the highest level of sequence identity with *C. elegans* (Gasser, R. B. *et al.* 2000). Hence one would predict that a lot of basic *C. elegans* biology will be conserved with the Strongylid nematodes and so this group of organisms represents a good place to initiate a comparative study. *Haemonchus contortus* is one of the better characterised and most economically important members of order *Strongylida*. Furthermore, comparisons of the sequence of predicted genes for the *C. elegans* genome reveals 42% of genes have homologues in

organisms from other phyla (Blaxter, M. L. 1998). What proportion of the remaining 58% of genes is truly specific to nematodes is presently unknown but such genes will be of particular interest to parasitologists. However, even for those genes with homologues outside the nematode phylum there will be many aspects of gene function that are nematode-specific. Hence an important research priority is to undertake a comparative analysis of gene functions and pathways both within the nematode phyla and between nematodes and other organisms.

1.1.3 To what extent is biology conserved across the phylum nematoda?

One of the key questions at present is the extent to which developmental mechanisms are conserved across the nematode phylum. This has important implications for the use of *C. elegans* as a model for other nematode species including parasitic nematodes. It has often been casually assumed that developmental processes in *C. elegans* are representative of most species of the phylum. However more recent research shows that there is much more diversity regarding developmental mechanisms between nematode species than has previously been appreciated (Sommer, R. J. *et al.* 1996). Several recent comparative studies between *C. elegans* and other free-living nematodes are very thought-provoking for those wishing to use *C. elegans* as a model of particular aspects of parasitic nematode biology and some examples of these are discussed next. The phylogenetic relationships of all species mentioned are shown on figure 1.1.

i) General features of embryogenesis in *C. elegans* and other nematodes

A number of studies on patterning and cell lineage during the embryonic development illustrate how basic features of development can vary between different nematode species. *Acrobeloides nanus* is another free-living soil nematode previously thought to belong to a sister group of *C. elegans* (Sudhaus, W. 1976), although the more recent phylogenetic trees produced by Blaxter place these two nematodes further apart (figure 1.1) (Blaxter, M. L. 1998). The nematode develops parthogenetically and its embryogenesis is longer than that of *C. elegans*. All individuals are female, the oocytes are activated without sperm and eggs are usually laid at the 1-cell stage. Embryogenesis in *A. nanus* takes 50-60 hours at 25°C (as opposed to 12 hours for *C. elegans*). Also a different order of cell divisions leads to a different spatial arrangement of cells, which is incompatible with the cell-cell interactions found in the early *C. elegans* embryo. A series of unequal cleavages leads to the separation of soma and germ line and the establishment of five founder cells AB, MS, E, C, D and the primordial germ line cell P_4 as in *C. elegans*. In *A. nanus* however, the germline division occurs much earlier (6-cell stage) than in *C. elegans* (24-cell stage). Furthermore, in *A. nanus* there is no reversal of anterior-posterior cleavage polarity in P_2 as is the case for *C. elegans* (Schierenberg, E. 1987); this leads to a different arrangement of cells and all variants merge into a single pattern prior to the onset of gastrulation. Finally, in contrast to *C. elegans*, *A. nanus* exhibits prominent nucleoli (zygotic rRNA synthesis) in all early blastomeres from the 3-cell stage onward. Indeed, when exposed to the α-amanitin, an inhibitor of RNA synthesis, *C. elegans* 1-6 cell embryos go on to perform a normal cleavage program but arrest as 120-150 cell embryos without any signs of tissue differentiation. In contrast, *A. nanus* embryos exposed to the same α-amanitin

treatment arrest after a few cleavages, mostly at the 5-cell stage, which coincides with the appearance of nucleoli as markers of rRNA synthesis. This indicates that *A. nanus* initiates zygotic transcription during the very first cleavage steps, while in *C. elegans* maternal gene products are sufficient to allow development beyond the establishment of individual cell lineages. Hence it can be seen many fundamental aspects of early embryogenesis are quite different between these two nematode species.

A recent study looked into the embryonic cell lineage of the marine nematode *Pellioditis marina* and *Halicephalobus* sp (Houthoofd, W. *et al.* 2003). They found that the embryonic lineage of *P. marina* more closely resembles that of *C. elegans* than does *Halicephalobus* sp, as expected from their phylogenetic relationship with *C. elegans* (figure 1.1). The lineage homology between *P. marina* and *C. elegans* is 95.5 %, while the fate homology is only 76.4%. The lineage homology of *Halicephalobus* sp is only 74.1 % and the fate homology and 57.6 % with *C. elegans*. Hence there are significant cell lineage differences in quite closely related nematodes. Variations also can be observed between the development of several tissues between *C. elegans* and *P. marina*; for example, in the muscle lineage of *P. marina*, there are differences in the cell deaths after the last division round compared to *C. elegans*. Interestingly, in spite of this, the number of body muscle cells is the same in both *P. marina* and *C. elegans* (81 cells) and they share a similar configuration. Another difference between *C. elegans and P. marina* development is in pharyngeal development. In *P. marina*, Abp contributes pharyngeal cells, while in *C. elegans* it does not. A P_2 signal prevents the ABp lineage from responding to an MS-derived pharynx-inducing signal to produce pharyngeal cells (Hutter, H. *et al.* 1994; Mango, S. E. *et al.* 1994; Moskowitz, I. P. *et al.* 1994). Hence either this P_2

signal is absent in *P. marina* or later signals induce secondary production of pharyngeal cells. Finally, comparisons of the cell lineages of *P. marina*, *Halicephalobus* sp and *C. elegans* have also revealed that there are two fundamentally different patterns of cell fate determination. In *Halicephalobus* sp. the cell lineage is predominantly monoclonal e.g. pharynx and body muscle cells are formed in large identical clonal blocks of sublineages. In *P. marina* and *C. elegans*, the cell lineage is predominantly polyclonal, i.e. equivalent lineage blocks exhibit numerous individual fate transformations.

These studies clearly illustrate that *C. elegans* does not necessarily reflect all the details of developmental processes in other nematodes, even in closely related ones such as other rhabditids.

(ii) Endoderm specification in *C. elegans* and other nematodes

The early embryogenesis of the nematode *Acrobeloides nanus*, and particularly the specification of its endoderm, was studied in detail by Wiegner and colleagues (Wiegner, O. *et al.* 1998; Wiegner, O. *et al.* 1999). They compared the specification of gut cell fate in *A. nanus* and *C. elegans*. As in *C. elegans*, the endoderm in *A. nanus* derives solely from the E cell, which results from the cleavage of EMS into E and MS. However, unlike in *C. elegans*, the two daughter cells of E (Ea and Ep) occupy a larger and more anterior area inside the *A. nanus* embryo. Antibody staining also revealed that the final number of gut cell in *A. nanus* was 22 (there are only 20 in *C. elegans*, (Sulston, J. E. *et al.* 1983)). It is also worth noting that two markers of *C. elegans* gut differentiation, the birefringent and auto-fluorescent granules are absent in *A. nanus*. Nevertheless, gut differentiation can be monitored by the appearance of endocytotic function which is visualised by the

uptake of fluorescently labelled transferrin (Wiegner, O. *et al.* 1998). In *C. elegans*, endocytosis is first observed at the 16E cell stage, whereas in *A. nanus* this is seen as early as the 2E cell stage. However, not all gut differentiation markers appear at different times in development between the two species; the intestinal valve, which can be observed after staining with 1CB4 monoclonal antibody, appears at the same developmental stage in both nematodes.

In *C. elegans*, gut specification requires inductive interaction between P2 and EMS which depends on a direct cell-cell contact and can be inhibited by removal of P_2 by laser ablation (Goldstein, B. 1993). This interaction results in EMS asymmetrically dividing to yield two somatic founder cells MS and E, the latter forming the complete intestine. The molecular basis of this induction is not fully understood as yet, although genes homologous to those found in the Wnt/Wingless signalling pathways are involved (Rocheleau, C. E. *et al.* 1997; Thorpe, C. J. *et al.* 2000).

In *A. nanus*, ablation experiments have revealed that neither P_2, P_3 nor P_4 were necessary for EMS to produce E and MS and that no induction between EMS and either AB or P1 was necessary either. Further ablation experiments in early *A. nanus* embryos also showed that gut forming potential was present in AB and P2 in the absence of EMS, indicating that in *A. nanus* embryos, the gut-forming potential is a default state which becomes restricted to EMS by a lateral inhibition from adjacent blastomeres (Wiegner, O. *et al.* 1998; Wiegner, O. *et al.* 1999). Hence although endoderm is one of the most highly evolutionary conserved tissues there are significant differences in its specification and development between nematode species.

iii) Vulva development in *C. elegans* and other nematodes

One aspect of post-embryonic development that has been compared in detail between several nematode species is that of the vulva (Sommer, R. J. *et al.* 1995; Sommer, R. J. *et al.* 1996; Jungblut, B. *et al.* 1998; Sommer, R. J. *et al.* 1998). This provides another example of how developmental processes can vary between closely related nematodes. Species in most genera form the vulva in the central body region as in *Caenorhabditis*. A study by Sommer and Sternberg looked at the cell lineages and pattern formation in the vulva equivalence group in rhabditids of the five genera *Oscheius*, *Rhabditella*, *Rhabditoides*, *Pelodera* and *Protorhabditis* (Sommer, R. J. *et al.* 1995). The cell lineages that form the vulva are divided into primary, secondary and tertiary and most lineage differences between species are observed for the secondary and tertiary cell fates. The tertiary cell fates differ in the number of cell progeny they produce: two in *Caenorhabditis* or *Rhabditella*, four in *Oscheius* and six in *Pelodera* and *Rhabditoides* (Sommer, R. J. *et al.* 1995). It is possible that the observed variety of tertiary cell fates occurs due to the minimal functional importance of the number of cells formed; the cells either fuse with the epidermal syncytium which already contains many nuclei, or undergo programmed cell death. The secondary lineage differs among four, six and seven progeny cells in *Oscheius*, *Pelodera* and *Caenorhabditis*, respectively (Sommer, R. J. *et al.* 1996). These cells all form part of the vulva and alterations in the number of cells might be tolerated as long as enough cells are present to generate a scaffold for the structure of the vulva. This interpretation is consistent with the ultrastructural analysis of the vulva in *Caenorhabditis*, indicating that extensive cell fusion occurs during fourth larval stage (Sommer, R. J. *et al.* 1996). In contrast, the primary cell lineage is conserved among all analysed rhabditids that form vulva in the central body region.

In all species, eight cell progeny are generated indicating that this lineage may be more constrained based on the cell interactions taking place during vulval morphogenesis in the L4 stage.

Comparative analysis of the *lin-39* gene function in vulval development between *C. elegans* and *P. pacificus* provides an illuminating example of how molecular function can vary between relatively closely related nematode species. The six cells of the vulval equivalence group (P3.p-P8.p) in *C. elegans* are initially equipotent but respond to several interacting signalling systems, including an inductive Epidermal Growth Factor (EGF) mediated signal from the gonadal anchor cell. This determines which cells adopt the primary, secondary and tertiary cell fates. The descendants of these cells undergo a complex morphogenesis to form the vulva, while the remaining ventral hypodermal cells (P1.p, P2.p and P9.p-P11.p) fuse with the rest of the hypodermal syncytium during L1 stage. The hox gene *lin-39* is a transcription factor that regulates *C. elegans* vulva formation at two separate points: first its expression is required in early larval development to prevent the cells from the vulval equivalence group fusing with the hypodermal syncytium. Secondly, *lin-39* acts downstream of the EGF-RAS-MAPK signalling pathway in the vulval precursor primary cell fate in the L4 stage and is required for the induction of the primary cell fate. Detailed analysis of the *P. pacificus lin-39* homologue has shown that its role has changed during evolution. The vulva precursor cells of *P. pacificus lin-39* mutants undergo apoptosis rather than fusing to the hypodermal syncytium as they do in *C. elegans lin-39* mutants. Hence the function of *lin-39* has changed from the inhibition of cell fusion in *C. elegans* to the inhibition of cell death in *P. pacificus*. Furthermore, in *P. pacificus*, *lin-39* is required only once during vulval development, to prevent apoptosis of the vulva precursor cells, but is not required for

the vulval precursor cells to respond to gonadal signal during vulval induction as is the case in *C. elegans*. Hence it can be seen that post-embryonic development events can vary at both the cellular and molecular level between different nematode species even though the final developmental outcome appears conserved.

1.1.4 Use of *C. elegans* to study parasite gene function

C. elegans has been used as a surrogate expression system to study the promoters of a number of parasite genes and this is discussed in detail in the introduction to chapter 5. However, there are as yet only three published examples of the functional expression of a parasite gene in *C. elegans*: the *H. contortus* tubulin gene *tub-1* (Kwa, M. S. G. *et al.* 1995), the *H. contortus* cathepsin-L gene *cpl-1* (Britton, C. *et al.* 2002) and the *Onchocerca volvulus* glutathione-S-transferase gene *gst-3* (Krause, S. *et al.* 2001).

The first use of *C. elegans* as a surrogate expression system was an approach to study the mechanisms of resistance against benzimidazole (BZ) anthelmintics in *H. contortus* (Kwa, M. S. G. *et al.* 1995). *C. elegans* was used as a heterologous expression system for parasite BZ-sensitive and resistant alleles and also mutagenised forms of the *H. contortus* β-tubulin *tub-1* gene. Constructs introduced in *C. elegans* were found to be stably transmitted and expressed and the transformants were tested for BZ susceptibility. *H. contortus tub-1* alleles carrying a mutation that changed the 200th amino residue from a Tyrosine to a Phenylalanine conferred susceptibility to thiabendazole when expressed in BZ-resistant *C. elegans*. This demonstrated that the *H. contortus* β-tubulin genes could function appropriately *C. elegans*.

More recently, homologues of a cathepsin-L protease from *C. elegans* have been isolated from the animal parasitic nematodes *H. contortus*, *Dictyocaulus viviparus*, *Teladorsagia circumcincta*, *Ancylostoma caninum* and *Ascaris suum* as well as from the plant parasitic nematode *Heterodera glycines* (Britton, C. *et al.* 2002). It was shown that *Hc-cpl-1* could compensate for *Ce-cpl-1* loss-of-function in *C. elegans*. The *Ce-cpl-1* gene in *C. elegans* was knocked out by RNAi and the *Ce-cpl-1* RNAi phenotype was rescued by expressing the *H. contortus* homologue *Hc-cpl-1* from a transgene under the control of the *Ce-cpl-1* promoter. This was possible because the RNAi did not inhibit the parasite transgene because the sequence homology was not sufficient for the *C. elegans* cpl-1 dsRNA to interfere with the *Hc-cpl-1* mRNA. This work clearly showed that the *H. contortus* cpl-1 gene could function appropriately when expressed in *C. elegans* and that these genes are functionally conserved in the two species.

Genes from more distantly related nematodes have also been studied using *C. elegans*. A gluthatione-S-transferase gene from *Onchocerca volvulus*, *Ov-GST1a* was ectopically expressed in *C. elegans* in order to determine the ability of *C. elegans* to carry out correct processing and post-translational modifications (Krause, S. *et al.* 2001). The parasite derived *Ov-GST1a* gene product was correctly processed in transgenic *C. elegans* and post-translational modifications such as signal peptide cleavage and N-glycosylation were performed appropriately. More recent work on *O. volvulus* GST genes has also shown that *Ov-GST3* was able to function appropriately in *C. elegans* (Kampkotter, A. *et al.* 2003). This parasite enzyme is induced by oxidative stress and thought to be involved in the defence of the parasite against the host's immune system. Transgenic lines of *C. elegans* expressing this parasite gene were found to be more resistant to both internal and external oxidative

stress than wild-type worms, again showing the ability of this gene to function in *C. elegans.*

One example of a gene which does not appear to function appropriately in *C. elegans* is the *B. malayi* homologue of the prolyl-4-hydroxylase gene (P4H) (Winter, A. D. *et al.* 2003). This gene has unusual properties: the recombinant *B. malayi* α-subunit of this enzyme, PHY-1, is an active P4H itself. In the heterologous host *C. elegans*, high levels of *Bm-phy-1* expression were localized to the hypodermal tissue. However, *phy-1* was not able to replace enzyme function in a *C. elegans phy-1* mutant. This could be due to a number of reasons, such as low levels of expression, missplicing of introns or low activity of the transgenic protein. It is also possible that the recombinant protein, although shown to be active by itself, may have higher activity *in vivo* when associated with unidentified partners and *C. elegans* might fail to provide these. It is also of importance to note that the *C. elegans* and the *B. malayi* PHY-1 polypeptides shared only 59 % identity (reflecting the distant evolutionary relationship between the two nematodes).

These different studies suggest that *C. elegans* can be a suitable model for heterologous gene expression for at least some parasitic nematode genes. More work needs to be done to assess the suitability of *C. elegans* as surrogate expression system for different types of gene from different species of nematode.

1.2 *Haemonchus contortus:* an important veterinary parasite and a close relative of *C. elegans*

1.2.1 Description and life cycle

Haemonchus contortus is a parasitic nematode of sheep that belongs to the order Strongylida (phylum Nematoda) and falls into the superfamily Trichostrongyloidea. This group of parasites is in Clade V, as defined by rRNA sequence based phylogeny, and is the closest clade of parasitic nematodes to *C. elegans* (Blaxter, M. L. 1998). It is this group of parasites that are best placed to utilise *C. elegans* as a model system and consequently *H. contortus* was chosen as the primary parasite on which to focus this work. Trichostrongylid nematodes are especially common and pathogenic in grazing ruminants but other domesticated animal species are also hosts. The abomasum and small intestine are the location of the parasitic stages for most trichostrongylid species. The life cycle of *H. contortus* is direct (figure 1.2) (Urquhart, G. M. *et al.* 1987). The females are prolific egg layers and the eggs are released on the pasture in the sheep faeces. They hatch to L1 on the pasture and the time of development to L3 can be from a few days to several weeks or months depending on temperature and humidity. The cuticle of the L2 larvae is not shed during the moult to L3 but is retained and is called the L3 sheath in infective larvae. L3 larvae are ingested and exsheathed in the abomasum where they moult twice to become adults. The parasitic L4 and adult stages move freely on the surface of the mucosa and feed on host blood. The prepatent period (time of development from ingested L3 to a mature gravid adult worm) is approximately three weeks but under appropriate conditions the L4 can undergo arrested development and survive for many months in the abomasal glands of the host before maturing to the adult worm (Urquhart, G. M. *et al.* 1987).

1.2.2 The veterinary importance of *H. contortus*

Haemonchus contortus worms can be up to 30 mm in length and live in the abomasum of sheep. Their buccal cavity is armed with a lancet, which they use to pierce the host mucosa and feed on blood. The white egg-filled uterus of the female spirals around the blood-filled gut, giving rise to the so-called "barber pole" appearance. Burdens of up to 50,000 worms can occur in a single host and it has been estimated that heavy burdens of *Haemonchus contortus* may remove one fifth of the circulating erythrocytes per day during the course of an infection lasting several months (Urquhart, G. M. *et al.* 1987). The pathogenic effects of *H. contortus* results from the inability of the host to compensate for blood loss resulting in anaemia and often death. Signs of haemonchosis are pallor mucous membranes, a haematocrit reading of less than 15 %, extreme weakness and shortness of breath. The parasite is best adapted to warm humid conditions and is a major sheep and goat pathogen in most tropical and sub-tropical regions of the world. However it can adapt to colder climates and has been reported as far north as the artic circle (Lindqvist, A. *et al.* 2001). Control of the infection in tropical and subtropical areas varies, depending on the duration and number of periods in the year when rainfall and temperature permit high pasture levels of *H. contortus* larvae to develop. However in many regions prophylactic anthelmintic treatments are given at 2-4 weekly intervals for much of the year. This has led to an extremely high degree of selection pressure being applied to parasite populations and the resultant development of anthelmintic resistance.

1.2.3 Anthelminthic resistance in *Haemonchus contortus* and other trichostrongylid nematodes

In nematodes of small ruminants, and particularly in *H. contortus*, anthelmintic resistance threatens livestock production in many parts of the world (Waller, P. 1997; Jackson, F. *et al.* 2000). Resistance to benzimidazoles (BZ), ivermectin (IVM), levamisole (LEV) and pyrantel is now found in *H. contortus* and several other species of trichostrongylid nematodes of sheep (Gopal, R. M. 1999; Sangster, N. C. *et al.* 1999). Resistance to all anthelmintic classes, except ML, is also widespread in cyathosome parasites in horses (Woods, T. 1998). Although anthelmintic resistance is not as widespread in cattle nematodes, there have been reports of resistance to BZ and LEV in *Ostertagia ostertagi*, *Cooperia oncophora* and *Trichonstrongylus axei* (Eagleson, J. S. *et al.* 1986; Geerts, S. 1987; Eagleson, J. S. 1992; Vermunt, J. J. 1995; Jackson, F. *et al.* 2000), and to IVM in *Cooperia* spp, *Trichostrongylus* spp. and *H. placei* in cattle (Coles, G. C. 1998; Prichard, R. K. 2001). So far there is no conclusive evidence of anthelmintic resistance in human parasitic nematodes. However there is growing concern that resistance could develop since anthelmintic usage has dramatically increased in recent years (Prichard, R. K. 2001).

Hence there is an urgent need for new drugs and vaccines for parasitic nematodes. This in turn will require a deeper understanding of their biology and the development of appropriate tools and resources for experimental studies.

1.2.4 *H. contortus* as a "model parasite"

In addition to its importance as a pathogen, *Haemonchus contortus* is an important experimental model of nematode parasitism. The human hookworms *Necator* and *Ancylostoma*, which infect over one billion people annually (Bundy 1997) are from the same taxonomic order (Strongylida) as *Haemonchus contortus* and share the same blood-feeding lifestyle. *H. contortus* is the most experimentally tractable parasite of the strongylid group. *Haemonchus contortus* has been at the forefront of parasitic nematode vaccine research and basic biological research on these organisms (Newton and Munn 1999, Yatsuda *et al* 2003). Its high fecundity and relatively large size facilitates the production of large amounts of parasite material and also offers potential for physiological studies. It is also one of the few parasitic nematodes for which it is possible to perform genetic crosses (LeJambre *et al* 2000).

The genomic resources of this parasite are the best developed of all parasites with the exception of the human filarial nematode *Brugia malayi*. There are over 20,000 ESTs available and the construction of an integrated HAPPY /BAC clone map and 5-fold shotgun-sequencing project have been initiated for the sheep nematode *Haemonchus contortus* (J. Gilleard, personal communication). Hence both its phylogenetic position and its experimental tractability make it one of the most useful parasitic nematode species to exploit *C. elegans* biology and resources.

1.3 Overview of GATA transcription factors in *C. elegans* and other organisms

The GATA transcription factors are a group of transcriptional regulators with important functions in the development of several tissues in *C. elegans,* most notably endoderm and hypodermis (Page, B. D. *et al.* 1997; Zhu, J. *et al.* 1997; Fukushige, T. *et al.* 1998; Gilleard, J. S. *et al.* 1999; Gilleard, J. S. *et al.* 2001; Fukushige, T. *et al.* 2003). They have been well characterised in *C. elegans* and there are a variety of tools available to study their function. Consequently they represent an interesting and amenable class of molecule in which to study conservation of function between *C. elegans* and parasitic nematodes. This comparative analysis has value in several different ways. Firstly, it should indicate the extent to which there is conservation of gene function and the regulation of particular pathways or processes between *C. elegans* and individual parasitic nematode species. Secondly, it may provide important information regarding functional polypeptide domains and regulatory elements Thirdly, it may allow the development of *C. elegans* as a surrogate expression system to directly study parasite gene function; something which is not possible in the parasite itself.

1.3.1 The vertebrate GATA transcription factors

GATA transcription factors are a class a DNA binding proteins that were first identified in vertebrates and are found in plants, fungi and metazoa where they play important roles in embryonic development and cell fate specification (Lowry, J. A. *et al.* 2000; Patient, R. K. *et al.* 2002). The vertebrate GATA factors are characterised by a highly conserved DNA-binding motif consisting of two zinc fingers of the motif Cys-X_2-Cys-X_{17}-Cys-X_2-Cys followed by a basic domain (figure 1.3). They bind to

the classic consensus sequence (A/T)GATA(A/G) -although this consensus sequence has recently been extended to (A/T)GAT(A/C)(A/G)(Newton, A. *et al*. 2001)- and can act as either transcriptional activators or repressors (Ko, L. J. *et al*. 1993; Omichinski, J. G. *et al*. 1993; Trainor, C. D. *et al*. 2000; Newton, A. *et al*. 2001). There are six family members in vertebrate species named GATA-1 to GATA-6. Structure function-relationships have been studied for vertebrate and yeast GATA factors and it has been determined that the C-terminal finger and adjacent basic domain is necessary for specific DNA binding in vitro (Omichinski, J. G. *et al*. 1993; Muro-Pastor, M. I. *et al*. 1999; Trainor, C. D. *et al*. 2000). The N-terminal and central portion of the GATA-1 DNA-binding domain makes specific contact within the major groove whereas the C-terminal portion makes site-specific interactions within the minor groove (Trainor, C. D. *et al*. 2000; Newton, A. *et al*. 2001)). The role of the N-terminal zinc finger has been investigated further in recent years and new roles for this domain have been uncovered. Until recently it was believed that the N-terminal zinc finger of GATA-1 did not interact with the DNA (A/T)GATA(A/G) motif, but instead was instrumental in recruiting co-regulatory proteins, such as Friend of GATA (FOG) (Tsang, A. P. *et al*. 1997). These results seemed to suggest that the regulation of gene expression by GATA-1 depends on DNA-binding mediated by the C-finger and recruitment of various cofactors by the N-finger (and additional domains within GATA-1, including the C-finger) (Newton, A. *et al*. 2001). However, the situation is in fact more complex and it has now been demonstrated that the N-finger of GATA-1 binds GATC motifs *in vitro* and that GATA-1 is capable of activating transcription via GATC elements (Newton, A. *et al*. 2001). An earlier study had also reported that the N-finger and C-finger of GATA-1 and GATA-2 influenced one another in DNA-binding (Trainor, C. D. *et al*. 2000).

Indeed it was showed that their activity and specificity were modified such that binding to certain sites was inhibited. The study also revealed that the N-terminal finger of GATA-1 bound to double DNA sites, i.e. two GATA consensus binding sites next to each other either on the same or opposite DNA strands, suggesting a possible influence on the association and activity of other factors involved in erythroid differentiation.

There are six vertebrate family members (GATA-1 to GATA-6), all of which have central, distinct and non-redundant roles in differentiation and tissue-specific gene expression in a large variety of tissues during development (Laverierre, A. C. *et al.* 1994; Morrisey, E. E. *et al.* 1997; Charron, F. *et al.* 1999; Lowry, J. A. *et al.* 2000; Molkentin, J. D. 2000; Shimizu, R. *et al.* 2001; Patient, R. K. *et al.* 2002). GATA-1, -2 and –3 are mostly expressed in hematopoietic cells. GATA-1 is a key activator of globin genes and binds to conserved GATA sites in their promoters and the locus control region (Orkin, S. H. *et al.* 1998; Ohneda, K. *et al.* 2002). Its expression characterises more differentiated states with transcripts being found in mature erythroid cells and megakaryocytes (Charron, F. *et al.* 1999). In contrast, GATA-2 expression is initially required for the proliferation of the hematopoietic precursors (Ohneda, K. *et al.* 2002) whilst GATA-3 expression is restricted to T lymphocytes and is required for T-cell differentiation (Samson, S. I. *et al.* 2003; Zhou, M. *et al.* 2003). GATA-1, -2 and –3 also have roles later in development in non-hematopoietic cells, namely the gonads for GATA-1 and the brain and kidney for GATA-3. GATA-4, -5 and –6 are all involved in endoderm and mesendoderm-derived tissue development including the heart, liver, lung, gonad, and gut (Laverierre, A. C. *et al.* 1994; Lyons, G. E. 1996Molkentin, 2000 #36; Morrisey, E. E. *et al.* 1997; Charron, F. *et al.* 1999; Zhang, H. *et al.* 2003). In these tissues, they

play a critical role regulating tissue-specific gene expression. GATA-4 transcripts are found in the heart, lung, liver and small intestine and also in the gonads of the adult mouse (Molkentin, J. D. 2000); GATA-5 is expressed in adult small intestine, stomach, bladder and lungs, whereas developmental expression is detected in the allantois, heart, outflow tract, lung bud, urogenital bridge, bladder and gut epithelium (Molkentin, J. D. 2000). Finally GATA-6 was shown to be expressed in the adult heart, aorta, stomach, small intestine and bladder and weakly in the liver and lungs. During embryonic and foetal development, GATA-6 mRNA is detected in the primitive streak, allantois, visceral endoderm, heart, lung buds, urogenital ridge, vascular smooth muscle cells and the epithelial layer of the stomach, small intestine and large intestine (Molkentin, J. D. 2000). Within the heart, the three GATA factors are differentially regulated throughout development with GATA-4 being the predominant transcript in cardiomyocytes at all stages. GATA-6 is also expressed in the precardiac mesendoderm at the late primitive streak stage and is later found in myocardial cells and also in the vascular smooth muscles (Morrisey, E. E. *et al.* 1996). Recent studies have also suggested another role for GATA-4, -5 and –6 in steroidogenesis (Tremblay. J. J. *et al.* 2003). Indeed, they observed that some steroidogenic gene regulatory regions contain GATA elements and hence that GATA factors, especially GATA-4 and GATA-6 which are also expressed in the adrenals and the gonads, could have a role in regulating steroidogenesis. *In vitro* experiments also revealed that GATA-4 is phosphorylated in steroidogenic cells and that phosphorylation levels are rapidly induced by camp (Tremblay. J. J. *et al.* 2003). Phosophorylation increases GATA-4 DNA-binding activity and enhances its transcriptional properties on multiple steroidogenic promoters. Finally it appears that all six GATA factors are involved in mammalian reproduction (LaVoie, H. A. 2003).

Indeed, GATA-4 plays a role in foetal male gonadal development by regulating several genes that in turn regulate Mullerian duct regression and the onset of testosterone production (LaVoie, H. A. 2003). GATA-2 expression appears to be sexually dimorphic; it is found to be transiently expressed in the germ cell lineage of the foetal ovary but not the foetal testis (LaVoie, H. A. 2003). In the reproductive system, GATA-1 is exclusively expressed in Sertoli cells at specific seminiferous tubule stages. In addition, GATA-4 and GATA-6 are localised primary to ovarian and testicular somatic cells and that GATA-2 and GATA-3 are expressed in pituitary and placental cells possibly regulating alpha-glycoprotein subunit gene expression (LaVoie, H. A. 2003). Hence it can be seen that the vertebrate GATA factors are involved in the development and maintenance of a multitude of different tissues.

1.3.2 GATA transcription factors in *C. elegans*

There are eleven GATA factors in *C. elegans*. Namely, *elt-1*, *elt-2*, *elt-3*, *elt-4*, *elt-5* (or *egl-18*, as it has been renamed) *elt-6*, *elt-7*, *end-1*, *end-3*, *med-1* and *med-2*. Like GATA transcription factors in vertebrates, they play important roles in the regulation and determination of cell-fate specification and differentiation. The vertebrate GATA factors are often placed in two groups (GATA-1, -2, -3 and GATA-4, -5, -6), based on their major sites of expression. Similarly *C. elegans* GATA factors can be grouped into two major tissue types; *elt-1*, *elt-3*, *elt-5 and elt-6*, which are involved in hypodermal development and *elt-2*, *elt-4*, *elt-7*, *end-1*, *end-3*, *med-1* and *med-2*, which are involved predominantly in endodermal (as well as mesodermal) development. Figure 1.3 illustrates the different domain structures of the *C. elegans* GATA factors and the typical vertebrate GATA factor structure. A

24

number of studies have examined the role of all these genes in *C. elegans* and the following provides a review of the current state of knowledge. *elt-1* was the first GATA factor to be isolated in *C. elegans* and was so named after its similarities with GATA-1; hence the name *erythroid-like transcription factor-1* (Spieth, J. *et al.* 1991). The hypodermal GATA factors will be discussed first and then the endodermal GATA factors. Spieth and colleagues observed that *elt-1* was expressed almost exclusively in embryonic stages and that it contained a region encoding two zinc finger domains very similar to the DNA binding domain of GATA-1 (Spieth, J. *et al.* 1991). ELT-1, like vertebrate GATA factors, contains two zinc finger domains (figure 1.3) and interestingly, it is the only *C. elegans* GATA factor to do so. A later study showed that *elt-1* is required for the production of epidermal cells (Page, B. D. *et al.* 1997). It was shown that epidermal cells changed fate to either neuronal or muscle cells in *elt-1* null mutants. A second hypodermal GATA factor, *elt-3*, was later identified and this is expressed in ventral and dorsal hypodermal cells, as well as hypodermal cells from the head and tail, immediately after they are formed but is not expressed in the lateral hypodermal cells (Gilleard, J. S. *et al.* 1999). It was also reported that *elt-1* was required for the formation of most but not all *elt-3*-expressing cells (Gilleard, J. S. *et al.* 2001). *elt-1* and *elt-3* have both been shown, by ectopic expression experiments, to be sufficient to activate a program of hypodermal differentiation (Gilleard, J. S. *et al.* 2001). However only *elt-1* is essential to hypodermal cell formation with *elt-3* appearing to have a redundant role (Page, B. D. *et al.* 1997; Gilleard, J. S. *et al.* 1999; Gilleard, J. S. *et al.* 2001). In contrast to ELT-1, ELT-3 only contains one GATA-type zinc finger. The other two hypodermal GATA factors are *elt-5* (*egl-18*) and *elt-6* (Koh, K. *et al.* 2001) and these have been shown to be involved in the development of the lateral epidermal cells (the seam

cells) and are required continuously for post-embryonic development of seam cells and for normal molting. Both ELT-5 and ELT-6 contain only one GATA-type zinc finger domain. Inhibition of *elt-5* through RNAi results in penetrant late embryonic (pretzel) and early larval lethality (early L1) with various phenotypes (Lpy, Unc, and slightly Dpy), whereas *elt-6* RNAi does not produce any obvious phenotype. After *elt-5* RNAi, seam cells do not differentiate properly, alae are absent and expression of seam-specific markers is absent. Double inhibition of *elt-5* and *elt-6* does not result in an enhanced phenotype over the *elt-5* RNAi phenotype. Interestingly, *elt-5* dsRNA down-regulates of both *elt-5* and *elt-6* in seam cells and these genes are expressed in an operon. Data from these experiments suggest that *elt-5* is essential and *elt-6* is redundant, however *elt-5* dsRNA defects can be rescued by expression of *elt-6*, implying that the two gene products can to some extent be interchangeable.

The seven remaining GATA factors, *elt-2, elt-4, elt-7, end-1, end-3, med-1* and *med-2* are all involved in endoderm specification and development (figure1.6) (Zhu, J. *et al.* 1997; Fukushige, T. *et al.* 1998; Zhu, J. *et al.* 1998; Maduro, F. M. *et al.* 2002; Fukushige, T. *et al.* 2003). *elt-2* has been studied in great detail and has been shown to be essential to endoderm formation (Fukushige, T. *et al.* 1998). ELT-2 is first expressed at the 2E cell stage and is expressed continuously throughout development and is restricted to the endoderm lineage. It contains one C-terminal zinc finger and an N-terminal zinc finger-like domain that has been termed a "pseudofinger" (figure 1.3). *elt-2* null-mutants arrest development as early L1 with an abnormal intestinal brush border, showing that *elt-2* is essential. Ectopic expression experiments demonstrated that *elt-2* is capable of activating a program of gut cell differentiation and can activate the expression of a number of terminal differentiation genes such as *ges-1* (Fukushige, T. *et al.* 1998; Maduro, F. M. *et al.*

2002). *elt-7,* a second endodermal GATA factor was more recently identified as acting redundantly with *elt-2* and was shown, through ectopic expression experiments, to be able to activate a program of gut cell differentiation even though *elt-7* loss of function through RNAi does not result in any abnormalities (Strohmaier and Rothman, personal communication). This suggests that *elt-2* and *elt-7* are a pair of genes acting in a concerted fashion, with *elt-7* not being essential and having possibly arisen by duplication of *elt-2. end-1, end-3, med-1* and *med-2* have also been shown to be sufficient to activate a program of endodermal differentiation by ectopic expression experiments. The two GATA factors acting upstream of *elt-2*, and most likely activating its expression, are *end-3* and *end*-1 which together with *dpr-1* (a nuclear receptor type transcription factor formerly known as *end-2*), form a region known as the EDR (endoderm-determining region). *end-1* and *end-3* are both expressed in the E blastomere shortly after its birth, indicating that the *end* genes are the earliest known zygotically expressed genes in the endoderm lineage (Zhu, J. *et al.* 1997). Both genes overlap in function and *end-1* can rescue the intestinal differentiation defects but not the embryonic lethality of the deficiencies in EDR-deficiency embryos, suggesting that *end-1*(+) activity can direct E-cell specification in EDR deficiency embryos (Zhu, J. *et al.* 1997). However *end-1* and *end-3* are not completely redundant for endoderm specification: while simultaneous elimination of the function of both genes leads to a large proportion of embryos without endoderm, if *end-3* function is depleted alone a small proportion of embryos also lack endoderm. Furthermore, a point mutation altering a residue in the zinc finger of END-3 leads to an impenetrant lack of intestine (Maduro, F. M. *et al.* 2002). The two GATA factors *med-1* and *med-2*, which act upstream of *end-1* and *end-3,* are both involved in mesoderm specification and regulate the specification of the EMS

blastomere into E and MS; they also appear to have mutually redundant functions. *med-1* and *med-2* are both activated by *skn-1*, marking the passage from maternal to zygotic control in mesendoderm specification (Maduro, M. F. *et al.* 2001). Both *med-1* and *med-2* are expressed exclusively in the EMS blastomere, promoting the specification of EMS into E and MS. Finally the last endodermal GATA factor is *elt-4*, although a functional role has yet to be established and it has not been shown to be essential (Fukushige, T. *et al.* 2003).

Hence in summary, the only *C. elegans* GATA factors that appear to have essential, non-redundant functions are *elt-1*, *elt-2* and possibly *elt-5*. Interestingly, most of the *C. elegans* GATA factors appear to be present as "functional pairs" of single-finger GATA factors; *med-1/med-2*, *end-1/end-3*, *elt-2/elt-7* and *elt-5/elt-6*. Only ELT-1 contains two GATA-type zinc fingers, a characteristic feature of the vertebrate GATA factors (Lowry, J. A. *et al.* 2000; Patient, R. K. *et al.* 2002). The C-terminal zinc finger of vertebrate GATA factors is involved in binding to a cognate GATA site (Ko, L. J. *et al.* 1993; Merika, M. *et al.* 1993; Omichinski, J. G. *et al.* 1993; Trainor, C. D. *et al.* 2000) and shares the most homology with the C-terminal finger of ELT-1 and the single fingers of the remaining *C. elegans* GATA factors (Lowry, J. A. *et al.* 2000; Patient, R. K. *et al.* 2002). However, recent research suggests that the N-terminal finger, as well as being involved in interactions with co-regulators, can also interact with DNA (Newton, A. *et al.* 2001). One possibility then is that the single GATA factor pairs achieve a higher promoter-binding specificity through heterodimeric interactions, perhaps necessitating the existence of pairs of partially redundant, single-finger GATA factors (Maduro, F. M. *et al.* 2002). Although direct evidence has yet to be produced to support this theory,

this would certainly explain the presence of pairs of partially redundant GATA factors

1.4 Endoderm development in *C. elegans*

A major focus of this thesis is the comparative analysis of the GATA transcription factor *elt-2* between *H. contortus* and *C. elegans;* a gene that has a key role in the regulation of endodermal differentiation and gut development in *C. elegans*. Consequently, *C. elegans* endodermal development is briefly reviewed here.

C. elegans embryogenesis has been described in detail and endoderm development in *C. elegans* is relatively simple (Sulston, J. E. *et al.* 1983; Sulston, J. E. *et al.* 1988). All gut cells are derived from the clonal expansion of a single blastomere, called the E-blastomere (Sulston, J. E. *et al.* 1983; Sulston, J. E. *et al.* 1988). The one-cell embryo (or P_0 cell) divides to give rise to two cells (AB and P_1). Each of these divides once more resulting in the 4-cell embryo (ABa, ABp, P_2 and EMS). The EMS blastomere then divides asymmetrically at the next cell division producing the E and MS blastomeres; section 1.3.3 has described the role of the GATA transcription factors *med-1* and *med-2*, regulated by the maternal factor SKN-1, in promoting the division of EMS into E and MS (Maduro, M. F. *et al.* 2001). The MS blastomere produces cells that will contribute to the formation of different tissues such as muscles and neurons (Sulston, J. E. *et al.* 1983; Sulston, J. E. *et al.* 1988). The E blastomere then divides clonally to give rise to gut cells (figure 1.4). Specification of the E blastomere fate requires a signal from the P_2 cell at the 4-cell stage, which polarises the EMS blastomere (Goldstein, B. 1993). This involves the

Wnt signalling pathway (Thorpe, C. J. *et al.* 2000). The E cell then divides on the ventral surface of the embryo producing two cells, Ea and Ep which migrate into the interior of the embryo during gastrulation. The subsequent divisions of the E-lineage occur after gastrulation is complete. At the end of the embryogenesis the final gut comprises of 20 cells. Endoreplication occurs during post-embryonic development leading to gut nuclei containing thirty-two times the haploid content (White, J. 1988).

The gene regulatory network controlling endoderm development appears to be conserved across metazoan phylogeny (Maduro, F. M. *et al.* 2002). Indeed, many genes involved in the control of *C. elegans* endoderm specification have clear homologues in other metazoan such as flies, Xenopus or even mammals (Thorpe, C. J. *et al.* 2000). Two maternal regulatory pathways specify the identity of EMS and its daughters (Figure 1.5). Both function within EMS itself: the first activates the genes that specify both MS and E fates, while the second, which makes the E-cell differ from MS, is part of a reiterative switching system that directs daughters of symmetric cell divisions to acquire differential transcriptional states. As previously described in section 1.4.1, the EMS cell divides into E and MS in the six-cell embryo in response to the signal from blastomere P_2. This signal involves the Wnt pathway, which results in the polarization of EMS so that it produces E and MS and also orients its mitotic spindle. The Wnt pathway involves several maternally expressed genes and loss-of-function mutants for these genes fail to produce endoderm; instead, the EMS gives rise to two MS-like daughters (Maduro, 2001 #101;Maduro, 2002 #96;Thorpe, 2000 #198}. It is thought that the *gsk-3* homologue *sgg-1* (a member of the glycogen synthase kinase family) is responsible for the control of endoderm induction and mitotic spindle orientation, as the other Wnt genes acting

downstream of *sgg-1* appear not to be required for mitotic spindle orientation (Schlesinger, A. *et al.* 1999). Orientation of the mitotic spindle does not require gene transcription in EMS, suggesting that Wnt signaling may directly target the cytoskeleton in a responding cell (Schlesinger, A. *et al.* 1999). The genes acting upstream of *sgg-1* are the *mom* genes *mom-1*, *mom-2* and *mom-5* (for <u>mo</u>re <u>m</u>esendoderm), while *lit-1* (<u>l</u>oss of <u>i</u>ntestine) and *pop-1* (<u>po</u>sterior <u>p</u>harynx-defective) act downstream of *sgg-1* (the pathway is summarised in figure 1.5). *mom-1* is maternally expressed in P_2 and it signals EMS for *mom-5* expression through *mom-2*. *sgg-1* is then activated in EMS by *mom-5* and polarizes the cell to establish endoderm-forming potential in the future E-cell. *mom-1*, *mom-2* and *mom-5* encode homologues of known components of Wnt pathways in other organisms. *mom-1* encodes a protein highly similar to *Drosophila* porcupine (*porc*), while *mom-2* is most similar to wingless (wg) and *mom-5* to frizzled (fz). *sgg-1* then activates the β-catenin homologue *wrm-1*, nuclear accumulation of which inhibits the maternal expression of *pop-1* in E (Thorpe, C. J. *et al.* 2000). The TCF/Pangolin homologue *pop-1* is an inhibitor of the gene cascade in the E-cell leading to endoderm specification. The APC-tumor suppressor-related gene called *apr-1* is also required for endoderm induction at the same level as *wrm-1*. *pop-1* is down-regulated only in the posterior daughter of EMS to permit endoderm fate and a mutation that eliminates maternally expressed *pop-1* results in both EMS daughters producing endoderm (Thorpe, C. J. *et al.* 2000). The two genes inhibited by POP-1 in the non-endoderm forming blastomere MS are the GATA transcription factors *end-1* and *end-3* which are activated in E by the maternally expressed GATA factors *med-1* and *med-2* (figure 1.6). The maternally expressed *skn-1*, a bZIP/homeodomain, activates expression of *med-1* and *med-2* in EMS, marking the switch from maternal to

zygotic control in mesendoderm specification (figure 1.6). *skn-1* mRNA is maternally contributed and found throughout all cells in the 4-cell embryo; however SKN-1 protein is translated asymmetrically, appearing at higher levels in P_1 descendants specifically (Bowerman, B. *et al.* 1992; Bowerman, B. *et al.* 1997). At the 4-cell stage maternal SKN-1 is present in EMS and P_2 but is required for the identity of, and functions only in, the EMS blastomere (Bowerman, B. *et al.* 1992; Bowerman, B. *et al.* 1997). *med-1* and *med-2* are a pair of unlinked, though nearly identical genes, encoding GATA transcription factors (see section 1.3.2). The regulation of the *med* genes by *skn-1* appears direct, as their promoters contain clusters of SKN-1 binding sites that are essential to reporter expression (Maduro, M. F. *et al.* 2001). At high levels, SKN-1 appears sufficient to activate *med* transcription, as shown by ectopic expression of *skn-1*, which results in widespread expression of the *med* genes and the activation of a widespread program of mesendoderm cell differentiation. The activation of the *med* genes by *skn-1* in P_2, sister of EMS, is inhibited by the expression of the maternally provided transcription factor PIE-1. The *med* genes, in turn activate two other GATA factors, *end-1* and *end-3* (described in details in section 1.3.2). The activation of zygotic *end-1* and *end-3* marks the switch from maternal to zygotic control in the endoderm lineage. These two genes activate the expression of another GATA transcription factor, *elt-2*, in the Ea and Ep cells, when they are ingressing towards the center of the embryo. *elt-2* is then expressed in all the subsequent E-lineage cells from the 2-E cell stage onwards. This gene is responsible for the activation of terminal differentiation products, such as the gut esterase gene *ges-1*, (Stroeher, V. L. *et al.* 1994; Fukushige, T. *et al.* 1998; Marshall, S. D. G. *et al.* 2001). Hence the early stages of endoderm differentiation are regulated by a cascade of GATA transcription factors.

1.6 Aims of the project

There are several main aims to this project. Firstly, to undertake a comparative analysis of GATA transcription factors between *C. elegans* and parasitic nematodes using available sequence databases and molecular biology approaches. Secondly, to test the extent of GATA factor functional conservation between *C. elegans* and parasitic nematodes and also the extent to which endoderm development is conserved. Thirdly, to explore *C. elegans* as a surrogate expression system as a tool to study parasite gene function and regulation using GATA factors as an example. The latter two aims are investigated in detail in the thesis using the *Haemonchus contortus* homologue of the *C. elegans* endodermal GATA factor *elt-2*.

Strongylida **Haemonchus**, *Ostertagia, Trichostrongylus, Nippostrongylus, Necator, Ancylostoma,*

Rhabditina **Caenorhabditis**, *Oscheius, Pellioditis*

Diplogasterida *Pristionchus*

Strongyloididae *Strongyloides*

IVa Panagrolaimidae *Panagrellus, Halicephalobus*

Cephalobidae *Acrobeloides*

IVb

Tylenchida *Meloidogyne, Globodera*

Spirurida *Brugia, Onchocerca, Dirofilaria*

Ascaridida *Ascaris, Toxocara*

Enoplida *Enoplus brevis*

Trichocephalida *Trichinella, Trichuris*

V
III
II
I

Infective stage

Figure 1.2: *H. contortus* **life cycle.**

Eggs pass out of infected sheep in faeces which hatch and undergo two moults to the L3 stage over a period of time depending on climatic conditions. The L3 larvae are the infective stage which are quiescent and free-living. These are ingested by the sheep grazing on the pasture. Once in the abomasum, infective L3 will undergo two more moults to reach the adult stage

1) Vertebrate GATA transcription factors

Vertebrate GATA −1 to-6 (all essential genes)

2) *C.elegans* GATA transcription factors

(A)

elt-1 (essential function)

(B)

elt-2 (essential function)

(C)
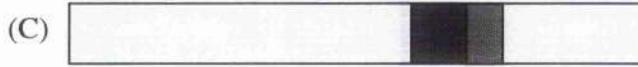
elt-3, elt-5, elt-6, med-1, med-2, elt-4, elt-7, end-1, end-3

"pseudo" zinc finger

GATA type zinc finger: $CX_2CX_{17}CX_2C$

Basic domain

Figure 1.3: Structure of GATA transcription factors in vertebrates and *C.elegans*.

All vertebrate GATA factors have the same structure with an N- and C-terminal zinc finger domain (1). However in *C.elegans* the eleven GATA factors fall into three categories. The first type has the same structure as the vertebrate GATA factors (A), the second type has one true C-terminal zinc finger domain and one degenerate N-terminal zinc finger domain (B), and finally, the third type only has one C-terminal zinc finger domain (C). The majority of *C.elegans* GATA factors fall in the last category .

(A)

(B)

(C)
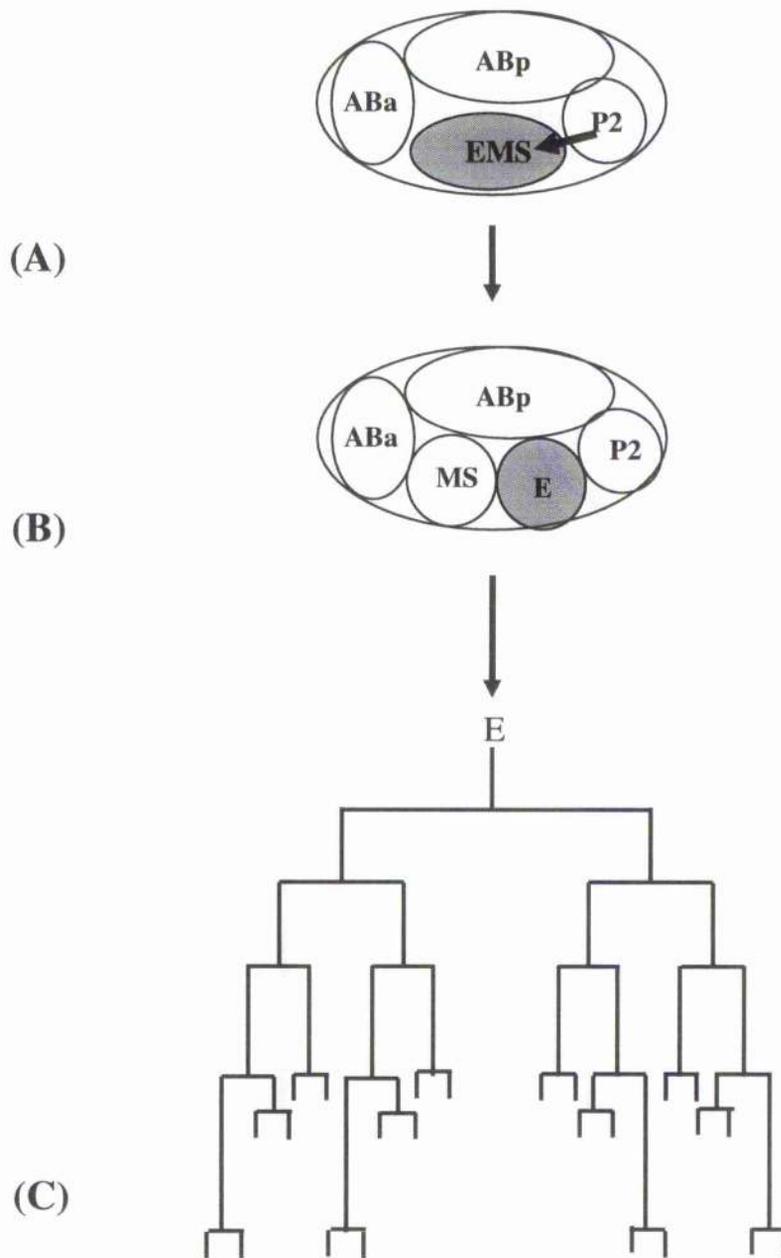
Figure 1.4: Initiation of the endoderm lineage in *C.elegans*.

P_2 signals EMS to divide into MS and E (A), resulting

in a 5-cell embryo (B). The E-cell then proceeds to 15 division

cycles, producing the 20 endodermal cells (C). Adapted from

Sulston, {Sulston, 1983 #94}.

Figure 1.5: Wnt signalling pathway in the early *C. elegans* embryo (adapted from Thorpe *et al*, 2000).

Wnt signals from P2 induce two responses in EMS. One is the polarization of endoderm potential in EMS such that POP-1 is downregulated in its daughter E, thereby permitting the specification of endoderm fate. Another response id the rotation of the centrosomal/nuclear complex in EMS, which requires the Wnt pathway only through *sgg-1*, but does not require the more downstream components *wrm-1*, *apr-1* or *pop-1*. *mom-1* is thought to be required for secretion of the *mom-2* ligand. *mom-5* encodes a predicted seven-pass trans-membrane protein receptor that presumably binds and responds to MOM-2.

mom-1   P$_2$

mom-2

mom-5   EMS

sgg-1   apr-1

wrm-1

spindle
orientation

pop-1   E

endoderm

38
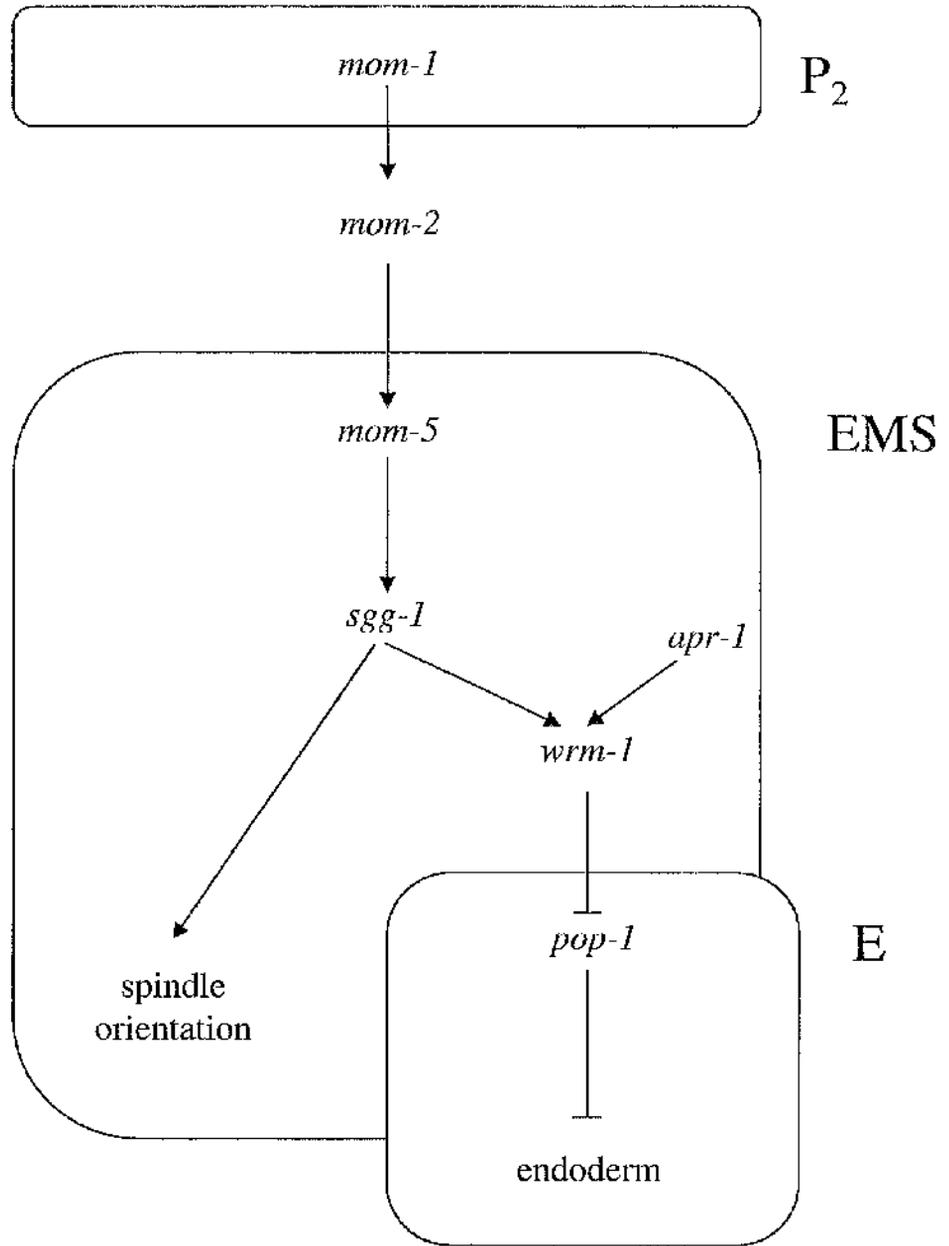
Figure 1.6: Gene pathway regulating the specification and development of endoderm in early *C. elegans* embryos.

The Wnt pathway is represented by one box and shown in more detail in figure 1.5. Maternally expressed genes are represented by green boxes and zygotically expressed genes by blue boxes. Red arrows indicate activation and black connectors repression. Genes acting as intermediate are shown in blue where they are thought to intervene.
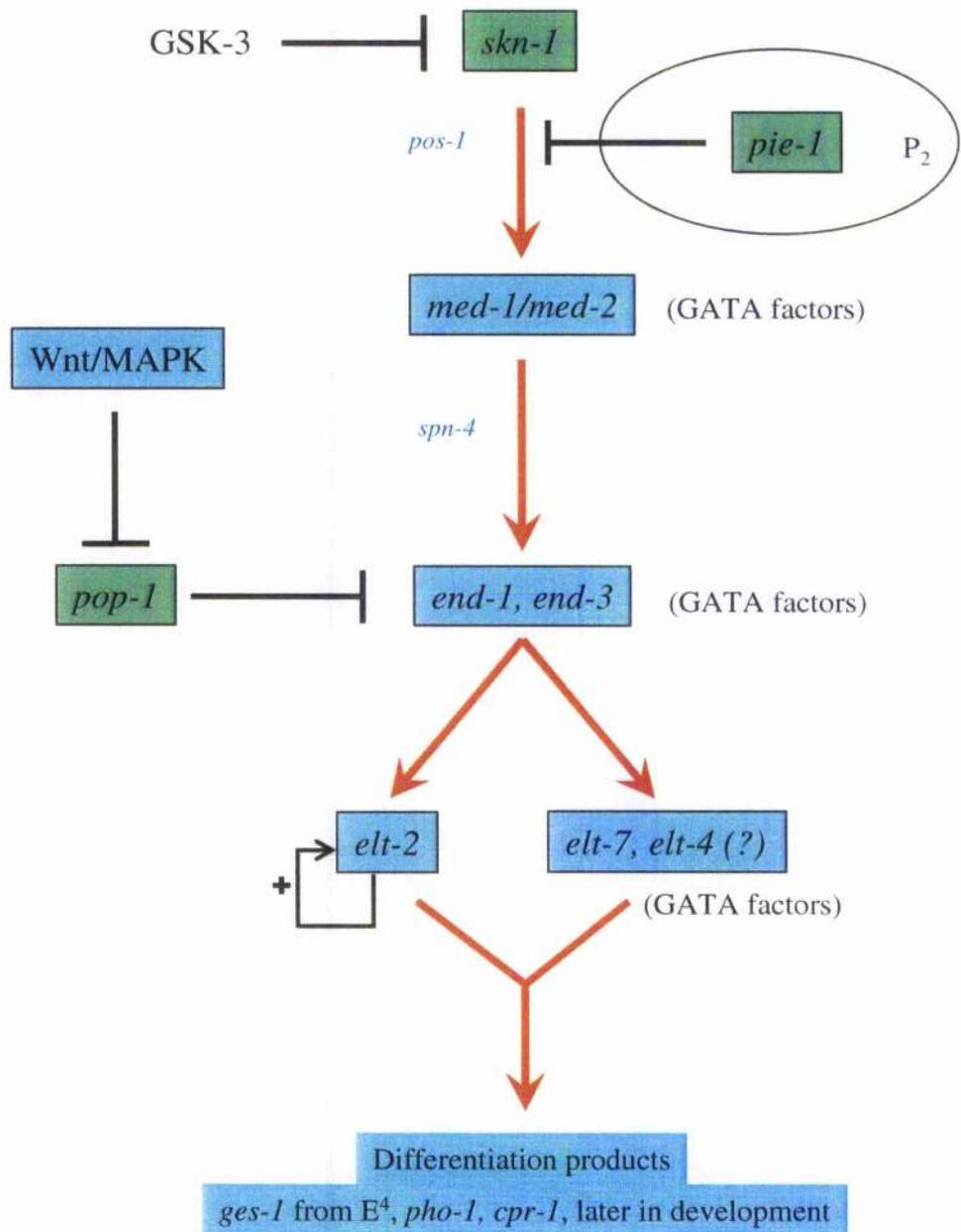
Chapter 2: Material and methods

2.1 *Caenorhabdtis elegans* methods

2.1.1 Culture and maintenance of *Caenorhabditis elegans*

C. *elegans* was grown and maintained on Petri dishes containing NGM agar (see appendix 1) and seeded with a lawn of the *E. coli* strain OP 50. (Lewis, J. A. *et al.* 1995; Stiernagle, T. 1997). Strains were acquired from the *Caenorhabditis* Genetics Center (CGC) unless otherwise stated.

Worm stocks were frozen using the method described by Stiernagle (Stiernagle, T. 1997). Worms were washed off plates with S buffer (see appendix 2), 0.6 ml pipetted into a cryovial and an equal volume of S buffer + 30 % glycerol was added and gently mixed. The vials were then placed into a small Styrofoam box which was then placed in the freezer overnight at -80°C. Stocks were thawed as required using the method described by Stiernagle (Stiernagle, T. 1997).

2.1.2 Transformation of *C. elegans*

The injection procedure used for transformation of DNA into *C. elegans* was as described in detail by Stinchcomb, Mello and co-workers (Stinchcomb, D. T. *et al.* 1985; Mello, C. C. *et al.* 1992). DNA was injected into the cytoplasm of the syncytial, mitotically active gonad of adult hermaphrodites. Following injection, the DNA concatenates to form extrachromosomal arrays that consist of multiple tandem copies of the introduced DNA (Stinchcomb, D. T. *et al.* 1985). These arrays behave

genetically in the same way as free chromosomal duplications and are inherited through the germline in a non-mendelian fashion. Generally a marker allowing recognition of the transgenic animals is co-injected with the DNA of interest to facilitate the identification of individual worms carrying the transgene in the F1 and F2 progeny. There are several such markers in common use and the one used in this work was the plasmid pRF4 which encodes the *su1006* dominant mutant allele of the cuticle collagen gene *rol-6* (Kramer, J. M. *et al.* 1990). This mutation confers a right roller phenotype that is easily identifiable under low power magnification.

Preparation of DNA for microinjection.

The quality of the DNA preparation used for injection is critical to the success of the transformation procedure. Contaminants such as bacterial endotoxins can be lethal to the injected worm and therefore high quality plasmid DNA preparations are required. Consequently, Qiagen Mini or Midi kits were used for preparation of DNA for microinjection. The concentration of the injected DNA is also critical to successful transformation. If the concentration is too high it can be lethal to the injected hermaphrodite or F1 progeny depending on the genes involved (Stinchcomb, D. T. *et al.* 1985). The injected DNA concentration also determines the size of the extrachromosomal arrays produced and the efficiency of the transmission of an array is dependant on its size (Mello, C. C. *et al.* 1995). If the concentration is too low then only very small arrays will form which are poorly transmitted through the germ line resulting in few transgenic F2 and stable transgenic lines. It is generally recommended that the total concentration of DNA injected is at least 100 ng/µl for the DNA to be able to form a stably transmitted extrachromosomal array (Mello, C.

C. *et al.* 1992). In this work, the DNA for injection was diluted in an injection buffer (see appendix 2) to achieve a final concentration of 100-400 ng/µl for the pRF4 marker plasmid and 5-10 ng/µl of the plasmid DNA of interest. If the concentration of either plasmid was reduced, pBlueScript II plasmid (Stratagene, USA) was added ensure the total DNA concentration remained constant so that the extrachromosomal arrays formed following injection were of adequate size. DNA preparations can contain contamination with particulate matter that can cause needle clogging and it is therefore important to remove this. Consequently, all DNA solutions were centrifuged at 20,000 rpm in a microfuge for 5 minutes and only the upper portion of the supernatant was used to make up the injection mixture. The injection mixture itself was then centrifuged at 20,000 g for 10 minutes and the top half of the supernatant was transferred into a clean tube for immediate use.

Pulling needles and loading injection mixture

Microinjection needles were made from borosilicate glass capillaries of external diameter 1.0mm and internal diameter of 0.78mm (Harvard Scientific, GC100TF-10) using a P-97 micropipette puller (Sutter Instrument Co). DNA samples were loaded by capillary action using a manually drawn glass pipette and filled needles were used immediately after loading. Needles were placed in a needle holder with a three dimensional micromanipulator system (Narashige MMO-202N) mounted on an Axiovert S100 inverted compound microscope with Normarski optics (Zeiss). Needles were opened by gently rubbing the tip on the agarose pad used for mounting the worms whilst observing through the x10 objective and simultaneously

applying pressure to the needle from a Nitrogen cylinder using a foot switch controlled regulator (Tritech Research).

Mounting worms for injection

Worms chosen for injection were young, well-fed and healthy hermaphrodites. They were immobilised for injection on dried agarose pads, which were made by placing 30-40 μl of melted 2% agarose onto a 22mm x 64mm coverslip. A 22mm x 22mm coverslip was then gently pressed on top of the agarose to flatten it and the two coverslips separated by gentle sliding. The agarose pads were left to dry at room temperature overnight before use. The worms were immobilised on the dried agarose pads by placing them in a drop of mineral oil with the dorsoventral axis of the worm parallel to the surface of the coverslip.

Microinjection Procedure

Worms immobilised on the agarose pads were positioned with respect to the needles under low magnification (x10) so that the needle is inclined at a 10-20° angle to the distal arm of the gonad. The opened needle was gently inserted through the cuticle into the distal arm. The needle was the pressurised and the filling of the gonad directly observed. Whenever possible, both arms of the gonad were injected in each worm. Injected worms were removed from the pads immediately after injection by placing a drop of M9 buffer (appendix 2) to the surface of the mineral oil to release the worms from the pad. The worms were then transferred into a drop of M9 on freshly seeded agar plates and left to recover at 20°C for several hours. After 2-3 hours, each worm that was still alive was transferred to a separate plate. This 2-3

hour recovery period allowed the worms to lay those eggs that were already fertilised prior to injection and so would have no chance of carrying a transgene.

Identification of transformants and establishing transgenic lines

Over a period 24 to 72 hours after injection, each plate containing a separate injected hermaphrodite (referred to as P_0) was examined for F1 progeny showing the right roller phenotype conferred by the pRF4 co-injection marker. Roller F1 progeny were picked and transferred onto individual plates. Any "roller" F2 progeny identified in the next generation was then picked singly onto fresh plates. Transgenic lines were founded from single F2s and lines from separate injected P_0 individuals were considered to be independent transgenic lines and given a formal strain designation.

2.1.3 RNA mediated interference (RNAi)

RNAi can be applied to *C. elegans* by a variety of approaches including direct injection of dsRNA into the hermaphrodite gonad, feeding worms on lawns of bacteria expressing dsRNA, soaking worms in dsRNA and expressing dsRNA from a transgene (Tabara, H. *et al.* 1998; Bosher, J. M. *et al.* 2000; Fjose, A. *et al.* 2001; Timmons, L. *et al.* 2001). The RNAi experiments presented in this thesis involve the feeding of bacteria expressing dsRNA to *C. elegans* and *H. contortus* (Kamath, R. S. *et al.* 2001).

Bacterial expression of dsRNA

The vector used to apply RNAi by feeding was the plasmid L440 (kindly provided by A. Fire, Carniegie Institute) which carries two T7 promoters on either side of the multiple cloning sites in opposite orientations (Timmons, L. *et al.* 2001). The DNA fragments from which the dsRNA was expressed were cloned into the multiple cloning site of the vector and the resulting plasmid transformed into HT115 (DE3), an RNase III-deficient *E. coli* strain with IPTG-inducible expression of T7 polymerase (Kamath, R. S. *et al.* 2001). This allowed transcription of RNA from the insert in the two opposite orientations to produce double stranded RNA.

Preparation plates used to feed dsRNA expressing bacteria to worms

IPTG and ampicillin were added to standard NGM agar to a final concentration of 2 mM and 25 µg/ml respectively and this was used to pour agar plates. The HT115 bacterial cells containing the appropriate dsRNA expressing plasmid were grown overnight in LB broth containing 50 µg/ml of ampicillin to produce a bacterial suspension. Several drops of this suspension were placed onto each plate, which were then left at room temperature overnight to produce a bacterial lawn. Plates could be stored for up to 7 days at 4°C before use.

Application of RNAi to worms by bacterial feeding

In order to study the effects of knocking down gene function during embryogenesis, dsRNA was fed to hermaphrodites and the subsequent effect on F1 progeny observed. Briefly, *C. elegans* L4 worms were harvested from OP50 plates,

washed in M9 to remove any OP50 bacteria and subsequently transferred onto a plate containing a lawn of dsRNA expressing bacteria and left to feed for 8 hours. This allowed time for any remaining OP50 to disappear from the worms' digestive tract, dsRNA expressing bacteria to be ingested and dsRNA to be taken up and pass to the gonad. After 8 hours the adult worms were transferred onto fresh RNAi plates (10-20 worms per plate) and left to lay eggs for 6 hours at the appropriate temperature. After this time all adults were removed and plates were left to incubate at the selected temperature overnight. The number of hatched and unhatched F1 eggs were then counted and the phenotype of any arrested embryos or abnormal larvae examined. The plates were then left at the selected temperature for several days and the development of the F1 larvae regularly examined.

In order to study the effects of knocking down gene function during postembryonic development, eggs or L1 larvae were placed on plates with bacterial lawns expressing dsRNA and their subsequent development observed. L1 larvae were prepared by a method modified from that described by Lewis and Fleming (Lewis, J. A. *et al.* 1995) in the following manner. Worms were washed from well-grown mixed stage plates in M9 buffer and pelleted by centrifugation in a microfuge at 6000 rpm for 2 minutes. The pellet was resuspended in a 0.25M KOH, 1-15 % hypochlorite solution and incubated for approximately 10 minutes until the adult worms and larvae had disintegrated (Lewis, J. A. *et al.* 1995). The resulting egg suspension was then centrifuged at 6000 rpm for 2 minutes and the pellet washed several times in M9 to remove any trace of the hypochlorite solution. The eggs were resuspended in 5 ml of M9 solution and then transferred into a Petri dish and left at 20°C overnight in order for L1s to hatch. In the absence of food the larvae remain

46

viable at the L1-stage for up to a week in M9 at 15°C. To initiate an RNAi experiment, the appropriate number of L1 larvae were pipetted in a volume of 20μl of M9 adjacent to a lawn of bacteria expressing dsRNA.

RNAi experiments in *H. contortus*

Post-embryonic RNAi was performed on *H. contortus* larvae using the same method as described above for *C. elegans*. In addition, the size of *H. contortus* larvae was measured during RNAi experiments in the following manner. Images of larvae were obtained using a Laborulux K compound microscope (Leica) and a Panasonic F15 CCD Video Camera connected to a PC monitor. The outline of each larva to be measured was traced on transparent acetate paper placed on the monitor screen. A slide carrying a scale in μm was placed on the stage under the same magnification and traced onto the same piece of transparent acetate paper. The length of each larval trace on the acetate paper measured in cm using a curvimeter (Carpentras (Donarier)/ Sharp, 15360). The trace of the μm scale was also measured in cm, which allowed the conversion of the larval trace measurement in cm to the actual size of the larvae in μm.

Statistical analysis of RNAi results

The data obtained from the RNAi experiments for both *C. elegans* and *H. contortus* were analysed using the StatPad package.

2.1.4. Ectopic expression experiments

Ectopic expression is a technique that involves examining the effect of forced expression of a gene of interest in at a time and/or place where it is not normally expressed. It has been widely used in *C. elegans* as described in more detail in chapter 4. This assay is particularly suited to the analysis of transcription factors, where the arrested embryos can be examined for the expression of potential downstream genes. The work presented in this thesis involves the use of the *C. elegans* ectopic expression vector pPD49-78 (Fire, A. *et al.* 1990; Fire, A. 1995). This vector allows the gene of interest to be placed under the control of a heat-shock promoter. Details of the vector and the cloning procedures are discussed further in chapter 4.

Ectopic expression constructs were transformed into the appropriate *C. elegans* strain by microinjection and transgenic lines established. The general method used for the ectopic expression experiments is now described. Around 50-60 young transgenic adult worms were placed into the first well of a two-well watch glass slide containing 50 µl of M9 buffer. The worms were dissected by shearing using two gauge 3 hypodermic needles (Microlance) under low magnification. This step released the embryos from the adults. One drop of hypochlorite solution (0.25 M KOII, 1-15 % hypochlorite, freshly mixed) was then added to aid the release of the eggs by dissolving the adult carcasses; however this step must be brief to avoid the hypochlorite damaging the embryos. Several drops of sterile 5 % BSA solution were added to neutralise the hypochlorite. A fine glass pipette was produced by manual pulling of a borosilicate glass capillary of 1.5mm outside diameter and 1.17mm internal diameter (Harvard Scientific, GC150T-15) over an ethanol burner flame.

This was used to harvest 1-4 cell embryos and transfer them to the second well of the watch glass containing 50 µl of M9 buffer and 10 µl 5% Bovine Serum Albumin (BSA). The embryos were washed three times to remove the remaining hypochlorite by pipetting and removing M9 buffer from the well. After the final wash, 50 µl of M9 and 10 µl of 5 % BSA were added to the well and the slide placed at 20°C for one hour to allow the embryos to develop to further. The slide was then placed at 33.5°C for 40 minutes to heat shock the embryos and induce expression of the transgene under the control of the heat shock promoter. Embryos were examined 16-24 hours later under 60-80 X magnification on a Stemi 6 Zeiss microscope and the number of arrested and developed embryos counted. Embryos were then mounted onto a fresh 2% agarose pad for examination by Normarski optics or under UV illumination for the expression of GFP markers at 600 X magnification. Alternatively, the embryos were fixed on slides for immunofluorescence studies.

2.2 Molecular biology techniques

2.2.1 Polymerase chain reaction (PCR)

Polymerase chain reaction (PCR) was performed using standard techniques (Sambrook, J. *et al.* 1989). Unless otherwise stated, the components of the final PCR reaction mixtures were: DNA template; primers (at 0.5 µM each); dNTPs (dATP, dCTP, dGTP, dTTP at 1mM each); Tris HCl pH 8.8 (45 mM), Ammonium sulphate

(1mM); MgCl$_2$ (4.5mM); 2-mercaptoethanol (6.7 mM); EDTA pH 8.0 (4.4 µM); BSA (113 µg/ml) and Taq polymerase (0.05U/µl). Unless otherwise stated, the standard thermocycling conditions for PCR used in this work were: [94°C, 3mins] for 1 cycle, [94°C, 1min; A°C, 30secs; 72°C, 1min/kb] for n cycles, [72°C, 10 mins] for 1 cycle. Where A is the annealing temperature chosen for each primer pair and n equals the number of cycles which was generally between 20 and 35 cycles depending on the particular experiment.

Where an amplified fragment was to be used in expression or functional studies, a proofreading thermostable DNA polymerase were used for PCR amplifications in order to minimise the chance of incorporating polymerase induced mutations. Herculase (Stratagene, USA) or Proofstart (Qiagen) was used for this purpose. For long range PCR, i.e. PCR aimed at amplifying DNA fragments of >3Kb, a mixture of Taq polymerase and the proof-reading enzyme Proofstart was used as described in the Proofstart manual (Qiagen). Proofstart was used with Taq polymerase at a ratio of 1:5 and the reaction mixture was made up as directed in the manual including the use of the "Q-solution". The thermocycling conditions were as described for the standard PCR protocol except that the extension was performed at 68°C for a time of 2 min/kb (rather than 72°C for 1 min/kb) and the denaturation time in each cycle was for 10 seconds.

2.2.2 Restriction digestion of DNA

Restriction digestions of plasmid, lambda or genomic DNA were performed using standard methods (Sambrook, J. *et al.* 1989). Incubations were performed for 2 to 8 hours using 1-10 units of restriction enzyme for each μg of DNA digested. Enzymes and their buffers were obtained from Invitrogen, QBio and New-England Biolabs.

2.2.3 Agarose gel electrophoresis

0.8-2 % (w/v) agarose gels were prepared by dissolving LE analytical grade agarose (Promega) in 1X TAE and adding ethidium bromide to a final concentration of 0.5 μg/ml. Gels were cast and run in 1X TAE using the Anachem horizontal agarose gel system. DNA samples and markers 1 kb ladder (Invitrogen) or *Hind*III digested λ DNA (Stratagene) were diluted in loading buffer and gels were electrophoresed at 60-100V until appropriate separation was achieved. DNA was visualised by examination of gels on a UV transilluminator (205-260 nm wavelength).

2.2.4 Purification of DNA from agarose gels

Prior to cloning, DNA fragments from PCR amplifications and restriction digests were purified from agarose gels using the QIAquick QXII gel extraction kit (Qiagen). DNA was purified from 1% TAE agarose gels using the manufacturers'

protocol and eluted from the solid matrix in 30 µl EB buffer (10 mM Tris-HCl pH 8.5).

2.2.5 DNA Cloning procedures

<u>TA cloning</u>

Routine cloning of PCR products was performed using a TA vector system (pCR2.1, Invitrogen.). This vector is linearised at the polylinker with overhanging thymidine residues that act as complementary ends for PCR products that have a single adenosine residue added by Taq polymerase during PCR. The ligations and transformations were performed as described in the manufacturers' protocols. Generally 50ng of vector and a molar ratio of 3:1 insert:vector were used. Ligations were transformed into One Shot INVαF'™ cells (Invitrogen) (genotypes are given in 2.2.17) using the manufacturers protocol. Briefly, 20 µl aliquots of cells were thawed on ice and 1µl of ligation reaction added. The cells were then incubated for 30 minutes on ice, followed by a heat-shock of 42°C for 45 seconds. The cells were then placed back on ice for 2 minutes before adding 200 µl of SOC medium and incubating for 1 hour at 37°C with constant agitation (225 rpm) on a horizontal shaker. Each batch of transformed cells were plated out on LB plates containing 100 µg/ml of ampicillin and coated with 50 µl of 20 mg/ml of 5-bromo-4-chloro-3-

indolyl-2-galactopyranoside (X-gal) using standard methods (Sambrook, J. *et al.* 1989). Plates were subsequently incubated overnight at 37°C.

General cloning procedures

The general strategy for cloning, unless otherwise stated, was for primers used for PCR amplifications to be designed with 5' tags containing appropriate restriction sites. Both recipient vector and insert fragment were then digested with one or two restriction enzymes and subsequently gel purified (as described above) prior to ligation. Generally, 10-30 ng of linearised vector in a final ligation volume of 10 μl with an insert:vector ratio of 3:1 was used. Ligation reactions were set up using T4 DNA ligase/5X buffer (New England Biolabs) following manufacturers instructions. The ligation reactions were incubated overnight at 15°C. 50-100 μl of XL-1 Blue or XL-10 Gold (Stratagene) supercompetent bacterial cells (genotypes are given in 2.2.17) were transformed with 1 μl of ligation reaction using standard procedures (Sambrook, J. *et al.* 1989).

2.2.6 Preparation of plasmid miniprep DNA

This protocol is based on the alkaline lysis method for DNA purification and was used for the routine plasmid preparations to screen transformation plates for plasmid clones containing the correct insert (Birnboim, H. C. *et al.* 1979). Single colonies were picked from transformation plates and grown overnight in 3 ml of LB broth containing the appropriate antibiotic(s), typically ampicillin (100 μg/ml). An

aliquot of the overnight culture was transferred to a 1.5 ml microfuge tube and the cells centrifuged at 13,000 rpm in a microfuge for 10 minutes and the supernatant discarded. The cells were resuspended in 150 µl of resuspension solution (see appendix 2). The cells were then lysed by addition of 150 µl of lysis solution (see appendix 2) and the tube was inverted several times to ensure complete lysis. 150 µl of neutralisation solution (see appendix 2) was then added and the tube inverted several times again. The lysed cells were centrifuged at 13,000 rpm in a microfuge for 20 minutes in order to pellet the insoluble cellular debris and the supernatant was transferred to a fresh microfuge tube and 1 ml of ethanol was added. The sample was mixed well, incubated at -20°C for 30 minutes to precipitate the DNA and centrifuged at 13,000 rpm in a microfuge for 30 minutes. The supernatant was discarded and the pellet washed in 500 µl 70 % ethanol. The sample was centrifuged again at 13,000 g for 10 minutes, the supernatant removed and the DNA pellet was air dried and resuspended in 30 µl dH$_2$O. The concentration of plasmid solution was calculated by comparison of 1 µl of the preparation on a 1% agarose gel alongside a known concentration of 1 kb DNA ladder marker (Invitrogen) and also by measuring the absorption of the solution in a spectrophotometer at 260 and 280 nm wavelengths.

2.2.7 Freezing glycerol stocks of bacterial cultures

Glycerol stocks were prepared by adding 0.5 ml of sterile freezing solution (see appendix 2) to 0.5 ml of an overnight bacterial culture and freezing at -70°C.

2.2.8 ABI sequencing

Double stranded sequencing reactions were carried out by a method modified from that of Sanger which uses fluorescently labelled ddNTPs chain terminators (Sanger, F. *et al.* 1977). Automated sequencing was performed on an ABI Prism 3100 DNA sequencer (Applied Biosystems). The reagents and protocols were as described in the Big Dye Terminator Sequencing kit version 3.0 (Applied Biosystems). Each sample contained 2 µl of Ready Mix®, 500 ng of template DNA and 0.32 µM of appropriate oligonucleotide primers (made up to a final volume of 10 µl with dH_2O). Cycle sequencing was performed in a 9600 ABI thermocycler using the standard conditions outlined in the Big Dye Terminator Sequencing kit version 3.0 protocol (Applied Biosystems).

On completion of thermocycling, unincorporated nucleotides were removed from samples using DTR Gel Filtration Cartridges (Edge Biosystems, 42453) and freeze-dried. Samples could be kept at -20°C in a dry form. Immediately prior to loading, samples were resuspended in Hi-Di formamide (provided with the Applied Biosystems sequencing kit) and loaded on a 96-well microtiter plate (Applied Biosytems). The plates were subsequently loaded into the ABI Prism 3100 DNA sequencer (Applied Biosystems). Sequence chromatograms were visualised using Chromas Version 2.23 software (Technelysium Pty, Ltd) and analysed using Vector NTI software version 8 (InfoMax).

2.2.9 Southern blotting

Southern blotting was performed using standard procedures (Sambrook, J. *et al.* 1989). Briefly, between 5 and 10 μg of genomic or λDNA were digested with the appropriate restriction enzyme, ethanol precipitated, washed and resuspended in dH_2O. The digested DNA was then separated by agarose gel electrophoresis using a 0.8% agarose gel. Following electrophoresis, the gel was immersed in 250mM HCl for 10 minutes with agitation at room temperature. It was then rinsed in dH_2O and immersed in denaturation solution (appendix 2) for 2 x 15 minutes, rinsed in dH_2O and immersed in neutralisation solution (appendix 2) for 2 x 15 minutes. DNA was then transferred to a Hybond N nylon membrane (Amersham Life Science) by the standard method using 20 X SSC (Southern, E. M. 1975). DNA was immobilised and cross-linked onto the nylon membrane by exposure to 120 mJ of UV radiation.

2.2.10 Random (High Prime) digoxigenin labelling and purification of DNA probes

DNA probes were labelled with digoxigenin for hybridisation experiments. The method described in the DIG High Prime labelling and detection starter kit II manual (Roche) was followed. Briefly, purified double stranded DNA (300 ng) was denatured by boiling in water for 10 minutes and then chilled on ice for 5 minutes. 5X DIG High Prime mix (containing random oligonucleotides, 1 U/μl Klenow polymerase, 1 mM dATP, 1 mM dCTP, 1mM dGTP, 1 mM dTTP, 0.35 mM alkali-

labile-digoxigenin-11-UTP) was added to a final concentration of 1X and the sample was incubated overnight at 37°C. The reaction was terminated with the addition of 2 µl 0.2 M EDTA, pH 8. Before addition to the hybridisation solution, the probe was denatured in boiling water for 10 minutes and chilled on ice for 5 minutes. Hybridisation solution containing the DNA probe could be stored at -20°C after use and reused several times after denaturation at 68°C for 10 minutes. Probe activity was assayed before use by probing a dot blot of different amounts (100, 10, 1, 0.1, 0.01 and 0.001 ng) of plasmid DNA containing an insert corresponding to the DNA sequence of the probe.

2.2.11 cDNA library screening

Plating of library and determination of titre.

The library used to isolate a full-length *Hcelt-2* cDNA clone was a λZap II cDNA library constructed from mRNA isolated from adult *H. contortus* gut cells (kindly supplied by Dr D Jasmer). The phage were plated out and the library screening was performed using standard techniques (Sambrook, J. *et al.* 1989). Host cells were prepared by growing XL-1 Blue cells (Stratagene, USA) overnight in LB broth supplemented with 10 mM $MgSO_4$ and 0.2% maltose. The cells were centrifuged at 6,500 rpm for 10 minutes and resuspended in 10 mM MgSO4 to achieve a final OD_{600} of 1. The cells could be stored for up to three weeks at 4°C.

In order to determine the titre of the library, a serial dilution was made from a 1×10^{-1} to 1×10^{-9} dilution in SM buffer (see appendix 2). 1 µl of each dilution was

added to 100 μl aliquots of cells which were incubated at 37°C for 20 minutes before

7 ml of pre-warmed "top" agarose (50°C), supplemented with 10 mM $MgSO_4$, were

added to the cells and the resulting suspension poured onto 90 mm LB agar plates

(appendix 2). The top agarose was allowed to solidify and the plates were inverted

and incubated at 37°C overnight. The number of plaques was counted on those plates

on which discrete plaques were clearly visible and the number of plaque forming

units (pfu) in the original library was calculated.

Library screening

In order to screen the library, 1μl of the appropriate dilution of library was

added to 100 μl of plating bacterial suspension and this was plated out onto 130 mm^2

plates by same method as described above but using 8 ml of top agarose. The plates

were incubated at 37°C for 8-10 hours and then at 4°C overnight prior to performing

plaque lifts. A total of 100,000 pfu of the adult gut-specific *H. contortus* λZAPII

cDNA library were screened (five 130 mm^2 plates, each containing 20,000 pfu of the

library). Duplicate plaque lifts were performed by placing Hybond-N filters

(Amersham Life Science) onto the pates and orientating them using needle stabs

through the filter. The first filter was left on the plate for 1 minute and the second

filter for 2 minutes. After removing from the plates, the filters were placed on two

layers of Whatman filter paper soaked in denaturation solution (appendix 2) for 5

minutes, then on Whatman filter paper soaked in neutralisation solution I (appendix

2) for 15 minutes and finally on Whatman filter paper soaked in 2X SSC (appendix

2) for 10 minutes. The DNA was fixed onto the filters by exposing them to 120 mJ of UV radiation in a UV cross-linker.

A 651 bp PCR fragment (PrHcelt2 5'; chapter 3) corresponding to the 5' end of the *Hc-elt-2* cDNA sequence was labelled with digoxigenin as previously described. Following denaturation, the probe was incubated with the filters at 42°C in hybridisation solution overnight (appendix 2). The filters were subsequently washed twice for 15 minutes in 2X SSC/0.1% SDS at 42°C and then twice in 1X SSC/0.1% SDS at 42°C for 30 minutes. The hybridised probe was detected by an alkaline phosphatase conjugated anti-digoxigenin antibody, which produces chemiluminescence in presence of a substrate, as described by the manufacturer (DIG High Prime labelling and detection starter kit II, Roche). After adding the chemiluminescent substrate to the filters, autoradiography was performed. Autoradiographs from the first and duplicate lifts were superimposed to align potential positive plaques and only plaques present on both filters were deemed positive. Each positive plaque was cored from the LB agar plate and placed into 1ml of SM buffer (see appendix 2) containing several drops of chloroform. This was then left overnight at 4°C to allow diffusion of the phage into the buffer.

A secondary and tertiary screen were conducted in a similar fashion on putative positive phage by plating out 5×10^2 pfu on 90 mm plates. At each screen, a single well-isolated positive plaque was cored out and transferred into SM buffer/chloroform to ensure purity of the plaque. A final plating and screening was performed on the isolated clone to ensure that 100% of plaques hybridised to the probe confirming the clonal nature of the phage isolate.

<u>*In vivo* excision of the pBluescript phagemid from the λZapII vector</u>

The λZAPII vector used for the construction of the cDNA library allows the excision of the cloned insert to form a circular phagemid to be used for further manipulations. The protocol used was essentially that from the λZAPII vector kit (Stratagene, USA). Phage in SM buffer (appendix 2) were used to inoculate a 200 µl culture of XL-1 Blue cells (Stratagene, USA, genotype given in section 2.2.17) at an OD_{600} of 0.8. > 1 x 10^6 pfu ExAssist helper phage were added to this and grown in LB broth (appendix 2) supplemented with 0.2 % maltose and 10 mM MgSO4 for 3 hours at 37°C. The culture was heated at 70°C for 20 minutes and then centrifuged at 1000g for 15 minutes. The supernatant, which contained the excised phagemid, was then transformed into freshly grown SOLR cells (Stratagene, USA, genotypes given in section 2.2.17) and plated on LB-ampicillin plates (appendix 2).

2.2.12 Genomic library screening

<u>Plating and screening the *H. contortus* genomic DNA library</u>

XL1-Blue MRA cells (Stratagene, USA) were grown and prepared in the same manner as XL1-Blue cells as described for cDNA library screening. Two *H. contortus* genomic libraries were screened; a λEMBL3 library (kindly supplied by Dr M. Roos, Lelystadt) and a λDASHII library (kindly provided by Dr Robin Beech, Montreal). The number of plaques to be screened in a genomic library depends on the average size of inserts in the library and on the size of the genome being

screened. The minimum number N of plaques to be screened to have a probability P of obtaining a positive clone can be estimated by the formula:

$$N= \ln (1-P)/ \ln (1-f),$$

where f is the ratio of insert size:genome size (Sambrook, J. *et al.* 1989). In this case the *H. contortus* genome is estimated to be around 1×10^8 bp and the average size of an insert in a λEMBL3 library is $10\text{-}15 \times 10^3$ bp. Hence for a value of P=0.99, N is between 30,698 (if inserts are of 15 kb) and 46,049 (if inserts are of 10 kb). Consequently, in order to maximise the probability of isolating one or more positive clones 100,000 plaques were screened. The library was plated out as described for the cDNA library, except that XL1-Blue MRA cells were used instead of XL1-Blue. Hybridisation with probe PrHcelt25' and phage isolation was performed as described for the cDNA library screening.

Isolation and preparation of phage λ DNA

λ phage DNA was isolated using the QIAgen Lambda DNA Midi kit following the manufacturers protocol. Briefly, the phage clone of interest was used to infect a mid-log phase liquid culture of the appropriate phage competent cells and the culture was grown at 37°C until lysis was visible (white strands observed floating in clarified culture medium). Chloroform was added to the culture to a final concentration of 2 % (v/ v) to enhance lysis and the culture incubated for a further 15 minutes at 37°C. The culture was centrifuged at 10,000 g for 10 minutes to remove bacterial debris and the supernatant was retained. The first buffer included with the kit (containing Rnase and DNaseI) was then added to the supernatant and incubated at 37°C for 30 minutes to eliminate bacterial DNA and RNA. The ice-cold second

buffer, containing PEG to facilitate precipitation of the phage particles, was added and the mixture was incubated on ice for 60 minutes and centrifuged at 10,000 g for 10 minutes. The supernatant was discarded and the pellet containing the phage particles was resuspended in the third buffer provided with the kit. After complete resuspension of the pellet a fourth buffer, provoking lysis of the phage particles, was added and the mixture was incubated at 70°C for 20 minutes and then cooled on ice. The fifth buffer, causing precipitation of the phage protein, was then added and gently mixed, and the suspension was then centrifuged at 15,000 g for 30 minutes at 4°C. The supernatant was transferred to a fresh tube and centrifuged again at 15,000 g for 10 minutes at 4°C to ensure a particle-free lysate. The lysate was applied to a previously equilibrated QIAgen-tip and was allowed to enter the resin by gravity flow. The QIAgen-tip was washed once with the washing buffer provided and the DNA was eluted in a fresh tube with the elution buffer provided. The DNA was precipitated by addition isopropanol at room temperature and centrifugation at 15, 000g for 30 minutes at 4°C. The supernatant was carefully discarded and the DNA pellet was washed with room-temperature 70% ethanol and centrifuged at 15,000 g for 10 minutes at 4°C. The pellet was allowed to dry and was resuspended in an appropriate volume of sterile dH_2O overnight at room-temperature to ensure complete resuspension of the high molecular weight DNA. The concentration of DNA was then measured with a spectrophotometer at 260nm, and the purity of the DNA was assessed by the OD260 nm/ OD280 nm ratio.

2.3. Protein and Immunological techniques

2.3.1 SDS-PAGE

SDS-Polyacrylamide gels were poured using standard methods (Sambrook, J. *et al.* 1989) with the Biorad minigel system; 12-15 % gels with a stacking gel was routinely used. Electrophoresis was carried out in Tris-glycine running buffer (see appendix 2) at 400mA for 40 minutes or until the loading dye front had reached the bottom of the gel. After electrophoresis, gels were either stained with Coomassie Brilliant Blue or used for Western blot analysis.

2.3.2 Western blot analysis

Western blot analyses were carried using standard methodology (Sambrook, J. *et al.* 1989) using the Biorad miniblot system. The Supersignal West Pico Chemiluminescent Substrate ®, (Pierce, Perbio, USA) was used for signal detection. Transfer onto Hybond-C nitrocellulose membrane (Amersham Life Sciences) was carried out at 100 mA for one hour in Transfer buffer (see appendix 2). After transfer, the nitrocellulose membrane was stained with Ponceau-S to visualise the proteins, subsequently rinsed in dH_2O and blocked for one hour or overnight at room temperature in 10 % milk in TBST (see appendix 2). The immunological detection was then carried out using incubation times and temperatures depending on the particular experiment as indicated elsewhere. Incubation with substrate for

chemiluminescent detection was as described in Pierce Perbio Supersignal West Pico Chemiluminescent Substrate ® kit handbook (Walker, G. R. *et al.* 1995). The membrane was then exposed to Kodak film for 10 seconds to 15 minutes and the film developed using standard procedures.

2.3.3.Production of parasite material for immunofluorescence experiments

Faeces from animals experimentally infected with *H. contortus* (Moredun Susceptible strain) were provided by Dr F. Jackson, Moredun Institute, Edinburgh. The faeces were stored and transported anaerobically to prevent egg development. Eggs voided in fresh faeces are generally at the 10-20 cell stage of embryogenesis and can be maintained without further development for up to 1 week at 15°C in anaerobic conditions. Anaerobic storage simply involves storing faeces in an air-tight container, which has be filled with faeces and topped up with water to exclude air. Eggs were extracted from fresh faeces by floatation. The faecal matter was diluted in approximately 2L of water and poured through a series of wet sieves of decreasing porosity (1 mm, 355 μm, 210 μm, 63 μm) and in each case the filtrate was retained. The final filtrate was collected into a bowl and poured carefully into a 38 μm sieve. The eggs retained on this sieve were thoroughly washed with water before rinsing into a small beaker. The solution was then used to half fill 15 ml test tubes which were then topped up with water and centrifuged at 2000 rpm for 10 minutes. The supernatant was discarded and the pellet was dislodged by vortexing. 0.05 g kaolin

powder was added and the pellet was resuspended in a saturated salt solution (appendix 2) so that only the very top of the tube was left free to aid mixing. The tube was inverted 4-6 times to mix and centrifuged again at 2000 rpm for 10 minutes. This procedure causes the eggs to float to the top of the saturated salt solution. The upper 1ml of solution was pipetted onto a 38 µm sieve. The eggs were rinsed thoroughly with water and then washed into a 60 mm Petri dish in 5ml of M9 solution. The eggs were then left to incubate at 15°C or 20°C for various lengths of time to allow the embryos to develop further.

2.3.4 Production of polyclonal antisera against Hc-ELT-2/His tag fusion proteins

Cloning in of *Hc-elt-2* fragments into the pQE30 vector

In order to produce bacterially expressed fusion proteins corresponding to different regions of the *H. contortus* ELT-2 polypeptide, two non-overlapping fragments of *Hc-elt-2* were cloned into the expression vector pQE30 (QIAexpressionist kit®, Qiagen). This vector enables the expression of a fusion protein that carries a 6x Histidine tag at the N-terminus. The pQE vectors are based on the T5 promoter transcription-translation system. They belong to the pDS family of plasmids (Bujard, H. *et al.* 1987), were derived from plasmids pDS56/RBSII and pDS781/RBSII-DHFRS (Stuber, D. *et al.* 1990) and are low copy plasmids. The optimised promoter-operator consists of phage T5 promoter and two *lac* operator sequences, which increase *lac* repressor binding and ensure efficient repression of

the powerful T5 promoter. There is a synthetic ribosomal binding site RBSII for high translation rates and a 6x HIS tag coding sequence 5' of the cloning region (for pQE30). There are multiple-cloning sites and translational stop codons in all reading frames for convenient preparation of expression constructs. The vector also has two strong transcriptional terminators: t_0 from phage lambda and T1 from the *rrnB* operon of the *E. coli* to prevent read-through transcription and ensure stability of the expression construct. A β-lactamase gene (*bla*) conferring ampicillin resistance at 100 μg/ml (Sutcliffe, J. G. 1979) has been included and finally there is a ColEI origin of replication (Sutcliffe, J. G. 1979). Expression of ORFs cloned into the pQE30 plasmid needs to be regulated so that expression can be induced when desired. The extremely high transcription rate from the T5 promoter can only be efficiently regulated and repressed by the presence of high levels of the *lac* repressor protein. The *E. coli* host strains (M15[pREP4]) used in the QIAexpress system contain the low-copy plasmid pREP4, which confers kanamycin resistance and constitutively expresses the *lac* repressor protein encoded by the *lacI* gene. Multiple copies of the pREP4 plasmid are present in the host strains and ensure the production of high levels of the *lac* repressor protein, which binds to the operator sequences and represses recombinant protein expression. The pREP4 plasmid is compatible with all plasmids carrying the ColE1 origin of replication and is maintained in *E. coli* in the presence of kanamycin at a concentration of 25 μg/ml. Expression of recombinant proteins encoded by pQE 30 is rapidly induced by the addition of IPTG, which binds to the *lac* repressor protein and inactivates it. Once the *lac* repressor is inactivated, the host cell's RNA polymerase can transcribe the sequences downstream from the promoter.

Two *Hc-elt-2* cDNA fragments were cloned into the PQE30 vector in order to produce recombinant fusion proteins. A PCR product, from nucleotide 167 to 819, was amplified from *Hc-elt-2* cDNA template (plasmid pBC12) with the primers Ann1 and Ann2 (appendix 2). This fragment was cloned into the *Hind*III and *Sst*I resctriction sites of the PQE polylinker by virtue of 5' tags containing these restriction sites on the Ann1 and Ann 2 primers respectively. The resulting construct was named pAC1 and was used to produce fusion protein HCELT2HISA. A second PCR fragment, from nucleotide 909 to nucleotide 1344 on the cDNA sequence, was amplified from *Hc-elt-2* cDNA template and cloned into the *Sph*I and *Hind*III restriction sites of the PQE polylinker by virtue of 5' tags containing these restriction sites primers respectively. The resulting plasmid was called pBC7 and was used to produce fusion protein HCELT2HISB.

Expression of Hc-ELT-2 6x His tag fusion proteins

Both recombinant proteins were purified and dialysed after expression in M15 cells (for genotype see section 2.4) following the procedure described in the QIAexpressionist handbook and outlined below.

The first stage was to optimise the induction procedure to maximise the yields of fusion protein. This involved induction of cultures by different concentrations of IPTG. The constructs pAC1 and pBC7 were transformed into the M15[pREP4] strain. Overnight cultures of each were grown in 25 ml of LB medium with kanamycin (12.5 μg/ml) and ampicillin (50 μg/ml) at 37°C with shaking. In order to produce cultures for induction of fusion protein expression, 100 ml of pre-

warmed (37°C) LB medium containing the same antibiotic concentrations were inoculated with 15 ml of overnight culture and incubated at 37°C with shaking until the OD_{600} reached 0.5-0.7. Expression of fusion protein was induced by adding IPTG to a final concentration of between 0.1 and 1mM. Prior to induction 1ml aliquots (pre-induced sample) were sampled from each flask. Induction was defined as t =0 min. The culture was then grown for an additional 5 hours at 37°C with shaking and duplicate 0.5 ml aliquots were taken at t =60 min, t =120 min, t =180 min and t =240 min and t =300. All samples taken from the cultures were centrifuged at 14,000rpm in a microfuge for 10 minutes and the bacterial pellet was resuspended in 50 μl of SDS-PAGE loading buffer and stored at -20°C until use. The samples were then examined by SDS-PAGE as described in section 2.2.14 to determine the optimum IPTG concentration and time of culture harvest to give maximal fusion protein yield. From these experiments an IPTG concentration of 1mM and a harvest time of 5 hours post-induction was determined as optimal for both fusion proteins.

The cultures were then repeated on a larger scale to prepare large amounts of fusion protein by using 100 ml of overnight culture to inoculate 1000 ml of pre-warmed culture media. Cells were harvested five hours post induction by centrifugation at 3,000 g for 20 minutes at 4°C. The fusion protein was then solubilised under denaturing conditions as follows. The bacterial pellet was transferred to a clean falcon tube, weighed and frozen at -70°C for 20 minutes. It was then thawed on ice for 15 minutes and resuspended in lysis buffer B (see appendix 2) at 5 ml/g of pellet. This was incubated at room temperature on a rocker for 60 minutes or until the solution was translucent, indicating the cell lysis was complete.

The samples were then centrifuged at 9,500 rpm for 30 minutes and the supernatant was transferred into a clean tube. A 10 μl aliquot was taken and mixed with 5μl SDS-PAGE loading buffer and stored at -20°C. The remaining supernatant was stored at -70°C. Successful expression and solubilisation was tested by analysis of pre and post-induction aliquots by SDS-PAGE.

Purification of expressed recombinant proteins under denaturing conditions

The system of purification involves binding the fusion proteins to a Ni-NTA column by virtue of their 6X His tags and, following washing, eluting them from the column in an a low pH elution buffer. The protocol outlined in the QIAexpressionist handbook was followed. Briefly, 4ml of the clear lysate, obtained as described above, was added to 1ml of 50 % Ni-NTA slurry and mixed gently on a rotary shaker at room temperature (200rpm) for 15-60 minutes. The resin-lysate mixture was then carefully packed into an empty 2ml plastic column using standard procedures. After leaving the column to settle for at least 1 hour, the column was washed with 2 x 4 ml of buffer C (appendix 2) and the wash fractions were collected for subsequent SDS-PAGE analysis. The recombinant protein was then eluted with 4 x 0.5 ml of buffer D (appendix 2), followed by 4 x 0.5 ml of buffer E and E$_1$ (appendix 2). All fractions were collected in separate tubes and used for SDS-PAGE analysis and stored at -20°C. SDS-PAGE analysis was used to determine which fractions contained maximal purity and yield of fusion proteins. Those fractions were then pooled together for use in immunisations.

<u>Generation of rabbit polyclonal antisera against the HCELT2HISA and HCELT2HISB fusion proteins</u>

The concentration of urea (8M) present in the eluted fusion proteins was too high for immunisation of rabbits and so the urea concentration was reduced to 2M by dialysis. Fusion protein was sent to Diagnostics Scotland for the immunisation of rabbits. Two rabbits were immunised with each fusion protein (see chapter 3 for details).

2.3.5 Immunofluorescent assay of parasite specimens

<u>Embryonic stages</u>

After extraction from fresh faeces, *H. contortus* eggs were incubated for a variable period of time from 4 to 12 hours at 15°C or 20°C in order to obtain the desired stage of development. Embryos were then fixed and prepared for immunofluorescence according to the method described for *C. elegans* embryos by Miller and Shakes (Miller, D. M. *et al.* 1995). Clean egg suspension was centrifuged at 5,000g for 2 minutes and resuspended in 50-100 µl M9 buffer. A small volume (5 µl) was pipetted onto a poly-L-lysine coated slide and a coverslip was applied. The slide was then placed onto a metal block packed in dry ice and left for 30 minutes to freeze. After this time the coverslip was flipped off with a scalpel blade to freeze-crack the embryos. The embryos were then fixed in 100% methanol for 5 minutes followed by 100% acetone for 5 minutes; they were then rehydrated by incubation in a series of solutions of decreasing percentage of methanol in dH_2O: 95%, 70%, 50% and 30%. After rehydration, the slides were placed in PBS-0.1 %Triton-X-100 overnight at 4°C as a pre-blocking step. The slides were subsequently dried with

tissue paper, carefully avoiding the area where the embryos are fixed, and 40 µl of blocking solution (PBS0.1% Triton-X-100, 10% normal donkey serum (Jackson Immunoresearch, USA) were applied to the embryos. The slides were incubated at room temperature in a humid chamber for 30 minutes. After blocking, the slides were washed three times for 10 minutes in PBS 0.1% Triton-X-100 and then dried with a tissue avoiding the sample area. The primary antibody was applied to the embryos (30-50 µl of whole rabbit anti-serum diluted 1:500 in PBS 0.1% Triton-X-100) and then incubated at room temperature in a humid chamber for 1½ hours (the optimum incubation conditions for each antibody was previously determined empirically). Following primary antibody incubation, the slides were washed three times for 10 minutes in PBS 0.1% Triton-X-100 at room temperature and carefully dried with a tissue avoiding the sample area. A Cy2 conjugated donkey anti-rabbit IgG secondary antibody (Jackson Immunoresearch, USA) was applied in a volume of 50ul at a dilution of 1:200 in PBS 0.1% Triton-X-100. The slides were incubated with the secondary antibody for 1 hour at room temperature in a humid chamber after which time they were washed twice for 10 minutes in PBS/0.1% Triton-X-100 at room temperature and once for 10 minutes in PBS 0.1% Triton-X-100, 100ng/ml DAPI to stain the nuclei. The slides were then dried and mounted for microscopic examination using DABCO (see appendix 2).

Larval and adult stages

Larval and adult stages were fixed, permeabilised and stained according to the procedure describe by Finney (Finney, M. *et al.* 1990). L1 and L2 stages were harvested from NGM plates by washing with 1ml of M9 buffer (appendix 2). The

larvae were collected into a 1.5ml Eppendorf tube and spun at 7,000 rpm for 2 minutes in a micro-centrifuge. They were washed three times in M9 buffer, resuspended in 100-150µl M9 and chilled on ice. Ice-cold 2x Ruvkun buffer (appendix 2) and 40% formaldehyde were added to a final concentration of 1x and 2% respectively and, after vigorous mixing, the tubes were placed on dry ice/ethanol. The frozen worm pellet was freeze-thawed three times and subsequently incubated on ice with occasional agitation for 1 hour. The worms were then washed twice in Tris-Triton buffer (appendix 2). After the final wash they were resuspended in Tris-Triton buffer, 1% β-mercaptoethanol and were incubated with mild agitation at 37°C overnight to reduce the disulfide linkages and permeabilise the highly cross-linked nematode cuticle. To complete the reduction reaction, the worms were washed once in 10 volumes of 1X BO$_3$ buffer (appendix 2), 0.01 % Triton and resuspended in 1X BO$_3$ buffer, 0.3 % H$_2$O$_2$, 0.01% Triton and incubated for 15 minutes at room temperature with agitation. The worms were then briefly washed again in 1X BO3 buffer, 0.01 % Triton and once for 15 minutes in AbB buffer (appendix 2). The worms were then stored in AbA (appendix 2) buffer at 4°C or used for the subsequent staining steps.

Worms were stained by incubating 25 µl of fixed worms in 1:500 dilution of the primary antibody in AbA buffer overnight at 4°C with gentle agitation. The worms were then washed for 3 hours on a rocker at room temperature in several changes of AbB buffer and then rinsed once in AbA buffer prior incubation in secondary antibody. Secondary antibody (as described above in immunoassay of embryonic stages) incubation was at a 1:200 dilution in AbA overnight at 4°C with

gentle agitation. After a final wash for 3 hours in AbB at room temperature, the worms were mounted in DABCO- 100ng/ml DAPI for observation.

2.4 Genotypes of bacterial strains used

SOLR strain[a]   e14⁻(McrA⁻)Δ(*mcr*CB-*hsd*SMR-mrr)171*sbcCrecBrec*j*uvr*C

*umu*C::Tn(Kan*r*)*lacgyr*A96*rel*A1thi-1*end*A1λ^R[F'proAB

*lac*1^{q}ZΔM15]su⁻(nonsuppressing)

XL-1 Blue   *rec*A1*end*A1*gyr*A96thi-

1*hsd*R17*sup*E44*rel*A1*lac*[F'*proAB* *lacI*^{q}

ZΔM15Tn*10*(Tet^r)]

INVαF'   *end*A1*rec*A1*hsd*R17(rk⁻mk⁺)*sup*E44*thi*-1*gyr*A96*rel*A1

Φ80*lac*ZΔM15Δ(*lac*ZYA-argF)U169λ⁻

XL-1 Blue MRA   D(*mcrA*)183Δ(*mcr*CB-*hsd*SMR-*mrr*)173*end*A1*sup*E44thi-1

*gyr*A96*rel*A1*lac*^{c}

XL-10 Gold   Tet^{r}D*(mcrA)*183D*(mcrCB-hsd*SMR-*mrr)*173*end*A1*sup*E44*thi*-

1*rec*A1

*gyr*A96*rel*A1*lac*Hte[F'*proABlacI*^{q}ZDM15Tn10(Tet^r)AmyCa

m^r]^a

Chapter 3: Isolation and characterisation of GATA transcription factor homologues from parasitic nematodes

3.1 Introduction

C. elegans is being increasingly used as a model for parasitic nematodes (Blaxter, M. L. 1998; Burglin, T. R. *et al.* 1998; Gasser, R. B. *et al.* 2000; Geary, T. G. *et al.* 2001; Hashmi, S. *et al.* 2001; Brooks, D. R. *et al.* 2002). Whilst *C. elegans* has tremendous advantages as an experimental system, the extent to which different aspects of biology are conserved with different nematode species is still poorly understood. Current comparative studies between *C. elegans*, and both free-living and parasitic nematodes have been reviewed in chapter 1. Transcription factors are central to regulatory mechanisms and often have essential roles in development. They are also increasingly being viewed as potential drug targets (Orlowski, R. Z. *et al.* 2002; Akiyama, Y. *et al.* 2003; Kim, R. *et al.* 2003). Consequently they are an important class of molecules for which to undertake comparative analysis between *C. elegans* and parasitic nematodes.

GATA transcription factors have been identified and isolated in many organisms, including yeast, plants, nematodes, flies and vertebrates (Lowry, J. A. *et al.* 2000; Patient, R. K. *et al.* 2002). The vertebrate family is the most intensively studied and there are six family members (GATA-1 to GATA-6), all of which have essential roles during vertebrate development. The vertebrate GATA factors have central, distinct and non-redundant roles in differentiation and tissue-specific gene expression in a large variety of vertebrate tissues and this has been reviewed in chapter 1(Laverierre, A. C. *et al.* 1994; Morrisey, E. E. *et al.* 1997; Charron, F. *et al.*

1999; Lowry, J. A. *et al.* 2000; Molkentin, J. D. 2000; Shimizu, R. *et al.* 2001; Patient, R. K. *et al.* 2002).

All six vertebrate GATA factors are defined by the presence of two characteristic zinc finger domains ($CX_2CX_{17}CX_2C$). The C-terminal finger and adjacent basic domain form the DNA binding domain and the N-terminal finger is involved in both DNA binding and protein-protein interactions (Figure 1.3).

In *C. elegans*, there are eleven GATA factors involved in the regulation of development of various tissues. They can be placed into two major groups based on the tissues in which they function: *elt-1, elt-3, elt-5 and elt-6*, are involved predominantly in hypodermal development and *elt-2, elt-4, elt-7, end-1, end-3, med-1* and *med-2* are involved predominantly in endodermal (as well as mesodermal) development. The role of each *C. elegans* GATA transcription factor has been reviewed in detail earlier in chapter 1, and so only a brief overview is given here.

Considering the hypodermal GATA factors first: ectopic expression experiments have shown that both *elt-1* and *elt-3* are sufficient to activate a program of hypodermal differentiation in early embryos (Gilleard, J. S. *et al.* 1999; Gilleard, J. S. *et al.* 2001). However only *elt-1* is essential to hypodermal cell formation with *elt-3* appearing to have a redundant role (Page, B. D. *et al.* 1997; Gilleard, J. S. *et al.* 1999; Gilleard, J. S. *et al.* 2001). *elt-5* and *elt-6* (Koh, K. *et al.* 2001) have been shown to be involved in the development of the lateral epidermal (seam) cells and are required continuously for post-embryonic development of seam cells and molting. Inhibition of *elt-5* through RNAi results in a penetrant late embryonic (pretzel) and early larval lethality (early L1) with loss of differentiated seam cells whereas *elt-6* RNAi does not produce any obvious phenotype. Overall, results suggest that *elt-5* is essential and *elt-6* is redundant, however the *elt-5* dsRNA

phenotype can be rescued by expression of *elt-6*, implying that there is some redundancy of the two genes products.

The seven remaining GATA factors, *elt-2, elt-4, elt-7, end-1, end-3, med-1* and *med-2* act in a cascade that regulates endoderm specification and development (Zhu, J. *et al.* 1997; Fukushige, T. *et al.* 1998; Zhu, J. *et al.* 1998; Maduro, F. M. *et al.* 2002; Fukushige, T. *et al.* 2003). Most of these genes, *elt-2, elt-7, end-1, med-1* and *med-2* have been shown to be sufficient to activate a program of endodermal differentiation by ectopic expression experiments. Only *elt-2* has a strictly essential role and *elt-2* null-mutant worms die as L1 larvae with severe malformation of the gut (Fukushige, T. *et al.* 1998). *end-3* and *end-1*, together with *dpr-1* (a nuclear receptor type transcription factor formerly known as *end-2*) form a region known as the EDR (endoderm-determining region, described in more detail in chapter 1). *end-1* and *end-3* are both expressed in the E blastomere shortly after its birth (Zhu, J. *et al.* 1997), indicating that the *end* genes are the earliest expressed known genes in the endoderm lineage. Both genes overlap in function and *end-1* can rescue the intestinal differentiation defects but not the embryonic lethality of EDR-deficiency embryos suggesting that *end-1*(+) activity can direct E-cell specification in the absence of *end-3* or *dpr-1* function (Zhu, J. *et al.* 1997). However *end-1* and *end-3* are not completely redundant for endoderm specification and *end-3* mutations lead to an impenetrant lack of endoderm in a minority of embryos (Maduro, F. M. *et al.* 2002). *med-1* and *med-2*, which are expressed exclusively in the EMS blastomere, promote the specification of EMS into E and MS and also appear to have mutually redundant functions. Of all the endodermal *C. elegans* GATA factors, *med-1* and *med-2* are expressed the earliest and mark the passage from maternal to zygotic control of mesendoderm specification (Maduro, F. M. *et al.* 2002). Although GATA factors *elt-*

4 and *elt-7* are expressed in endoderm, a functional role has yet to be established and neither gene is essential (Fukushige, T. *et al.* 2003). The way in which those *C. elegans* GATA factors act in a cascade to regulate endoderm differentiation and development is summarised in Figure 1.6. Hence in summary, the only *C. elegans* GATA factors that appear to have essential, non-redundant functions are *elt-1*, *elt-2* and possibly *elt-5*.

Although each of the 11 *C. elegans* GATA transcription factors have been characterised in some detail, there is no published information on GATA transcription factors from any other nematode species. Hence the extent to which individual family members are conserved in other nematode species is unclear. The identification of GATA transcription factors from other nematode species is generally problematic, since conserved sequence is limited to the zinc finger domains. Until very recently there was limited genomic sequence from nematode species other than *C. elegans*. However the *C. briggsae* genome sequence has recently been completed (Stein, L. D. *et al.* 2003). The genome is 102 Mb in size and around 19,500 proteins are predicted, 12,200 of which have clear *C. elegans* orthologues and 6,500 have one or more *C. elegans* homologues; around 800 *C. briggsae* genes have no detectable match in *C. elegans*. The completion of the genome of a species so closely related to *C. elegans* represents an important comparative genomic resource. The only parasitic nematode which currently has a genome-sequencing project is *Brugia malayi*. There is currently approximately a five-fold genome coverage although this is shotgun-based sequence and is yet to be assembled or annotated. These two genome projects now provide excellent resources for identifying GATA transcription factors in these two nematode species. In addition, EST projects from around 20 species of parasitic nematode are underway,

which aim to generate around 315,000 ESTs, including 235,000 at the Genome Sequencing Centre (GSC), Washington, St. Louis and 80,000 at the Sanger Centre and Edinburgh University. These are publicly available at http://ncma.cap.ed.ac.uk/nematode/ESTs/small_genomes/overview.html and http://www.nematode.net (Table 3.1) However, although these are generally very valuable resources, transcription factors tend to be expressed at low levels and so are poorly represented in EST databases. This is a particular problem as most of the parasitic nematode EST projects are relatively small (Table 3.1). Molecular biology approaches to identifying GATA transcription factors in parasitic nematode species are also potentially problematic. Heterologous probing of cDNA or genomic libraries by hybridisation with *C. elegans* GATA factor probes is unlikely to be successful. This is because comparative studies of GATA factors between closely related vertebrate species suggests there is little sequence conservation outside the DNA binding domain (Lowry, J. A. *et al.* 2000). Hence PCR based approaches are likely to be the most successful but even these present difficulties since there is likely to be very limited sequence conservation from which to design primers.

This chapter outlines a number of approaches to isolate GATA factors from parasitic nematodes based on database searching and molecular biology. It also presents the detailed molecular characterisation of a GATA factor isolated from the parasitic nematode *Haemonchus contortus*.

3.2 Results

3.2.1 Searching the *C. briggsae* genome for GATA transcription factors

The *C. elegans* GATA transcription factor family consists of eleven members, most of which have been characterised in some detail (Hawkins, M. G. *et al.* 1995; Page, B. D. *et al.* 1997; Zhu, J. *et al.* 1997; Fukushige, T. *et al.* 1998; Zhu, J. *et al.* 1998; Gilleard, J. S. *et al.* 1999; Gilleard, J. S. *et al.* 2001; Koh, K. *et al.* 2001; Fukushige, T. *et al.* 2003). The *C. briggsae* genome was searched as a first step to determine the extent to which each of these family members is likely to be conserved and identifiable in nematode species. Clearly if individual orthologues could not be identified in another nematode from the *Caenorhabditis* genus it is unlikely that orthologues could be identified in parasitic nematodes that are phylogenetically more distant. Blast searching (tBLASTn) (Altschul, S. F. *et al.* 1990) of the *C. briggsae* genome shortly before its completion identified eight separate GATA zinc fingers. The genomic sequence surrounding each of these zinc fingers was analysed by the Genefinder program to predict the structures of the encoding genes (by Dr J. Speith, University of Washington). The predicted polypeptide sequences of each of the eight *C. briggsae* GATA factors were aligned with all the eleven *C. elegans* GATA factor amino acid sequences using the ClustalX X program and a tree constructed by Neighbour-Joining using Paup 4.0 Beta software (Figure 3.1). Each of the *C. briggsae* predicted polypeptides was found to be more similar to an individual *C. elegans* polypeptide than to any of the other *C. briggsae* genes (except for cb4023 and cb3752). Hence the *C. elegans* GATA transcription factors *elt-1*, *elt-2*, *elt-3*, *elt-5*, *elt-7* and *end-3* have clearly discernable homologues in *C. briggsae* based on the comparison of the predicted full polypeptide

sequence. The two *C. briggsae* genes cb4023 and cb3752 are more similar to *med-1* and *med-2* than to any of the other *C. elegans* GATA factors. However their relationships to *med-1* and *med-2* is not one of simple orthology since they are more similar to each other than they are to either *med-1* or *med-2*. Hence it appears that *med-1* and *med-2* in *C. elegans* and cb4023 and cb3752 in *C. briggsae* have arisen by gene duplication subsequent to the separation of these species from a common ancestor. This is consistent with the published description of *med-1* and *med-2* which are two very similar *C. elegans* GATA factors (100% homology at the protein level and 98 % identical at the cDNA level) and it has been suggested they arose from a relatively recent gene duplication (Maduro, M. F. *et al.* 2001) At the time this sequence analysis was originally performed the *C. briggsae* genomic sequence was incomplete and so it seemed likely that the lack of *C. briggsae* orthologues of *end-1*, *elt-4* and *elt-6* was due to the incomplete nature of the *C. briggsae* sequence database. However recently the *C. briggsae* genome sequence was completed (Stein, L. D. *et al.* 2003) and so the complete sequence was searched again, using tBLASTn, with each of the *C. elegans* zinc fingers. No additional GATA factor homologues were found in this new search. Hence the only *C. elegans* genes that appear to lack a *C. briggsae* orthologue are *elt-6*, *end-1* and *elt-4*. Interestingly, the study of *elt-5* and *elt-6* revealed that *elt-6* is not an essential gene and that the two genes shared 46 % homology suggesting that one arose by duplication of the other (Koh, K. *et al.* 2001). This could have occurred after *C. elegans* and *C. briggsae* diverged from a common ancestor, explaining the absence of an *elt-6* homologue in *C. briggsae*. The absence of an *end-1* homologue in the *C. briggsae* genome is perhaps more surprising since *end-1* and *end-3* appear not to be entirely redundant; Although they overlap in function, and it has been shown that in *C. elegans*, *end-1* has an important role in

endoderm specification (Maduro, F. M. *et al.* 2002). Finally, in *C. elegans*, *elt-4* is located approximately 5 kb upstream of the *C. elegans elt-2* gene (Fukushige, T. *et al.* 2003). Approximately 12 kb of sequence upstream of cb4066 (*cb-elt-2*) was examined in detail and no homologue of *elt-4* was found in this region. The *C. elegans elt-4* function and origin was studied by Fukushige and colleagues (Fukushige, T. *et al.* 2003). They also searched for an *elt-4* homologue in *C. briggsae* and failed to identify an orthologue. Their analysis of the *C. elegans elt-4* gene sequence and function suggested that *elt-4* most likely resulted from the recent duplication of *elt-2* in *C. elegans* between 25 and 55 MYA, after *C. elegans* and *C. briggsae* diverged from a common ancestor and it does not appear to have a functional role in *C. elegans*. They went on to suggest this gene would most likely disappear from the *C. elegans* genome (Fukushige, T. *et al.* 2003).

3.2.2 Searching the *Brugia malayi* genome sequence for GATA transcription factors

The genome of the filarial nematode *B. malayi* is currently the object of a sequencing project (available at http://www.tigr.org/tdb/e2k1/bma/). The available sequence was searched for potential GATA factors when there was approximately 5-fold genome coverage. The sequence at this stage was not assembled and annotated and so the aim was simply to identify GATA type DNA binding domains for comparison with the *C. elegans* gene family. The objective was to investigate the feasibility of identifying homologues of individual *C. elegans* family members in a more distantly related nematode than *C.briggsae*. A Blast search (tblastn) was performed using the default parameters on the TIGR website using the DNA binding

domain of each of the *C. elegans* GATA factor family members. The parameters used were the same as those used for the *C. briggsae* search. In addition a number of other searches were performed using other Blast parameters (data not shown). A total of five GATA factor zinc fingers were found and these were named BmGATAa, BmGATAb, BmGATAc, BmGATAd and BmGATAe. It is interesting that only five family members could be identified in spite of 5-fold genome coverage being searched (Table 3.2). Each of these GATA fingers was identified multiple times on independent clones in the *B. malayi* database by the BLAST search (Table 3.2). Hence it seems likely that the *Brugia malayi* GATA transcription factor family is smaller than that of *C. elegans*. The available amino acid sequences of the five GATA factor DNA binding domains were aligned with the corresponding region of the eleven *C. elegans* family members using the ClustalX program and an unrooted Neighbour-Joining tree constructed (Figure 3.2). Interestingly, the tree supports the hypothesis that the three GATA factors in *C. elegans* that have essential functions have homologues in *B. malayi* (*elt-1, elt-2* and *elt-5/6*). Clear homologues are not identifiable for those genes that are redundant in *C. elegans*. Bootstrap values for the nodes which place BmGATAe with *elt-1*, BmGATAa with *elt-2/elt-4* and BmGATAc with *elt-5/elt-6* are 820, 903 and 975 respectively. It will be interesting to examine in future whether the function of these *B. malayi* genes reflect the homology relationships suggested by the sequence analysis.

3.2.3 Searching the parasitic nematode EST database for GATA transcription factors

The available parasitic nematode EST databases (http://nema.cap.ed.ac.uk and http://www.nematode.net) (Table3.1) were BLAST searched (tBLASTn) using the C-terminal zinc finger of each of the *C. elegans* GATA factors. EST clusters potentially encoding GATA transcription factors were identified in *Strongyloides ratti* (BI324036), *Meloidogyne arenaria* (BI747253) and *Ancylostoma ceylanicum* (BQ275651). The EST clones, kt148e08, rm37h01.y1 and pj42bo9.y1 were obtained from Dr J. McCarter (Washington University, St Louis) and the inserts fully sequenced. The translated ORF containing the GATA zinc fingers are shown in figure 3.3. The available amino acid sequences of the DNA binding domains of these predicted polypeptides were aligned with those of the eleven *C. elegans* family members using ClustalX and a Neighbour-Joining Tree was constructed (Figure 3.4). The *S. ratti* and *M. arenaria* predicted polypeptides are most similar to the *C. elegans* ELT-2 and ELT-4 polypeptides although the bootstrap support is low. Furthermore there is no zinc finger-like domain in an N-terminal position relative to the GATA zinc finger that corresponds to the "pseudo zinc finger" of the *C. elegans* ELT-2 polypeptide. Similarly the *A. ceylanicum* predicted polypeptide is most similar to that of *C. elegans* END-1 but the bootstrap support is very low. Hence the relationship of these parasitic nematode GATA factors identified in the EST databases to individual *C. elegans* family members is unclear.

3.2.4 Isolation, cloning and sequencing of a full-length cDNA encoding a *Haemonchus contortus* GATA factor

Blast searching of the *H. contortus* EST database using the GATA zinc finger amino acid sequence for each of the *C. elegans* GATA factors did not identify any putative GATA factor homologues. Consequently a molecular biology approach based on PCR using degenerate antisense primers was adopted. The DNA binding domain, and in particular the zinc finger domain, which is shared by all GATA transcription factors is the region of the polypeptide that is most highly conserved between species. Nested degenerate primers complementary to the sense strand encoding this region were designed and called GA2 and GA3 (figure 3.5A). These primers were designed to be most likely to amplify homologues of the three of *C. elegans* GATA factors, *elt-1*, *elt-2* and *elt-3* (figure 3.5B). *elt-1* and *elt-2* are of particular interest since they have essential roles in *C. elegans*. PCR was performed with primer GA2 and a T7 vector-specific primer using a λgt11 immature adult *H. contortus* cDNA library as template (kindly provided by D. Knox). Although no PCR product was visible from this amplification, the reaction was diluted and used as template in a second round hemi-nested PCR reaction using primer GA3 in conjunction with the same T7 primer. An amplicon of approximately 100 bp was produced and this was cloned into the pGM vector (Promega) and the insert of the resulting plasmid (pHaem1010) was sequenced. The sequence obtained was consistent with a GATA-type zinc-finger: the presence of sequence encoding amino acids TTL immediately 5' to the end of the GA3 primer and the presence of two cysteine residues in positions consistent with a $CX_2X_{17}CX_2$ zinc finger domain (figure 3.6). In order to isolate the 3' end of this transcript, hemi-nested PCR was performed with the sense primers Helt-8a (first round) and Helt-9a (second round) in

conjunction with a T3 vector specific primer using the same cDNA library as template (figure 3.7). A fragment of approximately 800 bp was amplified and cloned into the vector pGM (Promega) and the insert of the resulting plasmid (pY4.6) was sequenced. The insert of pY4.6 was found to overlap with the insert of pHaem1010 by 70 bp and was bounded by a polyA tail at its 3' extremity confirming it as the 3' end of the transcript (figure 3.7). In order to obtain the 5' end of the transcript, hemi-nested PCR was performed with antisense primers Hcelt-2/7 (first round) and Hcelt9 (second round) in conjunction with an SL1 sense primer using *H. contortus* adult cDNA as template (kindly provided by M. Roos) (figure 3.7). An amplicon of approximately 800 bp was cloned into vector pGM and the insert of the resulting plasmid (pBC12) was sequenced. The insert of pBC12 was found to overlap with the insert of pY4.6 by 50 bp confirming this was the 5' end of the same transcript (figure 3.7). Thus the full transcript from the SL1 trans-spliced leader sequence to the polyA tail was represented as three overlapping cloned PCR products (figure 3.7)

In order to isolate an intact cDNA clone, an adult *H. contortus* intestine λZAP II cDNA library ((Rehman, A. *et al.* 1998), provided by D.P. Jasmer, Pullman, WA, USA), was screened with a PCR amplified probe corresponding to the 5' end of the novel GATA factor cDNA sequence. The digoxigenin labelled probe (PCR amplified with primers Ann1 and Ann2) encompassed 652 bp of pBC12 sequence from nucleotide 121 to nucleotide 773 of the full-length cDNA sequence and was used to screen approximately 100,000 pfu as described in section 2.2.13. A single positive clone was isolated and the pBlueScript SK+ plasmid *in vivo* excised from the lambda ZAP II phage clone (see section 2.2.13) was subsequently named pACELT2. The insert was sequenced and was found to be 96.7 % identical with the consensus sequence of the three overlapping PCR fragments, pBC12, haem31010

and pY4.6 (figure 3.8). The 5' end of the PACELT-2 insert is 15 bp downstream of the SL1 splice acceptor site sequence (as determined by the pBC12 insert sequence) and the 3' end is bounded by a poly-A tail in the same position as in the pY4.6 insert. Although there are 44 single nucleotide polymorphisms between the pACELT2 insert and the consensus sequence of the overlapping PCR products the high level of identity between the pACELT-2 insert and the original overlapping PCR products confirms the identity of the transcript.

All except one of the SNPs between the pACELT-2 insert and the PCR fragments are synonymous substitutions suggesting they reflect sequence polymorphism rather than Taq polymerase-induced mutations introduced during the amplification of the PCR products. This interpretation is consistent with the high levels of sequence polymorphism previously reported for *H. contortus*, which has been ascribed to its very large effective population size (Blouin, M. S. *et al.* 1995). The different fragments of the transcript that have been sequenced are derived from different sources of parasite material: the inserts of the pHaem1010 and the pY4.6 plasmids were amplified from a λgt11 phage library from a *H. contortus* immature adult cDNA library as template (kindly provided by D. Knox); the insert of the pBC12 plasmid was derived from cDNA provided by Dr Marleen Roos; the insert of the pACELT-2 clone was derived from a lambda ZAP II cDNA library (provided by D. Jasmer). Furthermore each of these different parasite cDNA sources are derived from different laboratory isolates of *H. contortus* that were derived from different geographical locations. Hence it is likely that the SNPs are due to sequence variation between worms (or isolates) rather than suggesting the two sequences are derived from different genes in the same genome.

3.2.5 Sequence analysis suggests the *H. contortus* GATA factor is a *C. elegans* *elt-2* homologue.

The full polypeptide sequence of the *H. contortus* GATA factor is shown aligned to the *C. elegans* ELT-2 polypeptide in figure 3.9. There are a number of pieces of evidence to suggest the *H. contortus* GATA factor is a homologue of *C. elegans* ELT-2. Firstly, the amino acid sequence corresponding to the DNA-binding domain of the *H. contortus* GATA factor was compared to that of all eleven members of the *C. elegans* GATA family. The C-terminal DNA binding domain (zinc-finger and basic domain) was found to be more similar to that of *C. elegans* ELT-2 (70% homology or 51 out of 65 residues) than it was to any of the other *C. elegans* GATA transcription factor (alignments were generated using ClustalX and trees generated using PAUP4.0 beta) as shown on figure 3.10 The zinc finger of the *H. contortus* GATA factor is has 23 out of 25 residues identical to that of *H. contortus* ELT-2. Secondly, examination of the *H. contortus* GATA factor polypeptide sequence reveals a zinc finger-like domain that is 53 amino acids N-terminal to the GATA zinc finger. This in the same relative position, has the same arrangement of cysteine residues and shares 45.8% homology (11 out of 24 residues) with the "pseudo" zinc finger domain of *C. elegans* ELT-2 (figure 3.9). This is a domain unique to *elt-2* amongst the *C. elegans* GATA transcription factors, although its functional significance is not known. Thirdly, although there is very little sequence identity outside the DNA binding site, there are small regions of identity with *C. elegans* ELT-2 at the N and C-terminus, and these are not present in any of the other *C. elegans* GATA factors (figure3.9). Hence the GATA transcription factor isolated from *H. contortus* appears to be a homologue of the *C. elegans* *elt-2* gene. In the absence of a full genome sequence for *H. contortus* it is not possible to

definitively state that this is a true orthologue of *C. elegans elt-2*. However this sequence analysis and the results presented later in this thesis suggest they are likely to be orthologues and so this gene has been designated *Hc-elt-2*.

In order to determine whether Hc-ELT-2 is likely to be a functional GATA factor, its amino acid sequence was compared to that of *C. briggsae* ELT-2, *C. elegans* ELT-4, the yeast GATA factor AreA and F2B, a peptide from chicken GATA-1 (Omichinski, J. G. *et al.* 1993) as shown in figure 3.11. The structure-function relationship of GATA factors has been studied in detail using chicken GATA-1 as a model for other vertebrate GATA factors, as well as the yeast GATA factor AreA (Omichinski, J. G. *et al.* 1993; Ravagnani, A. *et al.* 1997; Manfield, I. W. *et al.* 2000; Trainor, C. D. *et al.* 2000; Newton, A. *et al.* 2001; Shimizu, R. *et al.* 2001). Of particular interest is the work from Omichinski and colleagues (1993), who established the minimum size of a GATA-1 derived peptide that retained the ability to bind the canonical DNA targets (A/T)GATA(A/G). They found that the C-terminal GATA zinc finger of the double finger motif was necessary but not on its own sufficient for sequence-specific interaction at the canonical GATA site (A/T)GATA(A/G). Basic amino acids located C-terminal to the C-terminal zinc finger were also required for tight binding; the minimal peptide able to bind with affinity and specificity to the canonical GATA site was named F2B (figure 3.11). More recent work on the GATA transcription factor AreA revealed that the universally conserved Leucine at the seventh position after the second cysteine of the GATA zinc finger was essential for the AreA activity (Ravagnani, A. *et al.* 1997). Swapping the Leucine for a Valine resulted in inability to activate some ArcA-dependent promoters, while some other AreA-dependent promoters were found to function more-efficiently than in a wild-type context. Hence work on both GATA-1

and AreA have helped to understand which residues are essential to DNA-binding within the zinc finger domain. The alignment in figure 3.11 shows that all the residues known to be important in GATA factor function are conserved in Hc-ELT-2.

Finally, the amino acid sequence was searched for Nuclear Localisation signals (NLS), since a GATA transcription factor would be expected to be nuclear localised. Christophe *et al.*, (2000) have described several types of NLS, including the classical basic types, following the bipartite consensus $(K/R)_2X_{10-12}(K/R)_4$ or the monopartite consensus $(K/R)_{4-6}$. Figure 3.12 shows the presence of two such NLS in HcELT-2 (Christophe, D. *et al.* 2000).

3.2.6 Expression pattern of the Hc-ELT-2 polypeptide

In order to determine the expression pattern of the endogenous Hc-ELT-2 polypeptide, a polyclonal antiserum against the Hc-ELT-2 polypeptide was generated. This was then used in immunofluorescence studies to localise the Hc-ELT-2 polypeptide in the parasite.

3.2.6.1 Expression of Hc-ELT-2 fusion proteins

Two Hc-ELT-2 recombinant 6xHisTag fusion proteins HCELT2HISA and HCELT2HISB were expressed in M15 bacterial cells as described in section 2.3.2.2. Neither polypeptide includes the DNA-binding domain and they do not overlap with each other (figure 3.13). Both polypeptides were expressed, solubilised and purified

on a Ni-NTA resin column as described in section 2.3 and shown in figure 3.13. Figure 3.13 shows each fraction obtained after wash and elution of the columns for protein HCELT2HISA separated on SDS-PAGE gels stained with Coomassie blue. The fractions with the highest concentration of eluted polypeptide (E3-E6) were pooled together, concentrated and dialysed to decrease urea concentration as described in section 2.3.2.2.

3.2.6.2 Production and analysis of Hc-ELT-2 specific antibodies

Hc-ELT-2 specific polyclonal antisera were produced by immunising rabbits with the two recombinant proteins HCELT2HISA and HCELT2HISB as described in section 2.4.2. Rabbits R1064 and R1104 were immunised with the HCELT2HISA fusion protein and rabbits R1077 and R1082 were immunised with the HCELT2HISB fusion protein. Booster immunisations were performed at 35, 60 and 86 days and serum samples were taken from each rabbit at 28, 53 and 79 days after the primary immunisation and a terminal bleed performed at 105 days. The immunoreactivity of each serum sample was tested by probing Western blots of the two fusion proteins and *H. contortus* adult extracts (figure 3.14). There was little immunoreactivity with the pre-bleed sera from any of the four rabbits with either of the fusion proteins or worm extracts. Although the Western blots are poorly resolved, they were sufficient to demonstrate a high level of specific immunoreactivity to the appropriate fusion proteins.

3.2.6.3 Immuno-localisation of HcELT-2

Mixed stage *Haemonchus contortus* embryos were harvested, freeze-cracked, fixed and immuno-fluorescence was performed as described in section 2.4. Whole antiserum was used as primary antibody at dilutions of 1/50, 1/200 and 1/500 and a Cy2 conjugated donkey anti-rabbit IgG secondary antibody (Jackson ImmunoResearch) was used at a 1/200 dilution (as per manufacturer's instruction). It was found that the optimal signal to background ratio was obtained at a primary antibody dilution of 1/500. Details of the methodology are given in section 2.3. The sera obtained from the third bleeds from four immunised rabbits (R1064, R1104, R1077 and R1082) were tested using these conditions and all produced the same expression pattern when used to stain mixed stage *H. contortus* embryos. The antiserum from rabbit 1064 was judged to be the most sensitive and so was used for subsequent detailed analysis. Negative controls in which the primary antibody was replaced by pre-immune antisera were performed at the same time using the same conditions. These showed no fluorescence confirming the specificity of the antibody staining.

H. contortus eggs were extracted from fresh sheep faeces as described in section 2.4, cleaned in 0.125 M hypochlorite solution and washed several times in M9. They were then immediately freeze-cracked on poly-L-lysine coated slides, methanol/acetone fixed and immunostained using antisera 1064 (see section 2.4 for methods). Staining with DAPI revealed that these embryos were at an early stage of development, with less than 30 cells in total (data not shown) and no specific staining with the antisera could be seen. In order to investigate the staining with the 1064 antisera at different times of embryonic development, embryos that had been harvested from faeces were allowed to develop for different periods of time at 20°C

prior to freeze-cracking, fixation and staining. Specific staining with the 1064 antisera was first seen when embryos were fixed 3-4 hours after harvesting at which time the earliest embryos contained a total of 30-40 cells, as determined by DAPI staining (figure 3.15). In these embryos nuclear localised staining was seen in just two cells. Figure 3.15 shows an *H. contortus* embryo stained with DAPI (panel A) and the same embryo stained with 1084 antisera (panel B). The two nuclei that stain with the antibody are larger than the rest of the nuclei in the embryo. This suggests they may correspond to the two daughter cells of the E-blastomere since in *C. elegans* these two E cells can easily be distinguished by large size of their nuclei (Sulston, J. E. *et al.* 1983). No early *H. contortus* embryos were observed in which a single cell stained with the 1064 antisera. Also the two cells that stain in these 30-40 cell embryos are centrally placed within the embryo. By analogy with *C. elegans* early embryonic development, this suggests that Hc-ELT-2 expression is first seen in the two daughter cells of E after they have migrated to the centre of the embryo during gastrulation. Nuclear localised staining with the 1064 antisera was detected in all subsequent stages of *H. contortus* embryogenesis from 8 cells in a lima bean-like stage to a total of 20 cells in comma stage, 1-fold and 2-fold embryos (Figure 3.16, Panels A-D). These cells are clearly endoderm due to their central position and organisation in the embryo. Post-embryonic stages were also fixed and stained with the R1064 polyclonal antiserum. Freshly harvested eggs were cleaned in 0.125 M hypochlorite solution and washed several times in M9 as previously described. Approximately 500 *H. contortus* eggs were placed onto standard NGM agar plates allowed to hatch overnight and incubated at 25°C. Larvae were harvested every 24 hours, fixed and stained as described in section 2.4 (Miller, D. M. *et al.* 1995). The development of *H. contortus* larvae on NGM plates is described in more detail in

Chapter 6. Larvae identified as L1 showed nuclear localised staining in 20 cells which, from their anatomical position, were clearly gut cells (figure3.16, panel F). In larvae judged to be L2 stage (figure 3.16, panel G), nuclear localised staining was observed in a larger number of gut cells (up to 30 cells). L3-stage larvae, derived either from NGM plates or faecal cultures showed no specific staining with the R1064 antisera (figure 3.16, panel H). This absence of could either reflect lack of Hc-ELT-2 expression or a lack of antibody penetration. *H. contortus* L3 larvae retain the L2 cuticle as a sheath and this could block antibody penetration. In order to investigate this possibility, L3 larvae were exsheathed in hypochlorite solution prior to fixation and staining but there was still no specific staining. L3 larvae were also freeze-cracked following exsheathment and the same methanol acetone fixation method used on embryos was performed but again no staining was seen. Hence, in spite of a variety of approaches no specific staining of L3 stage larvae was seen with the R1064 antisera.

Permeabilisation of whole adult worms is problematic due to their large size and so, for detailed immunolocalisation studies, it is necessary to embed and section them. However, during whole worm fixation (described in section 2.4) some adult worms break open and gut tissue is accessible to antibodies. Staining with R1064 antiserum was confirmed in this way by examination of gut tissue that had been sheared from the worm (figure 3.16, panel I). It can be seen that there is nuclear localised staining of a large number of nuclei present in the adult gut.

In addition to the strong staining seen in endodermal cells of L1 and L2 larvae, staining was also see in one cell either side of the pharynx; although this staining was faint, it was consistently seen in many larvae (figure 3.17). This aspect of the expression pattern is discussed further in chapter 5.

3.6.4 Western blot analysis of Hc-ELT-2 expression

Since Hc-ELT-2 expression could not be detected in L3 by immunofluorescence, Western blots of L3 and adult *H. contortus* extracts were probed with the 1064 HcELT-2 specific antisera. Details of the method used are in section 2.2.19. To maximise the solubilisation of L3 polypeptides, adults and larvae were boiled in SDS-PAGE loading buffer for 20 minutes. A Western blot of *H. contortus* adult and L3 extracts were probed with the R1064 third bleed antisera and also the pre-immune R1064 anti-sera as a negative control. A predominant band of approximately 50kDa was detected in the adult *H. contortus* extract by the third bleed R1064 antisera (Figure 3.18). This band was not detected by the R1064 pre-immunisation sera (data not shown). The Hc-ELT-2 polypeptide is predicted to be approximately 42kDa based on its amino acid sequence (417 amino acids). Although the band detected on the Western blot is slightly larger than this, its size is entirely consistent with the Hc-ELT-2 polypeptide. Polypeptides often show lower than predicted mobility on SDS-PAGE gels and indeed this has been found to be the case for the *C. elegans* ELT-2 polypeptide. However, no band was detected by the R1064 antisera in the *H. contortus* L3 extract. Observation of the Ponceau S staining of the membrane after transfer confirmed that the loading and transfer of L3 polypeptides was similar to that of the adult extract (data not shown). This was repeated several times and no convincing expression of Hc-ELT-2 was observed in L3 larvae. Hence, two independent methods (IFA and Western blot analysis) have failed to detect Hc-ELT-2 expression in *H. contortus* L3 larvae whereas expression was seen in all other stages examined. This is a surprising result since ELT-2 is expressed in all *C. elegans* stages including L3 larvae (Fukushige, T. *et al.* 1998). However the *H. contortus* L3 is a quiescent non-feeding stage and is considered to have some

similarities to the *C. elegans* dauer stage, which is a facultative quiescent, non-feeding stage (Hope, I. A. 1999). The expression of ELT-2 in *C. elegans* dauer larvae has not been reported in the literature and consequently the expression of an *elt-2::lac-Z* reporter gene in *C. elegans* dauer larvae was examined. Dauer larvae were made from the JM62 strain (kindly supplied by J. McGhee) which is a transgenic strain carrying a chromosomally integrated copy of an *elt-2::lacZ* reporter gene. X-gal staining was clearly visible in JM63 dauer larvae (figure 3.19). Although it is difficult to quantify lacZ staining, the intensity appeared to be less in the dauer larvae than for other the *C. elegans* larval stages of the JM63 strain (figure 3.19)

3.6 Discussion

The GATA zinc finger is a highly conserved motif that characterises GATA transcription factors in phylogenetically divergent species. The *C. briggsae* and *Brugia malayi* genomes, as well as the parasitic nematode EST databases, were searched for potential GATA transcription factors by BLAST searching with the GATA zinc finger of each of the *C. elegans* GATA transcription factors. In the *C. briggsae* genome, a total of 8 GATA factors were found, each of which (with the exception of cb4023 and cb3752) are clear homologues of individual *C. elegans* family members. Of all the *C. elegans* GATA factors, only *elt-4*, *elt-6* and *end-1* appear to have no homologues in *C. briggsae*. The search for parasite homologues of GATA factors yielded 5 GATA factors from the *B. malayi* genome, and one each from the EST databases of *S. ratti, M. arenaria* and *A. ceylanicum*. In the case of the

B. malayi genome, although the sequence available already provides a 5.4-fold coverage of the genome, only five GATA factors were identified. All of the putative GATA factor homologues from *B. malayi* were found on several independent clones consistent with the 5-fold genome coverage. Homologues of ELT-6, ELT-4 and END-1 also appear to be missing from the *C. briggsae* genome. Hence it appears that not all nematode species contain as many GATA transcription factors as *C. elegans*. This theory is further supported by the fact that some of *C. elegans* GATA factors appear to be redundant and that the essential family members ELT-1, ELT-2 and ELT-5 appear to be present in the *C. briggsae* and *B. malayi* genomes. It was a particular aim to identify GATA transcription factors from the parasitic nematode *H. contortus*. However Blast searching of the 16,000 currently available *H. contortus* ESTs did not identify any putative GATA factors. Screening of an *H. contortus* cDNA library by PCR with degenerate primers was more successful and a cDNA fragment encoding a putative GATA zinc finger was isolated. Subsequently a full-length cDNA clone was isolated which clearly encoded a GATA transcription factor. This gene was considered to be a homologue of the *C. elegans elt-2* gene for a number of reasons. Firstly, the zinc finger domain is more similar to that of ELT-2 than to any of the other *C. elegans* GATA factor members. Additionally, the DNA-binding domain of *C. elegans* ELT-2 is closer to that of the *H. contortus* polypeptide than it is to any of the other *C. elegans* GATA transcription factor family member. Secondly, the *H. contortus* polypeptide contains a zinc finger-like structure that has 50% identity to a domain that has been termed a "pseudo zinc finger" in *C. elegans elt-2*; this feature is unique to *elt-2* amongst all the *C. elegans* family members. Thirdly, there are small regions of homology at the N- and C-termini of the *H. contortus* polypeptide with similar regions of the *C. elegans* ELT-2 polypeptide but

not the other *C. elegans* GATA factors. Finally, the expression pattern of the polypeptide encoded by this gene was examined in *H. contortus* and revealed to be extremely similar to that of the *C. elegans* ELT-2 polypeptide. The above evidence strongly suggests the *H. contortus* gene is an orthologue of *C. elegans elt-2* although it is impossible to prove an orthologous relationship in the absence of a full *H. contortus* genome sequence. Consequently, this gene has been named *Hc-elt-2*.

The level of amino acid sequence identity between Hc-ELT-2 and Ce-ELT-2 outside the DNA-binding and pseudo zinc finger domains is only 35.7 %, and 26.9 % overall. This is relatively low compared to most other *H. contortus* polypeptides that have been compared with *C. elegans* homologues. For example Hc-CPL-1 (Cathepsin L), Hc-GBR2B and HcGBR2A (the Glutamate-Chloride-gated channels), Hc-UNC-18 (a syntaxin-binding protein) and Hc-STK (a Serine-Threonine protein Kinase) share 83%, 87%, 88%, 83% and 54% amino acid sequence identity respectively with their *C. elegans* homologues (Jagannathan, S. *et al.* 1999; Britton, C. *et al.* 2002; Nikolaou, S. *et al.* 2002; Shompole, S. *et al.* 2003). This lower overall level of sequence identity may be a reflection of the modular domain structure that typifies transcription factors compared to structural proteins and enzymes.

As might be anticipated from phylogenetic relationships the overall lack of sequence identity between Hc-ELT-2 and Ce-ELT-2 is strikingly less than between Ce-ELT-2 and Cb-ELT-2. Providing there is functional conservation of the *C. elegans* and the *H. contortus* polypeptides this overall lack of amino acid sequence identity might provide valuable information to identify key functional domains and residues of the ELT-2 polypeptide. For example, the functional significance of the "pseudo" zinc finger in the *C. elegans* ELT-2 polypeptide is unclear (Hawkins, M. G. *et al.* 1995; Fukushige, T. *et al.* 1998). Its conservation in the Hc-ELT-2

polypeptide might suggest it is of functional significance, particularly since its position in the Hc-ELT-2 polypeptide is very similar to its position in the *C. elegans* ELT-2 polypeptide relative to the DNA binding domain. Furthermore, this pseudo finger domain in ELT-2 is also in a similar position to the N-terminal GATA zinc finger domain in the vertebrate GATA factors (relative to the DNA binding domain), which has been shown to be important in both DNA binding and protein-protein interactions (Trainor, C. D. *et al.* 2000; Newton, A. *et al.* 2001). Even though the "pseudo" zinc finger of ELT-2 is not a GATA-type zinc finger ($CX_2CX_{17}CX_2C$) it could perform a similar role. Recently site directed mutagenesis of a cysteine residue in the pseudo zinc finger has been shown to remove the ability of the polypeptide to activate endodermal differentiation suggesting the functional importance of this domain (J. Smith and J. Gilleard, personal communication). The pseudofinger domain is also of interest because vertebrate GATA factors do not contain such a domain and so it is potentially a nematode-specific domain, although this domain appears to be missing from the putative ELT-2 homologues from *S. ratti* and *M. arenaria.*

Hc-ELT-2 antibodies were generated and used to examine the spatial and temporal expression of the polypeptide. At all stages of embryonic and post-embryonic development the Hc-ELT-2 polypeptide was localised to the nucleus of expressing cells. No evidence for cytoplasmic localisation was observed. This is consistent with the role of Hc-ELT-2 as a DNA binding protein and transcriptional regulator. The earliest expression during development was seen when *H. contortus* embryos consist of approximately 30-40 cells. At this stage HcELT-2 was detected in two cells and comparison with DAPI stained embryos suggests that the nuclei of these cells are the largest in the embryo at this point in development. In *C. elegans,*

elt-2 is first expressed in the two E cells (endoderm precursors) when the embryo has around 40-50 cells (Fukushige, T. *et al.* 1998). The nuclei of the *C. elegans* E cells are rounder and larger than cells from other lineages and so, by analogy, the large cells expressing HcELT-2 in 30-40 cell *H. contortus* embryos are likely to be the two endoderm precursor cells. Interestingly expression of Hc-ELT-2 has not been seen in a single cell in embryos earlier than the 30-40 cell stage suggesting it is not expressed in the single endoderm blastomere in *H. contortus*. Similarly, *C. elegans* ELT-2 is not expressed in the single E blastomere but is first seen in the 2E daughter cells. Hence the onset of Hc-ELT-2 expression in the *H. contortus* endoderm lineage is identical to that of ELT-2 in *C. elegans*. In addition, when Hc-ELT-2 expression is first seen in the 2E cells in *H. contortus* embryos (this terminology is used by analogy to *C. elegans* development) these cells are invariably towards the centre of the embryos. This suggests that Hc-ELT-2 expression initiates either during or after ingression of the 2E cells into the centre of the embryo during gastrulation. Again this is precisely the same as for the ELT-2 expression in *C. elegans* embryos (Fukushige, T. *et al.* 1998). These results suggest that not only is the initiation of expression of Hc-ELT-2 very similar to that of *C. elegans* ELT-2 but the early developmental events of endoderm specification and gastrulation are also be highly conserved between these two nematode species. The specification of the single E-cell in *C. elegans* and the onset of *elt-2* expression are regulated by a pair of GATA factors *end-1* and *end-3* (Maduro, F. M. *et al.* 2002). It would be interesting to examine the conservation of these genes and their expression patterns in *H. contortus* to further investigate the conservation of early endoderm specification between the two nematode species.

In *H. contortus*, the cell lineage has yet to be determined but the anatomical arrangement of the cells expressing the Hc-ELT-2 polypeptide during development is consistent with the *H. contortus* endodermal lineage being clonal as is the case in *C. elegans* (Sulston, J. E. *et al.* 1983; Sulston, J. E. *et al.* 1988). Furthermore, the number of endodermal nuclei observed in each stage of embryonic development in *H. contortus* is identical to that observed in the corresponding stages of *C. elegans*. Hence the *Hc-elt-2* specific antiserum has proven to be a very useful marker of endoderm development in *H. contortus*. It has demonstrated the high level of conservation of E-lineage development between *H. contortus* and *C. elegans*.

In spite of the similarity of the endodermal expression pattern of the Hc-ELT-2 polypeptide to that of the *C. elegans* ELT-2, a number of differences in the overall expression patterns were observed. Firstly, no expression could be detected in the L3 stage by either immunostaining or by western blotting. However since a positive control antibody was not available, it is still possible that Hc-ELT-2 is expressed in L3 larvae but was not detected due to difficulties in permeabilisation and protein solubilisation. However this seems unlikely particularly since the Ponceau S staining suggested protein extractions for western blots were effective. It would be valuable to perform RT-PCR on L3-stage larvae to determine whether *Hc-elt-2* transcripts are present, in order to confirm a lack of expression. One possible reason for the lack of Hc-ELT-2 expression in infective L3 larvae is that this is a metabolically and developmentally quiescent stage that survives for many months (and even years) on the pasture. Expression of an *elt-2::lacZ* reporter was examined in *C. elegans* dauer stage larvae, since this quiescent stage has been considered to be analogous to the infective L3 larvae of parasitic nematodes such as *H. contortus*. The *elt-2*::GFP reporter was found to be expressed in *C. elegans* dauer larvae albeit at

lower levels than other larval stages. Hence the apparent lack of Hc-ELT-2 expression in *H. contortus* L3 remains surprising.

A second apparent difference in the *H. contortus* and *C. elegans* ELT-2 expression pattern was that close observation of immunofluorescence staining of *H. contortus* L1 and L2 larvae with the R1064 antisera revealed some faint staining outside the endoderm. The nuclei of two cells either side of the pharynx stained very faintly but nevertheless consistently in several experiments. It is difficult to unequivocally identify these cells as there are many cells in this region and the cell lineage of *H. contortus* has not been completely mapped. However there position would be consistent with that of neuronal cells and this is discussed in more detail in chapter 5. By analogy to *C. elegans* and with respect to a recent anatomical study of *H. contortus* by (Ashton, F. T. *et al.* 1996; Li, J. *et al.* 2000), it is likely that these cells are amphid neurons. A list of *C. elegans* neurons can be found on Leon Avery's *C. elegans* server http://www.wormatlas.org/neuroimageinterface.htm. This extra-endodermal expression was a surprising result because the *C. elegans elt-2* gene expression is restricted entirely to the endoderm and no such additional expression in cells around the pharynx has been described.

In summary the expression pattern of Hc-ELT-2, a homologue of the *C. elegans* gut-specific GATA transcription factor *elt-2* is very conserved with that of *C. elegans elt-2* both spatially and temporally and only differs by additional low levels of expression in two cells either side of the pharynx and the absence of expression in *H. contortus* L3.

Nematode	Number of ESTs
Ancylostoma caninum	9331
Ancylostoma ceylanicum	10591
Ascaris suum	29962
Brugia malayi	3773
Dirofilaria immitis	4005
Globodera rostochiensis	5040
Haemonchus contortus	14014
Heterodera glycines	20110
Meloidogyne arenaria	5018
Meloidogyne chitwoodi	3186
Meloidogyne hapla	16303
Meloidogyne incognita	16501
Meloidogyne javanica	7565
Onchocercha volvulus	1230
Ostertagia ostertagi	6558
Parastrongyloides trichosuri	7915
Pratylenchus penetrans	1928
Strongyloides ratti	10760
Strongyloides stercoralis	11335
Toxocara canis	4370
Trichinella spiralis	10548
Trichuris vulpis	3063
Zeldia punctata	391

Table 3.1: Parasitic nematode species with current EST projects (adapted from http://www.genome.wustl.edu/est/).

The table shows the name of the species in alphabetical order and the number of ESTs currently sequenced for a each species.

Figure 3.1: Unrooted Neighbour-Joining tree showing the relationship of each of _C. elegans_ GATA factor to its homologue in _C. briggsae._

Full-length predicted polypeptide sequences were aligned in ClustalX and a boot-strapped Neighbour-Joining tree was constructed. The boot-strap values are out of 1000 replicates. The figure shows that each _C. briggsae_ potential GATA transcription factor is a clear homologue of a _C. elegans_ GATA factor, with the exception of cb3752 and cb4023,
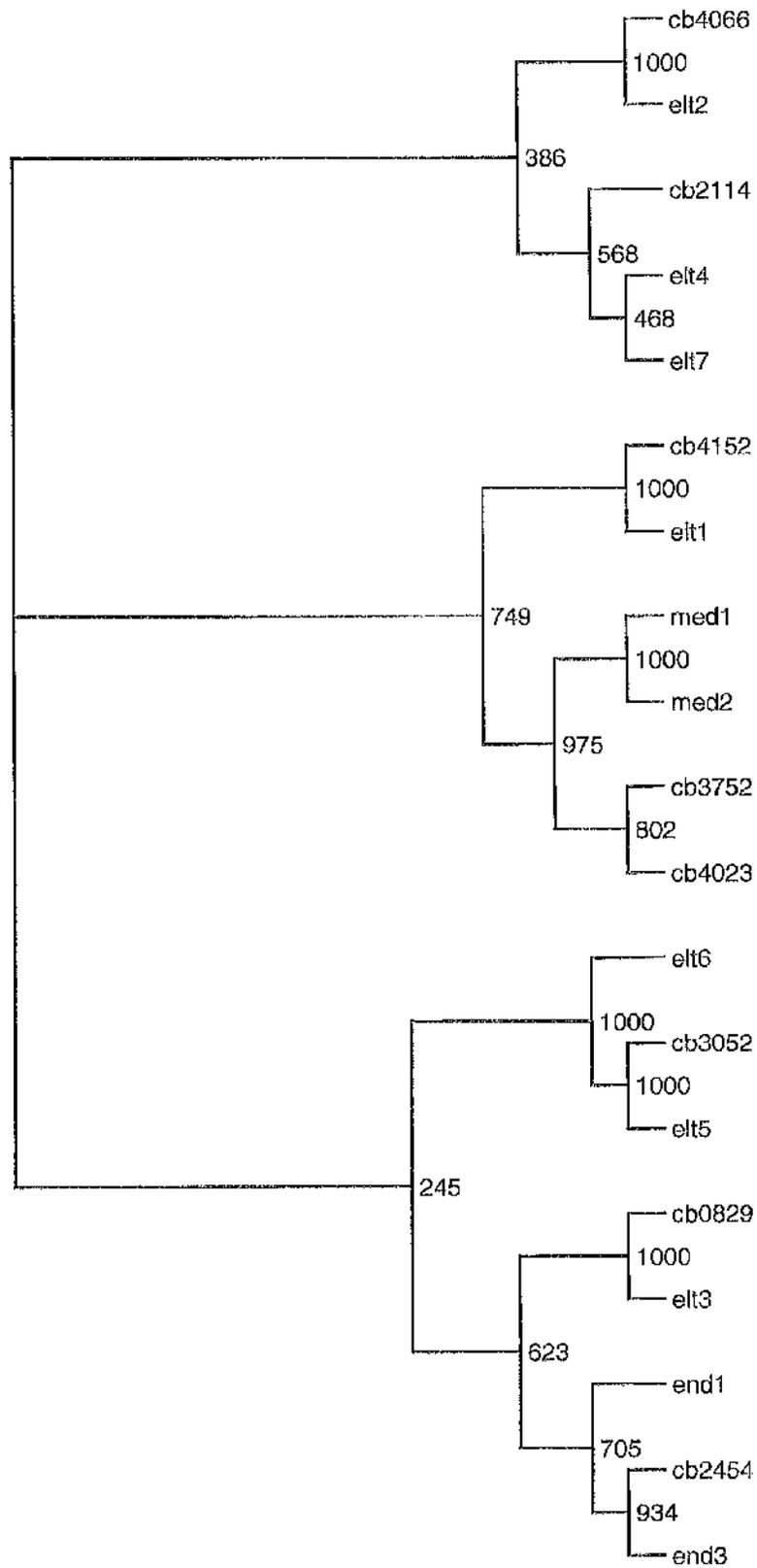
Table 3.2: List of clones containing sequence corresponding to potential homologues of *C. elegans* GATA factors identified in the *B. malayi* genome.

Each potential homologue identified by a BLAST search from the *B. malayi* genome was given a name BmGATAa to BmGATAc and the corresponding clones identified for each homologue are listed in the right-hand column.

Name of potential GATA factor	Name of clone in TIGR database
BmGATAa	1131478
	48M00611
	Brklj66tr
	BRNINME41TF
	BRMJF91TF
BmGATAb	1117125
	248M00036
BmGATAc	42M03163
	BRIUL64TR
	BRBAP12V
	BRLED61TF
	BRME59TR
	BRNN253TR
BmGATAd	42m03208
	BRGMP78TR
	BRIUN27TR
	BRJRF11TR
	BRLGA32TF
	BRML473TR
	BRM829TR
BmGATAe	Brbbg10tj
	BRHHB47R
	BRIGY69TR
	BRKWW51TF
	Brmyu88tf
	Brnih32tr

Figure 3.2: Alignment of *C. elegans* GATA factors zinc finger domain with those of potential GATA homologues from *B. malayi*

The unrooted Neighbour-Joining tree tree shows the relationship between the potential *B. malayi* GATA factors and the *C. elegans* GATA factors. The alignment was performed by ClustalX and the tree is an unrooted Neighbour-Joining tree based on the alignment shown. The boostrap values shown on each nodes are out of 1000 replicates.

```
elt1      -RTGIECVN--CRTNTTTLWR--RNGEGHPVCNACGLYFKLHKVE--RPITMK--KDGI-  50
BMGATAe   --TGIECVN--CKTNNTTLWR--RNSLGQPVCNACGLYHKLHNVSYFKPSIQKWNRNGRC  54
elt2      -RQGLVCSN--CNGTNTTLWR--RNAEGDPVCNACGLYFKLHHIP--RPTSMKKEGA---  50
elt4      -RKRLVCSN--CNGTNTTLWR--RKAEGDPVCNACGLYFKLHHVT--RPIPMKKNKK---  50
BMGATAa   RRQGLVCAN--CRGTNTTLWR--RDADGHPVC-------------------------    28
bmGATAd   --EGRECVN--CGAVQTPLWR--RDGTGHYLCNACGLYHKMNGQN--RPLVKPKKRQ---  49
elt6      -SSISKCSN--CSTIKTTAWR--RDLEGKLVCNACGLYYRLHRTH--RPVHMRKDFI---  50
BMGATAc   -PSMSRCSN--CLTTKTTAWR--RDQTGKLVCNACGLYYRLHRVN--FHYLLVT------  47
elt5      -PNAARCSN--CRTDKTTAWR--RDAEGKLVCNPCGLYYRLHKVR--RPIEMRKNHI---  50
end3      -HINSSCSN--CGCRETKLWR--RNEQGETECNPCNLYERVKGHK--RPQHLWNKPA---  50
BmGATAb   THINSVCSN--CGTKETTLWR--RSNMGAVECNACNLYYRKNNRP--RPITMSNKIRK--  52
elt3      -HQNSICSN--CKTRETTLWR--RNGEGGVECNACNLYFRKNNRK--RPLSLRKDGI---  50
end1      -HTNSVCSNPNCRTRETTLWR--RTDSGAIECNGCSLYFRKNGIQ--RPAELCRK-----  50
elt7      -IRDACCSH--CSTTTTTLWR--KNDEGNLECNACNLYYRHNKVK--RPLSLCKQKP---  50
med1      -KKSFQCSN--CSVTETIRWRNIRSKEG-IQCNACFIYQRKYNKT--RPVTAVNKYQ---  51
med2      -KKSFQCSN--CSVTETIRWRNIRSKEG-IQCNACFIYQRKYNET--RPVTAVNKYQ---  51
```
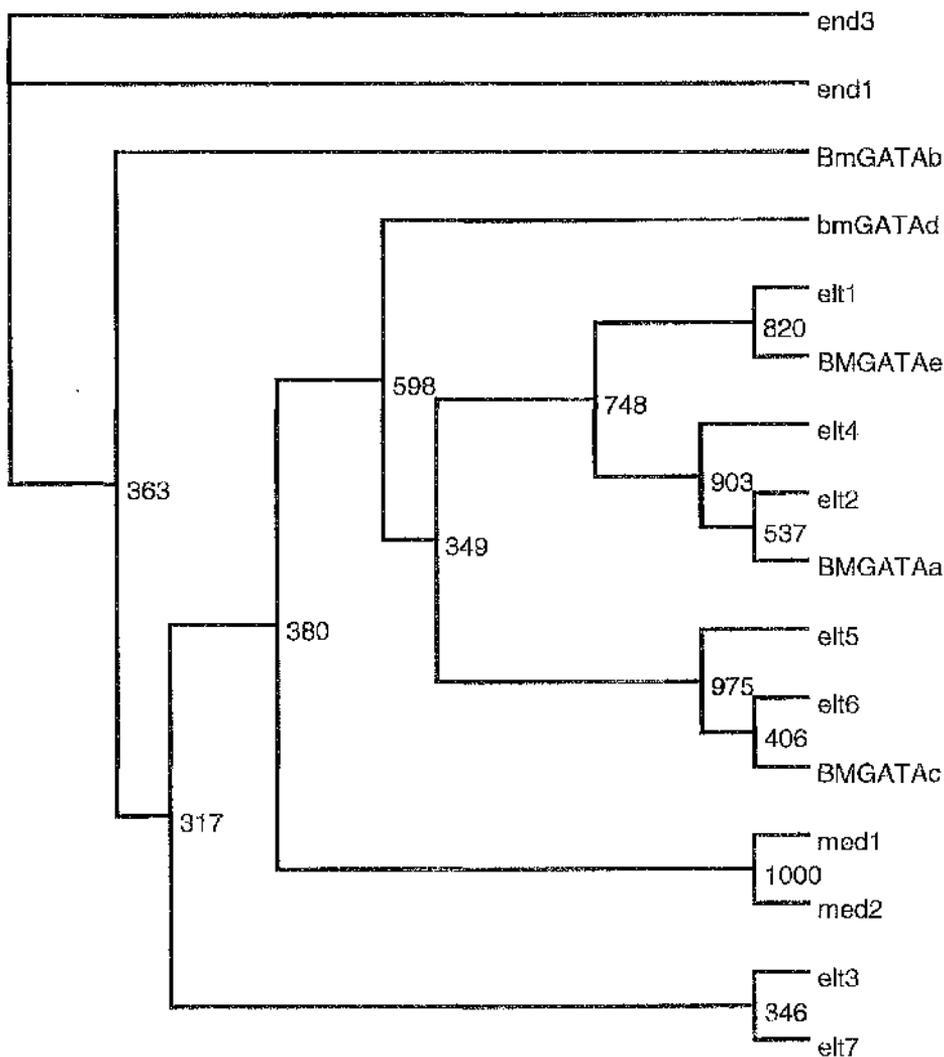
Figure 3.3: Translated open reading frames of EST clones

encoding potential GATA transcription factors.

Three EST clusters which encoded potential GATA transcription factors were identified by BLAST searching the parasitic nematode EST databases with the GATA zinc fingers of each of the *C. elegans* GATA factors. The inserts of a clone from each of these clusters was sequenced and a translation of the open reading frame containing a GATA zinc finger is shown . The GATA type zinc finger is underlined in each case.

S. ratti GF= kt148e08

M. arenaria GF= rm37h01.y1

A. ceylanicum GF= PJ42b09.yl

S. rattiGF
1   AANTTHYRAN DSNSSKILNM NNMQQQQSG TQVSISDSTE GSTKKTFQKK TITNNQKRQD
61   MKCNNCGGSN TTLWRRNAEG HPVCNACGLY YKLHGIQRPL SMKKDGDTQK RKRKPKSNNP
121 DNVRKGRQEI GQIQGLHQSG NHQPSHVIYT NQQSFNPSTQ NTVSVQISTN GLDNSYENTN
181 FNIDHFSRIS IQNQMEGPYM GIENQNLQQN DVTITKQDYI TGFGINTSNV IQAPMVQHKD
221 EINNENKGEQ ILVKDDHLIT TNDASENECI ASEFLRPRVP STVPENVEQM NVENISESYH
281 NTDNEENTLP NVESPDNKEE NLQKNLDNDD NLSISIEENN H

M. arenariaGF
1   PLIRLTIGDA FGYSICDNCA QPAALLHAQQ QQENAVDLMG NKDGQQQQFC ASEFLFPQPQ
61   QHQQPIATQL SPLNSGNFPT LPKVRGRKPN SGKRNVLKGG PTKTLPKPEK MDNISVMDES
121 SSDSSALNCP PPMPTQKRQN LICSNCNGTD TTLWRRNNNG EPVCNACGLY YKLHNIARPL
181 TMKKETVQTR QNKIFLILYF NNLFIKG

A. ceylanicumGF
1   MMTREHSLRD CLSDPYTDFF AKEXLQHIMD PHQIHFNVST TGSSKKLYED DHSTSFQCAA
61   PLFELPHSNP MNMLLYEMPP MGMYPSSSRS ELEVNTSAHL PTVFPIPSAT YLHQSHSPSE
121 YIPNGYSPPP QDPIYPPKPP SQHLSPLIKP KKRIQAVPCH SNSVCANCKT TETTLWRRAK
181 TGEIECNACN LYFXFERNSI GSI

Figure 3.4: ClustalX alignment and unrooted Neighbour-Joining tree of three potential GATA factor homologues of *C. elegans* identified in the parasitic nematode EST databases

The alignment was performed by ClustalX and the tree is an unrooted Neighbour-Joining tree based on the alignment. The bootstrap values shown on each node are out of 1000 replicants.

```
elt2        RQGLVCSN--CNGTNTTLWR--RNAEGDPVCNACGLYFKLHHIPRPTSMKKEGA-
elt4        RKRLVCSN--CNGTNTTLWR--RKAEGDPVCNACGLYFKLJHHVTRPIPMKKNKK-
meloid      RQNLICSN--CNGTDTTLWR--RNNNGEPVCNACGLYYKLHNIARPLTMKKETVQ
strongy     RQDMKCNN--CGGSNTTLWR--RNAEGHPVCNACGLYYKLHGIQRPLSMKKDGDT
elt1        RTGIECVN--CRTNTTTLWR--RNGEGHPVCNACGLYFKLHKVERPITMKKDGI-
elt5        PNAARCSN--CRTDKTTAWR--RDAEGKLVCNPCGLYYRLHKVRRPIEMRKNHI-
elt6        SSISKCSN--CSTIKTTAWR--RDLEGKLVCNACGLYYRLHRTHRPVHMRKDFI-
elt7        IRDACCSH--CSTTTTTLWR--KNDEGNLECNACNLYYRHNKVKRPLSLCKQKP-
elt3        HQNSICSN--CKTRETTLWR--RNGEGGVECNACNLYFRKNNRKRPLSLRKDGI-
ancylost    HSNSVCAN--CKTTETTLWR--RAKTGEIECNACNLYFSLNGTA-----------
end3        HINSSCSN--CGCRETKLWR--RNEQGETECNPCNLYERVKGHKRPQHLWNKPA-
end1        HTNSVCSNPNCRTRETTLWR--RTDSGAIECNGCSLYFRKNGIQRPAELCRK---
med1        KKSFQCSN--CSVTETIRWRNIRSKEG-IQCNACFIYQRKYNKTRPVTAVNKYQ-
med2        KKSFQCSN--CSVTETIRWRNIRSKEG-IQCNACFIYQRKYNFTRPVTAVNKYQ-
```
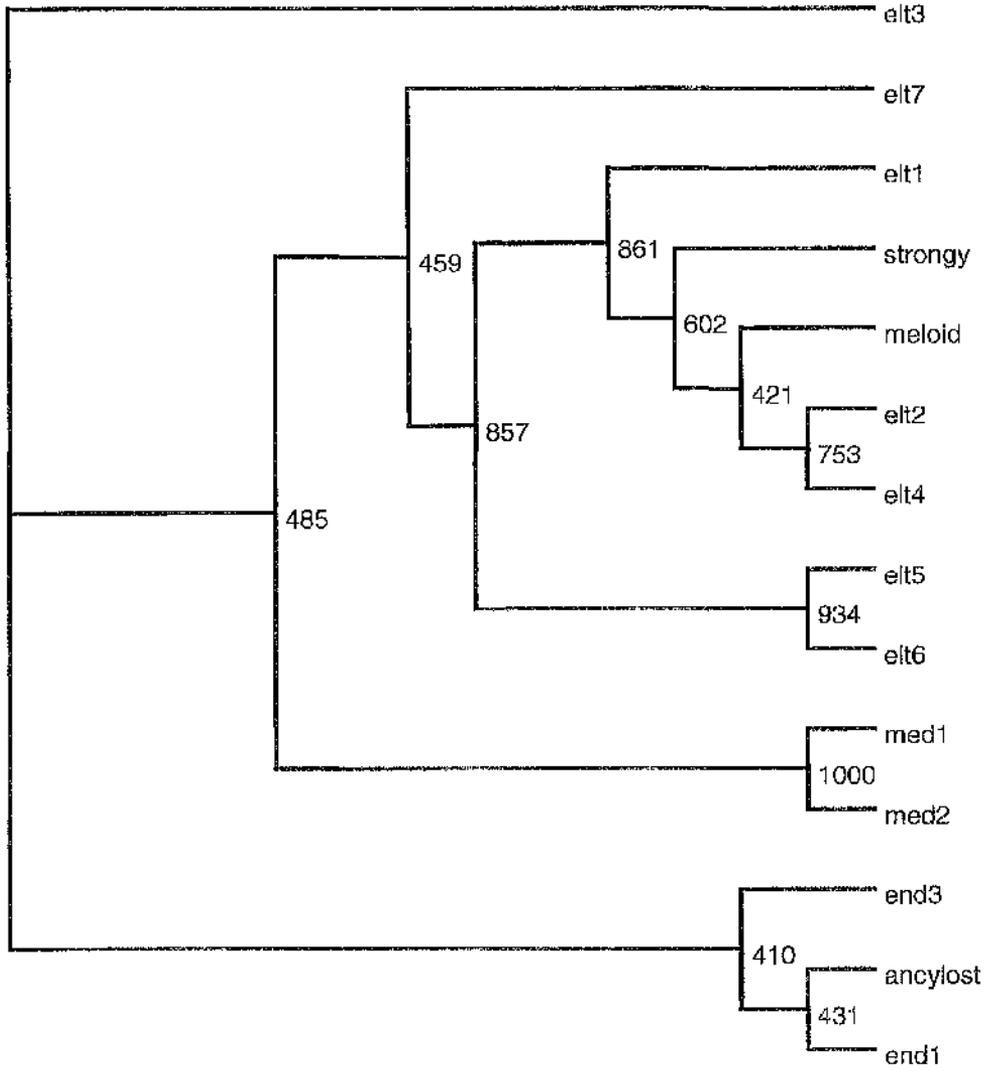
(A)

Primer name	Primer sequence	Protein sequence used for back translation
GA2	[G/A]TAIA[G/A]ICC[G/A]CAIGC[G/A]TT[G/A]CA	**CNACGLY**
GA3	ICC[T/C]TCI[G/C]C[G/A]TTIC[G/T]IC[G/T]CCA	**WRRN(A/G)EG**

(B)

```
ELT-1 CF    CVN--CRTNTTTLWRRNGEGHPV-CNACGLYFKLH
ELT-1 NF    CVN--CGVHNTPLWRRDGSGNYL-CNACGLYFKMN
ELT-2       CSN--CNGTNTTLWRRNAEGDPV-CNACGLYFKLII
ELT-3       CSN--CKTRETTLWRRNGEGGVE-CNACNLYFRKN
ELT-4       CSN--CNGTNTTLWRRKAEGDPV-CNACGLYFKLH
ELT-5       CSN--CRTDKTTAWRRDAEGKLV-CNPCGLYYRLH
ELT-6       CSN--CSTIKTTAWRRDLEGKLV-CNACGLYYRLH
MED-1       CSN--CSVTETIRWRNIRSKEGIQCNACFIYQRKY
MED-2       CSN--CSVTETIRWRNIRSKEGIQCNACFIYQRKY
END-1       CSNPNCRTRETTLWRRTDSGAIE-CNGCSLYFRKN
END-3       CSN--CGCRETKLWRRNEQGETE-CNPCNLYERVK
```

Figure 3.5: Primer sequence for the amplification of *H. contortus* GATA factors.

The table (A), shows the sequence of primers GA2 and GA3 and the amino acid sequence from which they were back-translated. The alignment (B), shows the position of GA2 and GA3 with respect to the DNA-binding domain of each *C. elegans* GATA factor. The three GATA factors most likely to be amplified with these primers are *elt-1*, *elt-2* and *elt-3*.

```
 1   CGGCACGAGGAAGCCTTCAGCAGCTCAAAATTCGCAACGG

41   AGGCAAGGCTTGGTTTGTTCAAACTGCGGTGGAACGAATA
                        C   S   N   C   G   G   T   N
81   caacgttgtggcgcagcaacgccgaggg
      T   T   L
```

Figure 3.6: 100-bp PCR amplicon obtained after a hemi-nested PCR using primers GA3 and T7 on diluted PCR product amplified with primers GA2 and T7.

Part of the translation is shown (bold letters) below the cDNA sequence and the sequence corresponding to the GA3 primer is underlined twice. The CSNC and TTL sequences are characteristic of the C-terminal zinc finger of several *C. elegans* GATA factors.

Figure 3.7: Sequences of the three overlapping cloned PCR products covering the full-length cDNA sequence of a novel *H. contortus* GATA factor.

Panel 1 shows the sequence of insert pBC12, panel 2 shows the sequence of the amplicon haem31010 and panel 3 shows the sequence of insert pY46. The sequence of pBC12 overlapping haem31010 is in red, the sequence of pY46 overlapping haem31010 is in blue and the sequence common to all three inserts is shown in purple.

```
  1 gaattcggtt taattaccca agtttgagag gtggagcaga cggctggaag ctgtgattct
 61 cgaaaagaaa gtggagttac tgtgttcaaa tggagaattc tcactgatag ccagccgttg
121 gcctattcag cagctgtgga ctgactgtct gacagacctc gcatgaccct tcaacttccc
181 actcagcctg ctgaaaagga agaaaaagat gtgcaacagt taactcgtat gaaagtcgaa
241 cccgatctac ctccgtcttt ccgacaatcc gtcgtcattg aaccgcagcc agcagcagtt
301 gatgtacgac cgaattattc cattatttcc ggatatccgc ctccaccgct tgaaactctg
361 tcgccgtaca gtgtagaaag tataactggt atggggatac cgtcatacac caataattat
421 atctatagct atcaacaacc taatctcaat ttccagccag ctcagtcttc ggctataaat
481 ggaaatatca tcttccatcc ttatccatca acgactactc attatataga caccacaaat
541 aatacatttg atctcaatca gcactatttc ccacctgcgc cttcattgga atgcttgaaa
601 tgtggcgaat catgcggtga caatcactca ataactggcg gatatgtatg tgacaactgt
661 tatcttggaa tggagaagcc aacatatact ttatatcctc ccacatcgat ggagccccgg
721 cataccgttg aagcgatccc cgtctctaag tcaagaggcg gctcatcaca taagaaacct
781 tcagcagctc aaaattcgca acggaggcaa ggcttggttt gttcaaactg cggtgaacga
841 atacaacgtt g
```

(1) pBC12 insert: 5' end of GATA factor

```
  1 cggcacgagg aagccttcag cagctcaaaa ttcgcaacgg aggcaaggct tggtttgttc
 61 aaactgcggt ggaacgaata caacgttgtg gcgcagcaac gccgaggg
```

(2) haem31010: 100 bp amplicon

```
  1 ggaggcaagg cttggtttgt tcaaactgcg gtggaacgaa tacaacgttg tggcggagaa
 61 atgccgaggg agaaccagtt tgcaacgcct gtgggttata ttacaagctt cacaatgttc
121 aacggccacc aacaatgaag aaggacggac agctacaaac caggaaacga aaagcaaaat
181 cggatggaac caatactggt aagaagcggg accggtctag taattatacc cagtcaaccc
241 aagcgattcc agatcggagt accactacgt atcagccagc attcagttcg atcggcttct
301 ccaatcaaat agatagcagt tattctcaaa tgaacggcta ttcctcctcc tggcagagct
361 ctggtggtct cggcacatca ggctccgact ccagccactc atccgctttc cattccacct
421 atccctcacc gtatgcgcca tcaatcccga aaatatccga ggctcgacat atgggcgatg
481 atgagacgca agccgccacc cagggtttgg tagaagatag tgctgaaaga tcaccacatg
541 tttgaccagt agaaaggaaa agaggaagtt cttacaaaaa tggccacgct tggcaagctc
601 agacgatacc atgatgtttg aaatctaccc ctataccatt gaaatgtgtc atgatgaata
661 attttattt cctttattaa tggcatttcg aatacctgtt tgtgcctgga tttttattga
721 tgtgtttatt cgggttgccg attacttttt ggtaaacgaa taaagatttt tagcaaaaaa
781 gaaaaaaaaa aa
```

(3) pY4.6 insert: 3' end of GATA factor

Figure3.8: Alignment of the *Hc-elt-2* sequence assembled from the three overlapping fragments pBC12, pY46 and Haem30100 (*Hcelt-2*A) and the sequence from pACELT2 insert (*Hcelt-2*B).

The ATG is highlighted in black, a sequencing error for *Hcelt-2*A had transformed the ATG in GTG. The poly-A tail is highlighted in grey. Discrepencies in the sequence are shown in white font over darker gray background.

```
Hcelt-2A  gaattcggtt taattaccca agtttgagag gtggagcaga cggctggaag ctgtgattct

Hcelt-2A  cgaaaagaaa gtggagttac tgtgttcaag tggagaattc tcacgatag ccagccgttg
Hcelt-2B                                  a tggagaattc tcacgatag ccagccgttg

Hcelt-2A  gcctattcag cagctgtgga ctgactgtct gacagacctc gcatgaccct tcaacttccc
Hcelt-2B  gcctattcag cagctgtgga ctgactgtcc gacagacctc gcatgaccct tcaacttccc

Hcelt-2A  actcagcctg ctgaaaagga agaaaaagat gtgcaacagt taactcgtat gaaagtcgaa
Hcelt-2B  actcagcctg ctgaaaagga agaaaaagac gtacaacagt taactcgtat gaaagtcgaa

Hcelt-2A  cccgatctac ctccgtcttt ccgacaatcc gtcgtcattg aaccgcagcc agcagcagtt
Hcelt-2B  cccgatctac ctccgtcttt ccggcaatcc gtcgtcattg aaccgcagcc agcagcagtt

Hcelt-2A  gatgtacgac cgaattattc cattatttcc ggatatccgc ctccaccgct tgaaactctg
Hcelt-2B  gatgtacgac cgaattattc cattatttct ggatatccgc ctccaccact tgaaactctg

Hcelt-2A  tcgccgtaca gtgtagaaag tataactggt atggggatac cgtcatacac caataattat
Hcelt-2B  tcgccgtaca gtgtagaaag tataactggt atggggatac cgtcatacac aaataattac

Hcelt-2A  atctatagct atcaacaacc taatctcaat ttccagccag ctcagtcttc ggctataaat
Hcelt-2B  atctatagct atcaacaacc taacctcaat ttccagccag cccagtcttc ggctatcaat

Hcelt-2A  ggaaatatca tcttccatcc ttatccatca acgactactc attatataga caccacaaat
Hcelt-2B  ggaaatatca tctttcatcc ttatccatca acgactactc attatataga caccacaaat

Hcelt-2A  aatacatttg atctcaatca gcactatttc ccacctgcgc cttcattgga atgcttgaaa
Hcelt-2B  aatacatttg atctcaatca gcactattt ccacctgcac cttcattgga atgcttgaaa

Hcelt-2A  tgtggcgaat catgcggtga caatcactca ataactggcg gatatgtatg tgacaactgt
Hcelt-2B  tgtggcgaat catgcggtga caatcactca ataactggcg gatatgtatg tgacaactgt

Hcelt-2A  tatcttggaa tggagaagcc aacatatact ttatatcctc ccacatcgat ggagccccgg
Hcelt-2B  tatcttggaa tggacaagcc tacatatact ttatatcctc ccacatcgat ggagcccgg

Hcelt-2A  cataccgttg aagcgatccc cgtctctaag tcaagaggcg gctcatcaca taagaaacct
Hcelt-2B  cataccgttg aagcgatccc cgtatccaaa tcaagaagcg gctcatcaca taagaaacct

Hcelt-2A  tcagcagctc aaaattcgca acggaggcaa ggcttggttt gttcaaactg cggtggaacg
Hcelt-2B  tccgcagcac aaaattcgca acggaggcaa ggtttggttt gttcaaactg cggtggaaca

Hcelt-2A  aatacaacgt tgtggcggag aaatgccgag ggagaaccag tttgcaacgc ctgtgggtta
Hcelt-2B  aatacaacat tgtggcggag aaacgccgag ggagaaccgg tttgcaacgc ttgtgggtta

Hcelt-2A  tattacaagc ttcacaatgt tcaacggcca ccaacaatga agaaggacgg acagctacaa
Hcelt-2B  tattacaagc ttcacaatgt tcaacggcca ccaacaatga agaaggacgg acagctacaa

Hcelt-2A  accaggaaac gaaaagcaaa atcggatgga accaatactg gtaagaagcg ggaccggtct
Hcelt-2B  accaggaagc gaaaagcaaa atcggatgga acgaatactg gtaagaagcg agaccggtct

Hcelt-2A  agtaattata cccagtcaac ccaagcgatt ccagatcgga gtaccactac gtatcagcca
Hcelt-2B  agtaattata cccagtcaac acaagcgatt ccagatcgga gtaccactac gtatcagcca

Hcelt-2A  gcattcagtt cgatcggctt ctccaatcaa atagatagca gttattctca aatgaacggc
```

```
Hcelt-2B   gcattcagtt cgatcggctt ctccaatcaa atagatagca gttattctca aatgaacggc

Hcelt-2A   tattcctcct cctggcagag ctccggtggt ctcggcacat caggctccga ctccagccac
Hcelt-2B   tattcctcct cctggcagag ctccggtggt ctcggcacat caggctcaga ctccagccac

Hcelt-2A   tcatccgctt tccattccac ctatccctca ccgtatgcgc catcaatccc gaaaatatcc
Hcelt-2B   tcgtctgctt tccattccac ctatccctca ccgtatgcgc catcaatccc gaaaatatcc

Hcelt-2A   gaggctcgac atatgggcga tgatgagacg caagccgcca cccagggttt ggtagaagat
Hcelt-2B   gaggctagac atatgggcga tgatgagacg caagccgcca cccagggttt ggtagaagat

Hcelt-2A   agtgctgaaa gatcaccaca tgtttgacca gtagaaagga aaagaggaag ttcttacaaa
Hcelt-2B   agtgctgaaa gatcaccaca tgtttgacca gtagaaagaa aatcaggaag ttcttagaaa

Hcelt-2A   aatggccacg cttggcaagc tcagacgata ccatgatgtt tgaaatctac ccctatacca
Hcelt-2B   aatagccacg ctcggcaagc tcagacgata ccatgatgtt tgaaatctac ccctatacca

Hcelt-2A   ttgaaatgtg tcatgatgaa taattttat ttcctttatt aatggcattt cgaatacctg
Hcelt-2B   ttgaaatgtg tcatgatgaa ttatttttat ttttctttga aatggcattt cgaataactc

Hcelt-2A   tttgtgcctg gatttttatt gatgtgttta ttcgggttgc cgattacttt ttggtaaacg
Hcelt-2B   tttgtgcctg gatttttatt gattgtttta ttcgggttgc cgattacttt ttggtaaacg

Hcelt-2A   aataaagatt tttagcaaaa aagaaaaaaa aaaa
Hcelt-2B   aataaagatt tttagcaaaa
```

Figure 3.9: Alignment of the Hc-ELT-2 and Ce-ELT-2 polypeptides

The consensus sequence is shown in blue and regions of conservation are highlighted in grey. The zinc finger domain, (double underline) and basic domain (boxed) form the DNA-binding domain, which is the most conserved region, followed by the pseudo finger domain (single underline). Finally two short stretches of amino acids at either end of the protein are highly conserved.

CEELT-2 MDNNNDVNGWEMES QPMGGERLPIQMDPEQNFSQISEFPRMKIINDVASPIERQSVTISG TNN
HCELT-2 MENSIDSGHAYSAAWTISDRERMILQLPTQPAE KEEKDVQQLTRMKVEPDLPPSFRQSVVIEPQPAVDVRENYSITSGY S
         MEN H N AWA S P LLPIQ D E L L RMKID VI

CEELT-2 YEPKVEIVISFFHIGIDMSNFGMLDQTIMQPFYPLMSGLPVNTLGIFSG YINSIYDKRSLMDPSIPTTNIPSTYPIVAPTYE
HCELT-2 PPPIPETLSPYSVESITGVEIRSYINMVTYSYQQPNINFQPAQSSAINQNITFHPYPSITTHYIDITINIFXINQHYFPPARSLE
         P LEILS F I   N I F   G F   D S T I Y ARS E

CEELT-2 CVKCSQSCGFAGMKANGEMCVNCSIPKCITYSPPVAYSISIGQPPTIETPSEQPTAKTAKQSKKSSSNRGHGSASRRQEMC RRQEMC
HCELT-2 CLKCGFSGGD NSFTIGGVVDNCYLGMVDKPFYXLYPPTSMEPRHVEA IPVSKSRGGSSHIKPSAQNSQ SN RRQEMC
         CLKC SCG   AI GG MC NC   S L TSL LE P AK SS K SA SN RRQEMC
         =

CEELT-2 SNDNGINITLWRRNAEGHDPVCNAG GHYPKLHHIPRPTISMKKEGALQTIRRKSKSG DSSTPSTISRAPERKFEPASSSTEKAQRSSN
HCELT-2 SNGGINITLWRRNAEGHPAVDNAG GLYKICHNVQRPPIMKKDG QLQTIRRKAKSIGINIG KRPDRSSNMTQSITQALPD
         SNC GINITLWRRNAEGHDPVQNAG GHYPKLH I RP SVKKDG LQIRRKRAKS S T   KK DRAS T Q

CEELT-2 RRAGSAKADRELSTPAVAAATATMVSFADLYPVSSAAVILPDTYSNYYQANTAATAGTMMVENLQNVVZAAINQJGLRPADNI
HCELT-2 R SITTYQPAFRSSIGFSNQIIDSYSGMVGYSSSMQSFFGLGTISGSDSHSSAFHSI YPSPYAPSIPK
         R   AT TY   S S Q   WNSAA G   F AT Y S   PA

CEELT-2 QVHMPVQDDEIKAAARDLEAVDGDS
HCELT-2 ISFEARHGLDEIQAATQEEMEDSAEXSHV
         M DDET AA L AD

112

Figure 3.10: Neighbour-Joining tree showing the relationship of the novel *H. contortus* GATA transcription factor to each of *C. elegans* GATA transcription factors.

The putative DNA binding domain of the *H. contortus* GATA factor was aligned with of the eleven *C. elegans* GATA factors using ClustalX. A Neighbour-Joining tree was constructed and the boot-strap values shown are out of 1000 replicates. The GATA zinc finger domain is underlined.

```
end1       HTNSVCSNPNCRTRETTLWR--RTDSGAIECNGCSLYFRKNGIQRPAELCRK---
end3       HINSSCSN--CGCRETKLWR--RNEQGETECNPCNLYERVKGHKRPQHLWNKPA-
elt2       RQGLVCSN--CNGTNTTLWR--RNAEGDPVCNACGLYFKLHHIPRPTSMKKEGA-
elt4       RKRLVCSN--CNGTNTTLWR--RKAEGDPVCNACGLYFKLHHVTRPIPMKKNKK-
haemGATA   RQGLVCSN--CGGTNTTLWR--RNAEGEPVCNACGLYYKLHNVQRPPTMKKDGQL
elt1       RTGIECVN--CRTNTTTLWR--RNGEGHPVCNACGLYFKLHKVERPITMKKDGI-
elt5       PNAARCSN--CRTDKTTAWR--RDAEGKLVCNPCGLYYRLHKVRRPIEMRKNHI-
elt6       SSISKCSN--CSTIKTTAWR--RDLEGKLVCNACGLYYRLHRTHRPVHMRKDFI-
elt3       HQNSICSN--CKTRETTLWR--RNGEGGVECNACNLYFRKNNRKRPLSLRKDGI-
elt7       IRDACCSH--CSTTTTTLWR--KNDEGNLECNACNLYYRHNKVKRPLSLCKQKP-
med1       KKSFQCSN--CSVTETIRWRNIRSKEG-IQCNACFIYQRKYNKTRPVTAVNKYQ-
med2       KKSFQCSN--CSVTETIRWRNIRSKEG-IQCNACFIYQRKYNETRPVTAVNKYQ
```
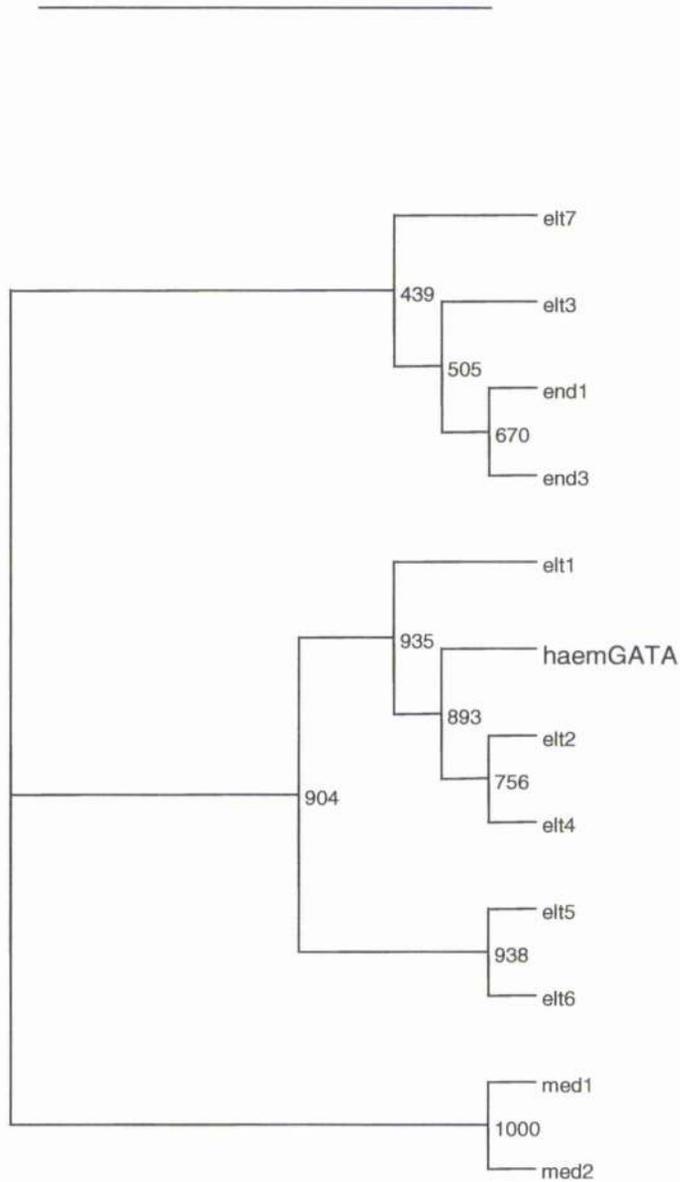
Figure 3.11: Alignment of Hc-ELT-2, Ce-ELT-2, Cb-ELT-2, Ce-ELT-4, yeast ArcA and the synthetic peptide F2B.

This figures shows the conserved residues between all four peptides in the C-terminal zinc finger region. Highlighted in blue are the residues conserved in three or more peptides and in purple the residues from HCELT-2 that are conserved in the other peptides. Residues involved in maintaining the three-dimensional structure of the DNA-binding domain are indicated with open circles and residues involved in DNA contact are indicated with solid circles. The underlined residues represent the highly conserved α-helix that inserts into the DNA major groove. The roles of these residues was established using the peptide F2B {Omichinski, 1993 #259}, and as indicated by the shaded groups of amino acids are very conserved between all five represented peptides. The last line indicates the consensus between the six DNA-binding domain.

```
CeELT-2    MDNNYNDN-----LVCSNCNGTNTTLWRRNAEGDPVCNACGLYFKLHHIPRPTSMKKEGALQTRKRKSKS-GD-

CeELT-4    MDNNYLDASHRKRLVCSNCNGTNTTLWRRKAEGDPVCNACGLYFKLHHVPRPIPMKKNKKHAVLPAPGISKL--

CbELT-2    MDTTYQEN-----LVCSNONGTNTTLWRRNAEGDPVCNACGLYFKLHHIARPTSMKKEGALQTRKRKSKT-GE-

HcELT-2    MENSH-DSQ----LVCSNCGGTNTTLWRRNAEGEPVCNACGLYYKLHNVQRPPTMKKDGQLQTRKRKAKSDG—

ARE  A     ----MKNGEQNGPTTCTNCFTGTIPLWRRNPEGQPLCNACGLFLKIHGVVRPLSLKTDVIKKRNRNSANS----

F2B        ------KRAGTVCSNCQTSTTLWRRSPMGDPVCNACGLYYKLHQVNRPLTMRKDGIQTRNRKVS------

Consensus  ---------CSNC----T-LWRR---G-P-CNACGLY-K-H---RP----
```

MENSHDSQPLAYSAAWTTSDRPRMTLQLPTQPAEKEEKDVQQLTRMKV

EPDLPPSFRQSVVIEPQPAAVDVRPNYSIISGYPPPPLETLSPYSVES

ITGMGIPSYTNNYIYSYQQPNLNFQPAQSSAINGNIIFHPYPSTTTHY

IDTTNNTFDLNQHYFPPAPSLECLKCGESCGDNHSITGGYVCDNCYLG

MDKPTYTLYPPTSMEPRHTVEAIPVSKSRSGSSHKKPSAAQNSQRRQG

LVCSNCGGTNTTLWRRNAEGEPVCNACGLYYKLHNVQRPPTMKKDGQL

QTRKRKAKSDGTNTGKKRDRSSNYTQSTQAIPDRSTTTYQPAFSSIGF

SNQIDSSYSQMNGYSSSWQSSGGLGTSGSDSSHSSAFHSTYPSPYAPS

IPKISEARHMGDDETQAATQGLVEDSAERSPHV

Figure 3.12: Nuclear localisation signals in Hc-ELT-2.

The degenerate zinc finger is highlighted in blue and the GATA zinc finger in yellow. The nuclear localisation signals (NLS) are circled in red. The NLS were identified as described by Christophe and colleagues (2000).

Figure 3.13: Expression and purification of Hc-ELT-2 recombinant peptide HCELTHISA on a Ni-NTA column.

Panel a shows the position of the polypeptides in respects to the Hc-ELT-2 protein. The recombinant polypeptides HCELT2HISA and HCELT2HISB contain a 6xHis tag and so were purified on Ni-Agarose columns. Panel B shows a Coomassie stained 12.5 % SDS-PAGE gel of bacterial lysates and column eluates of the purification of polypetide HCELT2HISA. Bacterial cells were induced by 1 mM IPTG after 2 hours of culture. The extracts of whole cell lysates were taken immediately before induction (pre-induced) and after 4 hours of culture after induction (post-induced). The column was eluted with three different buffers (D, E and E_1, QIAExpressionist, QIAgen). Samples of each 1 ml elution fraction are shown on the gel.

(A)

HCELT2HISA   HCELT2HISB
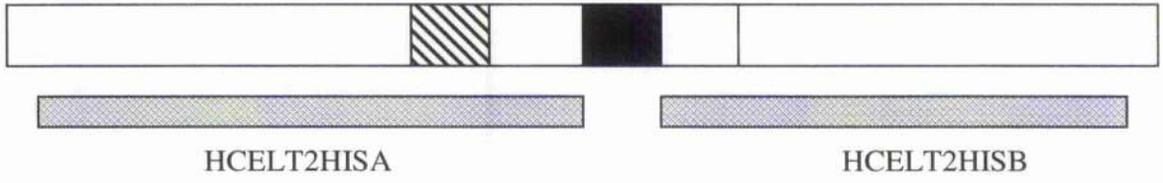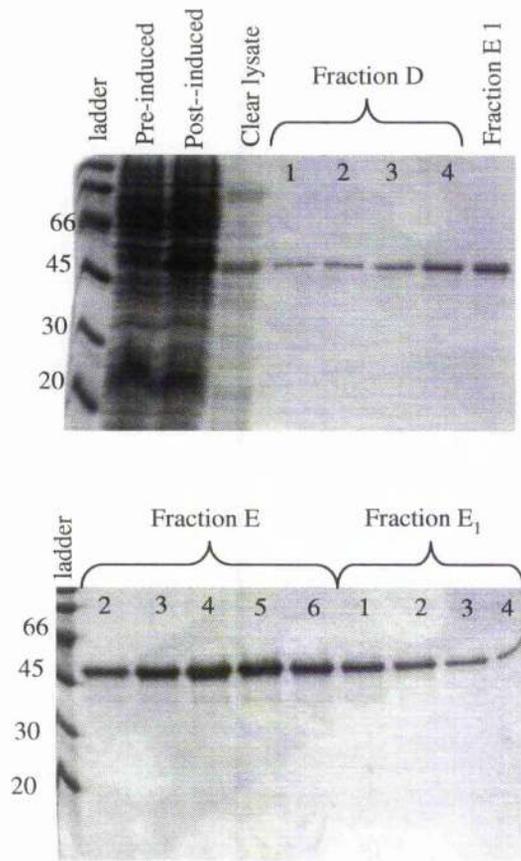
(B)

Figure 3.14: Western blots of *Haemonchus contortus* worm homogenate (WH) and fusion protein (FP) probed with antisera from immunised rabbits

Rabbits 1077 and 1082 were immunised with fusion protein HCELT2HISB and rabbits 1064 and 1104 with fusion proteins HCELT2HISA. Bleeds were taken at 28, 53 and 79 days after the primary immunisation.

WH= worm homogenate and FP= fusion protein.

In panel A, the fusion protein (FP) is HCELT2HISBand in panel B, the fusion protein (FP) is HCELT2HISA.

The antisera from rabbits R1077, R1064 and R1104 all show strong immunoreactivity to the appropriate fusion protein and the antiserum from rabbit 1082 showed weak immunoreactivity. None of the antisera showed strong immunoreactivity to worm homogenate (WH) on these blots

(A)

	Pre-immune 1/200	First bleed 1/200	Second bleed 1/200	Third bleed 1/200	Third bleed 1/1000
	WH FP	WH FP	WH FP	FP	FP

Rabbit
1077

	Pre-immune 1/200	First bleed 1/200	Second bleed 1/200	Third bleed 1/200	Third bleed 1/1000
	FP WH	FP WH	FP WH	FP	FP

Rabbit
1082

(B)

	Pre-immune 1/200	First bleed 1/200	Second bleed 1/200	Third bleed 1/200	Third bleed 1/1000
	FP WH	FP WH	FP WH	FP	FP

Rabbit 1064

	Pre-immune 1/200	First bleed 1/200	Second bleed 1/200	Third bleed 1/200	Third bleed 1/1000
	FP WH	WH FP	FP WH	FP	FP

Rabbit 1104

117

Figure 3.15:Staining of an early *H. contortus* embryos Hc-ELT-2 the specific polyclonal antiserum R1064 (A) and DAPI (B).

The embryo contains approximately 30-40 cells based on the DAPI staining.

Figure 3.16: Immuno-localisation of He-ELT-2 in *Haemonchus contortus* embryonic (A-F) and post-embryonic (G-I) stages with polyclonal rabbit anti-He-ELT-2 antibodies.

Specimens were stained with rabbit polyclonal antisera R1064. Panel A: 4E-cell stage of development; Panel B: early embryo staining in 8 nuclei; Panel C: lima bean embryo; Panel D: early comma stage embryo; Panel E: 1-fold embryo; Panel F: pretzel stage embryo; Panel G: L2 stage larva; Panel H: L3 stage larvae; Panel I; adult gut specimen.
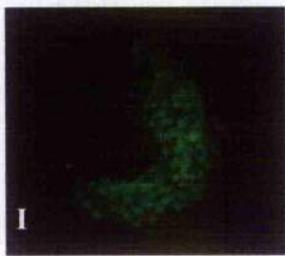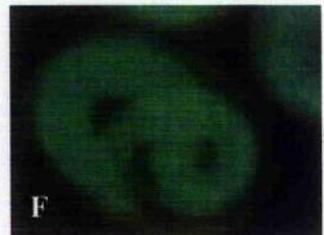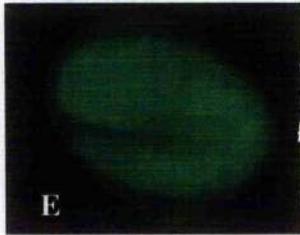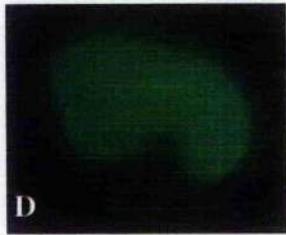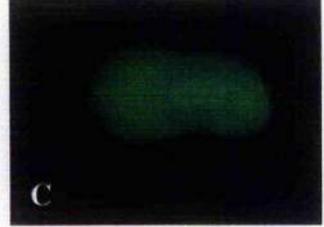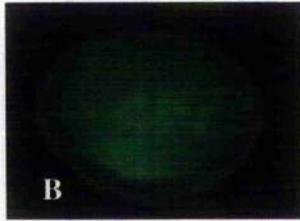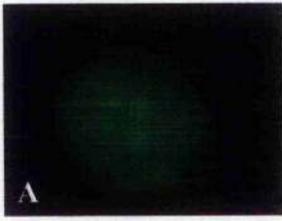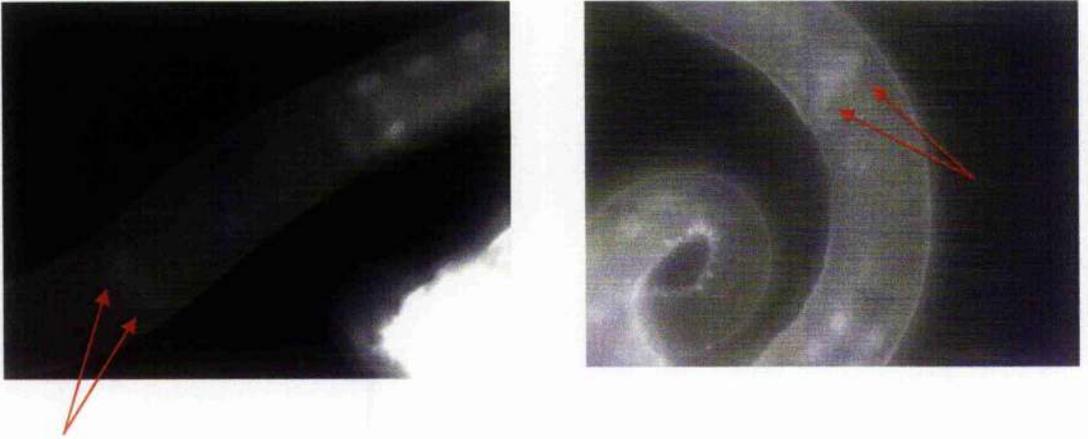
Figure 3.17: Staining of L1 *H. contortus* larvae with polyclonal antiserum R1064.

Staining of two different larvae is shown. In addition to the staining of endodermal nuclei, faint staining of two nuclei adjacent to the pharyngeal bulb can be seen.
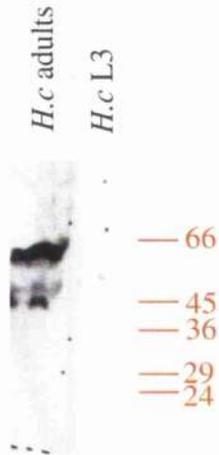
Figure 3.18: Western blot analysis of Hc-ELT-2 expression in *H. contortus* adult and L3 stages

The blot containing adult and L3 extracts was probed with R1064 antiserum Hc-ELT-2 is expressed in adults but not in L3-stage larvae.

| Dauer larvae | L3 larval stage |
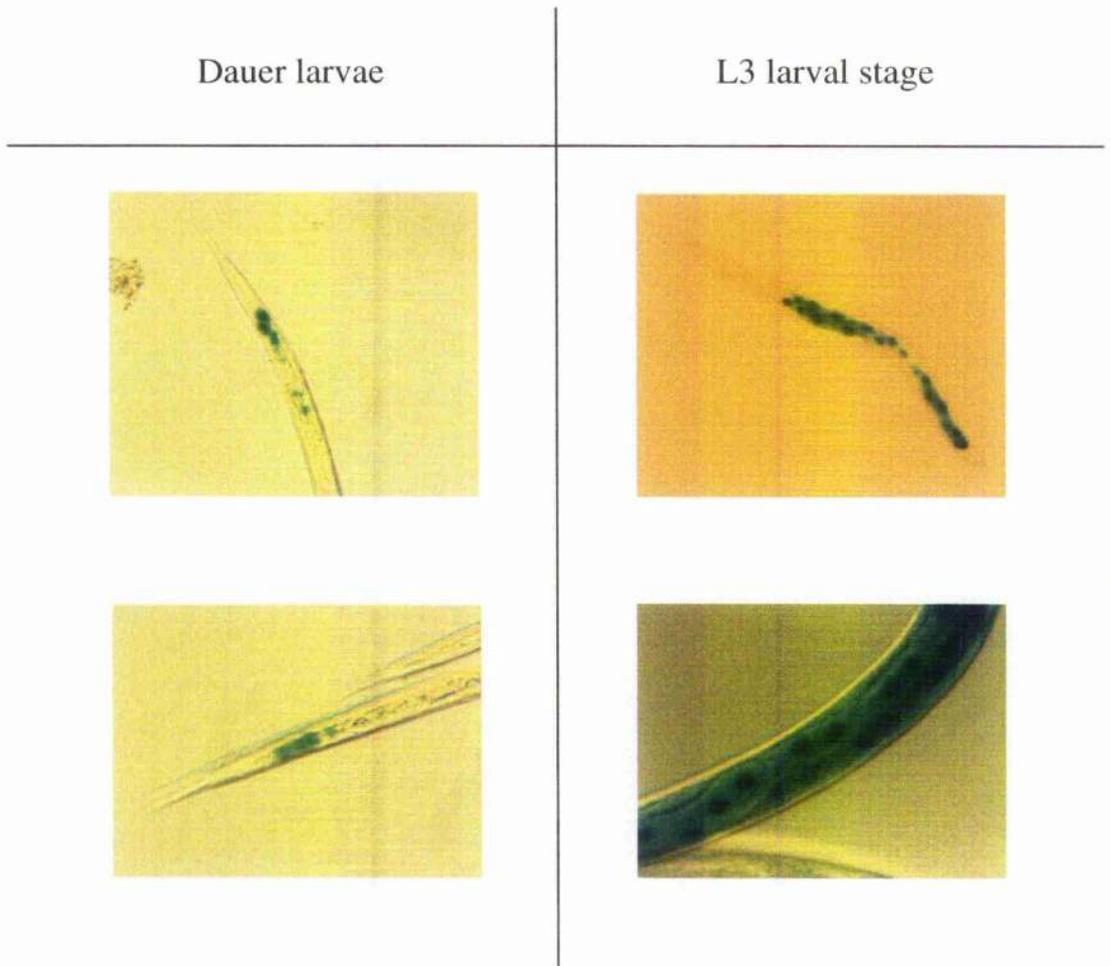
Figure 3.19: Expression of *C. elegans elt-2::lacZ* in L3 larval and dauer stages of *C. elegans*.

X-Gal staining of dauer and L3 larvae of the JM62 strain is localised to the endodermal nuclei in both cases, however the intensity of staining is less intense in the dauer stages.

Chapter 4: Functional analysis of the *Haemonchus contortus elt-2* gene using *C. elegans* as a heterologous expression system

4.1 Introduction

There are a number of ways in which *C. elegans* can potentially be used as a surrogate system in which to assay parasite gene expression and to test functional conservation. The most widely adopted approach used to study gene function in *C. elegans* is to rescue a mutant phenotype (Zhu, J. *et al.* 1997; Fukushige, T. *et al.* 1998). This is a very stringent test of gene function since a transgene needs to reproduce most, if not all, aspects of both gene function and regulation in order to achieve phenotypic rescue. This approach has already been successfully used to study parasite gene function: functional conservation of the *H. contortus* β-tubulin isotype I gene (*tub-1*) was demonstrated by transgenic rescue of a mutation in the *C. elegans ben-1* gene (Kwa, M. S. G. *et al.* 1995). However, phenotypic rescue experiments can be very variable in terms of efficiency and are often very sensitive to copy number. Indeed, the transgenic rescue of the *C. elegans elt-2* mutant phenotype using a wild type copy of the gene is reported to be very inefficient (J. D. McGhee, personal communication). Consequently, an alternative approach was considered to be desirable in order to investigate the function of the *H. contortus elt-2* gene using *C. elegans* as a surrogate expression system. Ectopic expression approaches have been widely used to study several of the *C. elegans* GATA transcription factors, including *elt-2,* and so it was decided to develop this approach to investigate the function of the *H. contortus elt-2* gene.

Ectopic expression is the forced expression of a gene in a spatial or temporal context that is different to its endogenous expression. This is an extremely powerful

approach to study gene function and is widely used to study regulatory molecules such as transcription factors and components of signal transduction pathways in *C. elegans* (Fukushige, T. *et al.* 1998; Zhu, J. *et al.* 1998; Gilleard, J. S. *et al.* 2001; Fukushige, T. *et al.* 2003; Uchida, O. *et al.* 2003). The ectopic expression of such molecules can result in cell fate changes or abnormal development of the organism which can yield important information regarding developmental function. Furthermore the activation, or repression, of specific downstream genes can be directly examined. The work of Haun and colleagues (1998) provides a nice example of this approach; ectopic expression was used to demonstrate the ability of the zebrafish homeobox gene *nkx2.5* to activate the expression of a pharyngeal muscle-specific gene in *C. elegans* (Haun, C. *et al.* 1998). The zebrafish *nkx2.5* gene was ectopically expressed in the body wall of transgenic *C. elegans* using the body wall muscle-specific *unc-54* promoter. This resulted in the expression of *myo-2* in the body wall muscle of transgenic worms. Since *myo-2* is normally only expressed in pharyngeal muscle, this demonstrated that the zebrafish *nkx2.5* could activate *myo-2* in *C. elegans* in a similar way to *ceh-22*, a *C. elegans* homeobox gene involved in pharyngeal muscle development. Since *nkx2.5* is involved in zebrafish cardiac muscle development, the authors suggested that this provided evidence of an evolutionary relationship between nematode pharyngeal muscle and vertebrate cardiac muscle development.

Ectopic expression has been widely used in a range of model organisms (e.g. sea urchin, *Drosophila melanogaster*, *C. elegans*), particularly during early embryonic development. A wide range of methodologies are used from the direct injection of mRNA in sea urchin embryos (Gross, J. M. *et al.* 2003; Kenny, A. P. *et al.* 2003) to more sophisticated systems allowing temporal and spatial specific

expression in *Drosophila* (Kosman, D. *et al.* 1997; Argiropoulos, B. *et al.* 2003). The ideal system is one which allows the inducible expression of the gene of interest in a range of different tissues. In the case of genes that have essential regulatory functions, inducible expression can be important since widespread ectopic expression can often result in lethality. In *C. elegans*, a set of vectors is available that allow the ectopic expression of genes using a number of constitutively active tissue specific promoters (Fire, A. *et al.* 1990; Fire, A. 1995). However, the only current approach to inducible ectopic expression in *C. elegans* is to use vectors utilising *C. elegans* heat-shock promoters to regulate the expression of the gene of interest. The promoters from two different heat-shock genes have been used for this purpose: *hsp16-2* and *hsp16-41*. The activity of these promoters has been studied in detail by examining their ability to direct the expression of lacZ reporter genes in transgenic *C. elegans* (Stringham, E. G. *et al.* 1992). It was found that both these promoters were inactive at 25°C but active above 30°C. Both promoters drove lacZ expression in a wide range of tissues but showed some differences in tissue specificity. The *hsp16-2* promoter predominantly drove expression of beta-galactosidase in muscle and hypodermis, whereas the *hsp-16-41* promoter was more efficient in intestine and pharyngeal tissue (Stringham, E. G. *et al.* 1992). Two vectors have been specifically designed to allow these promoters to be used to drive inducible ectopic expression of any gene of interest in *C. elegans*: pPD49-78 and pPD49-83 (Fire, A. *et al.* 1990; Fire, A. 1995). pPD49-78 contains the *hsp16-2* promoter and pPD49-83 contains the *hsp16-41* promoter. The approach to using these vectors is to clone the coding sequence of the gene of interest into the polylinker downstream of the heat shock promoter. The resulting plasmid is then transformed into *C. elegans* by microinjection to produce transgenic lines carrying the ectopic expression vector on

extrachromosomal arrays. These lines can be used directly or alternatively the array can be chromosomally integrated using gamma irradiation. The resulting transgenic lines then allow expression of the gene of interest in response to heat shock in the pattern of tissues described above (Stringham, E. G. *et al.* 1992). Integrated transgenic lines are labour-intensive to produce but allow stable transmission of the array since it is then part of the animal's genome. This allows for comparison between independent lines and the expression of the array is also very consistent between animals and experiments. In contrast, lines carrying a non-integrated extrachromosomal array can show some variation in their transmission within a line between experiments and several independent lines will not necessarily have the same transmission rate which can make comparisons between lines difficult.

Ectopic expression using the *C. elegans* heat shock promoters has been used to study several *C. elegans* GATA transcription factors: *elt-2* (Fukushige, T. *et al.* 1998), *end-1* (Zhu, J. *et al.* 1998), *elt-1*, *elt-3* (Gilleard, J. S. *et al.* 2001), and more recently, *elt-4* (Fukushige, T. *et al.* 2003). Ectopic expression of all of these genes during early embryogenesis, except for *elt-4*, results in embryos arresting as a ball of several hundred cells with no visible signs of morphogenesis. Ectopic expression of *elt-2* and *end-1* results in most blastomeres in the early embryo being transformed to endodermal cell fates, whereas ectopic expression of *elt-1* and *elt-3* results in transformation into hypodermal cell fates. Interestingly, these GATA transcription factors are the only *C. elegans* genes to date that have been shown to have this capability; i.e. transforming the fates of multiple early blastomeres to a single cell type. These GATA transcription factors only appear capable of doing this during early embryogenesis, whilst the embryonic blastomeres are still pluripotent (Zhu, J. *et al.* 1998; Gilleard, J. S. *et al.* 2001).

This chapter describes the use of this approach to examine the functional conservation between *Hc-elt-2* and *Ce-elt-2* and to explore the use of *C. elegans* as a heterologous *in vivo* system to study the parasite gene function. Specifically, I set out to examine the ability of *Hc-elt-2* to activate a program of gut cell differentiation when ectopically expressed in early *C. elegans* embryos. The construct pPD49-78 was chosen for these experiments because I wished to ectopically express the *Hc-elt-2* gene in many tissues outside the gut. The *hsp-16-2* promoter is the most suitable one to achieve this since it is most active in hypodermal and muscle cells whereas the *hsp-16-1* promoter in vector pPD49-83 is most active in gut cells (Stringham, E. G. *et al.* 1992). *Hc-elt-2* cDNA was cloned in the vector pPD49-78 in two different ways, one in which a synthetic intron was included and another without, as described in section 24.2.1.1 (Fire, A. *et al.* 1990; Stringham, E. G. *et al.* 1992). Several independent transgenic lines were made carrying these two constructs on extrachromosomal arrays and the effect of ectopic expression of the *H. contortus elt-2* gene was examined.

4.2 Results

4.2.1 Experimental approach to ectopic expression of *Hc-elt-2* in *C. elegans*

4.2.1.1 Cloning *Hc-elt-2* cDNA coding sequence into the ectopic expression vector pPD49-78

The ectopic expression vector used was pPD49-78, part of a set of vectors designed by Dr. A. Fire (Fire, A. 1995). It is derived from the canonical vector

pPD49.26 (figure 4.1). In order to produce pPD49-78 a fragment containing the *hsp-16-2* heat shock promoter (Stringham, E. G. *et al.* 1992) was cloned into multiple cloning site I (MSCI) (Fire, A. 1995). The promoter directs robust somatic expression only at temperature of 32-33.5°C, but some expression can be observed in embryos at 29°C; however this promoter is not active at the normal growing temperature (20-25°C) for *C. elegans*. Hence the pPD49-78 vector provides a method of inducible expression in response to heat-shock.

The pPD49-78 vector contains a synthetic intron sequence which is believed to enhance the levels of expression since it has been shown that the presence of introns often increases the level of transgene expression (Brinster, R. L. *et al.* 1988; Buchman, A. R. *et al.* 1988). Indeed, detailed experiments have shown that the expression levels of reporter constructs increases with the number of introns included in the coding sequence (Fire, A. *et al.* 1990; Fire, A. 1995). The enhancement of expression seems to be independent of the precise sequence of the introns concerned but seems to reflect the effect of splicing itself. Consequently, most *C. elegans* expression constructs are designed such that there is at least one synthetic intron in the coding sequence. Indeed the standard set of reporter gene vectors contains 12 synthetic introns. However it is also known that heat shock inhibits the splicing process in a number of systems (Yost, H. J. *et al.* 1986; Muhich, M. L. *et al.* 1989; Miriami, E. *et al.* 1994). Hence it was difficult to anticipate the effect of the presence of a synthetic intron on the expression of a transgene that was to be induced by heat shock. Consequently it was decided to make two types of ectopic expression for *Hc-elt-2* construct, one with the synthetic intron and one without, and to use these in parallel experiments.

Primers (appendix 2) were designed to amplify the *H. contortus elt-2* cDNA sequence from sequence encoding the methionine residue at position +24 to the stop codon and *Sma*I (Ann5), *Nhe*I (Ann7) and *Kpn*I (Ann6) restriction sites were engineered at the 5' end of each primer as appropriate (figure 4.1). At the time of cloning, when the primers were designed and engineered, the only available sequence suggested that the ATG encoding the Methionine at the +24 position was the start codon, and so the ectopic expression constructs encode a slightly truncated version of the full-length polypeptide which lacks the first 23 amino acids of the N-terminus. Construct pAC7 contains this *Hc-elt-2* cDNA fragment cloned into the *Sma*I and *Kpn*I restriction sites of the MCS1 and multiple cloning site II (MCSII) polylinkers of pPD49-78 respectively such that the synthetic intron is removed (Figure 4.1 C). Construct pAC9 contains this *Hc-elt-2* fragment cloned into the *Nhe*I and *Kpn*I restriction sites of the MCSII polylinkers of pPD49-78 such that the synthetic intron is retained (Figure 4.1D).

Transgenic lines were made by injecting the constructs pAC7 and pAC9 into the following *C. elegans* strains to obtain stable transgenic lines: N2, JM86, JM73 (both kindly provided by J. McGhee, Calgary), PD4251, and JR667 (see table 4.1 and appendix 2 for description of strains). Briefly, JM86 carries an integrated reporter construct *ifb-2::GFP* and JM73 carries an integrated reporter construct *elt-2::GFP*, hence both these reporter genes result in GFP expression in the 20 gut cell nuclei of wild type embryos. Strains PD4251 and JR667 carry a *myo-3:GFP* reporter construct and a pMF1::*GFP* reporter construct, which are muscle and seam cell markers, respectively.

4.2.1.2 Detection of ectopically expressed HcELT-2 in transgenic *C. elegans*

Expression of Hc-ELT-2 from the transgenes on extrachromosomal arrays in the different transgenic lines was examined by Western blotting as described in section 2.3.2. Worms from transgenic lines were prepared as described in section 2.3.2 and the resulting protein extracts were separated using 10% SDS-PAGE gels which were then Western blotted and probed with the Hc-ELT-2 specific polyclonal rabbit serum R1064.

The expression of the transgene in four independent transgenic lines was examined. Two lines, AC3 and AC9, contained the ectopic expression construct pAC7 on extrachromosomal arrays and two other lines, AC11 and AC13, contained the construct pAC9 (table 4.1). Plates containing mixed stage worms from each of these lines, as well as plates of N2 controls, were heat shocked at 33.5°C for 2 hours as described in section 2.1.4. Adult worms were then harvested two hours after the end of the heat shock period, transferred directly to SDS-PAGE sample buffer and boiled to produce protein extracts. Control plates which were not subjected to heat shock were set up in parallel for each of the strains. The protein extracts were separated on 10% SDS-PAGE gels with samples from heat shocked worms and non-heat shocked controls loaded side by side. The gels were then Western blotted and probed with the anti-Hc-ELT-2 polyclonal antiserum R1064. A number of intense bands were observed in all tracks including the controls (data not shown). Since the Hc-ELT-2 antiserum was not affinity purified, the presence of cross-reactive antigens from OP50 bacteria present in the gut lumen of worms was suspected. This hypothesis was supported by probing an OP50 bacterial extract on a Western blot with the R1064 Hc-ELT-2 specific antisera and finding a number of intense bands of similar size to those seen in the worm homogenates (data not shown). Consequently

the experiment was repeated with worms that had been starved overnight prior heat-shock to allow residual OP50 in the gut lumen to be digested. This was achieved by picking 50 transgenic worms onto a fresh non-seeded NGM plates and incubating overnight at 20°C prior heat-shock. This to proved effective in removing the cross-reactive bands thought to be of OP50 origin since the R1064 Hc-ELT-2 specific antisera no longer detected any polypeptides in the non-heat shocked and N2 control tracks (figure 4.2A, Tracks 1, 2, 4, 6, 8 and 10). In contrast, a major band of approximately 45 kDa was detected in extracts from each of the four transgenic lines following heat shock (Figure 4.2A, tracks 3, 5, 7 and 9). The polypeptide encoded by the pAC7 and pAC9 transgenes consists of 394 amino acids which would be predicted to have a molecular weight of approximately 40 KDa. Although the 45-kDa polypeptide induced by heat shock in the transgenic lines is slightly larger than this the mobility of polypeptides separated by SDS-PAGE electrophoresis is not simply a function of molecular weight. Polypeptides often show lower mobility than predicted from their molecular weight alone. Indeed, this has been reported for the *C. elegans* ELT-2 polypeptide (Hawkins, M. G. *et al.* 1995). Hence the 45kDa band is consistent with the size of the Hc-ELT-2 polypeptide predicted to be expressed from the pAC7 and pAC9 transgenes. Tracks 1 and 2 of figure 4.2A show this band is not present in extracts from either heat shocked or non-heat shocked N2 worms demonstrating that the anti-Hc-ELT-2 polyclonal antiserum does not cross-react with endogenous *C. elegans* polypeptides including molecules induced by heat shock. Hence this experiment shows that the Hc-ELT-2 polypeptide is expressed is each of the transgenic lines in response to heat shock and it also confirms the specificity of the antiserum. The absence of expression in the non-heat shock controls suggests there is no detectable "leaky" constitutive expression of the transgene in the absence

of heat-shock induction. In this particular experiment it appears that lines AC3 and AC9, which carry the pAC7 transgene (no synthetic intron), exhibit higher expression levels than lines AC11 and AC13, which carry the pAC9 transgene (with synthetic intron). However, on other occasions where similar experiments were performed, the levels of expression were similar in all four lines or sometimes higher in AC11 and AC13. Hence it appears that although there is some variation in the relative levels of expression between experiments there are no consistent differences between each of the four lines. Hence the presence or absence of a synthetic intron in the transgene appears to make no discernable difference in the level expression in these experiments. It is important to emphasise that all the lines used in these experiments are non-integrated and therefore variations in extrachromosomal array transmission may account for the minor differences observed in the Western blots between experiments.

Next, the dynamics of Hc-ELT-2 expression from the transgene in response to heat shock were studied. Mixed stage plates containing worms from lines AC3 and N2 controls were subjected to heat-shock for 2 hours at 33.5°C. Worms were then harvested immediately after the heat-shock (designated 0 hours) and at 2, 4 and 8 hours after the end of the heat-shock period. Extracts were prepared, separated on SDS-PAGE gels, Western blotted and probed with the R1064 Hc-ELT-2 specific antiserum (see section 2.3.2). Figure 4.2B shows that Hc-ELT-2 expression is detected in the AC3 transgenic line following heat shock but not in the heat shocked N2 controls. Expression of the Hc-ELT-2 polypeptide in AC3 is visible immediately after the heat shock period (Figure 4.2B, track 1). The Hc-ELT-2 polypeptide is still detectable at 2 hours and 4 hours after heat shock but appears to be at a lower level than immediately after heat shock. At 8 hours after heat shock expression can no

longer be detected. Hence it appears that the expression of HcELT-2 from the transgene occurs very rapidly after heat shock. However the polypeptide disappears between 4 and 8 hours after expression. This suggests a reasonably high turnover of the Hc-ELT-2 protein which is not unexpected for a transcription factor. For example, the transcription factor IκB, inhibitor of the transcription factor NFκB has a half-life of 5 minutes (Li, X. *et al.* 1999). Another possibility is that the *H. contortus* polypeptide is not as stable within the *C. elegans* cellular environment, since folding and post-translational modifications might be different in *C. elegans*. Indeed there are several examples in the literature where proteins expressed in heterologous system do not undergo appropriate post-transcriptional modifications (Legname, G. *et al.* 1995; Linton, K. J. *et al.* 2002). The Hc-ELT-2 polypeptide expressed from the pAC7 transgene in *C. elegans* was compared to the endogenous polypeptide in *H. contortus* by probing Western blots with the R1064 Hc-ELT-2 specific polyclonal antiserum. It was only possible to obtain sufficient material from *H. contortus* L3 and adult stages for Western blotting. Extracts from *H. contortus* L3 and adult worms and from AC3 mixed stage worms harvested immediately after heat shock were separated by SDS-PAGE and probed with R1064 polyclonal antisera. A polypeptide of approximately 50kDa was detected in adult worms (figure 4.2C track 1), which was not detected by pre-immune serum (data not shown). No polypeptide could be detected in the L3 extracts (figure 4.2C, track 2) in spite of the fact the protein loading was similar to that of the adult worms as detected by Coomassie staining of a duplicate gel and Ponceau S staining of the Western blot. Furthermore subsequent blots with higher loading of L3 extract also failed to detect the Hc-ELT-2 polypeptide. This is consistent with the immunofluorescence assays as described in chapter 3, suggesting Hc-ELT-2 may not be expressed in the L3 stage. The

polypeptide detected in the AC3 extract (figure 4.2.C, track 3) was 45 kDa as in previous experiments. The discrepancy in size between the endogenous Hc-ELT-2 polypeptide and that expressed from the transgene is not unexpected as the *Hc-elt-2* cDNA fragment used in the ectopic expression constructs lacks the sequence encoding the first 23 amino acids of the polypeptide.

In summary, the Hc-ELT-2 polypeptide appears to be expressed at easily detectable levels from transgenes in four independent transgenic lines. This expression is entirely inducible by heat shock and there appears to be no leaky expression from the transgene in the absence of heat shock. The inducible expression appears to be robust and although there is minor variation in the relative levels of expression between experiments it appears to be very repeatable. The onset of expression and the maximum levels of expression occur very rapidly after heat shock suggesting the transgenic lines can be used to activate Hc-ELT-2 expression in a temporally precise manner throughout *C. elegans* development.

4.2.2 Effects of ectopic Hc-ELT-2 expression during early embryogenesis

The details of the ectopic expression protocols are described in section 2.1.4 and the method is shown schematically in figure 4.3. Following heat shock and overnight post-heat shock recovery at 20°C, the embryos were examined with the dissecting microscope and the number of arrested and developed embryos counted. They were then mounted on 2% wet agarose pads for examination of markers under UV illumination or Normaski optics. The number of nuclei expressing GFP in each embryo was scored when appropriate.

In the first experiment performed on embryos, a 40-minute heat shock at 33.5°C was applied 1 hour after the 2-cell stage of development (see section 2.1.4 and figure 4.3) and then the embryos were allowed to develop for 16 hours or overnight at 20°C. Since it has been shown earlier that maximal levels of Hc-ELT-2 are expressed immediately following heat shock (section 4.2.2 and figure 4.2C) this protocol is expected to result in Hc-ELT-2 being ectopically expressed in *C. elegans* embryos at between the 30–40 cell-stage. Most cells are pluripotent at this stage and hence might be expected to be sensitive to the effects of an ectopically expressed transcription factor. After a period of 16 hours after heat shock, counts were made of the number of embryos which had developed normally, the number that had arrested at the comma stage or later (late arrest), and the number which arrested without showing any signs of morphogenesis (early arrest). First, a transgenic line containing the ectopic expression vector pAC7 in an N2 background (line AC3), a transgenic line containing pAC9 in an N2 background (line AC15) and N2 worms as a negative control were examined. The number of embryos that arrested development in the N2 control line was 20 % showing that the heat shock regime used resulted in relatively small background of "non-specific" death (table 4.2). The vast majority of those 20 % of arrested embryos in the control lines had undergone morphogenesis and arrested relatively late in development at the comma stage or later (late arrest). In contrast, the proportion of embryos arresting in the AC3 and AC15 transgenic lines was 79.1 % and 70.0 % respectively, much higher than for the control lines (table 4.2). A Fisher's exact test was carried out on the data presented in table 4.2 and it revealed that the difference in the proportion of arrest between AC3 and AC15 was not significant (p=0.5284); however the difference observed in the proportions of embryos arresting early between AC3 and N2 and AC15 and N2 was very significant

with p<0.0001 and p=0.0010, respectively. The majority of these embryos arrested without any visible signs of morphogenesis (table 4.2 and figure 4.4). These embryos appeared to consist of a fairly homogenous ball of cells with no obvious signs morphogenesis (figure 4.4). In many cases the appearance as a "ball of cells" is extremely uniform (figure 4.4, panels C and D) but in some embryos there are a number of vacuoles present suggesting some cell death and degeneration (figure 4.4, panel A and B). These arrested embryos show no visible sign of elongation, tissue formation or organogenesis. No embryo with the "ball of cells" phenotype were seen in the N2 controls. Arrested embryos with the "ball of cells" phenotype from the two transgenic lines AC3 and AC15 were methanol/acetone fixed (see section 2.3.5) and stained with DAPI, a DNA stain which allows the visualisation of cell nuclei (figure 4.5). The majority of the cell nuclei were relatively large which is an appearance more consistent with either gut cells or hypodermal cells rather than other cell types. It was not possible to accurately count the relatively large number of densely packed cell nuclei in individual embryos using conventional fluorescence microscopy. Although this is possible with confocal microscopy (Gilleard and McGhee, 2001.), it is extremely time-consuming and was not considered justified. However, some very approximate nuclei counts were performed on the early arrested embryos and the number of nuclei was typically between 200 and 300 in each arrested embryo.

4.2.3 Examination of endodermal marker expression in embryos arrested by ectopic Hc-ELT-2 expression

Previous studies have shown that ectopic expression of the *C. elegans elt-2* gene in early *C. elegans* embryos results in the reprogramming of most, if not all,

cell lineages to adopt endodermal cell fates (Fukushige, T. *et al.* 1998). The "ball of cells" arrest phenotype produced in response to the ectopic expression of *Haemonchus contortus elt-2* described above is consistent a similar transformation of cell fate by the *H. contortus elt-2* gene. In order to investigate this, the expression of four different endodermal markers were examined in the *C. elegans* embryos arrested as a "ball of cells" by ectopic *Hc-elt-2* expression. The markers examined were autofluorescent gut granules, birefringent gut granules, an *ifb-2::gfp* marker and an *elt-2::gfp* marker. Each of these markers are expressed only in endoderm with no expression in other tissues or cell types at any stage of the development of wild type worms. Birefringent gut granules are first visible at 4-6 hours after fertilisation, when the embryo has 200 cells and E has divided into 8-16 gut primordium. The birefringence is visible under polarized light and the autofluorescence under UV illumination using FITC filter sets (Laufer, J. S. *et al.* 1980). These granules are confined to the cytoplasm of endodermal cells. They can be easily distinguished from GFP fluorescence in transgenic lines on the basis of colour since GFP fluorescence is a much a deeper green. The *ifb-2::GFP* marker consists of a chromosomally integrated nuclear localised GFP reporter gene construct (transgenic line JM86 kindly provided by J. McGhee). The *ifb-2::GFP* is a reporter construct which utilises the promoter region for the *ifb-2* gene, which encodes an intermediate filament protein. This intermediate filament protein was originally localised by the monoclonal antibody MH33 which stains the gut brush border in *C. elegans* and is expressed from the lima bean stage onwards (Bossinger, O. *et al.* 2004). In this line the construct carries a nuclear localisation signal (NLS) for GFP, and so GFP is not localised to the brush border like MH33 staining but is localised in the gut nuclei. The *ifb-2::GFP* is expressed in a maximum of 20 gut nuclei in this line. The *elt-*

2::GFP marker consists of a chromosomally integrated nuclear localised GFP reporter construct under the control of the *C. elegans elt-2* promoter (transgenic line JM73 kindly provided by Prof. J.D. McGhee). This *elt-2::GFP* marker is expressed in endodermal nuclei from the 2E cell stage and throughout subsequent development and so again is expressed in a maximum of 20 gut cells. The expression of each of the four endodermal markers in wild type comma stage embryos is shown in figure 4.6, panels A-D. It can be seen that the autofluorescent and birefringent granules are confined entirely to the cytoplasm of endodermal cells, whereas the two nuclear localised GFP markers are expressed only in the 20 endodermal cell nuclei.

The effect of ectopic *Hc-elt-2* expression on the distribution of autofluorescent and birefringent granules was examined in transgenic lines AC3 and AC15. These are N2 worms carrying extrachromosomal arrays containing the pAC7 and pAC9 ectopic expression constructs respectively (table 4.1). Embryos arrested by ectopic expression of *Hc-elt-2* in early embryos using the protocol in figure 4.3 resulted in widespread expression of these granules throughout the whole of the arrested embryos (Figure 4.6, panels E and F and I and J). Because these are cytoplasmic markers it was not possible to determine the number of cells expressing these markers. In order to examine the expression of the two endodermal nuclear localised GFP markers in embryos arrested by ectopic *Hc-elt-2* expression it was necessary to produce new transgenic strains: The pAC7 and pAC9 constructs were each separately microinjected into transgenic strains JM86 and JM73, which carry chromosomally integrated copies of the *ifb-2::GFP* and *elt-2::GFP* reporter genes respectively. This resulted in transgenic line AC9, AC11, AC6 and AC14, which are described in table 4.1. *Hc-elt-2* was ectopically expressed in embryos from these transgenic lines at 1 hour after the 2-cell stage using the method described in figure

4.3. Embryos showing the "ball of cells" arrest phenotype were mounted on agarose pads and examined by fluorescence microscopy. Widespread expression of both the *ifb-2::GFP* and the *elt-2::GFP* reporter genes was seen throughout these embryos (figure 4.6, panels G and K, and H and L, respectively). The nuclear localisation of these two GFP reporter genes allowed the approximate number of cells expressing the GFP marker to be determined. For most of the arrested embryos examined both the *ifb-2::GFP* and *elt-2::GFP* markers were expressed in many more than the 20 endodermal cells seen in wild type embryos. In many arrested embryos the majority of cells in the embryo showed GFP marker expression (figure 4.6, panels G and H, and K and L). Hence it appears that the majority of cells in embryos showing the "ball of cells" arrest phenotype in response to ectopic *Hc-elt-2* expression express a range of different endodermal markers. Also, from the results of the DAPI staining described earlier, the majority of these cells have a nuclear size typical of endodermal cells (section 4.2.2).

4.2.4 Examination of non-endodermal marker expression in embryos arrested by ectopic *Hc-elt-2* expression

The expression of two non-endodermal reporter gene markers was examined in embryos arrested by ectopic *Hc-elt-2* expression. The integrated transgene ccIs4251 contains a *myo-3::GFP* nuclear localized reporter gene that is expressed in 81 body wall muscle cells in wild type embryos (Okkema, P. G. *et al.* 1993). The integrated transgene wIs51 contains a pMF1::GFP nuclear localized reporter gene that is expressed in the 20 lateral seam (hypodermal) cells in wild type embryos (WormBase web site, http://www.wormbase.org., Relase WS98). Seam cells are a

type of epidermal cell that play an organizing role in embryonic and larval morphogenesis, contribute to larval growth by generating additional epidermal and nervous tissue post-embryonically and regulate the body form during formation of the dauer larvae (Koh, K. *et al.* 2001). At the end of embryogenesis, there are ten pairs of lateral seam cells. The seam cell marker *pMF-1::GFP* is expressed from the early comma stage onwards ((Koh, K. *et al.* 2001)). The pAC9 *Hc-elt-2* ectopic expression plasmid was injected into two different transgenic strains, PD4251 and JR667 containing the *myo-3::GFP* and *pMF-1::GFP* chromosomally integrated reporter genes respectively. Transgenic lines AC16 and AC18 were established (table 4.1). Heat shock was performed on embryos from these lines at 1 hour after the 2-cell stage of development as for the experiments performed on the endodermal markers (figure 4.3). This resulted in the "ball of cells" embryonic arrest phenotype in a high proportion of embryos, consistent with the transmission rates of each of the lines: 66.3 % (n= 103) of arrest for AC16 and 75.9 % (n=29) for AC18). In order to examine the expression of the non-endodermal markers, embryos with the "ball of cells" phenotype were mounted on agarose pads examined by fluorescence microscopy. The identification of those embryos with a widespread distribution of autofluorescent granules allowed confirmation of the arrest phenotype. The number of nuclei expressing GFP in these arrested embryos were counted (table 4.3). A maximum of 4 cells were found to express GFP in arrested embryos from line AC18 (*myo-3::gfp*) and some embryos did not express the marker at all (figure 4.7). Hence the arrested embryos express this muscle marker in far fewer cells than the 81 present in wild type embryos. Similarly, a maximum of 4 cells were found to express GFP in arrested embryos from line AC16 and some embryos did not express the marker at all. This is fewer than the 20 seam cells present in wild type embryos

(figure 4.7). Hence in contrast to the widespread expression of endodermal markers seen in the embryos arrested by ectopic *Hc-elt-2* expression there is very restricted expression of two non-endodermal markers. Hence the majority of cells expressing endodermal markers in these arrested embryos do not co-express either of the two non-endodermal markers examined. This suggests the majority of cells in *Hc-elt-2* arrested embryos have endodermal-like fates.

4.2.5 Ectopic expression of *Hc-elt-2* at different times of embryonic development

The ectopic expression experiments described above have involved the application of heat shock to embryos at one hour of development after the 2-cell stage at 20°C. This means that the maximal levels of Hc-ELT-2 polypeptide will occur at approximately the 30-40 blastomere stage. This stage was chosen since the majority of cells are likely to be pluripotent and so susceptible to the effects of an ectopically expressed transcription factor. A number of experiments were performed to determine the effect of ectopic expression of *Hc-elt-2* at later stages of development, when many cell lineages are already committed to specific fates. Embryos were harvested at one, two and four hours after the 2-cell stage and a 40-minute 33.5°C heat shock was applied to induce *Hc-elt-2* expression, using the transgenic line AC13. This line was chosen as the JM86 background contains the *ifb-2::GFP* marker, which allows for GFP expressing cells to be counted. Figure 4.8 shows the percentage of AC13 embryos that arrest when heat-shock was applied after different periods of incubation at 20°C.

The number of embryos that arrested decreased as the duration of incubation at 20°C prior to heat shock increased (figure 4.8A). The percentage of AC13

embryos that showed the "ball of cells" early arrest phenotype was 85.2%, 71.4% and 60.0% when heat shock was applied after 1 hour, 2 hours and four hours, respectively. These results show that if Hc-ELT-2 expression is induced later during embryonic development, when most cells are committed to particular cell fates, it has less effect on embryonic development. Figure 4.8B shows the number of nuclei in which GFP was expressed in the embryos arrested by ectopic *Hc-elt-2* expression. After one hour of incubation prior to heat-shock, 10 out of 19 (52.6 %) of arrested embryos of line AC13 show widespread expression of GFP (i.e. in more than 40 nuclei). However after two and four hours of incubation at 20°C prior to heat-shock, only 4 out of 21 (19 %) and 2 out of 17 (11.8 %), embryos showed widespread expression, respectively.

Over the course of these ectopic expression experiments described above, a small number of fully developed L1 larvae were seen that expressed the endodermal *ifb-2::GFP* marker in nuclei outside of the gut. Although these larvae were observed several times during experiments both with lines AC11 and AC13, they were insufficiently frequent to determine if their occurrence was increased with a longer incubation period prior heat shock. The additional nuclei expressing GFP outside the gut of L1 larvae were consistently observed in the pharyngeal region (figure 4.9). These additional cells are localised in the terminal bulb of the pharynx as well as in a more anterior position. In addition, the pharynx of these larvae appeared deformed in some cases (data not shown). This ectopic expression of endodermal markers in fully developed larvae has not previously been reported in response to ectopic expression of the *C. elegans elt-2* gene (T. Fukushige and J. McGhee). However, it has been observed on several occasions with experiments performed with non-integrated lines carrying a *C. elegans elt-2* ectopic expression construct (Dr J. Smith and J Gilleard,

personal communication). Hence, this it does not appear to be a feature specific to the parasite gene.

4.2.6 Comparison of *Hc-elt-2* and *Ce-elt-2* ectopic expression in early embryos.

The ability of ectopic *Hc-elt-2* expression to transform *C. elegans* embryos into a ball of endodermal-like cells is qualitatively very similar to the results of ectopic expression experiments performed with the *C. elegans elt-2* gene (Fukushige, T. *et al.* 1998) This suggests a high degree of functional conservation of the gene between the two different nematode species. An experiment was performed to investigate whether the *H. contortus* ELT-2 polypeptide functioned with a similar efficiency as the *C. elegans* ELT-2 polypeptide in the ectopic expression assay. Transgenic lines which contained extrachromosomal arrays carrying a *C. elegans elt-2* ectopic expression construct were obtained from Dr J. Smith and experiments were performed in parallel with lines carrying the *H. contortus elt-2* ectopic expression constructs. Four lines were used: AC11 and AC9 carrying *Hc-elt-2* ectopic expression constructs (with and without the synthetic intron respectively), and JAS9 and JAS42, carrying *Ce-elt-2* ectopic expression constructs (with and without the synthetic intron, respectively, both these lines kindly provided by Dr J. Smith). All of these lines were derived by microinjection of the appropriate ectopic expression construct into JM86 worms and hence carry the caIs32 integrated transgene containing the endodermal specific *ifb-2::GFP* reporter construct. Although the ectopic expression assay only lends itself to limited quantification, it was anticipated that major quantitative differences in the ability of the *H. contortus* and *C. elegans* ELT-2 polypeptides to activate endodermal gene expression might be apparent on

comparison of these lines. The main limitation of this approach was that the ectopic expression transgenes were present on extrachromosomal arrays and consequently the four different lines to be compared had different transmission rates, although they were similar (around 60%). Also the ectopic expression technique is relatively labour intensive which limits the number of embryos that can be examined However we attempted to gain some assessment of the relative efficiency of *Hc-elt-2* and Ce-*elt-2* function by scoring the number of GFP positive nuclei in arrested embryos in each of the lines. Following ectopic expression, embryos were placed in one of three categories: embryos in which GFP is expressed in 20 cells or less; embryos in which GFP was expressed in 21-40 cells and embryos in which GFP was expressed in more than 40 cells. Occasional embryos that had arrested very early with a total cell were considered dead and not included in the analysis. Table 4.4 shows the number of nuclei expressing GFP in arrested embryos after ectopic expression of *C. elegans elt-2* and *Hc-elt-2*. Statistics were not applied to these figures, as the sample size was very small. The proportion of embryos with more than 40 nuclei expressing GFP was greater in both lines ectopically expressing the *C. elegans elt-2* gene compared to both lines ectopically expressing *H. contortus elt-2* gene; JAS9 and AC11 carry equivalent constructs (with intron) and JAS42 and AC9 carry equivalent constructs (without synthetic intron). Some qualitative differences between embryos resulting from *Hc-elt-2* ectopic expression and *Ce-elt-2* expression were also observed. A greater proportion of embryos arrested by ectopic *Hc-elt-2* expression in both AC11 and AC3 lines contained vacuoles (figure 4.4, panel C and D) than was the case for the embryos arrested by ectopic *C. elegans elt-2* expression in lines JAS9 and JAS42 (figure 4.4, panels A and B). This was reflected in the distribution of *ifb-2::GFP* expressing nuclei which tended to be less homogenous in some (but not all) of the

embryos arrested by Hc-ELT-2 ectopic expression compared to embryos arrested by *C. elegans* ELT-2 ectopic expression (figure 4.10). This may suggest that a greater number of cells in the embryos arrested by *Hc-elt-2* ectopic expression die and degenerate than in embryos arrested by ectopic *C. elegans elt-2* expression.

Hence, within the limitations of the experimental design discussed above, the results suggest ectopic expression of *Hc elt-2* is not quite as effective in activating widespread expression of *ifb2::GFP* as is ectopic expression of *C. elegans elt-2*.

4.2.7 The effect of the presence or absence of a synthetic intron on the efficacy of the *Hc-elt-2* ectopic expression construct

The results of the experiment described in the previous section and in table 4.4 also give some information regarding the effect of the presence the synthetic intron on the functional efficiency of the ectopic expression constructs.

The proportion of embryos expressing *ifb-2::GFP* in more than 40 cells is greater for the lines containing ectopic expression constructs which include the synthetic intron (AC11 and JAS9) than those which contain a construct which do not include a synthetic intron (AC9 and JAS42). This suggests that the presence of the synthetic intron increases the efficiency of the transgene expression. Interestingly, this was not reflected by the levels of transgene expression as determined by the Western blot experiments in section 4.2.2.

4.3 Discussion

Ectopic expression of the *C. elegans elt-2* and *end-1* genes in early *C. elegans* embryos results in developmental arrest as a ball of 200-300 cells with no signs of morphogenesis (Fukushige, T. *et al.* 1998; Zhu, J. *et al.* 1998). The majority of cells in these arrested embryos express gut-specific markers and do not express non-gut specific markers. Furthermore, the nuclei of most cells in these arrested embryos have a large and rounded appearance characteristic of gut cell nuclei (Zhu, J. *et al.* 1998). This suggests that the majority cells in these arrested embryos have undergone a program of gut cell differentiation and there is an absence of other cell types. Consequently it has been concluded that both the *C. elegans elt-2* and *end-1* GATA transcription factors are capable of reprogramming blastomeres in the early *C. elegans* embryo to direct them to adopt an endodermal cell fate. Two hypodermally expressed *C elegans* GATA transcription factors, *elt-1* and *elt-3* have been studied using this approach and their ectopic expression also leads to developmental arrest with a transformation of many cell lineages into hypodermal cell fates (Gilleard, J. S. *et al.* 2001). Hence several *C. elegans* GATA factors have the ability to transform the fates of pluripotent blastomeres in early *C. elegans* embryos in a highly tissue specific manner. The work outlined in this chapter was aimed at determining whether the *Haemonchus contortus elt-2* homologue could transform the fates of *C. elegans* blastomeres in a similar manner as the *C. elegans elt-2* gene itself and whether functional specificity was conserved.

The *Haemonchus contortus elt-2* cDNA was placed under the control of the *hsp16-2* heat shock promoter by cloning into the ectopic expression vector pPD49-78 with and without the presence of a synthetic intron. The resulting constructs were injected into the syncytial gonad of *C. elegans* hermaphrodites to generate transgenic

146

strains. Western blotting experiments demonstrated that there was no constitutive expression of the *Haemonchus contortus* ELT-2 polypeptide in these strains but the application of a 2-hour 33.5°C heat-shock induced detectable expression of the parasite polypeptide. Maximal levels of expression were detected immediately after the heat shock period and declined to undetectable levels after 8 hours at 20°C. Ectopic expression of the *H. contortus* ELT-2 polypeptide in early *C. elegans* embryos resulted in a very similar embryonic arrest phenotype to that seen in response to ectopic expression of *C. elegans* ELT-2. Arrested embryos consisted of a ball of 200-300 large round cells the majority of which expressed four independent endoderm specific markers but failed to express a muscle cell or a seam cell specific marker. These results suggest that ectopic expression of *Hc-elt-2* in early *C. elegans* embryos is sufficient to activate a program of gut cell differentiation in the majority of early embryonic blastomeres. This is a striking result when one considers the relatively low level of amino acid identity between the *C. elegans* and *H.contortus* ELT-2 polypeptides. There is only 26.8% overall sequence identity between the *C. elegans* and *H.contortus* ELT-2 polypeptide sequences. This sequence identity is largely confined to the DNA binding domain (82% identity) with very little sequence conservation outside this region. The ectopic expression approach described in this chapter has recently been used to undertake a structure-function analysis of both the *C. elegans* and *H. contortus* ELT-2 polypeptides (Smith, J. and Gilleard, J. S., personal communication). The principle of these experiments was to assay the effect of particular deletions, domain swaps and single residue mutations on the ability of the *C. elegans* and *H. contortus* ELT-2 polypeptides to activate endodermal differentiation in the ectopic expression assay. The objective was to determine the extent to which the molecular function of ELT-2 was conserved between the two

species and to identify key functional domains and residues in the parasite polypeptide. These experiments have so far revealed that the same effect on function was obtained with identical deletions and mutageneses performed on the *C. elegans* and *H. contortus* polypeptides, suggesting a high degree of conserved function at the molecular level. Several other interesting points have also emerged from these experiments: although the DNA binding domain is the only region of extensive polypeptide sequence identity between the two species, the deletion analysis has shown that it is not sufficient to activate endodermal differentiation. This was also confirmed by domain swap experiments where the ELT-2 DNA binding domain has been replaced with that of other non-endodermal GATA factors, hence these results suggest that the tissue-specific functionality of ELT-2 is determined by regions outside the DNA binding domain (Smith and Gilleard, personal communication). Since there is only limited sequence identity between the *C. elegans* and *H. contortus* ELT-2 polypeptides outside the DNA binding domain, residues in the small conserved regions were mutated in an attempt to identify the key functional domains. For instance, the pseudo zinc finger region, present in both *C. elegans* and *H. contortus* ELT-2 polypeptides, appears to be essential for ELT-2 to activate endodermal differentiation, as shown by ectopic expression of the chimeric construct lacking this region (Smith, J. A. and Gilleard, J. S., personal communication). In contrast, disruption of several other small domains conserved between the *C. elegans* and *H. contortus* polypeptides had little effect on the ability of the polypeptide to activate endodermal differentiation suggesting that some unconserved residues may be involved in the conservation of the tertiary structure between the *C. elegans* and *H. contortus* ELT-2 polypeptides. The experiments described in this chapter and the subsequent structure-function experiments illustrate how the expression of a parasite

gene in *C. elegans* can allow detailed analysis of its function which would not be possible in the parasite system itself.

Comparison of the effect of ectopically expressing *H. contortus elt-2* with the effect of ectopically expressing *C. elegans elt-2* in early *C. elegans* embryos suggests that *Hc-elt-2* does not function quite as efficiently in early *C. elegans* blastomeres as does the *C. elegans elt-2* gene itself. This is reflected in the number of embryos arresting with widespread expression of the *ifb-2::GFP* marker (more than 40 cells) and also in the uniformity of the distribution of the *ifb-2::GFP* expressing cells throughout the arrested embryos. In a proportion of the embryos arrested by *Hc-elt-2* expression vacuoles were present which were not generally observed in embryos arrested by *C. elegans elt-2* expression. These vacuoles presumably are the result of cell death and degeneration, suggesting that ectopic expression of *Hc-elt-2* in some cells in the *C. elegans* embryo results in cell death as opposed to a cell fate change to endoderm. This may reflect some limitations of the Hc-ELT-2 polypeptide to function in some cellular environments in the *C. elegans* embryo.

Ectopic expression of *Hc-elt-2* during later embryogenesis generally failed to produce arrested embryos with the "ball of cells" phenotype. The majority of embryos developed normally, but when arrested embryos were observed they appeared very irregular and with large vacuoles with irregular expression of the markers. This suggests that in order to activate a program of gut cell differentiation, *Hc-elt-2* needs to be ectopically expressed in the early stages of embryogenesis, when most lineages are not yet committed. When expressed in later stage of embryogenesis, as most cells are committed to their fate, *Hc-elt-2* fails to reprogram other blastomeres and most embryos will proceed to develop normally. This has also

been found to be the case for ectopic expression of *elt-1*, *elt-3* and *end-1* (Zhu, J. *et al.* 1998; Gilleard, J. S. *et al.* 2001).

During the course of ectopic expression experiments, occasional larvae were seen with clusters of cells outside the gut region that expressed the *ifb-2::GFP* gut-specific marker. One explanation of this is that these larvae originated from embryos with a mosaic distribution of the extra-chromosomal array resulting in only some lineages being exposed to the *Hc-elt-2* ectopic expression. However, this would not explain the additional expression of the marker outside the gut always being in the pharyngeal region. A more likely explanation is that some lineages remain sensitive to the effects of ectopic expression for longer than others.

Much of the pharyngeal tissue is derived from the MS blastomere, which is the sister blastomere of E. Hence it is possible that some of the MS-derived lineages that give rise to pharyngeal tissue remain susceptible to the effects of *Hc-elt-2* ectopic expression for longer than other lineages. Interestingly, promoter analysis of the endodermal gene *ges-1* has shown that cells of the pharynx that are derived from the MS blastomere have a latent capacity to express endodermal markers (Kennedy, B. P. *et al.* 1993).

In conclusion, the results discussed in this chapter have shown that the novel *H. contortus* GATA transcription factor *Hc-elt-2* is functionally conserved with *C. elegans elt-2* and can activate a program of gut cell differentiation in early *C. elegans* embryos, almost as efficiently as *C. elegans elt-2* itself. This work has provided the groundwork for the use of ectopic expression as a functional assay of parasite gene function.

Figure 4.1: Schematic diagram of canonical vector pPD49-26 (A), ectopic expression vector pPD49-78 (B), and ectopic expression constructs pAC7 (C) and pAC9 (D).

Panel A: the vector pPD49-26 has three multiple cloning sites, MSCI, MCSII and MSCIII. There is a synthetic intron between MSCI and MSCII and the 3'UTR of *unc-54* has been inserted between MSCII and MSCIII.

Panel B: The ectopic expression vector pPD49-78 (B) was derived from pPD49-26 by cloning the heat-shock promoter *hsp16-2* within MSCI. This vector allows for the cloning of cDNA downstream of the hsp16-2 promoter and/or the synthetic intron.

Panel C and D: pAC7 (C) and pAC9 (D) carry *Hc-elt-2* cDNA cloned under the control of hsp16-2 promoter. In pAC7 the cDNA has been cloned between MSCI and MSCII, removing the synthetic intron and in pAC9 the cDNA has been cloned within MSCII downstream of the synthetic intron .

(A) pPD49-26

(B) pPD49-78

(C) pAC7

(D) pAC9
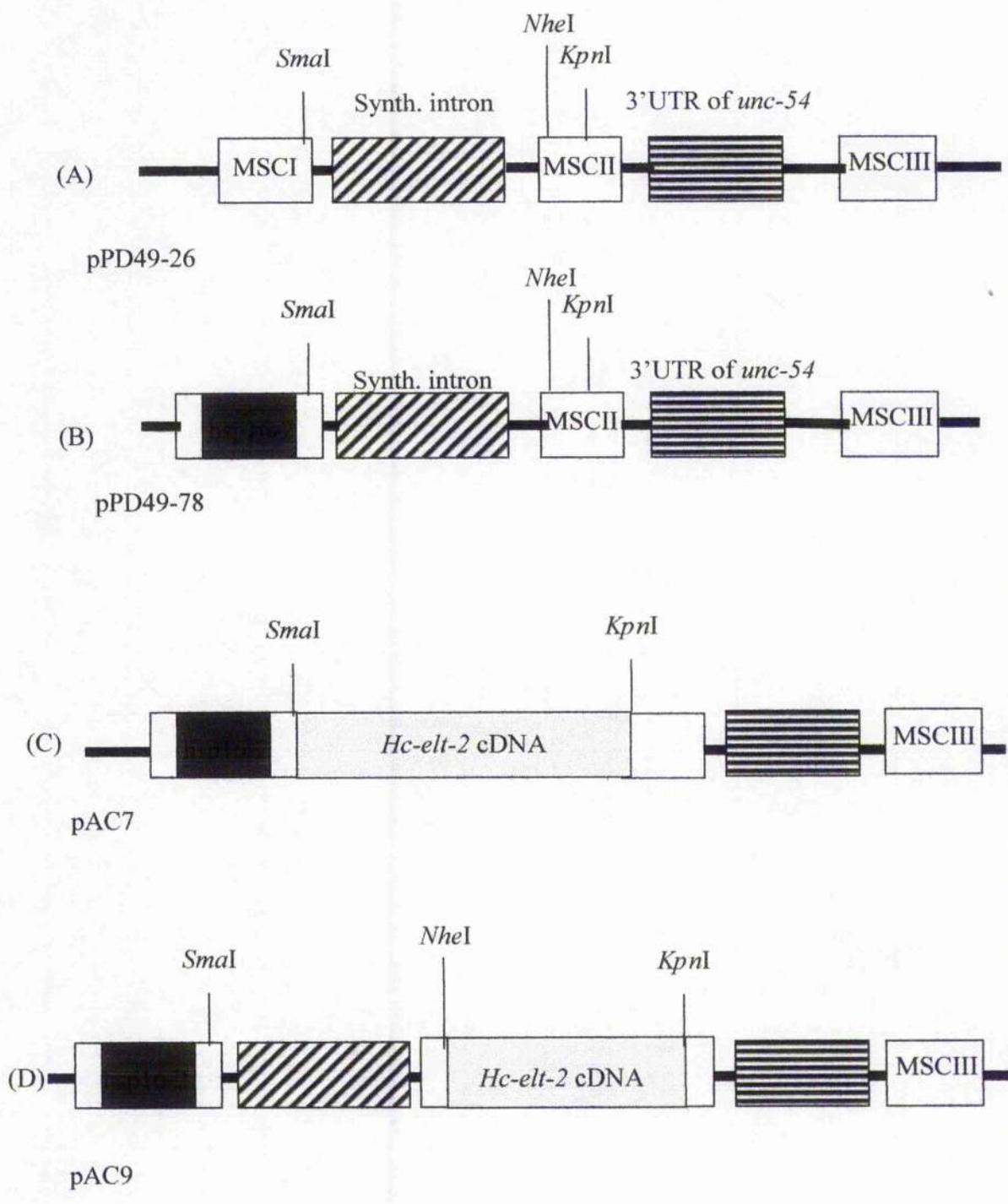
Strain name	Description
AC3	N2 injected with pAC7 (10 ng/µl) and rol-6 (400 ng/µl); non integrated.
AC6	JM73 injected with pAC7 (10 ng/µl) and rol-6 (400 ng/µl); non integrated.
AC9	JM86 injected with pAC7 (10 ng/µl) and rol-6 (400 ng/µl); non integrated
AC11	JM86 injected with pAC9 (10 ng/µl) and rol-6 (400 ng/µl); non integrated
AC13	JM86 injected with pAC9 (10 ng/µl) and rol-6 (400 ng/µl); non integrated
AC14	JM73 injected with pAC9 (10 ng/µl) and rol-6 (400 ng/µ l); non integrated
AC15	N2 injected with pAC9 (10 ng/µl) and rol-6 (400 ng/µl); non integrated
AC16	PD4251 injected with pAC9 (10 ng/µl) and rol-6 (400 ng/µl); non integrated
AC18	JR667 injected with pAC9 (10 ng/µl) and rol-6 (400 ng/µl); non integrated
JAS9	JM86 injected with *elt-2* cDNA under the control of *hsp-16-2*, with intron
JAS42	JM86 injected with *elt-2* cDNA under the control of *hsp-16-2*, without intron
JM73	unc-119(e2498::Tc1) III; caIs20 *elt-2::gfp*
JM86	unc-119(e2498::Tc1) III; caIs32 *ifb-2::gfp*
JR667	*unc-119*(e2498::Tc1)III; wIs51 *pMF-1::gfp*
PD4251	*dpy-20*(e1282)IV; ccIs4251 *myo-3::gfp*

Table 4.1: List of transgenic strains used for ectopic expression of *Hc-elt-2*.

Figure 4.2: Expression of Hc-ELT-2 in *C. elegans* transgenic lines.

Worm extracts were separated by SDS-PAGE electrophoresis, Western

blotted and probed with R1064 Hc-ELT-2 specific antiserum.

Panel A: worm extracts were prepared for SDS-PAGE after heat-shock for 2

hours (+) or in the absence of heat-shock (-).

Panel B: worm extracts were prepared for SDS-PAGE from AC3 and N2

worms at 0, 2, 4 and 8 after heat-shock for 2 hours.

Panel C: extracts were prepared for SDS-PAGE from *H. contortus* adults

(track 1), *H. contortus* L3 (track 2) and the *C. elegans* transgenic line AC3

immediately after heat-shock for 2 hours (this gel was also presented in

figure 3.18).

(A)

	N2		AC3		AC9		AC11		AC13	
	+	-	+	-	+	-	+	-	+	-

1 2 3 4 5 6 7 8 9 10

45kDa ━

(B)

	AC3				N2			
	0h	2h	4h	8h	0h	2h	4h	8h

66

45

1 2 3 4 5 6 7 8

30

20

(C)

Hc adults   AC3

1   2   3

66

45

36

30

24

20
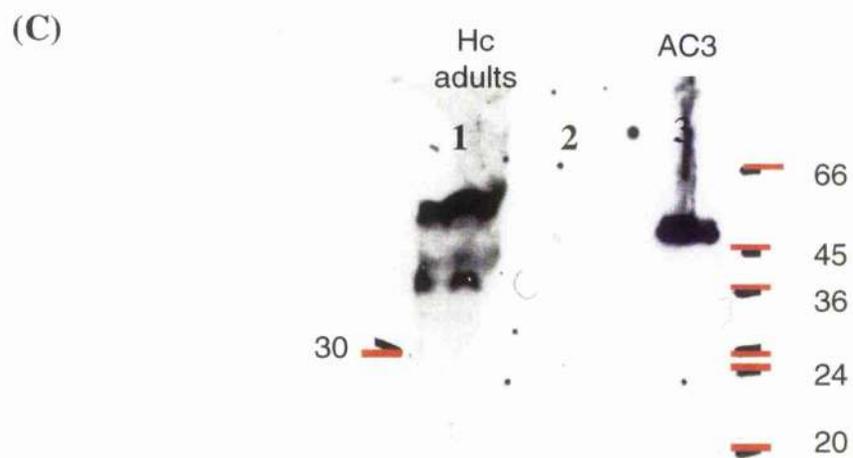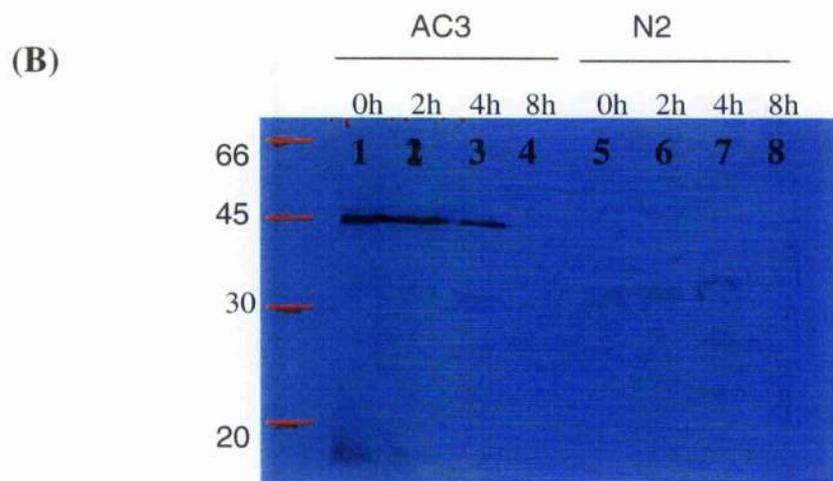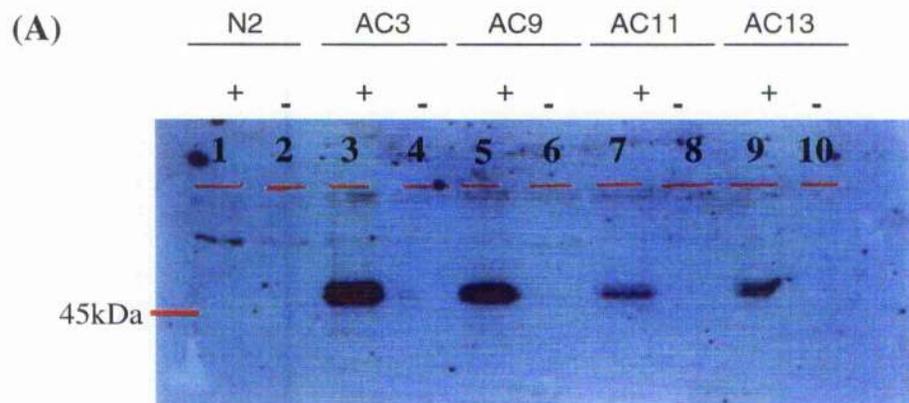
Figure 4.3: Outline of ectopic expression experiment methodology

The figure shows the different steps performed for to the preparation of the embryos for ectopic expression experiments.

50-60 young gravid adults

50 μl M9

Two-well watch glass slide
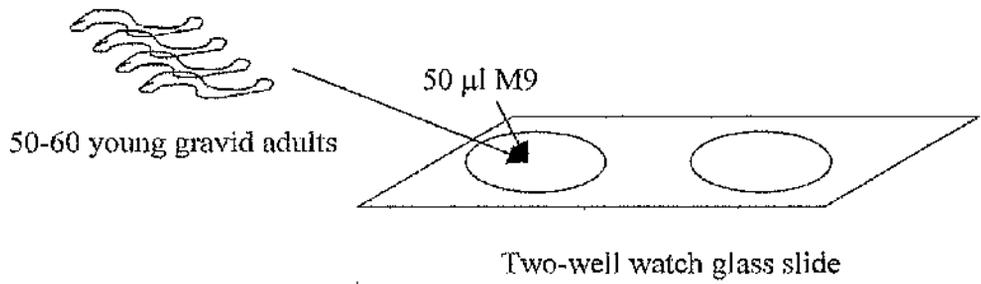
Dissect worms in watch glass slide with two 30-gauge needles and add one drop of 1-15% hypochlorite and complete dissection. Add 50 μl 5% BSA to neutralise hypochlorite.

Transfer 1-4-cell embryos to second well

50μl M9 + one drop BSA

Wash embryos by removing liquid carefully and adding fresh M9. Repeat 3 times and add BSA to final volume
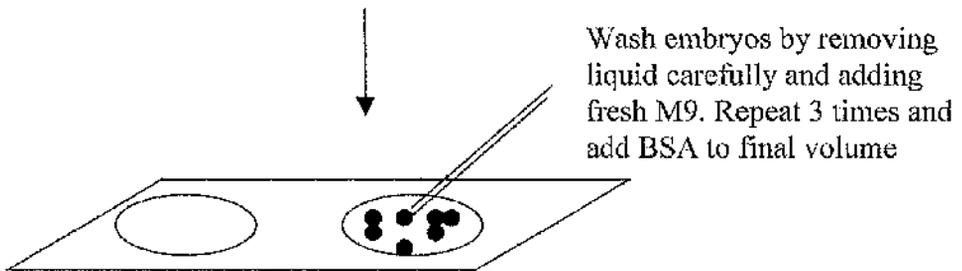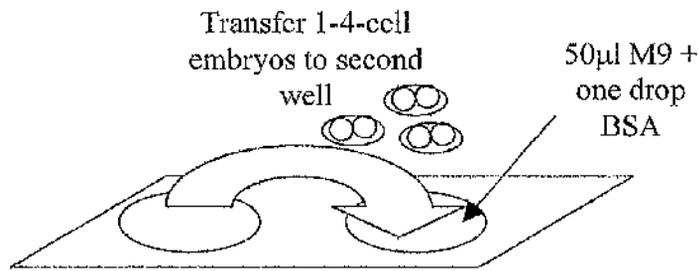
Incubate watch glass slide for 1h at 20°C

Incubate watch glass slide for 40 min at 33.5°C

Transfer watch glass slide at 20°C overnight; mount on agarose pads for examination

154

	AC3		AC15		N2	
	N	%	N	%	N	%
arrested	34	79.1	14	70.0	5	20.0
hatched	9	20.9	6	30.0	20	80.0
total	43	100.0	20	100.0	25	100.0

Table 4.2: Effect of ectopic expression of Hc-ELT-2 on the development of two lines compared to an N2 control

Line AC3 carries an *Hc-elt-2* ectopic expression construct without intron and AC15 carries an *Hc-elt-2* ectopic expression construct with intron.
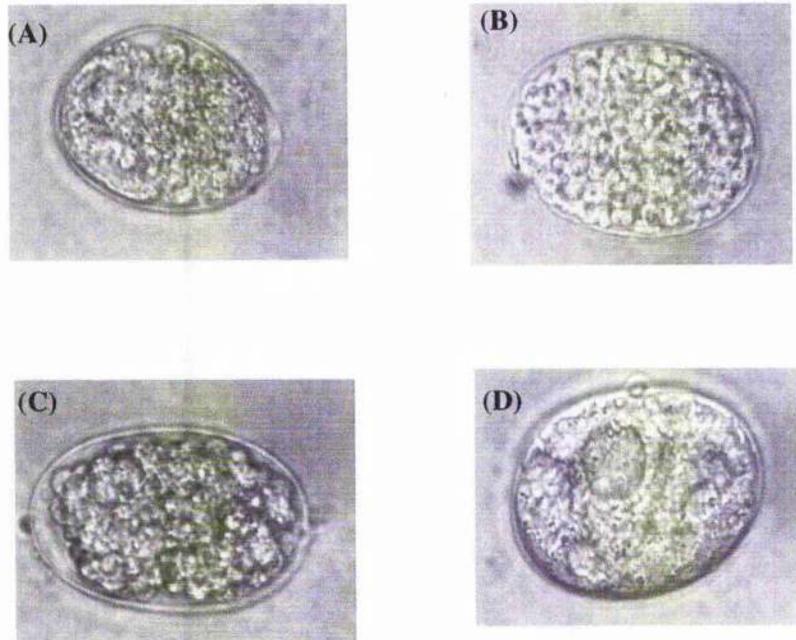
Figure 4.4: Images of arrested *C. elegans* embryos after ectopic expression of *H. contortus elt-2* during early embryogenesis.

Panel A: arrested embryo from line JAS9; it carries a transgene containing *C. elegans elt-2* ectopic expression construct with synthetic intron.

Panel B: arrested embryo from line JAS42; it carries a transgene containing *H. contortus elt-2* ectopic expression construct with synthetic intron.

Panel C: arrested embryo from line AC11; it carries a transgene containing the *C. elegans elt-2* ectopic expression construct without intron.

Panel D: arrested embryo from line AC3; it carries a transgene containing the *H. contortus elt-2* ectopic expression construct without intron.
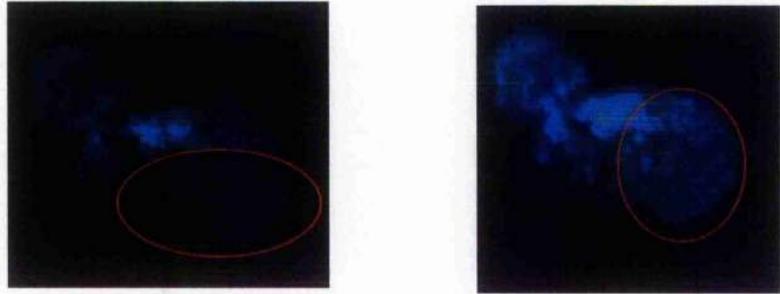
Figure 4.5: Arrested embryos showing the "ball of cells" phenotype, stained with DAPI.

Both embryos are from line AC4, which contains pAC7, the *Hc-elt-2* ectopic expression construct lacking the synthetic intron

Figure 4.6: Expression of gut-specific markers in arrested *C. elegans* embryos after ectopic expression of *H. contortus* elt-2 in early embryogenesis.

The expression of four different endoderm specific markers was examined: autofluorescent granules (panels A, E and I); birefringent granules (panels B, F and J); *ifb-2::GFP* nuclear localised reporter gene (panels C, G and K); *elt-2::GFP* nuclear localised reporter gene (Panels D, H and L).

-Panels A, B, C and D show the wild-type expression pattern of these markers.

-Panels E, F, G and H show the expression of these markers in embryos arrested by ectopic *Hc-elt-2* expression from ectopic expression construct pAC9 (with synthetic intron).

-Panels I, J, K and L show the expression of these markers in embryos arrested by ectopic *Hc-elt-2* expression from ectopic expression construct pAC7 (without synthetic intron).
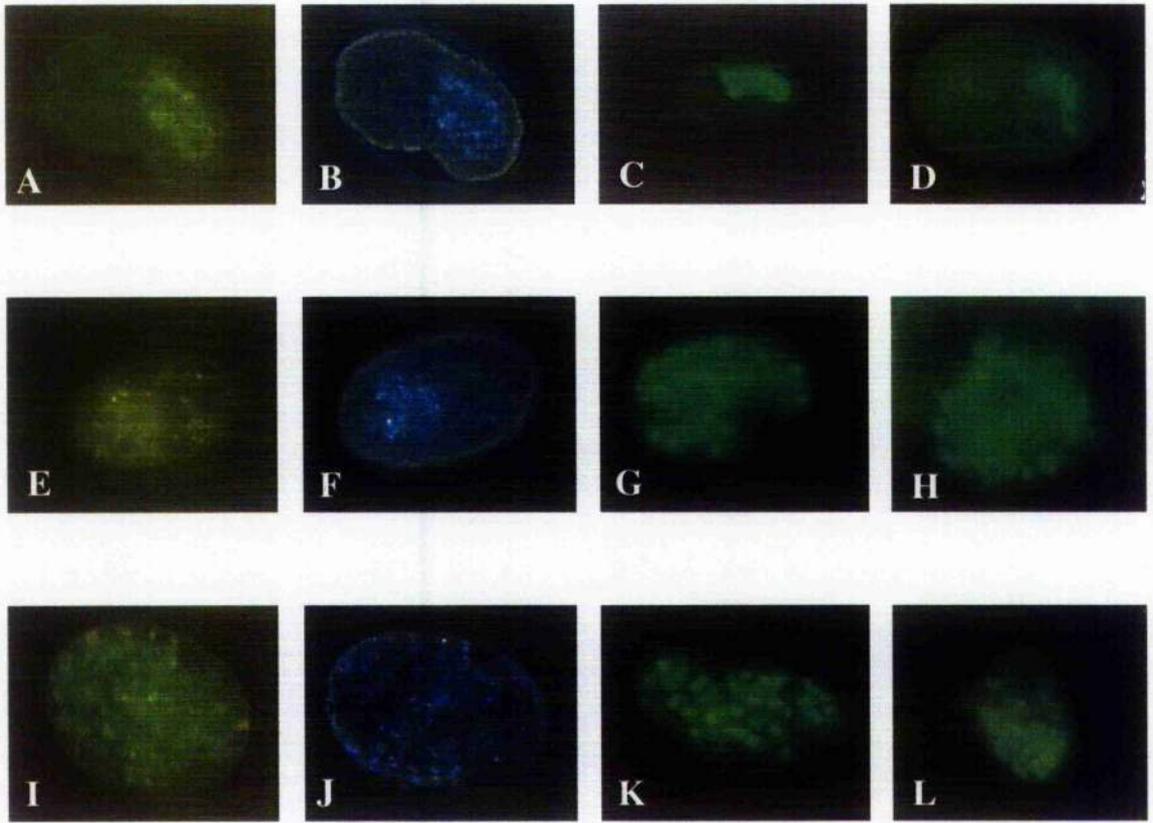
N° of nuclei expressing GFP	AC16	N° of nuclei expressing GFP	AC18
20	0	80	0
10-19	0	20-79	0
<10	10	10-19	1
		<10	7
Total	10		8

Table 4.3: Number of GFP expressing nuclei in arrested embryos of strains AC16 and AC18 after ectopic expression of *Hc-elt-2*.

The number of GFP expressing nuclei was counted in embryos arrested by ectopic expression in transgenic lines AC16 (*pMF-1::GFP*) and AC18 (*myo-3::GFP*). The columns 1 and 3 of the table show the categories of embryos in terms of number of GFP expressing nuclei. Columns 2 and 4 show the number of embryos in each category for lines AC16 and AC18, respectively. Wild type AC16 embryos contain a maximum of 20 GFP expressing nuclei at the end of embryogenesis, and wild type AC18 embryos contain a maximum of 81 GFP expressing GFP at the end of embryogenesis.

myo-3::gfp   *pMF-1::gfp*

(A)   (B)

WT (81 nuclei)   WT (20 nuclei)

(C)   (D)

AC16 (*hcelt-2*)   AC18 (*hcelt-2*)
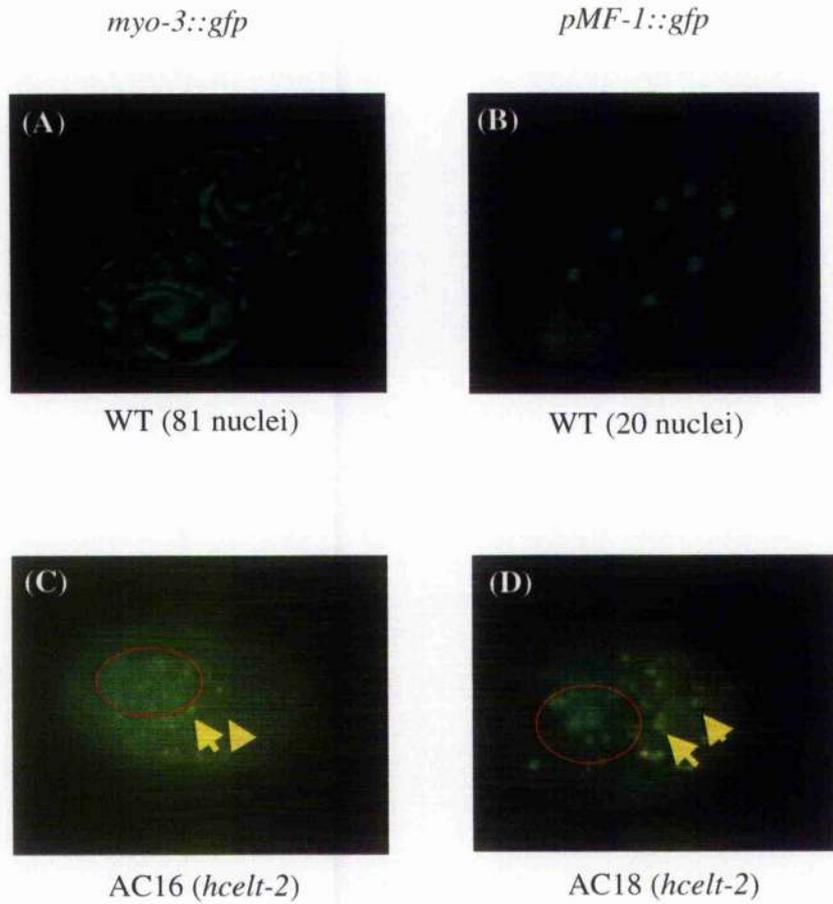
Figure 4.7: Expression of non endodermal markers in *C.* *elegans* embryos arrested by ectopic expression of *H.* *contortus elt-2* during early embryogenesis.

Images A and B show wild type expression of *myo-3::gfp*, a muscle cell marker and *pMF-1*::gfp, a seam cell marker, respectively. Images C and D show expression of the green non endodermal GFP markers (arrowheads) in arrested embryos showing widespread expression of yellow auto-fluorescent granules (circled).

160

Figure 4.8: Ectopic expression of *Hc-elt-2* at different times of embryonic development.

AC13 2-cell embryos were harvested and left at 20°C for either 60, 120 or 240 minutes prior to heat shock at 33.5°C for 40 minutes.

Panel A: shows the percentage of embryos that arrest "early" (no sign of morphogenesis), "late" (visible signs of morphogenesis) and develop normally.

Panel B: shows the percentage of embryos that express the *ifb-2::GFP* marker in less than 20 cells, 20 cells, 21-40 cells or more than 40 cells.

Early arrest

Late arrest

Normal development

(A)

Number of nuclei expressing *ifb-2::gfp:*

<20

20

21-40

>40
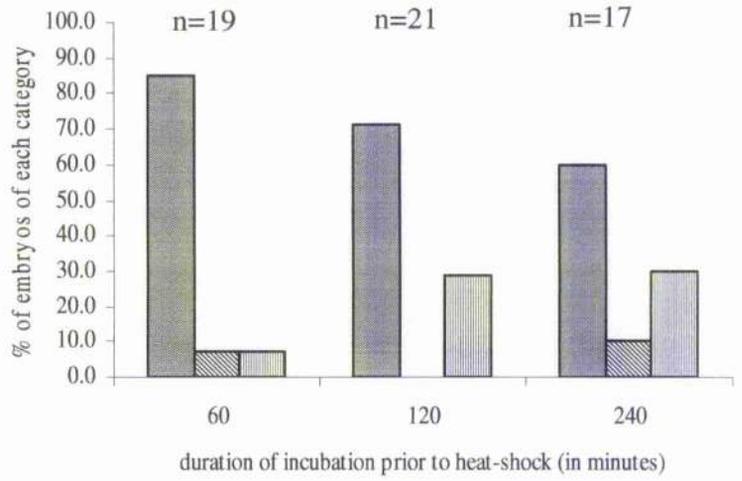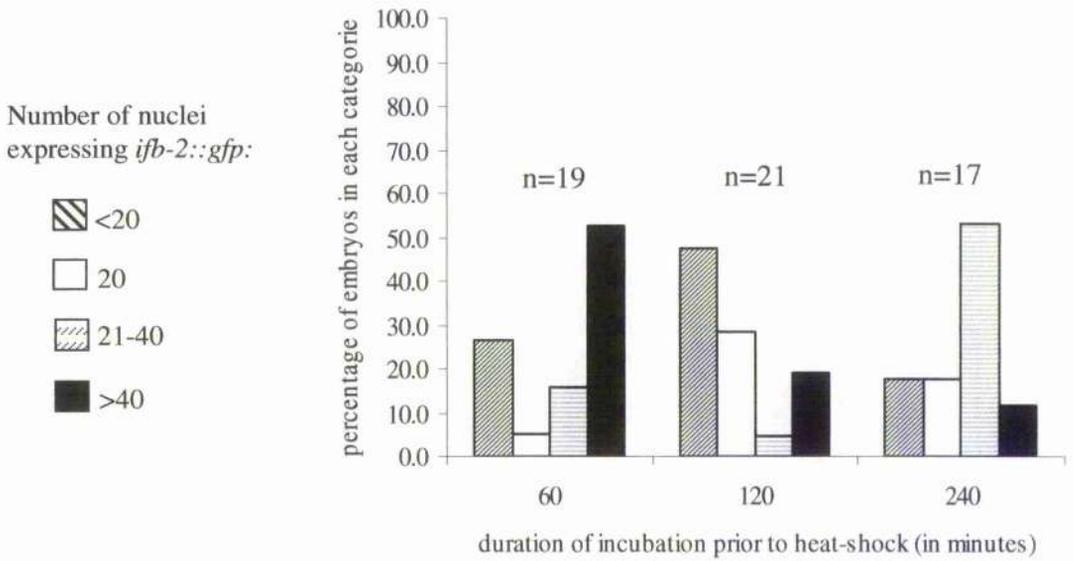
(B)
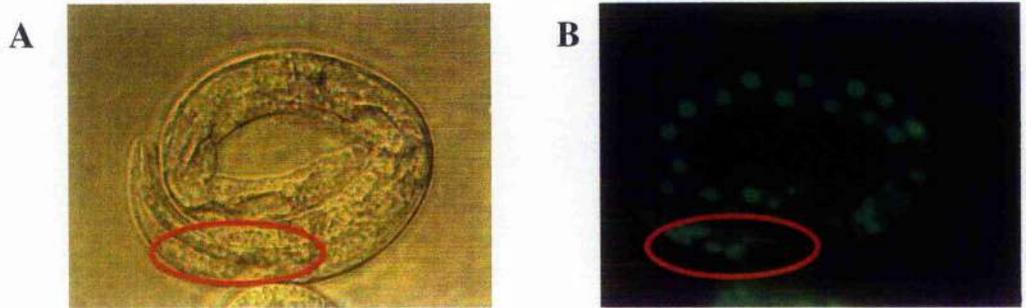
Figure 4.9: Expression of *ifb-2::GFP* transgene in the pharyngeal region of L1 larvae in response to ectopic *Hc-elt-2* expression.

The same AC13 larva is shown in panels A and B. It developed following a 40 minutes heat shock applied 1 hour after the 2-cell stage of development.

Panel A: Normaski

Panel B: Fluorescence

	Hc-elt-2 lines		Ce-elt-2 lines	
	AC9	AC11	JAS42	JAS9
20	13 (54.2%)	5 (16.7%)	6 (19.4%)	1 (6.7%)
21-40	8 (33.3%)	5 (16.7%)	4 (12.9%)	0
>40	3 (12.5%)	20 (66.7%)	21 (67.7%)	14 (93.3%)
total	24	30	31	15

Table 4.4: Comparison of *Hc-elt-2* and *Ce-elt-2* ectopic expression.

Ectopic expression of *Hc-elt-2* (lines AC9 and AC11) and *Ce-elt-2* (JAS9 and JAS42) was induced by a 40 minute heat shock at 33.5°C, 1 hour after the 2-cell stage. All the transgenic lines contain the chromosomally integrated *ifb-2::GFP* endodermal marker. The figures in the table show the number (and percentage in brackets) of embryos in each of four categories based on the number of nuclei showing *ifb-2::GFP* expression. The categories are less than 20 cells, 20 cells, 21-40 and more than 40 cells expressing GFP.

Four transgenic lines were examined (table 4.1 for details):

AC9 carries *Hc-elt-2* ectopic expression construct without synthetic intron

AC11 carries *Hc-elt-2* ectopic expression construct with synthetic intron

JAS42 carries *Ce-elt-2* ectopic expression construct without synthetic intron

JAS9 carries *Ce-elt-2* ectopic expression construct with synthetic intron

Figure 4.10: Selection of embryos arrested by *Ce-elt-2* or *Hc-elt-2* ectopic expression showing distribution of nuclei expressing the *ifb-2::GFP* endodermal marker.

Embryos were arrested by ectopic expression of either *Ce-elt-2* (Panels B, D and F) or *Hc-elt-2* (Panels A, C, E and G) induced by a 40 minute 33.5°C heat shock one hour after the 2-cell stage of development. All lines contained the chromosomally integrated *ifb-2::GFP* containing transgene (caIs32). Panel A: AC11; Panel B: JAS9; Panel C: AC9; Panel D: JAS42; Panel E: AC9; Panel F: JAS42; Panel G: AC11.
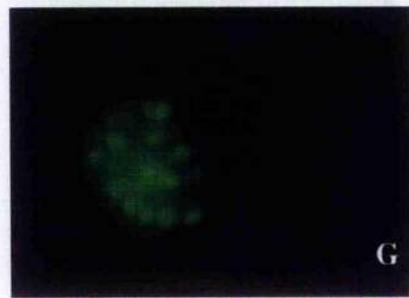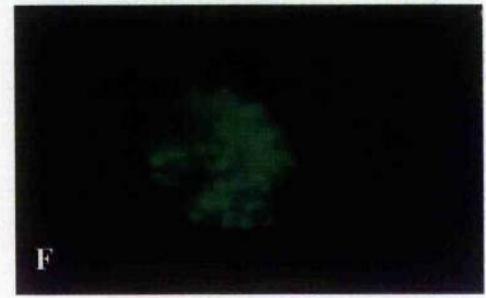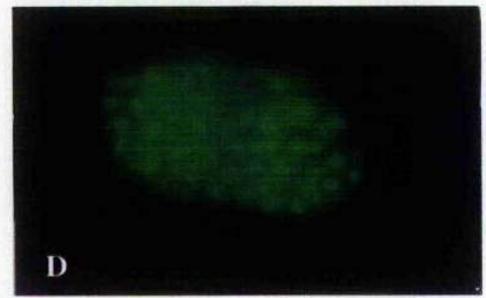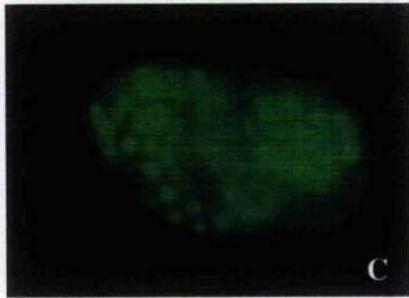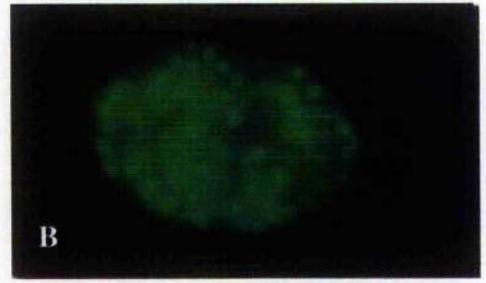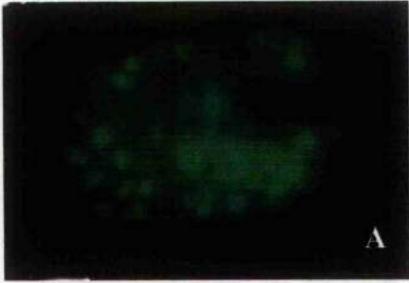
Chapter 5: Investigation of the regulation of *Hc-elt-2* using *C. elegans* as a heterologous system

5.1 Introduction

5.1.1 The use of *C. elegans* to study parasite gene regulation

The study of gene regulation in parasitic nematodes has been limited by the lack of transformation systems for these organisms. Although in recent years there have been great advances in DNA transformation techniques available for many biological systems, there are still only a few examples of successful application to parasitic nematode species. Transient transfection has now been achieved for several parasitic nematode species using microinjection and particle bombardment techniques (table 5.1). The microinjection of DNA into the nematode *Strongyloides stercoralis* resulted in reporter gene expression in gonadal tissue, close to the region of needle penetration and in progeny in the uterus during early embryogenesis (Lok, J. B. *et al.* 2002). Particle bombardment of adult *Brugia malayi* resulted in hypodermal and cuticular expression (Higazi, T. B. *et al.* 2002). When a transgene was introduced in *Brugia malayi* both by microinjection and biolistic approaches, differences in their expression pattern could be observed, most probably due to the deposition of DNA in different tissues by the two methods (Higazi, T. B. *et al.* 2002). It was observed that the microinjected adult expressed GFP from the reporter construct within the intra-uterine embryos, whereas the adults transfected through bombardment showed a punctuate pattern of fluorescence that localized to the hypodermal layer underlying the cuticle (Higazi, T. B. *et al.* 2002). However, there is at least one example where transgene expression was observed at a distance from the bombardment impact site: transgene expression in *Litomosoides sigmodontis* was

distant from the gold particle used to introduce the DNA, suggesting that the plasmid constructs may be transported throughout the entire organism (Jackstadt, P. *et al.* 1999). All the examples discussed above are of transient transfection. The only example where there has been transmission of a transgene through more than one generation was reported by Hashmi and colleagues (1995). They observed that microinjection of the nematode *Heterorhabditis bacteriophora* resulted in around 7% of the F1 progeny exhibiting *lacZ* expression and this persisted, although at reduced frequency, for several generations. The frequency of transmission started to decrease after the third and fourth generation, suggesting that the DNA, as in *C. elegans* is not integrated in the genome and thus tends to be lost in the population overtime (Hashmi, S. *et al.* 1995).

One of the main factors that make both the accomplishment and application of DNA transformation difficult in parasitic nematodes is the lack of *in vitro* culture systems. This difficulty, together with the practical limitations of transient transfection, limits the degree of experimental sophistication that can be currently achieved using DNA transformation in parasitic nematode species. For example, *H. contortus* cannot be maintained *in vitro* beyond the L4 stage. At present, DNA transformation has not been successfully applied to *H. contortus* and stable transfection would require the infection of sheep with transformed infective L3 to enable the transformed animals to reproduce and complete the life cycle. Even if successful, this would result in a labour intensive and expensive transformation system difficult to maintain in the laboratory.

Consequently, there has been a lot of interest in using *C. elegans* as a surrogate expression system to study parasite gene regulation (Britton, C. *et al.* 1999;

Krause, S. *et al.* 2001; Gomez-Escobar, N. *et al.* 2002; Boag, P. R. *et al.* 2003). *C. elegans* has previously been used as a heterologous expression system for the study of promoter regions of genes from a number parasitic nematode species. Upstream sequences from a *H. contortus* gut pepsinogen gene (*pep-1*), an *H. contortus* cysteine protease gene (*AC-2*) and the *Ostertagia circumcincta* cuticular collagen gene (*colost-1*) were cloned into the *C. elegans lacZ* reporter vector pPD95-03 (Britton, C. *et al.* 1999). These reporter gene constructs were then transformed into *C. elegans* and the *lacZ* expression was examined and, in the case of *pep-1* and *AC-2*, compared to the wild-type expression pattern of each gene. In the case of the *O. circumcincta* gene *colost-1*, the expression pattern of the gene is not known but it is predicted to encode a cuticle collagen, and so is expected to be expressed in the hypodermis. The spatial expression of lac Z reporter gene under the control of each of these three promoters correlated with the expected expression of these genes in the parasite. However the exact timing of expression did not, suggesting that regulatory mechanisms influencing the timing of expression may have evolved more rapidly than those controlling spatial expression of structural genes.

In a different study, Gomez-Escobar and colleagues (Gomez-Escobar, N. *et al.* 2002) fused the 5' flanking sequences of the *B. malayi alt* genes to a β-galactosidase reporter construct. It was found that the constructs were expressed in *C. elegans* gut cells; expression was more predominant in the posterior gut cells than in anterior gut cells from larval stages through to adulthood. In contrast, the ALT proteins in *B. malayi* are found packaged in the oesophageal gland in the later phase of growth in the mosquito and in the infective L3 larvae (Gomez-Escobar, N. *et al.* 2002). This shows that although the *C. elegans* transcriptional machinery can recognise the *alt* promoters, the transgene expression does not correlate with the

stage- and tissue-specific regulation that is observed for the polypeptide in the parasite itself. Two other *Brugia* promoters have also been studied using *C. elegans* as a heterologous expression system with more encouraging results. The *B. malayi phy-1* 5' flanking sequence was shown to direct hypodermal GFP reporter gene expression in all life-cycle stages which matched the expression profile in the non parasite gene (Winter, A. D. *et al.* 2003). Also, 540 bp of 5' flanking sequence upstream of *B. pahangi hsp-83* gene directed CAT reporter gene expression in transgenic *C. elegans* in response to heat-shock, suggesting that the *B. pahangi* promoter responded in the appropriate manner to *C. elegans* trans acting factors directing the heat-shock factor response (Thompson, F. J. *et al.* 2001). *C. elegans* was also used as heterologous expression system to study the promoter region of the *Onchocerca volvulus* gluthatione-S-transferase gene *Ov*-GST1a by transforming an *Ov*-GST1a::GFP reporter construct in *C. elegans* (Krause, S. *et al.* 2001). The timing of the expression pattern of GFP in *C. elegans* correlated with that of *Ov*-GST1a in *O. volvulus*. In *O. volvulus*, GST1a is mostly expressed in the body wall and GFP in *C. elegans* was found to be expressed largely in the hypodermis, which is broadly consistent with this observation. However, some GFP expression was also found in the isthmus and terminal bulbus of the pharynx which has not been reported for GST expression in the parasite.

These studies suggest that for clade V nematodes (e.g. *H. contortus*), cis-regulatory elements directing spatial expression may be functionally conserved, whereas those directing temporal expression may be less well conserved. For nematodes belonging to Clades more phylogenetically distant to *C. elegans,* such as *B. malayi*, the situation is less clear. Spatial expression seems to be conserved for some genes but not for others.

5.1.2 Regulation of the *C. elegans elt-2* gene

Although the regulation of the expression of the *C. elegans elt-2* gene has not been extensively studied there is some information available. Firstly there is another GATA factor, *elt-4*, located 5336 bp upstream *elt-2* and thus the 5' regulatory regions for *elt-2* are thought to be located within the 5336 bp intergenic region (figure 5.8). A 5432 bp fragment covering the region directly downstream the stop codon of *elt-4* and including the first 96 nucleotides of *elt-2* (encoding the first 32 amino-acids of ELT-2) was used to direct the expression of an *elt-2::GFP/lacZ* reporter construct in transgenic *C. elegans* (Fukushige, T. *et al.* 2003). Reporter gene expression in the transgenic embryos was consistent with the expression pattern of ELT-2 as previously established by immuno-localisation; the polypeptide is first detected in the 2E-cell stage and restricted to the gut through all stages of embryogenesis and post-embryogenesis. Several deletion constructs of this promoter region were also used to drive reporter gene expression (J. McGhee, personal communication). These experiments revealed the presence of an enhancer element lying approximately 3.5 kb upstream of the ATG, which was found to be both necessary and sufficient to establish expression of *elt-2* in early gut development (Fukushige and McGhee, personal communication). They also reported that this domain contained numerous consensus GATA sites (WGATAR) and SKN-1 (ATT(G/A)TCAT) sites. Ectopic expression of *end-1* and *elt-2* in worms transformed with the reporter construct containing this enhancer element revealed that *end-1* can activate the reporter construct whereas *elt-2* does not (T. Fukushige, personal communication). This suggests that this element is required for END-1 to activate expression of *elt-2* but unlikely to be required for *elt-2* auto-regulation and

hence not required to maintain *elt-2* expression (Fukushige and McGhee, personal communication).

This chapter describes the cloning and characterisation of the *H. contortus* *elt-2* genomic locus and comparative sequence analysis with *C. elegans*. Experiments to test the ability of the *H.contortus elt-2* 5' flanking sequence to direct the expression of a GFP reporter gene in transgenic *C. elegans* are presented. Finally, attempts to rescue the *elt-2* RNAi phenotype with the *H.contortus* genomic locus are also described.

5.2 Results

5.2.1 Isolation and characterisation of a λEMBL3 clone encompassing the *Hc-elt-2* genomic locus

5.2.1.1 Library screening

A λ EMBL3 *H. contortus* genomic library (provided by Dr M. Roos) was screened in order to isolate a clone containing the *Hc-elt-2* genomic locus. Approximately 100,000 pfu were screened by hybridisation with probe Hcelt2 5' using methods described in section 2.2.14. The Hcelt2 5' probe corresponds to 651 bp of the 5' end of the *Hc-elt-2* cDNA sequence from nucleotide 122 to 773 (figure 5.1). Three clones that hybridised to the probe on the primary screen were picked and subjected to a secondary screen and just one clone remained positive. This clone was purified by isolating a single hybridising plaque on a tertiary screen and was designated GC1. Phage DNA was prepared using a QIAgen Lambda Midi kit as described in section 2.2.15 for restriction mapping experiments.

5.2.1.2 Restriction mapping of GC1

A restriction map of clone GC1 was determined by undertaking a series of single and double restriction digests. Following restriction digestion, between 2 and 5 µg of phage DNA was separated on 0.8% agarose gels which were stained with ethidium bromide. The sizes of individual restriction fragments were determined by plotting the migration distance (x) against the log of the size of fragments (y) of the marker used on those gels (data not shown). Some of the gels were then Southern blotted and probed with one of two non-overlapping cDNA probes: the previously described probe Hcelt2 5' probe and a second non-overlapping 683 bp probe, Hcelt23', which corresponds to the 3' region of the cDNA sequence from nucleotide 785 to 1468 (figure 5.1). A *SalI* restriction digest released the insert from the vector arms to produce fragments of 1.4, 1.5, 5, 6.8, 9 kb and a large fragment above 20 kb. The 9 and >20 kb fragments correspond to the size of the vector arms and the sum of the sizes of the remaining fragments suggests the clone insert size is 14.7 kb. A series of single digests were performed with six different 6 bp cutter restriction enzymes and one 8bp cutter restriction enzyme (*NotI*) in order to identify several enzymes that cut inside the insert between 1 and 4 times but did not cut within the vector arms. On this basis *XbaI*, *NotI* and *SalI* were chosen for mapping purposes and a series of single and double digestions were performed. The results of these are presented in table 5.2. These digests were used for Southern blot analysis. Duplicate sets of digests were Southern blotted and each hybridised with one of the two probes described above using a final wash strength of 1X SSC as described in section 2.2.11. The blot obtained after hybridisation with each probe for GC1 is shown in figure 5.3 and the hybridising fragments are summarised in table 5.2. A number of the digests shown on the figure are partial, resulting in some additional bands. The

table documents the products considered to be the results of full digestion based on a number of blots These results were used to determine a restriction map of the GC1 insert (figure 5.3).

The restriction map on figure 5.3 shows the position of the restriction sites and fragments that hybridised to the 5' and/or 3'probe. This map allowed the determination of the orientation of the gene within the λ clone. The 5' probe hybridised to fragments close to the left arm of the vector, while the 3' probe hybridised to fragments further away from the left arm. Hence, the gene orientation 5' to 3' is within a left arm to right arm orientation of the vector. The map, combined with the hybridisation pattern of each fragment also gave an indication of which fragments to subclone to obtain the full genomic sequence of the *Hc-elt-2* genomic locus. It was decided to clone three overlapping fragments, shown on figure 5.3 as thick block lines below the restriction map. The first fragment, corresponding to the 2.4 kb *SalI/NotI* fragment, was cloned into vector pBSII (Stratagene, USA) as described in section 2.2.5. The remaining two fragments proved difficult to clone as restriction fragments, and so they were PCR amplified with primers Ann1 and Ann2 and primers Ann9 and Ann14, respectively (primers sequences in appendix 3 and position on figure 5.3). Both PCR products were cloned into the pCR2.1 vector from the TA cloning kit (Invitrogen Life Sciences, USA) as described in section 2.2.5. The *SalI/NotI* fragment cloned into pBSII produced plasmid pGC1P, and the plasmid containing the PCR products amplified with Ann1/Ann2 and Ann9/Ann14 were designated pGC11N and pGC123, respectively.

5.2.1.3 Sequence analysis of the *Hcelt-2* genomic locus

The inserts of plasmids pGC1P, pGC11N and pGC123 were sequenced on both strands and assembled to provide a contiguous sequence of 6082 bp which encompassed the full *Hc-elt-2* genomic locus of *Hc-elt-2* (see appendix 5.1 for full sequence). The position of introns and exons was established by alignment of the cDNA sequence against the genomic sequence (see figure 5.4). A total of 12 introns and 13 exons are present in the *Hc-elt-2* gene. All exon/intron boundary sites are consistent with the GT/AG rule (Breatbnach, R. *et al.* 1981), whereby all introns start by the nucleotides GT and end by the nucleotides AG. However, the 5' splicing sites did not perfectly matched that of the *C. elegans* consensus splice site as described by Blumenthal (Blumenthal, T. *et al.* 1997). This consensus follows the sequence GURAGUUU for the first eight nucleotides of the intron. This exact consensus was not found on any intron, however 7 of the 12 introns had 6 out of 8 nucleotides matching this consensus (see table 5.4), while the remaining five introns had 4-5 nucleotides out of 8 matching the consensus. Nevertheless, 2 introns carried the exact 3' splicing site (UUUUCAG for the last seven nucleotides of the intron) as described by Blumenthal (Blumenthal, T. *et al.* 1997) and 6 introns had 6 out of 7 nucleotides matching the consensus for the 3' splice site (see table 5.4).

The size of introns varied from 54 to 1705 bp and that of exons ranged from 77 to 196 bp. A greater number of introns is present in *Hc-elt-2*, and they are generally of much larger size than those found in *C. elegans elt-2*. Consequently the *Hc-elt-2* genomic locus is much larger (6082 bp) than that of the *C. elegans elt-2* gene (2201 bp). The *H. contortus* and *C. elegans* ELT-2 polypeptide sequences were compared in chapter 3. Except for the DNA-binding domain and the pseudo zinc finger domain, the two polypeptides share very little amino acid sequence homology

and so it is difficult to compare the positions of most of the introns between them. However, the DNA-binding domain and pseudo-finger are highly conserved between the two genes and can be precisely aligned. This reveals that intron 7 in *Hcelt-2* interrupts the gene between precisely the same amino acid residues as intron 6 in *C. elegans elt-2*, as shown in figure 5.5. This conservation of intron position supports the hypothesis that this gene is the true orthologue of *C. elegans elt-2*. The 5' breakpoint of the *Sal*I-*Not*I fragment is only 418 bp upstream of the ATG of the *Hcelt-2* gene. Since this fragment is adjacent to the left arm of the vector, this represents the extent of the 5' flanking sequence present in the GC1 lambda clone. Hence, it was considered necessary to isolate further upstream sequence in order to investigate the ability of the *Hc-elt-2* promoter to direct reporter gene expression in *C. elegans*. Consequently, a second *H.contortus* genomic library was screened to isolate another clone containing a genomic fragment overlapping with the GC1 genomic insert.

5.2.2 Isolation and characterisation of a λDASHII genomic clone encompassing further 5' flanking sequence of the *Hc-elt-2* gene

5.2.2.1 Library screening

A λDASHII *H.contortus* genomic library (kindly provided by R. Beech) was screened as described in sections 2.2.12, with probe Hcelt25'. Three putative positive genomic clones were isolated from a first round of screening and two of these remained positive after a third round of screening. Phage DNA was prepared form the two clones, designated clone 4 and clone 5, as described in section 2.2.12.

5.2.2.2 Analysis of phage clones and restriction mapping

A series of single restriction digests were performed on clones 4 and 5 with a range of 6-bp cutter enzymes in order to identify enzymes that cut within the insert but not the vector arms. Enzymes that had previously been used for the GC1 restriction mapping were included so that the restriction pattern of the two new clones could be compared for a possible overlap with GC1. The enzymes selected on this basis were *Bam*HI, *Eco*RI, *Not*I and *Xba*I. Restriction digests of clones 4 and 5, run on 0.8% agarose gels, were identical to each other (Figure 5.6A). Consequently, only clone 4 was chosen for further analysis and it was designated GC2. The single digests using *Xba*I and *Not*I produced a restriction pattern very similar to that of GC1 digested with the same enzymes, and restriction digests of GC1 and GC2 with *Xba*I were run side by side for comparison as shown on figure 5.6B. The figure shows (red asterisk) that two fragments of approximately 3 and 5kb are present in both *Xba*I digests. The gel was Southern blotted and probed with the cDNA fragment Hcelt23' (as described in section 2.2.9). The blot (figure 5.6C) shows that the 3kb *Xba*I fragments of both GC1 and GC2 hybridise to probe Hcelt23', confirming that it is a common fragment to the two clones. Using this information the GC2 clone was aligned to the restriction map of GC1 (figure 5.3).

The alignment of GC2 with GC1 suggests it contains additional 5' flanking sequence. This region was PCR amplified using a vector specific sense primer (corresponding to sequence in the λ DASHII left arm) in conjunction with the *Hc-elt*-2 cDNA specific antisense primer GC2400 (figure 5.3B and appendix 3 primer sequence) and cloned into vector pCR2.1 (TA cloning kit, Invitrogen Life Sciences, USA) to produce plasmid pAC12.

5.2.2.3 Analysis of *Ce-elt-2* and *Hc-elt-2* 5' flanking sequences

The pAC12 insert was sequenced in its entirety and was found to be 3765 bp in length. It included 380 bp of coding sequence and 418 bp of 5' flanking sequence already obtained from the GC1 clone. Hence the GC2 clone contains a total 3384 bp of 5' flanking sequence upstream of the initiator ATG start codon of the *Hc-elt-2* gene. The 3384 bp of *Hc-elt-2* 5' flanking sequence was aligned with 5336 bp of *C. elegans elt-2* flanking sequence using the GCG FASTA alignment program, which failed to reveal significant regions of homology.

In the case of *C. elegans elt-2*, it has been suggested that a stretch of 850 bp located 3.5 kb upstream of the initiator ATG containing several clusters of GATA and SKN1 consensus binding sites acts as an enhancing element (T. Fukushige and J. McGhee, personal communication). However, when I analysed the 5336 bp of 5' flanking sequence of *C. elegans elt-2*, I could not find any SKN-1 sites as described by Pal and colleagues (Pal, S. *et al.* 1997), who defined the SKN-1 consensus site as ATTGTCAT or ATTATCAT. However, 23 sites corresponding to the simplified SKN-1 consensus RTCAT (where R can be A or G) were found over the 5336 bp, corresponding to an overall frequency of approximately 4 sites/kb. The *C. elegans elt-2* flanking sequence was also searched for GATA sites and a total of 83 GATA/C sites were found in the same region between *elt-2* and *elt-4* (that is 15 sites/kb overall), but only 31 of these corresponded to the accepted WGATAR consensus (that is 5 sites/kb overall). The enhancer element reported by Fukushige and colleagues is located between −3865 and −3016 bp relative to A of the *C. elegans elt-2* ATG. This region was searched for GATA sites and SKN-1 sites (RTCAT). A total of 8 SKN-1 sites and 8 GAT(A/C) sites were found, 5 of which are canonical

WGATAR sites. These figures appear low, however they are equivalent to a frequency of approximately 10 sites/kb for SKN-1 sites and GAT(A/C) and around 6 sites/kb for WGATAR sites. Hence the frequency of SKN-1 sites in the 849 bp element is twice that of the SKN-1 sites in the whole of the 5' flanking sequence, suggesting that this could be a significant cluster. However, the frequency of WGATAR sites is similar between the element and the rest of the sequence (5 and 6 sites/kb respectively). Therefore, although this region has been shown to be of functional significance, there is no obvious clustering of GATA factor binding sites within it.

The 3384 bp of *Hc-elt-2* upstream sequence was analysed in a similar manner and the results are summarised in figure 5.7 and 5.8. A total of 17 SKN-1 sites (simplified consensus RTCAT) could be identified, corresponding to a frequency of around 5 sites/kb, which is similar to that found for the entire *C. elegans elt-2* 5' flanking sequence. A total of 44 GAT(A/C) sites were found (a frequency of 13 sites/kb) and of those, 21 were classical WGATAR sites (a frequency of 6 sites/kb), which again similar to that found in the *C. elegans elt-2* flanking sequence. Two potential clusters of GATA sites were identified. The first one includes 6 GATA sites in a stretch of 180 bp between 2834 and 2983 bp upstream the ATG (a frequency of 33 sites/kb, more than two-fold greater than compared to the rest of the sequence) and the second one is composed of 5 GATA sites in a 70 bp region between 434 and 783 bp upstream the ATG (a frequency of 71 sites/kb, more than five-fold greater than compared to the rest of the sequence). Furthermore all the sites in these two clusters, except one, correspond to the consensus WGATAR (figure 5.7 and 5.8). Interestingly, there is only one SKN-1 site in each of those two clusters.

Hence over half of the WGATAR sites in the whole 3384 bp *Hc-elt-2* sequence are located in these two clusters.

5.2.3 Assay of *Hc-elt-2* promoter activity using *C.* elegans as a surrogate expression system

5.2.3.1 Fusion of *Hc-elt-2* regulatory regions to the reporter vector pPD95-75

The aim of this part of the work was to investigate the ability of *Hc-elt-2* 5' flanking sequences to direct expression of a reporter gene in transgenic *C. elegans*. This was to investigate the presence of *H. contortus* regulatory elements in the 3.3 kb of 5' flanking sequence present in the GC2 clone and also test the extent to which *elt-2* regulation is conserved between *C. elegans* and *H. contortus*. The first approach to bring a GFP/lacZ reporter gene under the control of the *Hc-elt-2* regulatory elements was to clone the *Hc-elt-2* flanking sequence (from -3384 bp to +1842 bp relative to ATG) into the MCS of the *C. elegans* reporter gene vector pPD96-04. This vector is one of a set of vectors designed by A. Fire and colleagues, which includes an open reading frame for both lacZ and GFP and a multi-cloning site allowing for cloning regulatory regions upstream these ORFs (Fire, A. *et al.* 1990; Fire, A. 1995). Attempts to clone a 5226 bp fragment containing the 5' flanking sequence, the first exon, the first intron and the first two nucleotides of the second exon into the MCS of the pPD96-04 reporter vector to produce a translational fusion were unsuccessful. Consequently, a different approach was taken: instead of cloning the desired fragment into the pPD96-04 vector, it was decided to use a PCR-fusion-based approach to produce a reporter gene fusion with vector pPD95-75 (also from of the Fire vector kit). This approach entails fusing two primary PCR products

by a second PCR reaction with a set of nested primers (Hobert, O. 2002). One of the primary PCR fragments corresponds to the putative promoter region of the gene under study and the other primary PCR fragment corresponds to the linearised vector. The two primary PCR fragments are made to overlap by virtue of a 5' tag on the downstream primer used to amplify the first fragment, which is complementary to the 5' end of the second primer. The approach is shown schematically and explained further in figure 5.9. The PCR product consisting of the final fused reporter fragment is simply diluted to a concentration of 20-50 ng/μl and used directly for microinjection in *C. elegans* with no further purification. This technique has the advantage of being rapid, as it does not involve any ligation or cloning steps and there is no DNA purification step. Also, because there is no cloning step, it is possible to fuse large fragments using this method by adjusting the PCR conditions for long-range PCR. In this particular case the ProofStart enzyme kit (QIAgen, USA) was used according to the manufacturer's guidelines for long range PCR. Figure 5.9 also shows a gel with all three PCR products, confirming that they are of the expected size. The final reporter gene fusion is shown in figure 5.9, panel C.

5.2.3.2 Expression of GFP in transgenic *C. elegans*

The fusion PCR product obtained as described in figure 5.9 was injected in N2 worms and two transgenic lines, JAS94 and JAS95, were produced. Both lines were examined and the same expression pattern was observed for both. Mixed stage worms were washed off plates and examined by microscopy under UV illumination on 2% agarose pads. GFP expression was not seen in the early stages of embryogenesis and expression was first observed in the gut of 2-fold embryos (Figure 5.10, Panel A). The GFP polypeptide encoded by the ORF in pPD95-75

vector does not include a nuclear localisation signal (NLS) and so GFP expression is localised to the cytoplasm. GFP expression in the cytoplasm of the gut was seen throughout subsequent development in three-fold embryos L1, L2, L3, L4 and adult stages. The level of expression was highest in two-fold embryos, three-fold embryos and L1 larvae, and was markedly reduced in subsequent developmental stages. Not all worms in the JAS94 and JAS95 transgenic lines showed expression and the expression pattern was variable between worms. This is a typical observation because of the way in which extrachromosomal arrays are transmitted in *C. elegans*, with between 10 and 90% of progeny inheriting the array (Mello, C. C. *et al.* 1992) as well as random mitotic loss of arrays (Fire, A. 1995). Some animals showed expression mainly in the posterior gut, while others expressed GFP mainly in the anterior gut and others showed expression throughout the whole gut (Figure 5.10, panels B, C and D).

In addition to the expression pattern described above, GFP expression was seen in several cells outside of the gut. A number of cells in the second pharyngeal bulb showed GFP expression from the L1 to the adult stage (Figure 5.11, panel E). These were difficult to definitively identify but appear to be pharyngeal muscle cells. GFP expression was also seen in four cells adjacent to the pharynx in the L1 to adult stages and also two cells of the tail region (Figure 5.11, E, F and G). These cells are clearly all neurones since their axonal processes are visualised by the GFP. The location of a variety of neurons from these regions was examined on the Wormatlas web page (http://www.wormatlas.org/neuroimageinterface.htm), which provided information to select candidates for the identity of these cells expressing GFP. The four cells expressing GFP in the pharyngeal region are in a similar position to amphid neurons and their axons could be clearly seen to extend to the anterior of the

worm to near the oral opening. There are two pairs of amphid neurons ADF (left and right) and ADL (left and right). These two pairs of neurons constitute the likeliest candidates for the identity of the four cells adjacent to the pharynx. In the tail, there are several pairs of phasmid neurons that could correspond to the position of the two cells expressing GFP. These neurons are either PHA (left and right) or PHB (left and right).

In summary, these results indicate that the *Hc-elt-2* promoter region used in the reporter construct is capable of directing expression of GFP in *C. elegans* transgenic worms. The spatial expression corresponds broadly to the expression pattern of the parasite gene in that the gut represents the major expressing tissue. However, it appears that the temporal pattern of expression of the reporter gene differs from that of the endogenous *Hcelt-2* gene in the parasite since the transgene appears not to be expressed until the two-fold stage of embryogenesis Also several neuronal cells also express GFP in the transgenic worms. How this relates to the expression of the endogenous *C. elegans* and *H. contortus elt-2* genes is discussed in detail below.

5.2.4 Attempt to assay *H.contortus elt-2* function and regulation by rescue of the *C. elegans elt-2* RNA-mediated interference phenotype

A stringent test of gene regulation and function that is commonly used in *C. elegans* is to rescue a mutant phenotype with a transgenic copy of the gene to be tested. It is generally considered that in order for a transgene to successfully rescue the mutant phenotype, not only must all the essential aspects of gene function be fulfilled but the gene must also be expressed in the appropriate temporal and spatial

pattern. Hence if the rescuing transgene consists of a genomic copy of the gene all the key regulatory elements must be present for rescue to be achieved. Hence for a genomic copy of a parasite gene to rescue a *C. elegans* mutant phenotype, not only must the polypeptide be sufficiently conserved, but all the key regulatory elements must be present and conserved as well. The reporter gene experiments in the previous section suggest the GC2 genomic clone contains *Hc-elt-2* regulatory elements sufficient to direct appropriate gut expression in post-embryonic stages of *C. elegans*. Hence it seems possible that the GC2 clone might rescue a post-embryonic RNAi phenotype of the *C. elegans elt-2* gene. An RNAi phenotypic rescue approach has been previously successfully used to examine the function of the *H.contortus cpl-1* gene (Britton, C. *et al.* 2002). In this case, the *H. contortus* homologue of the *C. elegans cpl-1* gene was cloned under the control of the *C. elegans cpl-1* gene promoter region and the resulting construct was transformed in *C. elegans*. The transgenic line was then submitted to *Ce-cpl-1* RNAi and the results were compared *Ce-cpl-1* RNAi applied to wild-type *C. elegans*. Britton and co-workers found that the *H. contortus cpl-1* genecould indeed rescue the RNAi phenotype, demonstrating that the gene was successfully expressed and sufficiently conserved to replace *C. elegans cpl-1* loss of function. The effectiveness of this approach relies on the fact that a very high degree of sequence homology is required for a dsRNA molecule to interfere with an endogenous transcript. Hence, in general, the degree of sequence identity between *C. elegans* genes and their parasite orthologues is insufficient to allow dsRNA corresponding to the *C. elegans* gene to interfere with the transcript produced by a parasite orthologue transgenically expressed in *C. elegans*. This means that a parasite gene can be expressed efficiently

from a *C. elegans* transgene even in the presence of dsRNA corresponding to and interfering with the *C. elegans* orthologue.

elt-2 null mutants arrest development and die as L1 larvae with a severely deformed gut demonstrating that *elt-2* is an essential gene (Fukushige, T. *et al.* 1998). More recently, *elt-2* was included in the list of genes with an observable phenotype in the genome wide RNAi screen carried out by Ahringher and colleagues (Kamath, R. S. *et al.* 2003). The phenotype was described as developmental arrest at the L1 stage, and animals appeared clear i.e. lacked structure and *elt-2* RNAi worms were also reported to have an uncoordinated (Unc) phenotype. This phenotype was further confirmed by a second study using the RNAi hypersensitive *rrf-3* mutant strain (Simmer, F. *et al.* 2003). As a first step to develop an RNAi phenotypic rescue approach to test conservation of regulation and function of the *H. contortus elt-2* gene, the *C. elegans elt-2* RNAi phenotype was investigated in more detail.

5.2.4.1 Production of construct for RNA-mediated interference of *C. elegans elt-2*

A 632 bp fragment PCR was amplified with primers Ceelt2III and Ceelt2IV from *C. elegans elt-2* cDNA template (appendix 3 for primer sequence and figure 5.12) was cloned into the RNAi feeding vector L440 (Timmons, L. *et al.* 2001). The two primers were designed so that they carried 5' tag restriction sites for *Nhe*I and *Pst*I, respectively. The PCR product and L440 vector were both double digested with these enzymes and ligated as described in section 2.2.5. The resulting construct, pAC13, was then transformed into HT115 cells to generate a bacterial strain for the *C. elegans elt-2* RNAi feeding experiments. RNAi was performed, as described in section 2.1.3, on both embryonic and post-embryonic stages of two different *C.*

elegans strains: N2 and Nl2099, a *rrf*-3 mutant that is hyper-sensitive to the effects of RNAi (Simmer, F. *et al.* 2003) (appendix 2 for strain details)

5.2.4.2 RNA-mediated interference of *C. elegans elt-2*

(i) Application of *elt-2* RNAi to *C. elegans* throughout embryonic development

Two strains were used in these experiments: N2 and NL2099, an *rrf-3* mutant hypersensitive to RNAi. Two sets of RNAi plates were prepared, the first one was seeded with HT115 transformed with empty L440 as a negative control and the second set was seeded with HT115 transformed with construct pAC13 (subsequently referred to as RNAi-treated). Late L4 to young adult hermaphrodites were picked and placed onto RNAi plates and allowed to feed for 8 hours at 20°C to ensure that all the subsequently laid eggs had been adequately exposed to dsRNA. The adult hermaphrodites were then transferred to fresh RNAi (or control) plates (10 adults/plate) and left to lay eggs for 6 hours at 20°C. After this time, all adult worms were removed and the plates were left to incubate for 30 hours at 20°C. The number of hatched larvae and unhatched embryos were then counted on all plates and the sum of two sets of plates was used for statistical analysis (figure 5.12). On both the *elt-2* RNAi plates and the negative control plates for both the N2 and NL2099 worms the majority of embryos hatched (figure 5.12). The number of worms failing to hatch for strain NL2099 was just significantly different between RNAi plates and negative control plates (Fisher's exact test: $p < 0.0461$), but was not significantly different for strain N2 between negative control and RNAi plates (Fisher's exact test: $p=0.6702$). Hence *elt-2* RNAi does not appear to cause detectable embryonic lethality.

The plates were then left to incubate for a further 24 hours (hence 54 hours after being laid) to allow the larvae to develop and the number of L4 larvae and the number of L1-L3 larvae were counted (Figure 5.13) There is a very marked difference between the negative control and *elt-2* RNAi treated worms. 94.6 % of NL2099 larvae and 96.5 % of N2 larvae failed to reach the L4 stage after RNAi treatment, while for the negative controls, only 2.2 % of NL2099 larvae and 1% of N2 failed to reach the L4 stage over the same period of time. This is highly significant (Fisher's exact test: p<0.0001 for both N2 and NL2099) and suggests that inhibition of *elt-2* by RNAi has a dramatic effect on the development of the worms. The worms failing to develop to L4 after 55 hours of development appeared to be a range of stages from L1 to L3 but no attempt was made to count individual stages. This RNAi phenotype is entirely consistent with the *elt-2* null mutant phenotype in which development is arrested at the L1 stage with a very abnormal gut and evidence of feeding problems (Fukushige, T. *et al.* 1998). The effects observed in these experiments with *Ce-elt-2* RNAi is likely to be due to variable penetrance of the RNAi effect producing a wide a range of loss-of-function phenotypes from L1 to later arrest. The arrested worms appeared thin and starved in appearance, which is also consistent with abnormal feeding or gut development. The differences in the effect of *elt-2* RNAi between NL2099 and N2 are very slight and not significantly different (Fisher's exact test: p=0.2555). In this case the effect of *Ce-elt-2* RNAi is so strong on both strains that the hyper-sensitivity of NL2099 is not apparent. Figure 5.16 and 5.17 show views of negative control and *Ce-elt-2* RNAi plates at 55 hours of development to illustrate the phenotype. Hence *elt-2* RNAi applied during embryogenesis does not cause embryonic lethality but results in arrested and retarded larval development.

(ii) Application of *elt-2* RNAi to *C. elegans* by feeding during post-embryonic development

Several hundred hatched L1 larvae, were placed on *elt-2* RNAi and negative control plates (prepared as in the previous section) which were incubated at 20°C and subsequently observed after 48-50 hours (when worms were expected to have reached L4 stage) and then subsequently every 12-16 hours. After 48-50 hours, less than 10% of larvae on negative control plates had failed to reach the L4 stage for both strains (figure 5.14). The small proportion of larvae that had not reached L4 is probably reflection of the imperfect synchronicity of larval development due to differences in the time it takes individual larvae to start feeding on the bacterial lawn. In contrast, on the *elt-2* RNAi plates, more than 98% of N2 larvae and 100 % of the NL2099 larvae had failed to reach L4 stage at 48-50 hours of development. These differences between negative control plates and RNAi plates are highly significant (Fisher's exact test: p<0.0001 for both N2 and NL2099), but the difference in the proportion of larvae failing to reach L4 between the N2 and NL2099 RNAi plates is not significant (Fisher's exact test: p=0.0679). The retarded larvae on the RNAi treated plates appeared very thin and transparent with a lack of visible structure under the dissecting microscope and the NL2099 larvae seem to be more severely affected in this respect. This is consistent with the "clear" phenotype previously reported by Kamath and colleagues for this gene (Kamath, R. S. *et al.* 2003). After 75 hours, the majority of RNAi treated larvae have not yet reached the L4 stage (data not shown). After 96 hours, only a minority of RNAi treated larvae (14 % and 10 % for NL2099 and N2 respectively, data not shown) had not reached the size of L4 larvae, although the worms that were L4 or older did not appear normal or healthy and lacked internal structure. The *Ce-elt-2* RNAi treated larva also

appeared to move much slower than the L4 stage larvae present on the negative control plates. When plates were left for a further two days, a number of the worms did reach the adult stage on *elt-2* RNAi plates, but they appeared to have died and degenerated. The abnormal larvae showed a variety of abnormalities and phenotypes, throughout the course of the experiment such as dumpy and/or uncoordinated phenotypes (figure 5.15). A number of abnormal larvae were mounted on agarose pads and examined with Normaski optics and under UV illumination (figure 5.16). The gut lumen appeared present although deformed, and interestingly in all the abnormal larvae examined gut granules were still present although they appeared less dense and less evenly distributed throughout the animal's gut than in negative control larvae. This suggests that although a gut is formed, there are some abnormalities in its structure. The severity of the phenotype varies between larvae, which is likely to be due to variable penetrance of the RNAi effect between larvae as commonly observed for RNAi phenotypes particularly when the feeding method is used (Timmons, L. *et al.* 1998; Timmons, L. *et al.* 2001). Those worms that managed to develop as viable adults on *elt-2* RNAi plates often presented an egg laying defect phenotype (Egl), with some embryos having developed to the larval stage and moving inside the adult. In contrast, no Egl animals could be observed on the negative control plates. Hence the *C. elegans elt-2* gene appears to have an essential post-embryonic function in larval development.

5.2.4.4 Attempt to rescue the *C. elegans elt-2* RNAi phenotype by transgenic expression of *Hcelt-2*

Phage clone GC2 contains a 15.5 kb fragment encompassing the entire *Hcelt-2* genomic locus and includes 3.3kb of 5'flanking sequence and approximately 3.5

kb of 3' flanking sequence (Figure 5.3). In order for this clone to rescue the *C. elegans elt-2* RNAi phenotype, both the regulation and function of *H. contortus* gene needs to be conserved with that of *C. elegans elt-2* and the necessary cis-regulatory elements must be present. The reporter gene studies described above have shown that the GC2 clone contains *Hcelt-2* regulatory sequences that direct post-embryonic but not embryonic expression in the gut. Hence, it was decided to determine whether the GC2 clone could rescue the post-embryonic *C. elegans* RNAi phenotype.

The GC2 phage clone was injected into the gonad of NL2099 (*rrf-3* mutant) hermaphrodites at a final concentration of 5ng/μl along with the transformation marker plasmid pRF4 at a final concentration of 400 ng/μl. A single transgenic line was obtained which was designated AC19. *elt-2* RNAi was then applied to this transgenic line in order to test the ability of the GC2 containing transgene to rescue the RNAi phenotype. Two *elt-2* RNAi plates and two negative control plates were set up for both the AC19 strain and the NL2099 control strain by adding several hundred L1 hatchlings. The plates were incubated at 20°C for 48 hours and checked every 12-16 hours. The number of worms that had reached the L4 stage and the number that were still L1—L3 were counted and the number of roller and non-roller worms were also counted.

Figure 5.17 show the percentage of AC19 and NL2099 worms that failed to reach L4 at 48 and 72 hours of development on *elt-2* RNAi plates. The negative control plates are not shown on the graph for the clarity of the figure, but the percentage of worms failing to reach L4 on negative control plates were found to be 0 %. After 48 hours incubation on *elt-2* RNAi plates, only 94.2 % of NL2099 larvae and 100% of AC19 larvae had failed to reach the L4 stage. In contrast, on the negative control plates 0% of NL2099 and 0 % of AC19 worms had failed to reach

the L4 stage at 48 hours. This result confirms the penetrance of the *elt-2* RNAi in this experiment. After 72 hours 36.5 % of NL099 worms have failed reach the L4 stage and 23.5 % of AC19 worms have failed to reach L4. Although this is a small difference, it is statistically significant by Fishers exact test (p= 0.0035). This result tentatively suggests that the AC19 worms may be less affected by the *elt-2* RNAi with fewer worms failing to reach the L4 stage by 72 hours of development, which in turn would suggest that the transgene in the AC19 worms is providing some partial rescue of the *elt-2* RNAi phenotype. If the GC2 containing transgene does rescue the *elt-2* RNAi phenotype we would expect the transgenic worms in the AC19 line to be less affected by the *elt-2* RNAi than the non-transgenic worms in the same line. Hence we would predict that AC19 worms with a roller phenotype would have some growth/survival advantage over than wild type worms when grown on *elt-2* RNAi plates. The percentage of roller animals of the AC19 strain on *elt-2* RNAi and negative control plates is shown in figure 5.18. The percentage of roller phenotype worms on the negative control plates remains constant throughout the experiment, at, 22%, 26.3% and 27.2% at 48, 72 and 96 hours respectively. In contrast, the percentage of AC19 RNAi-treated roller phenotype worms increases over time; From 32.2 % at 48 hours, it reaches 64.1% at 72 hours incubation and 67.7 % after 96 hours. The difference in the proportion of roller worms between AC19 and NL2099 after 72 hours and 96 hours incubation is statistically significant (Fishers exact test, p<0.0001 for both time points). Hence worms carrying the GC2 containing transgene do appear to be at a growth advantage on *elt-2* RNAi plates. The results of the rescue experiment are preliminary evidence that there is some potential rescue of the RNAi phenotype by the GC2 clone. However the experiment

needs to be repeated several times in order to draw any firm conclusions regarding this.

5.6 Discussion

The purpose of this chapter was to investigate the extent to which the genomic locus and regulatory elements of the *elt-2* gene were conserved between *C. elegans* and *H. contortus*. Genomic library screening isolated two *Hc-elt-2* genomic clones. The first one, GC1, carried the full genomic locus of *Hc-elt-2* but only included 418 bp of 5' flanking sequence. The second clone, GC2, overlapped with GC1 and also encompassed the full *Hc-elt-2* genomic locus but extended further upstream and included a total of 3384 bp of 5' flanking sequence. Comparison of the genomic structure of the *Hcelt-2* gene with the previously sequenced *Ce-elt-2* gene revealed that *Hc-elt-2* contains more introns, which on average were of a larger size. Hence the *Hc-elt-2* genomic locus occupied 6082 bp of genomic sequence in total, compared to 2201 bp for the *C. elegans elt-2* gene. Similar observations have been reported previously in the case of the cathepsin L genes from *H. contortus*, *D. viviparus* and *C. elegans*, for which it was found that the parasite genes had more introns (9) than their *C. elegans* homologue (3) (Britton, C. *et al.* 2002). However only the first intron of each parasite *cpl-1* gene was found to be much larger than the introns in *C. elegans*, whereas the others were of smaller size. Interestingly, in *Hc-elt-2*, the first intron is also the largest one. The study of *H. contortus* and other parasite genes such as *gru-1* and *pep-1* have revealed that it is a common feature of *H. contortus* genes to have more introns than their homologues in *C. elegans* and

that introns also appear to be larger in *H. contortus* genes (J. Smith personal communication). For example, the *H. contortus* gene *gru-1* a homologue of the *C. elegans* β-tubulin gene *ben-1*, has 9 introns compared to only 4 in the *C. elegans* gene (Kwa, M. S. G. *et al*. 1993) and a homologue in the Strongylid nematode *Cylicocyclus nassatus* has 8 introns. It is difficult to precisely compare the intron positions of the *C. elegans* and *H. contortus elt-2* genes because of the lack amino acid sequence homology between the two polypeptides. However, the two introns (6 and 7 for *Hcelt-2* and 5 and 6 for *Ce-elt-2*) that are in the highly conserved DNA binding domain region have precisely conserved positions relative to the amino acid sequence. Typically, orthologous genes tend to have precisely conserved intron positions but in this case it is more difficult to establish this because of the low amino acid sequence homology. However, the fact that the only two introns that are in conserved regions of the polypeptide have precisely conserved positions is consistent with *Ce-elt-2* and *Hc-elt-2* being true orthologues.

Several approaches were used to try and identify potential regulatory element in the *Hc-elt-2* sequence. First, the 5' flanking sequence of the *Hc-elt-2* gene was compared to that of *C. elegans elt-2* but there were no striking regions of homology. GATA transcription factor and SKN-1 binding sites have been implicated in the regulation of the *C. elegans elt-2* gene a 600 bp region located 3.5 kb upstream the *C. elegans elt-2* ATG (Fukushige and McGhee, personal communication). Deletion experiments and cloning of this 600 bp fragment in reporter gene constructs have shown that it was both necessary and sufficient to direct expression of GFP (Fukushige and McGhee, personal communication). The details of this putative *C. elegans elt-2* enhancer have not been published and so the precise number GATA sites are not recorded, however its location has been communicated by T. Fukushige

as an 850 bp stretch (not 600 as previously reported) covering a zone from −3865 to −3016 bp upstream the ATG. A search to identify potential GATA and SKN-1 sites clusters revealed that this 80 bp region does contain several SKN-1 binding sites, with a frequency higher than that of the rest of the flanking sequence; in contrast the frequency of GATA binding sites is no higher in the 850 bp region than it is in the rest of the 5' flanking sequence. The analysis of the 5' flanking sequence of *Hc-elt-2* revealed that it contained a high number of GATA factor binding sites (44 GAT(A/C) and 21 of those of the canonical type WGATAR). These GATA sites appear to form two clusters: one covered the region from −3019 to −2839 and the second, closer to the ATG, covered a region from −504 to −434. None of these GATA clusters contained SKN-1 sites in a higher proportion than in the rest of the sequence; however, one of the GATA site cluster identified on the *Hc-elt-2* promoter region could correspond to the 850 bp region from *C. elegans elt-2* identified by Fukushige and colleagues. It is located at −2839 bp and hence is very similar in location to the *C. elegans elt-2* enhancer element and could play a similar role. However functional analysis and deletion experiments would be needed to establish its role as an enhancer. The presence of the two clusters of GATA binding sites suggests that this 3.3 kb 5' flanking region may be involved in the regulation of *Hc-elt-2* expression, as subsequent results from gene reporter constructs have shown.

A GFP reporter construct was made which was a translational fusion within the second exon of the *Hc-elt-2* gene and included 3386bp of 5' flanking sequence. Two independent transgenic lines carrying this reporter on extrachromosomal arrays were generated and the expression pattern determined. The major component of the expression pattern was gut suggesting that key tissue specific regulatory elements of the *H. contortus* gene were able to direct expression appropriately in transgenic *C.*

elegans. However in addition to this endodermal expression some pharyngeal muscle cells and neuronal cells either side of the pharynx and in the tail region also expressed the reporter gene. Although definitive identification is not possible the likeliest identity of the neurons are the pairs of amphid neurons ADF (L+R) and ADL (L+R) and the phasmid neurons PHA and PHB. Interestingly, amphid and phasmid neurons are functionally related (they are all chemosensory neurons) in *C. elegans* and the gene *osm-3* has been found to be expressed in the amphid ADF and ADL and the phasmids PHA and PHB (Tabish, M. *et al.* 1995).Hence it is possible that these neurones could share some common mechanisms of gene regulation. These non-endodermal aspects of the expression pattern were a surprise as the *C. elegans elt-2* gene is expressed only in endoderm. There are essentially three explanations for this additional expression of the *Hc-elt-2* reporter gene. Firstly, there could be important regulatory elements missing from the construct. It is extremely important to include all regulatory elements of a gene in a reporter construct because an incomplete promoter region can lead to either incomplete or ectopic expression depending on whether positive or negative regulatory region are missing (Aamodt, E. J. *et al.* 1991). However whilst this is always possible, there is no evidence from promoter deletion experiments of repressor elements upstream of the *C. elegans elt-2* gene and so if this were the explanation for the neuronal expression it would represent a significant difference in the regulation of the *H. contortus* and *C. elegans elt-2* genes. Secondly, it is possible that the additional components of the expression pattern is not a true reflection of the parasite *elt-2* gene but is an artefact of the heterologous expression system; that is the *H. contortus* regulatory elements are not sufficiently conserved to direct the *C. elegans* transcriptional apparatus in a fully appropriate fashion. Thirdly, it is possible that the

additional components of the expression pattern are a true reflection of the endogenous expression of the *Hc-elt-2* gene in the parasite. There is some evidence to support this for the expression in at least some of the neuronal cells: the Hc-ELT-2 specific antisera detects expression in two cells adjacent to the pharynx in *H. contortus* larvae in a position consistent with the posterior pair of amphid cells (see chapter 3). This staining is very weak but is very repeatable, and recent work has shown that the structure and function of amphid neurons in *H. contortus* are very conserved with that of *C. elegans* (Ashton, F. T. *et al.* 1999; Li, J. *et al.* 2000; Li, J. *et al.* 2000; Li, J. *et al.* 2001). Hence it is possible that the *Hc-elt-2* gene has an additional role in neuronal development, which *Ce-elt-2* does not have in *C. elegans*.

Another difference between the expression patterns of the reporter construct and of the gene in the parasite was that they differed in the timing of their onset. The reporter construct is only expressed in the 2-fold stage of embryogenesis while endogenous *elt-2* gene in both the parasite and *C. elegans* is expressed when the embryo is only 30-40 cells, a much earlier stage of embryogenesis. Similar to the situation discussed for the spatial expression, this difference could be due to the absence of a key regulatory element on the construct. However it is possible that the regulatory elements involved in temporal control of gene expression are not well conserved between *H. contortus* and *C. elegans*. Similar observations have previously been reported several other *H. contortus* promoters used to drive reporter gene expression in transgenic *C. elegans* (Britton, C. *et al.* 1999). Although the spatial expression of three *H. contortus* promoters (*AC-2*, *pep-1*, and *colost-1*) was conserved in *C. elegans*, the temporal expression pattern was different. In the case of *pep-1* and *AC-2* as the onset of expression of the transgene in *C. elegans* was much earlier than in the endogenous gene in the parasite .In the case of *colost-1*, although

the parasite gene expression pattern was not known its closest *C. elegans* homologue, *col-12*, is expressed at every developmental stage. In contrast the *colost-1* promoter only directed reporter gene expression in the L4 stage of transgenic *C. elegans*.

The RNAi phenotype of the *C. elegans elt-2* was studied in some detail. These experiments showed for the first time that *elt-2* has an essential role during post-embryonic development in addition to its role during embryogenesis. The phenotype was suggestive of an essential role in post-embryonic development of the *C. elegans* gut. The post-embryonic RNAi phenotype was then used as a means to assess the extent to which the *Hcelt-2* gene is functionally conserved with *C. elegans*. Although the ectopic expression experiments in Chapter 5 have shown that the parasite gene is sufficiently conserved to activate endodermal differentiation in *C. elegans* embryos, there could be other aspects of *elt-2* function that are not conserved. Phenotypic rescue represents a very stringent test of gene function. The ability of the genomic clone GC2 to rescue the *C. elegans elt-2* RNAi phenotype was investigated. These experiments provided preliminary evidence that the *H. contortus elt-2* gene may at least partially rescue the *C. elegans elt-2* RNAi phenotype. However the experiment does need to be repeated to validate this conclusion. The fact that the rescue is only partial could reflect imperfect functional conservation or alternatively could be simply due to the limitations of the experimental approach. Transgenic rescue is notoriously sensitive to levels of expression from transgenes. If more time were available, it would be interesting to transform this construct at a number of different concentrations to determine if different transgenic lines with different transgene copy numbers would rescue more efficiently. It would also be interesting to attempt to rescue the *elt-2* RNAi phenotype with *Hc-elt-2* cDNA

cloned under the control of *C. elegans elt-2* promoter. This might lead to more efficient transgene expression and more precise spatial and temporal regulation of expression. A similar experiment was carried out by Britton and colleagues with the *H. contortus* gene *cpl-1* (Britton, C. *et al.* 2002). They obtained "full" rescue of the *C. elegans cpl-1* RNAi phenotype the *H. contortus cpl-1* DNA under the control of *C. elegans cpl-1* promoter. Unfortunately, this experiment was not carried out for *H. contortus elt-2* due to lack of time.

Figure 5.1: Position of cDNA probes Hcelt25' and Hcelt23' on the *Hc-elt-2* cDNA sequence.

The probes are represented by thick block lines and the primers used to amplify the probe fragments are marked by arrows. The position in the cDNA sequence of *Hc-elt-2* is indicated by the numbers and arrows, A from ATG being postion +1. The probes do not overlap and hence can be used in conjunction for mapping purposes. The dark shaded box represents the DNA binding domain and the hatched box represents the "pseudo" zinc finger.
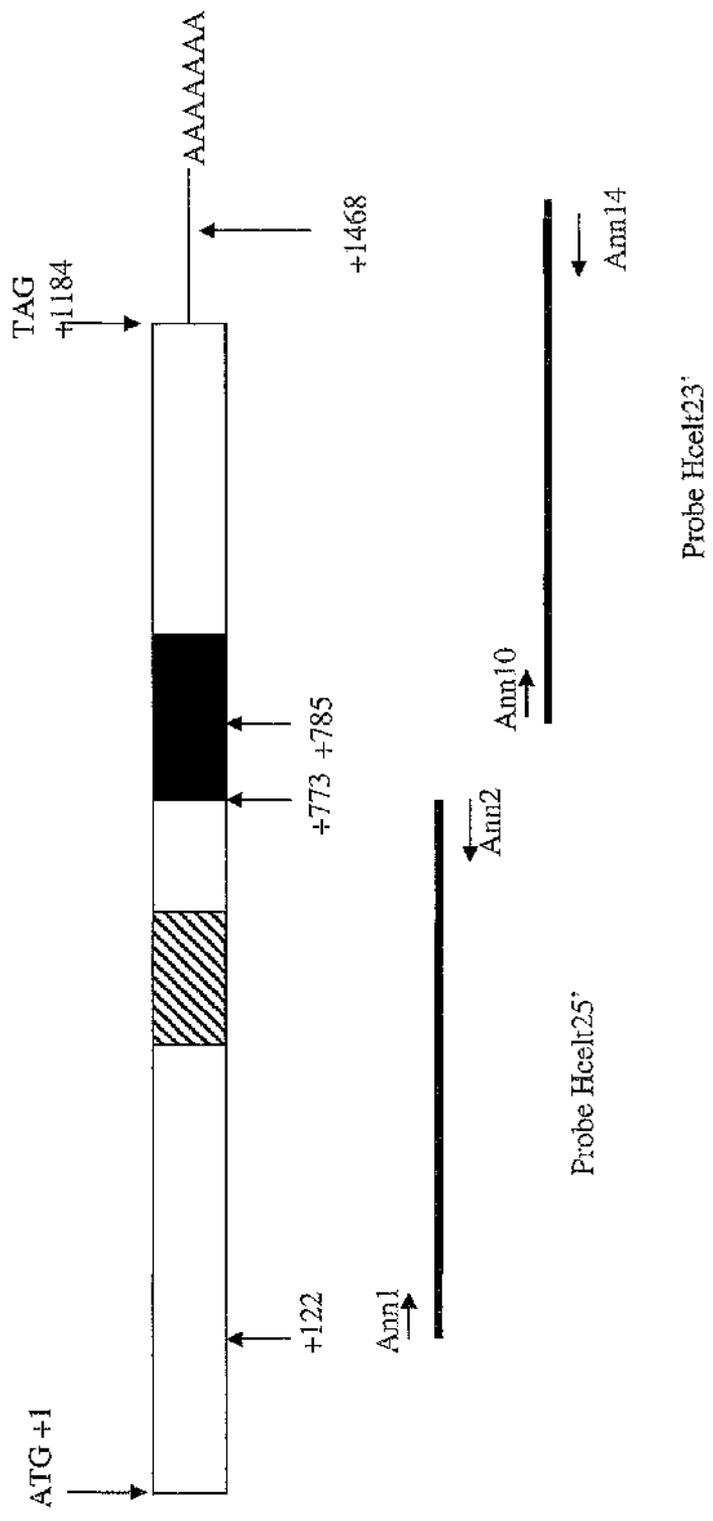
Table 5.2: Restriction digest of GC1

The table shows the number and size of restriction fragments (in kb), resulting from single and double digests of the λEMBL3 clone GC1. The sizes were approximated from a number of different independent digests and gels.

Digests	SalI	XbaI	XbaI + SalI	NotI + SalI	NotI	NotI + XbaI
Number of fragments*	4	2	7	5	2	3
Size of fragments (in Kb)	1.4 1.5 5 6.8	3 6	1.4 1.5 1.6 2.1 2.3 2.8 3.1	1.4 1.5 2.4 4.5 5	≈20 (too high on gel to be determined)	0.7 2.3 6

* excluding the two fragments corresponding to λ arms.

Figure 5.2: GC1 fragments hybridising to probes Hcelt25' and Hcelt23'

Southern blots of single and double restriction digests of the GC1 λ clone were hybridised with cDNA probes Hcelt25' (Panel A) and Hcelt23' (panel B). The hybridisation was performed at 42°C in the hybridisation solution provided with the DIG High Prime labelling and detection starter kit II (Roche), and the final wash stringency was at 42°C in 1X SSC, 0.1% SDS.
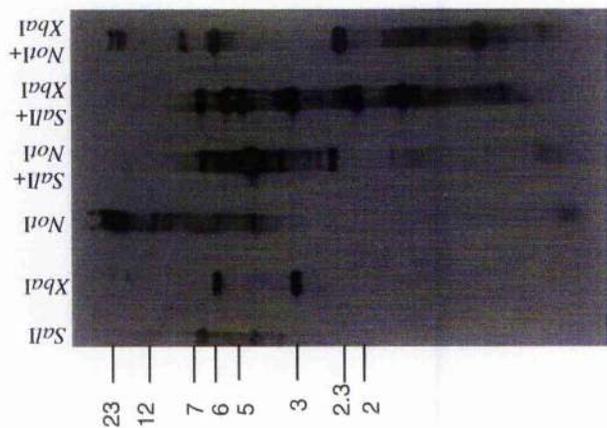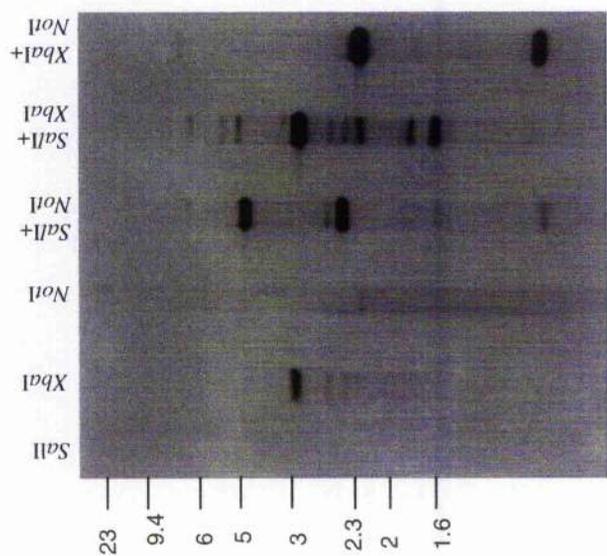
Figure 5.3: Southern blots of GC1 restriction digests probed with Hcelt25' (A) and Hcelt23' (B).

	Digests	SalI	XbaI	Not I	NotI + SalI	XbaI + SalI	NotI + XbaI
PrHcelt2 5' probe	Number of fragments	1	1	2	2	2	2
	Size (in kb)	6.8	3	>20	2.4 5	3.1	0.7 2.3
PrHcelt2 3' probe	Number of fragments	2	2	2	2	3	3
	Size (in kb)	6.8 1.5	3.1 6	>20	2.4 4.5	1.5 2.1 3.1	0.7 2.3 6

Table 5.3: Hybridising fragments of GC1 to probes Hcelt25' and Hcelt23'.

The table summarises the restriction fragments that hybridise to either one or both probes as established by Southern blot analysis (figure 5.2 A and B). *XbaI/SalI* digest on the blot shown on figure 5.2A and 5.2B is partial, resulting in additional hybridising fragments. Also the 6.8 kb *SalI* fragment failed to hybridise on this particular blot (the first track of panel A). The sizes of the hybridising fragments shown in the table were determined from a number of independent experiments.

Figure 5.3: Restriction map of GC1 (A) and GC2 (B).

The restriction sites for the enzymes used are marked by straight lines with the first letter of the enzyme (e.g. X for *Xba*I). A cross marks the fragments hybridising to Hcelt25' probe and a star marks fragments hybridising to Hcelt23' probe. The thick, overlapping block lines below GC1 represent the fragments cloned for sequencing and the primers used for their amplification when required are marked by arrows.

X = XbaI
S = SalI
N = NotI

＋ Hcelt25' probe

＊ Hcelt23' probe

Lambda EMBL3 right arm

Coding region

ATG

ann1 →
ann2 →
ann8 →
ann14 →

Total insert GC1 = 14.9 kb

Lambda EMBL3 left arm

(A)

Lambda DASHII right arm

←→ 1000 bp

ATG

GC2400

Total insert GC2 = 15.5 kb

Lambda DASHII left arm

(B)

5' flank = 3.3 kb

202
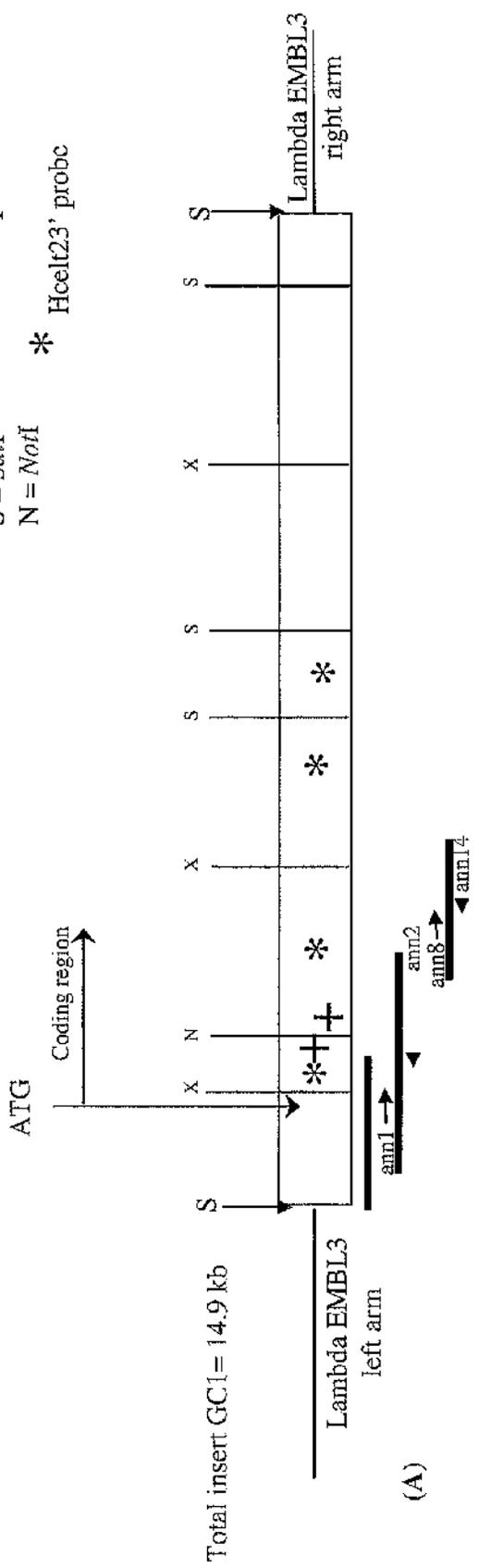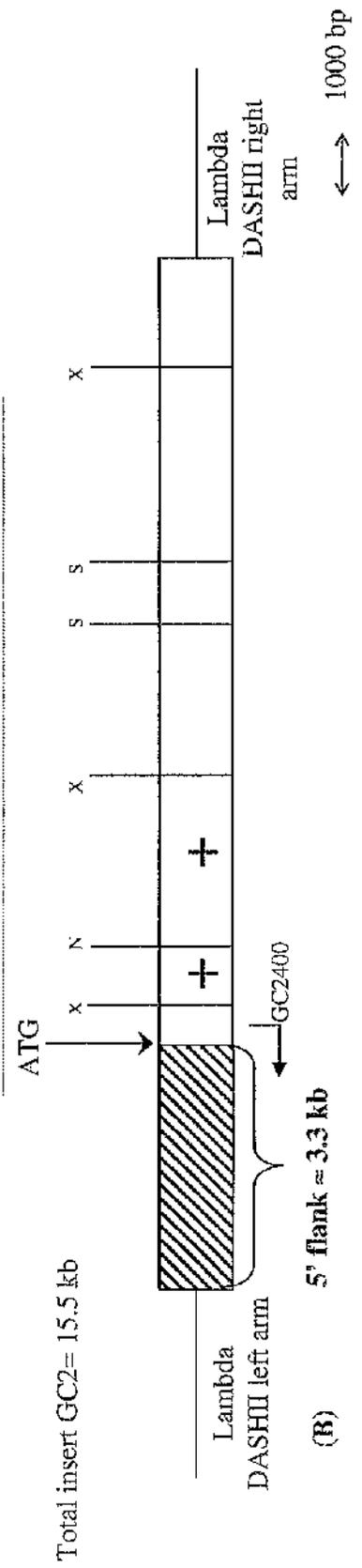
Figure 5.4: Genomic structure of *Hc-elt-2* and of *Ce-elt-2*.

The boxes represent exons and introns are marked by lines in between the exons. The pseudo zinc finger domain is represented by the hatched box and the DNA-binding domain is represented by the black boxes. The size of the exons in bp is shown above the figure and the size of the introns in bp below the figure. The A represents the start codon and the asterisk represents the stop codon.

Figure 5.5: DNA-binding domain of Ce-ELT-2 and Hc-ELT-2.

The arrow heads mark the position of the intron with conserved position between the two genes.

H.contortus

196 101 150 78 128 98 97 77 94 122 79 94 155

A

1705 282 54 148 74 581 78 97 61 842 239 452

C. elegans

156 86 127 141 169 98 374 65 341

A

50 43 44 94 45 46 164 158
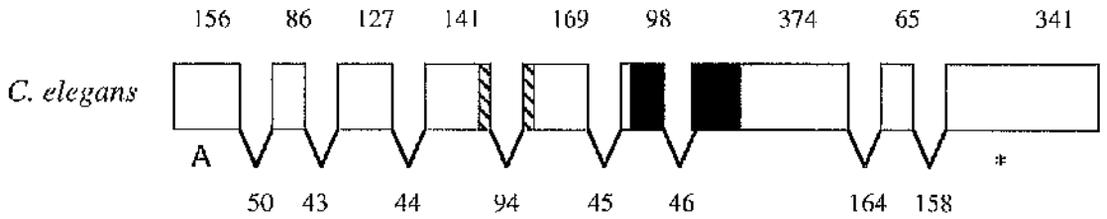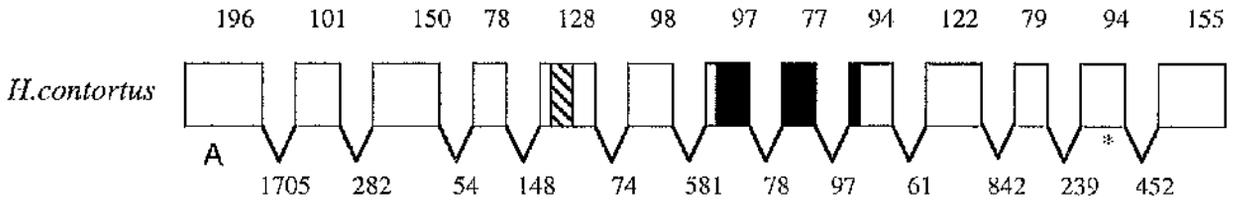
```
                 201                                              250
HCELT2   (196)   PTYTLYPPTSMEPRHTVEA---TPVSKSRSGSSHKKPSAAQNSQ----RR
ELT2     (183)   YSPPVAYSTSLGQPPILEIPSEQPTAKIAKQSSKKSSSSNRGSNGSASRR


                 251                            ▼                  300
HCELT2   (239)   QGLVCSNCGGTNTTLWRRNAEGEPVCNACGLYYKLHNVQRPPTMKKDCQL
ELT2     (233)   QGLVCSNCNGTNTTLWRRNAEGDPVCNACGLYFKLHHIPRPTSMKKEGAL
                                              ▲

                 301                                              350
HCELT2   (289)   QTRKRKAKSDGTNTC--------KKRDRSSNYTQSTQAIPDR--------
ELT2     (283)   QTRKRKSKSGDSSTPSTSRARERKFERASSSTEKAQRSSNRRAGSAKADR
```

Table 5.4: 5' and 3' splice sites of each of the *Hc-elt-2* intron:

The nucleotides highlighted in grey are conserved with the *C. elegans* consensus splice sites, as described by Blumenthal and colleagues (1997). The *C. elegans* consensus is shown at the top of each column.

	first 8 nucleotides of intron	last 8 nucleotides of intron
CONSENSUS	GTRAGTTT	TTTTCAG
INTRONS		
1	GTGCGTTC	CTTTCAG
2	GTAAGTAA	TTTGCAG
3	GTTTGTGC	ATTTTAG
4	GTATGGCT	TTTTTAG
5	GTAAGGAT	TTTCCAG
6	GTAAGTCA	TAATCAG
7	GTGTGTCT	TTTTTAG
8	GTGAGTCT	TGTAGAG
9	GTAAGAAC	ATTTTAG
10	GTATCCTA	ATTTCAG
11	GTATTTCG	TTTTCAG
12	GTGAAATT	TTTTCAG

Figure 5.6: Analysis of λ DASHII clones 4 and 5

Panel A: A series of single digests on clones 4 and 5 showing they are identical.

Panel B: Comparison of GC2 (clone 4) and GC1. Agarose gel shows *Xba*I digests of GC1 and GC2 and a *Not*I digest of GC2. An asterisk marks homologous fragments between digests fragments

Panel C: Southern blot of gel in panel B probed with Hcelt25'.

(A)

(B)

(C)

Figure 5.7: Schematic representation of *C. elegans elt-2* (A) and *Hc-elt-2* (B) genomic organisation.

Diagram (A) shows the position of the *elt-4* and *elt-2* genes in *C. elegans* and the 849 bp enhancer element on *elt-2* 5' flanking sequence (marked by double arrows) as determined from personal communication (T. Fukushige). GATA sites are represented by *. The enhancing element would therefore have eight GATA sites. Diagram (B) shows the *Hc-elt-2* gene and its 5' flanking sequence. Two clusters of GATA sites containing 6 GATA sites and 5 GATA sites respectively are shown.

(A)

(B)
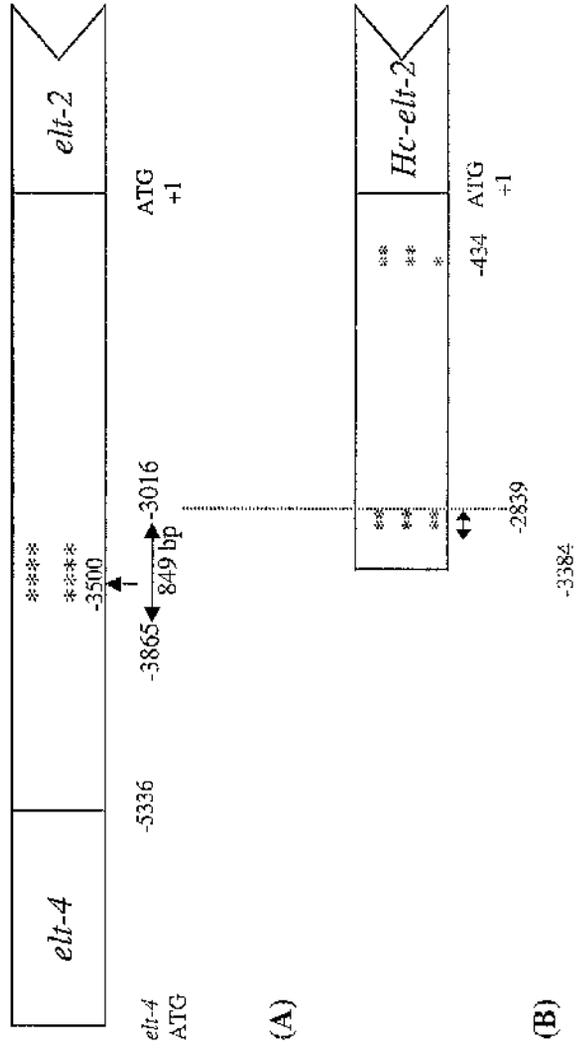
Figure 5.8: Clusters of consensus GATA binding sites identified in the 5' flanking sequence of *H. contortus*.

The GATA sites in each cluster are represented in blue for those on the sense strand, with the traditional consensus WGATAR circled in the same colour. The GATA sites present on the anti-sense strand are shown in pink, with those corresponding to WGATAR circled. Finally only one SKN-1 site was identified in each cluster and is shown in yellow. In cluster 2 it overlaps a GATA site and is therefore partly shown in red.

GATA CLUSTER 1:

5′ ACACACCAATGGGCTATCTGCACAATATTTTTGATTAGACATATCTATACAGCA

TCTATCTATCTCTAGACGACTATCAGGGGCCAATATGTTTTGAGTGATAAATTTCC

CCGTTTCCCCCGTAGAACCATGACAGCACCCACAGCTAAAGTTGGAATTCAAAAAA

TTTGGATTTAATTAAAAACTTTAGATAACTGGCT 3′

GATA CLUSTER 2

5′ CCTAATTGCGAATAAAATTGAAAACTGACACTGATAGGTTATCATACCTAAAGG

CCTTATACCTCGCTGGTCTTGATAGCTTCAGACTGATAAATTTATCT 3′

Figure 5.9: PCR fusion as an alternative to cloning for reporter constructs (adapted from Hobert 2002).

The approach consists of amplifying the promoter (or complete gene) under study by PCR with two primers A and B. Primer A sits at the 5' end of the fragment of interest and primer B at the 3' end. Primer B also has an overhang that is complementary to a region of the multiple cloning site sequence of pPD95-75. This is designed in such a way that the ORF of the gene interest can fuse in frame with the GFP ORF. In a second PCR, the GFP ORF and 3' UTR sequences of linearised vector pPD95-75 were amplified with primers C and D that lie at the multi-cloning site and the 3'UTR, respectively. After both the vector and the fragment of interest have been PCR-amplified, the products were quantified by agarose gel electrophoresis and diluted to 10-50 ng/µl and used as template in the next round of PCR. The products obtained for the two PCR reactions are shown adjacent to the diagrams in panels A and B, respectively. For the next round of PCR, a second set of primers, A* and D*was designed; Primer A* is nested with respect to primer A and primer D* nested with respect to primer D. The third PCR reaction using primers A* and D* on the first two PCR products as template produces an amplicon that is a fusion of these two products (panel C). The final fused product is shown on an agarose gel adjacent to the diagram in panel C.

(A)

(B)

(C)

PCR #1

PCR #2

PCR #3

A

B

C

D

A*

D*

1.6 kb

~5kb

2kb

1.6 kb

7kb

1.6 kb

PCR #1

PCR #2

PCR #3

208
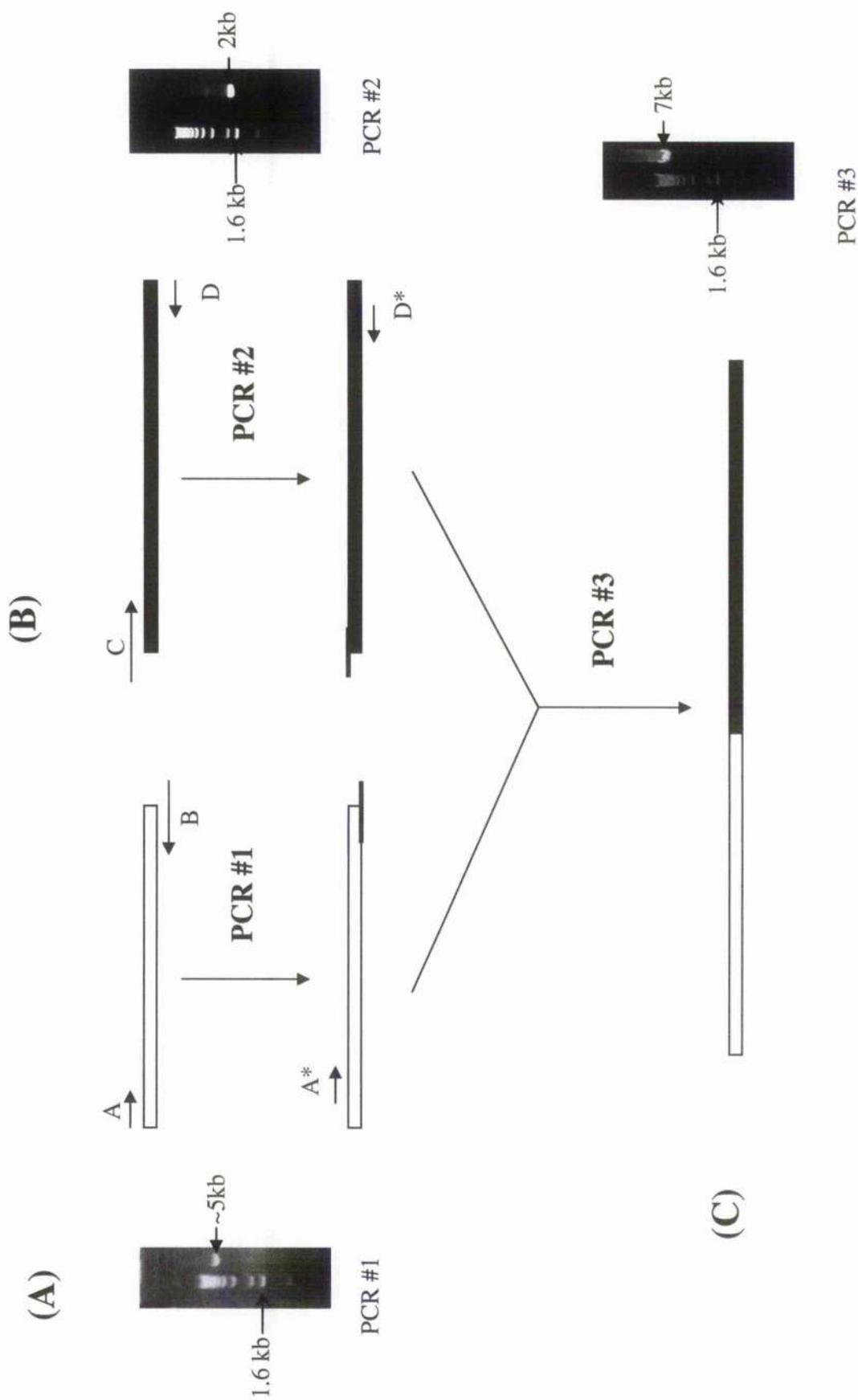
Figure 5.10: Expression of *Hc-elt-2::GFP* reporter gene fusion in transgenic *C. elegans* (expression is not nuclear-localised).

Expression of GFP is first detected in 2-fold embryos (pretzel stage) in the gut (Panel A) and in can then be observed in post-embryonic stages. Panels B, C and D show expression of GFP in L1 (green), superimposed on the endoderm-specific autofluorescent granules (yellow). Panels E and F show expression of GFP in four cells, thought to be amphid neurons (red arrows), while panel G shows expression of GFP in two tail neurons, thought to be phasmids, indicated by red arrows.
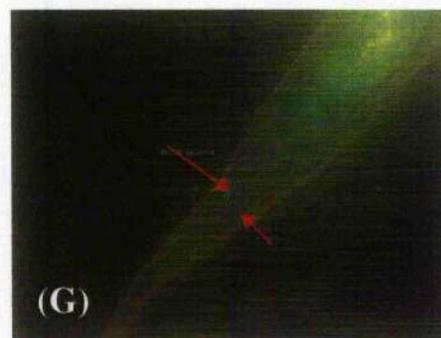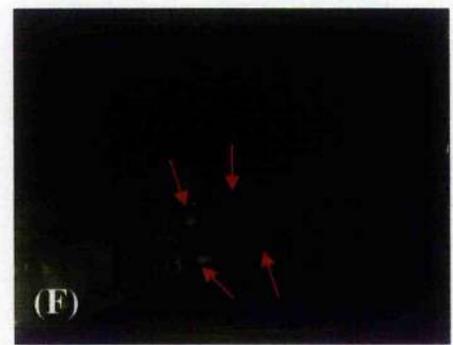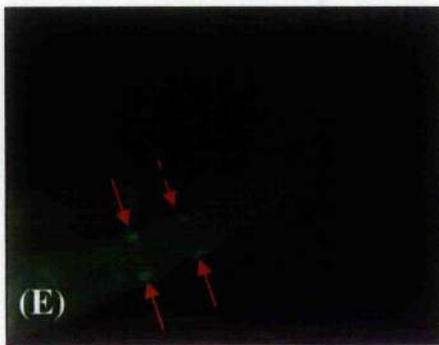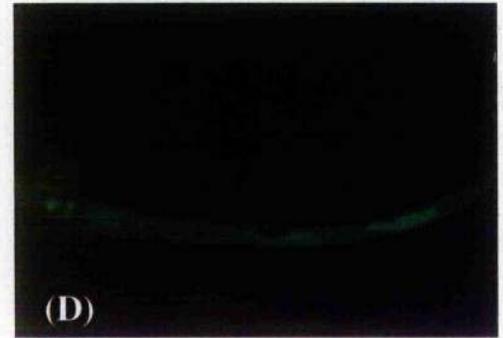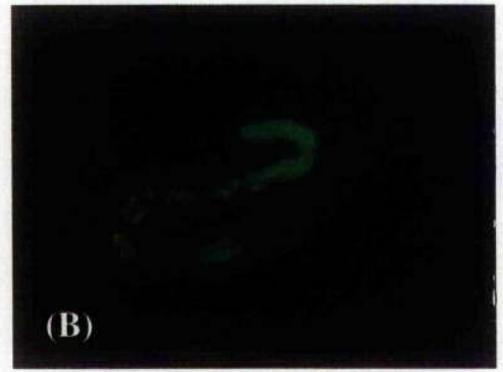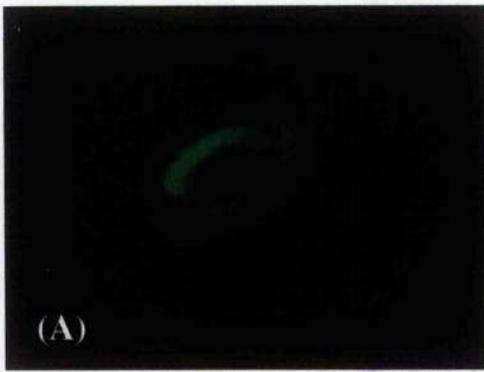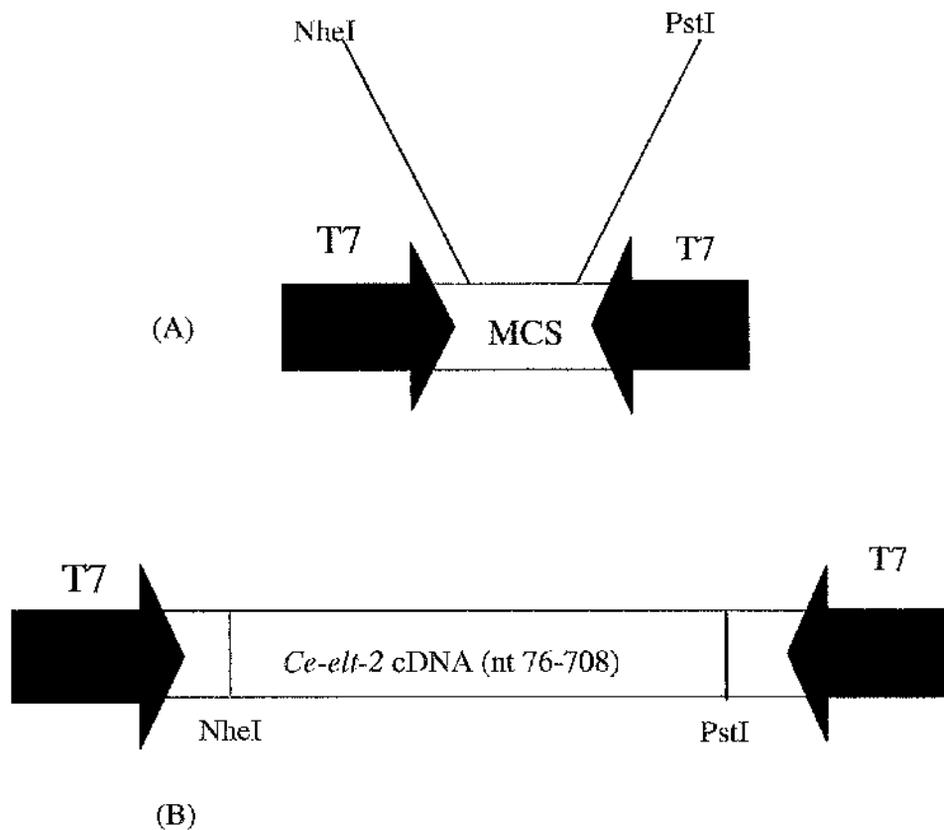
Figure 5.11: Cloning strategy to induce *elt-2* RNAi in *C. elegans.*

Diagram (A) shows the RNAi feeding vector L440, with two T7 promoters orientated in opposite directions either side of a multiple cloning site (MCS). This allows the transcription of double stranded RNA (dsRNA). Diagram (B) shows the construct after cloning of *C. elegans elt-2* cDNA within the restriction sites *Nhe*I and *Pst*I. This construct was designated pAC13 and transformed into the feeding strain HT115.

Figure 5.12: Percentage of *C. elegans* embryos hatching on *Ce-elt*-2 RNAi and negative control plates.

Embryos laid by adults worms on *Ce-elt-2* RNAi and negative control plates were left overnight at 20C and the number of hatched embryos counted. The figures in the table are the sum of two plates (embryos from 10 adults per plate).

■ NL2099

☐ N2
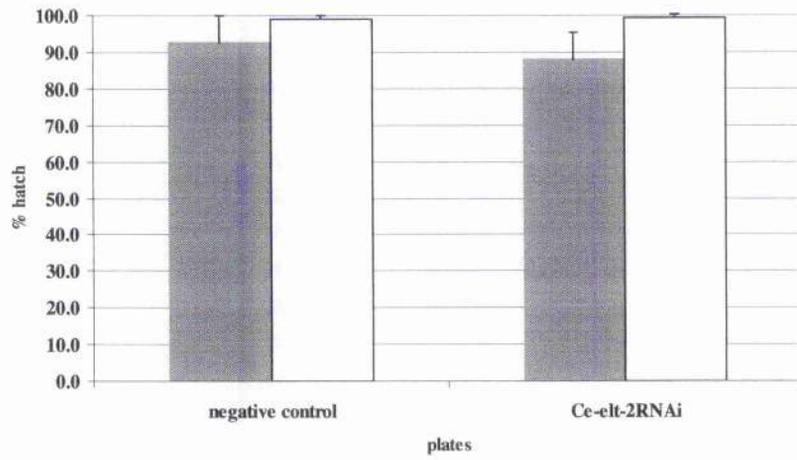
	NL2099		N2	
	Negative control	RNAi plates	Negative control	RNAi plates
Unhatched embryos	25	38	3	2
Hatched embryos	317	276	263	297
Total	342	314	266	299
% hatch	92.7	87.9	98.9	99.3

Figure 5.13: Percentage of larvae failing to reach L4 stage after 55 hours of incubation following *Ce-elt-2* RNAi treatment through embryogenesis.

The graph and table show the results from a continuation of the experiment shown in figure 5.13. The plates were left for a further 24-30 hours at 20°C and the number of L4 larvae and the number of L4 larvae not yet L4 were counted. The figures in the table are derived from the sum of two separate set of plates

■ NL2099

□ N2
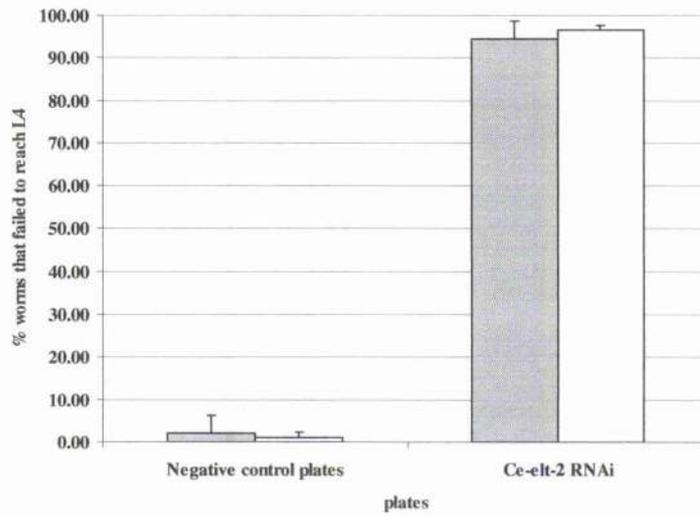
	NL2099		N2	
	Negative control	RNAi plates	Negative control	RNAi plates
Larvae < L4	8	279	4	356
L4	361	16	411	13
Total	369	295	415	369
% of <L4	2.2	94.6	1	96.5

Figure 5.14: Percentage of larvae failing to achieve the L4 stage after post-embryonic exposure to *elt-2* dsRNA for 50 hours at 20°C and comparison to negative control plates, for two strains, NL2099 and N2. Percentages are obtained from the figures in table above. In each experiment, three plates were set up and the number of animals on each plate was summed up. NL2099NEG and N2NEG refer to the set of negative control plates set up for strain NL2099 and N2, respectively and NL2099POS and N2POS refer to the RNAi plates set up for strain NL2099 and N2, respectively.
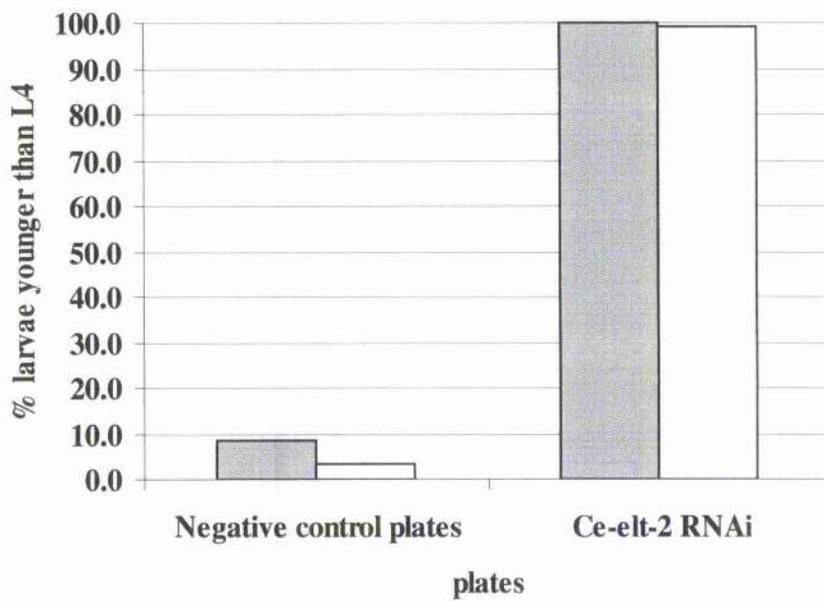
	<L4	L4 or older	total	% <L4
NL2099NEG	27	294	321	8.4
NL2099POS	466	0	466	100
N2NEG	32	917	949	3.4
N2POS	562	5	567	99.1

Figure 5.15: Larval arrest phenotype examples produced by *Ce-elt-2* RNAi during embryogenesis.

The panels show photomicrographs of larvae on plates from experiment shown in figure 5.13.

Panel A: Low power view of *Ce-elt-2* RNAi plate

Panel B: low power view of negative control plate

Panel C, D and E: Examples of arrested larvae on *Ce-elt-2* RNAi plates
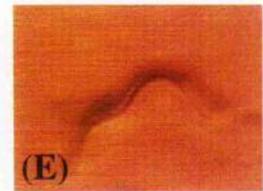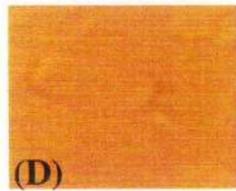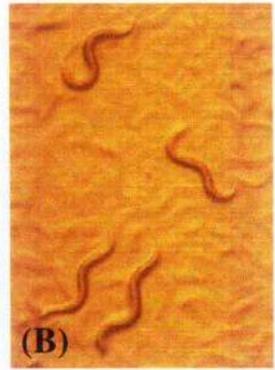
Figure 5.18: Examples of *elt-2* RNAi phenotypes after post-embryonic exposures to ds *elt-2* RNA.

(A) shows a larvae after 72 hours incubation on RNAi plates at 20°C under Normarski optics, the outline of the gut can be observed and the lumen appears very narrow, the structure of the intestine appears slightly disorganised. (B) and (C) show sections of larvae after 72 hours incubation on RNAi plates at 20°C under Normaski light; the internal structure of these larvae is very disorganised and corresponds to the "clear" phenotype previously described (Kamath *et al.*, 2003). (D) shows an adult worm after post-embryonic exposure to ds*elt-2* RNA; the animal present an "Egl" phenotype (some eggs have hatched within the worm) and the internal structure cannot be easily identified. (E) shows a larvae after 72 hours incubation on RNAi plates at 20°C under Normaski light; the larvae is abnormally small and present a slightly dumpy phenotype. (F) shows section of a larvae after 72 hours incubation on RNAi plates at 20°C under 450nm UV illumination and shows the distribution of the auto-fluorescent gut granules.

%larvae failing to reach L4

time in hours
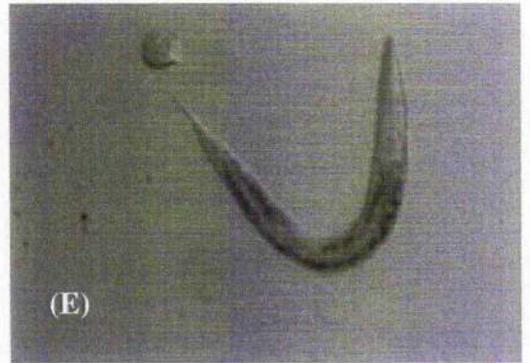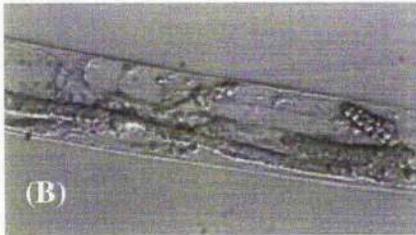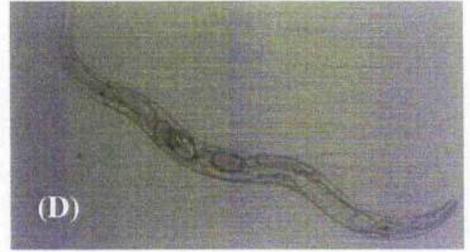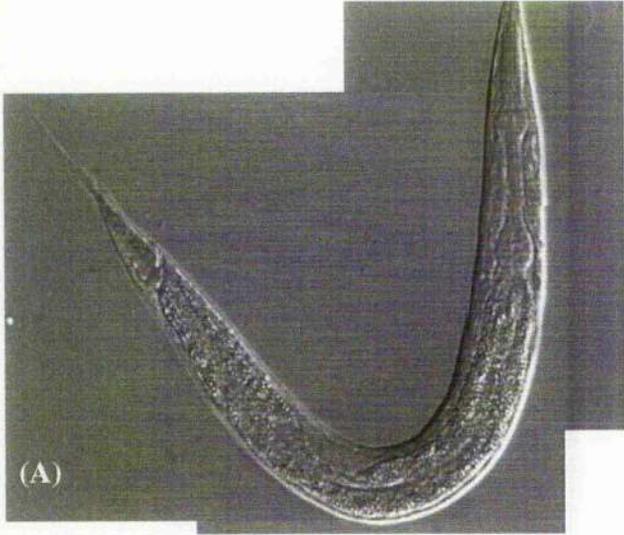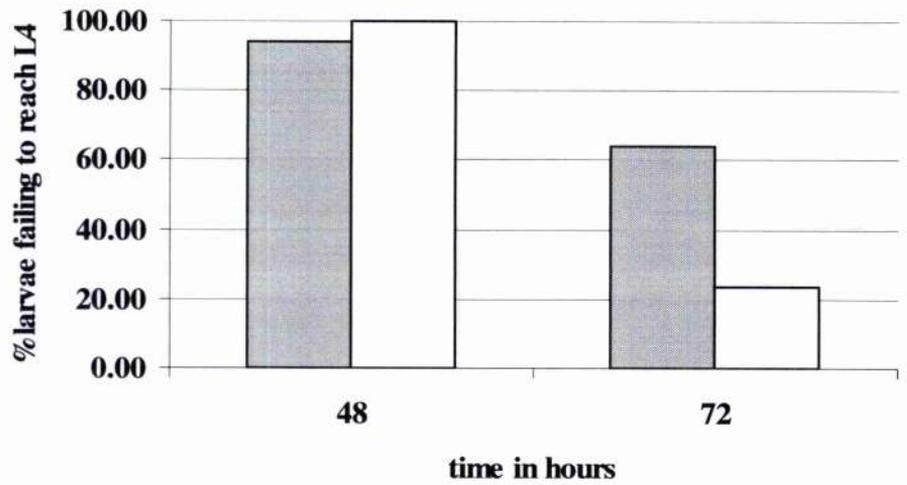
■ NL2099

☐ AC19D

	NL2099		AC19D	
	48 hours	72 hours	48 hours	72 hours
Larvae < L4	99	58	245	40
L4	6	33	0	130
Total	105	91	245	170
% worms < L4	94.2	63.7	100	23.5

216

Figure 5.17: Percentage of larvae failing to reach L4 stage after exposure to *elt-2* RNAi for 48 and 72 hours post-embryonically for two strains, NL2099 and AC19D.

The graph shows that both strains arrest in the same proportion after 48 hours, however after 72 hours the proportion of arrest for the rescue strain AC19D is less than that of strain NL2099. Percentages were calculated from figures shown in table above (figures are the sum of duplicate experiments)

Figure 5.18: Percentage of roll phenotype animals after exposure to *elt-2* RNAi for 48, 72 hours and 96 hours post-embryonically, and on negative control plates for strain AC19D.

The graph shows that after 72 and 96 hours incubation the percentage of roller worms exposed to post-embryonic *elt-2* RNAi in the strain apparently increases compared to animals on the negative control plates. Percentage were calculated from figures shown in table above (figures are the result of sum of duplicate experiment)
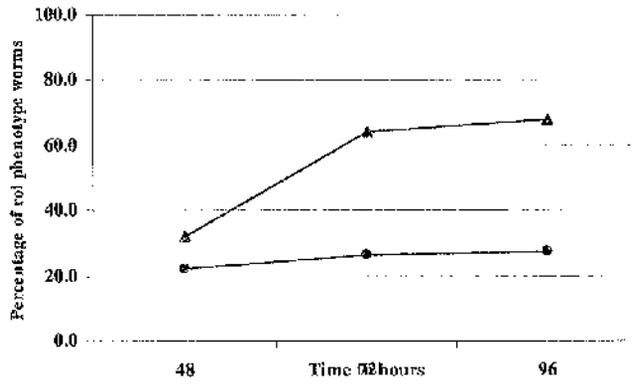
—▲—AC19D on *Ce-elt-2* RNAi plates

—●—AC19D on negative control plates

	AC19D RNAi			AC19D negative control		
	48 hours	72 hours	96 hours	48 hours	72 hours	96 hours
Rollers	102	109	107	138	151	89
Total	317	170	158	627	575	327
% rollers	32.2	64.1	67.7	22.0	26.3	27.2

217

Chapter 6: Attempts to apply RNA-mediated interference (RNAi) to the *H.contortus elt-2* gene

6.1 Introduction

The principal and application of RNA-mediated interference (RNAi) in *C. elegans* have been previously described in Chapter 5. RNAi is a powerful tool that allows for rapid investigation of gene function by creating a loss-of-function phenotype (Hunter, C. P. 1999; Bosher, J. M. *et al.* 2000; Fraser, A. G. *et al.* 2000; Maeda, I. *et al.* 2001). RNAi has been demonstrated in a wide range of organisms including protozoa, invertebrates and mammalian cells and has revolutionised the study of gene function [Fjose, 2001 #143; Grishok, A. *et al.* 2002; Kamath, R. S. *et al.* 2003). Helminth parasites are generally not amenable to genetic analysis and functional studies of gene products have been limited by available technologies and largely guided by predictions from primary sequence (Ash, C. 1999; Gasser, R. B. *et al.* 2000; Hashmi, S. *et al.* 2001). This has posed a considerable problem for parasitic nematode research and will continue to be the limiting factor in applying functional genomic approaches as more sequence becomes available from these organisms. (Williams, S. A. 1999; Allen, J. E. *et al.* 2000; Daub, J. *et al.* 2000; Hoekstra, R. *et al.* 2000; Lizotte-Waniewski, M. *et al.* 2000; Maizels, R. M. *et al.* 2000). The problem is particularly marked for genes with no clear sequence homologies in other organisms. But even genes for which homologues are clearly identifiable, there may be different or additional roles, particularly in the context of parasitism. Gene knock-out techniques have not been developed for parasitic nematodes due to the intrinsic difficulty in genetic manipulations and the lack of methods for heritable

transformation (Hashmi, S. *et al.* 2001; Brooks, D. R. *et al.* 2002). However some progress has recently been made in transient transformation of DNA and RNA by ballistic transfer (Davis, R. E. *et al.* 1999; Jackstadt, P. *et al.* 1999).

Recently, there has been a lot of interest in developing RNAi as an approach to study gene function in parasitic nematodes. The knock-down of acetyl-cholinesterase gene expression in *Nippostrongylus brasiliensis* by RNAi was the first successful example of this (Hussein, A. S. *et al.* 2002). dsRNA was introduced by overnight soaking, a technique that has been widely used in *C. elegans*. In *N. brasiliensis*, secreted acetylcholinesterases (AchE) are encoded by three separate genes expressed by fourth stage larvae (L4) and adult parasites. AChE B and AchE C are 90 % identical in amino acid sequence, whereas AChE A is 63-64 % identical to the others. In addition, *N. brasiliensis* expresses at least one other gene encoding a non-secreted (neuromuscular) AChE. In this experiment, adult worms were recovered from the jejunum of infected rats at different times post-infection and incubated in a solution of dsRNA (at a minimum concentration of 1mg/ml) for 16 hours at 37°C. Following incubation, the worms were washed in PBS and cultured *in vitro* for up to six days. The culture medium was changed daily, filtered and assayed for protein and AChE activity. Incubation in dsRNA corresponding to a 1799bp full-length AchE B cDNA fragment resulted in an 80% suppression of secretion of total AChE activity on the first day of culture, but there was a relatively rapid return to control levels over the following four days. When a smaller dsRNA fragment of 240 bp corresponding to the 5' coding sequence of AChE B was used, AChE expression was suppressed by over 90% and this suppression persisted throughout the 6 days of culture. This suggests that in this parasite species smaller dsRNA fragments are more

efficient, perhaps reflecting a better uptake by the parasite. Although incubation of the parasite in AchE B dsRNA resulted in suppression of all three secreted AchEs (presumably due to their high level of homology), parasite motility was unaffected by this treatment, suggesting that the secreted enzymes do not perform additional functions in the parasite neuromuscular system. RNAi has also been applied to two plant parasitic nematodes, *Heterodera glycines* and *Globodera pallida* (Urwin, P. E. *et al.* 2002). These two species are not amenable to microinjection and do not feed orally until after they have infected plants, making it impractical to apply RNAi by injection or feeding in the early larval stages. In this study, oral uptake was stimulated by the use of the neuroactive compound octopamine and pre-parasitic second stage juveniles were soaked in dsRNA targeting a cysteine protease gene before the larvae were used to infect plants. This did not result in a decrease of the number of parasites establishing infection but changed the female/male ratio from 3:1 to 1:1 by 14 days post-infection. Similar exposure of *H. glycines* to dsRNA corresponding to a polypeptide with homology to C-type lectins did not affect sexual fate, but 41% fewer parasites were recovered from the plants. Treatment with dsRNA corresponding to the major sperm protein in both species (MSP) had no effect on either parasite development or sexual fate over 14 days. Northern blot analysis showed lower transcript abundance for the MSP and C-type lectin genes in J2 (the pre-parasitic, second juvenile stage) plus a later inhibition for MSP transcripts when males developed sperm at 15 days post-infection.

Most recently, it has been shown that RNAi can be used to investigate gene function in the filarial parasitic nematode *Brugia malayi* (Aboobaker, A. A. *et al.* 2003). In this case, dsRNA was also delivered by means of soaking. Adult female

parasites were placed in PCR tubes whose lid tops had been cut off and replaced by dialysis membrane (15 kDa molecular weight cut off) and which contained culture media with dsRNA. These tubes where then placed in beakers containing 100 ml of standard culture medium preheated at 37°C. The nematodes were then checked every 2 hours to assess parasite health (assessed by motility) and microfilariae release. Incubation of the adult parasites with 300bp fragments of FITC-labelled dsRNA from *tub-1*, a β-tubulin gene, showed that there was uptake of dsRNA by adult worms in culture. The maximum uptake was reached after soaking for 18 hours and soaking for longer periods of time did not increase the levels of fluorescence. FITC staining in the adult female appeared to be strongest at the mouth and vulva which are openings presenting the least resistance to the entry of macromolecules. Fluorescence was also observed throughout the nematode, including the gonad. However, similar experiments revealed that there was no uptake of dsRNA by microfilariae whether the protective sheath was present or not. The effect of RNAi on three *B. malayi* genes was examined: *Bm-ama-1*, encoding the RNA polymerase II large subunit, *Bm-tub-1*, encoding β-tubulin and *Bm-shp-1*, encoding the microfilarial sheath-protein. Using the dialysis culture system, Aboobaker and colleagues found that RNAi of *Bm-ama-1* and *Bm-tub-1* resulted in reduction of transcript levels and subsequent death of adult female nematodes. They also found that RNAi of *Bm-shp-1* resulted in a reduction of transcript levels and the production of short abnormal microfilariae in culture.

In summary, there are already successful examples of the application of RNAi in several parasitic nematode species. The simplicity and flexibility of application makes RNAi a potentially powerful approach to study gene function in

parasitic nematode species. The free-living stages of *H.contortus* are potentially amenable to the application of RNAi by both feeding and soaking and this chapter describes a preliminary attempt to achieve this.

6.2. Results

6.2.1 Culture of *H. contortus* free-living stages on agar plates

H. contortus is an obligatory single host parasite from the infective L3 to the adult stage (Georgi, J. R. *et al.* 1990; Anderson, R. C. 1992). Adult worms lay eggs in the host abomasum which pass out in the faeces and are deposited on pasture. The L1 hatches from the egg and develops to L2 and then L3 in faeces on the pasture. The L1 and L2 are feeding stages which are thought to feed on bacteria and organic matter. The infective L3 stage, which retains the ecdysed L2 cuticle as an outer sheath, is a long-lived quiescent, non-feeding stage. Hence the L1 and L2 stages of *H. contortus* are the stages most likely to be amenable to the application of RNAi by virtue of being free-living and feeding stages. It was considered possible to culture the parasite from the L1 to L3 stage on bacterial lawns on agar plates in a similar way to standard *C. elegans* culture. This would then allow the application of RNAi to the free-living stages of *H. contortus* using essentially the same techniques that are used in *C. elegans*. Consequently, initial experiments were performed to see if *H.*

contortus L1 larvae would develop to the L3 stage on HT115 seeded NGM agarose plates and to determine the growth rates under these conditions.

The stage of *H. contortus* free-living larvae can be determined by a combination of length and morphological features {Anderson, 1992 #196;Soulsby, 1965 #250). L1 are 340-350 μm in length with an attenuated, sharply pointed tail and they moult to the L2 stage when they have reached 400-450 μm in length. The typical size of an L2 is not reported in the literature but the infective L3 is 754-756 μm in length. Since there is no growth after the moult to L3 due to the retention of the L2 cuticle as a sheath, the L2 size range must be between approximately 450 and 750 μm.

Anderson and co-workers performed a study on the development of *H. contortus* L3 in faecal cultures in vitro {Anderson, 1992 #196}. The optimum temperature for development of *H. contortus* embryos and larvae was found to be between 20-30°C. When cultured at 26°C, first-stage larvae hatched from eggs 14-17 hours after extraction from fresh faeces and started to feed around one hour after hatching. The first lethargus and moult occurred 10-12 hours after hatching and the second at 40 hours after the first moult, i.e., 50-52 hours after hatching. Infective larvae were present after 65 hours at 26°C (see figure 6.1). This data provides a guideline for growing *H. contortus* larvae on agar plates

H. contortus eggs were isolated from faeces using the floatation method described in section 2.3.3 The resulting egg suspension was then treated with a 0.18 M hypochlorite solution in order to eliminate bacterial contaminants before washing three times in M9 and resuspending in 50-100 μl of M9. The number of eggs per μl

was determined and the appropriate volume was deposited on standard HT115 seeded NGM plates so as to obtain between 100-300 eggs/plate. The plates were then incubated at 25°C and examined every 24 hours for several days. Three independent experiments were conducted and 10 larvae were picked from the plates each day and mounted on a 2% agarose pads containing 0.1 % sodium azide and examined by microscopy to determine their size and morphological features (as described in section 2.1.2). The larvae appeared motile on the agar plates and moved with a sinusoidal motion similar to *C. elegans*. They grew on plates to reach the size of a normal L3 (700-800 μm) for 72 to 96 hours of incubation (figure 6.2). After 96 hours, a separated cuticle was clearly visible on most of the larvae, suggesting the final L2-L3 moult takes place as normal and the L2 cuticle is retained as an L3 sheath. There was some variation in the growth curves between the three experiments, which may reflect variation in the time to hatching after being placed on the agar plates and also some inaccuracy in the method of length measurements.

6.2.2 Application of RNAi to *H. contortus* L1

6.2.2.1 Application of RNAi by feeding

The application of RNAi by feeding is widely used in *C. elegans* and has been used in a genome wide screening project (Kamath, R. S. *et al.* 2001; Kamath, R. S. *et al.* 2003). Furthermore RNAi of the *C. elegans* elt-2 gene by feeding *elt-2*

dsRNA to *C. elegans* L1 larvae results in a highly penetrant phenotype in which development is significantly retarded (described in detail in chapter 5). Consequently the method used for RNAi by feeding in *C. elegans* was adapted to apply RNAi to the *Hc-elt-2* gene in *H. contortus* L1 and L2 larvae. A 651 bp fragment corresponding to the 5' end of the *Hc-elt-2* cDNA sequence was PCR amplified with primers ANN1 and ANN2 from pACELT2 and cloned in *SstI* and *HindIII* sites of the polylinker of the *C. elegans* feeding vector L440 (figure 5.11 shows vector and appendix 3 shows primer sequences). This fragment corresponded to a similar region of the *Hc-elt-2* gene as was previously used for RNAi of the *C. elegans elt-2* gene. The resulting plasmid, pAC8, was transformed into the bacterial feeding strain HT115 and the resulting bacterial strain used to seed NGM agar plates. Two sets of RNAi NGM agar plates were prepared in duplicate: the first set was seeded with a lawn of HT115 bacteria transformed with the pAC8 construct and the second set of control plates was seeded with HT115 transformed with the "empty" L440 vector. 100-300 freshly extracted *H. contortus* eggs were placed on each plate and the plates were incubated at 25°C for 96 hours. Every 24 hours ten larvae were picked from each plate and mounted on 2% agarose pads for length measurement and detailed examination. The results of two independent experiments shown in figure 6.3. The *Hc-elt-2* RNAi treated larvae appeared to be smaller than the larvae on control plates at 24 hours incubation (figure 6.3A). However no such difference exist at 48 hours and only small differences exist at 72 and 96 hours of development. Although statistical analysis using Fisher's exact test gives P-values of less than 0.05 in some cases, the lack of consistent trend throughout and between experiments suggests the *Hc-elt-2*RNAi has little or no effect on larval growth. There were no other obvious

differences in morphology and motility between the larvae exposed to *Hc-elt-2* RNAi and negative controls.

The lack of an observable *Hc-elt-2* RNAi phenotype could be due either to a failure to inhibit *elt-2* expression or due redundancy of function of the *Hc-elt-2* gene. In order to test the effectiveness of *elt-2* RNAi in reducing *elt-2* expression levels, *H. contortus* larvae that had been subjected to *Hc-elt-2* RNAi were stained with the R1064 Hc-ELT-2 specific polyclonal antiserum (see section 2.3.5 for details of fixation and staining). Larvae were harvested from *Hc-elt-2* RNAi and control plates at 48 and 72 hours and duplicate slides were stained with the R1064 antisera. Although there was clear staining in the *Hc-elt-2* RNAi worms at 48 hours poor staining of the control worms made comparisons difficult (data not shown). The staining of the control slides was more consistent at 72 hours and the number of nuclei that stained with the antibody were counted in 10 randomly chosen worms from each slide (figure 6.4). There was no statistical difference between the numbers of nuclei staining in the *elt-2* RNAi worms compared to the negative controls (Figure 6.4). Similarly there was no obvious difference in the proportion of worms showing staining or the intensity of the staining (data not shown). This experiment was repeated several times with similar results. Consequently it appears that the application of *elt-2* RNAi has not resulted in a detectable reduction of Hc-ELT-2 expression.

6.3 Discussion

The experiments described in this chapter have shown that it is possible to culture *H. contortus in vitro* from eggs to L3 stage on standard NGM agar plates seeded with a lawn of HT115 bacteria. After 96 hours, the eggs have hatched and developed into third stage larvae, as defined by length measurements and the presence of the retained L2 cuticle. Once the larvae had reached the L3 stage they became very sluggish and tend to curl up on themselves, perhaps a reflection of the slower metabolism on entering the L3 stage, a quiescent, non-feeding stage. The period of incubation required for the larvae to reach the third larval stage in the experiments described above is longer than that previously reported in the literature for faecal cultures at 26°C, where they reach the L3 stage 60 hours after the eggs have been extracted from fresh faeces (Anderson, R. C. 1992). This difference could be explained by the differences in the conditions of incubation on agar plates compared to faecal culture, such as oxygen and carbon dioxide concentrations. Also faecal cultures may contain a variety of nutrients not present in the NGM agar plates. Also the *E. coli* HT115 strain may not provide an optimal food source compared to bacteria present in faecal cultures. The time taken to reach the L3 stage varied somewhat between experiments. The reason for this is not clear but there could be differences in the condition of eggs between batches and subtle differences in the conditions between experiments. Consequently, if growth rates are to be used to measure the effects on RNAi it is not possible to reliably compare between different experiments on different days. There are a limited number of phenotypes that are easily detectable in nematodes due to their simple body plan and this is one of the

limitations of examining the effects of RNAi on either *C. elegans* or parasitic nematodes. Therefore, accurate measurements of *H. contortus* larvae at different times of development during culture on RNAi plates and control plates were obtained to determine if *Hc-elt-2* RNAi had a detrimental effect on the growth of the *H. contortus* free-living stages. Two experiments were performed to assay the effects of *Hc-elt-2* RNAi on *H. contortus* L1 to L3 development and no obvious phenotype or difference in growth rate was observed. Clearly, although *C. elegans elt-2* has an essential post-embryonic function (Chapter 5) and is functionally highly conserved with the *Hc-elt-2* gene (Chapter 4) it is possible it does not have an essential role in *H. contortus*. Consequently, the lack of a detectable phenotype in response to *Hc-elt-2* RNAi could be due to either a lack of RNAi efficacy in *H. contortus* L1- L3 or to the *Hc-elt-2* gene not having an essential function. In order to distinguish between these possibilities the expression pattern of the Hc-ELT-2 polypeptide was examined by immunofluorescence with R1064 anti-Hc-ELT2 antisera on larvae harvested from *Hc-elt-2* RNAi and negative control plates. Although this is not a precisely quantitative assay, there was no obvious change in the expression pattern or levels of expression in the RNAi-treated worms relative to controls. These RNAi experiments have subsequently been repeated (using the same dsRNAi fragment) both by feeding and soaking and the *Hc-elt-2* expression levels measured by semi-quantitative RT-PCR (Britton and Murray, personal communication). Consistent with the results described in this chapter, there was no detectable reduction in *Hc-elt-2* transcript levels in response to RNAi. Hence it appears that the application of RNAi by feeding or soaking has failed to "knock down" *Hc-elt-2* expression.

Finally, attempts have been made to apply RNAi to several other *H. contortus* genes, which are predicted to have a visible phenotype based on the knowledge of its *C. elegans* homologue (Britton and Murray and Gilleard, personal communication). These are homologues of the *C. elegans unc-87, let-2,* and *cpr-1* genes, which were identified from the *H.contortus* EST dataset. In each case, no visible phenotype or growth defect could be detected when RNAi was applied by feeding and soaking (Britton, Murray and Gilleard, personal communication).

Hence the *H. contortus* L1 to L3 stages have so far been refractory to RNAi applied to four independent genes expected to have essential larval functions. Since the methods used were exactly those that are highly effective on *C. elegans* L1-L3 larvae. This may suggest there is a significant difference between the susceptibility of *H. contortus* larvae and *C. elegans* larvae to RNAi. This is not necessarily surprising as the free-living nematode *C. briggsae* is refractory to RNAi by feeding but is susceptible by microinjection (Descotte V. and Montgomery M, 2003 worm meeting, personal). The work of Hussein and colleagues (Hussein, A. S. *et al.* 2002) on RNAi in *Nippostrongylus* also revealed that larger dsRNA fragments (1799 bp) corresponding to the full-length cDNA for AChE B gene were less efficient in suppressing the production of AChE and the effect was more short lived compared to the effect of a shorter fragment of 240 bp. This phenomenon has not been reported for *C. elegans* and the reason for this phenomenon in *N. brasiliensis* is not known. Nevertheless it would be worthwhile to explore this avenue in *H. contortus* and try shorter dsRNAi fragments in this species to examine their efficacy to that reported in this chapter. Interestingly, Britton and Murray (personal communication) have examined the uptake of FITC-labelled dsRNA by *H. contortus* L1; eggs were

hatched overnight and L1 stage larvae were soaked in fluorescently labelled dsRNA for 20-24 hours with lipofectin (with or without spermidine at 1.5mM). Fluorescence microscopy showed that the larvae had ingested the fluorescent RNA as fluorescence could be seen in the gut although the presence in other tissues was unclear. Hence *H. contortus* L1 stage larvae do ingest dsRNA and there appears to be uptake into the gut but it is not clear whether the dsRNA is then absorbed through the gut into other tissues and if it stays intact after having been absorbed. Hence further experiments to examine the fate of ingested dsRNA would yield valuable information to understand the lack of effect of RNAi observed thus far.

Figure 6.1: Schematic diagram of *H. contortus* larval development in faecal culture at 26°C as described in the literature (Anderson, 1992)

The diagram shows the length of time required for the eggs to hatch after extraction from fresh faeces and the time required for each moult to proceed at 26°C. Hatching and the two moults, from L1 to L2 and from L2 to L3 are shown in red on the time scale. According to this, it takes between 60 and 65 hours for eggs extracted from faces to reach the L2 to L3 moult.

Extraction of
eggs from
faeces

hatch

L1/L2

L2/L3

0   4   24   48   72   96

4 hours
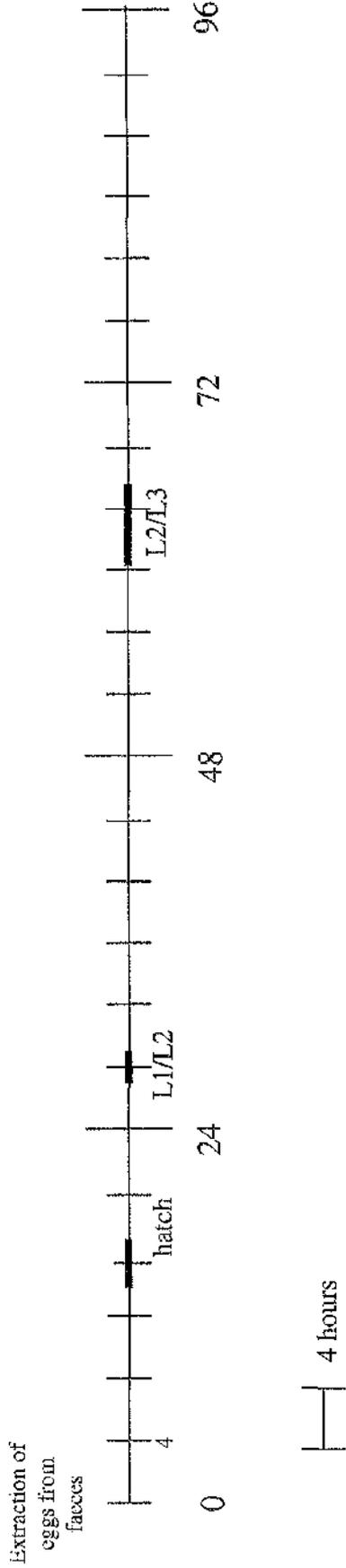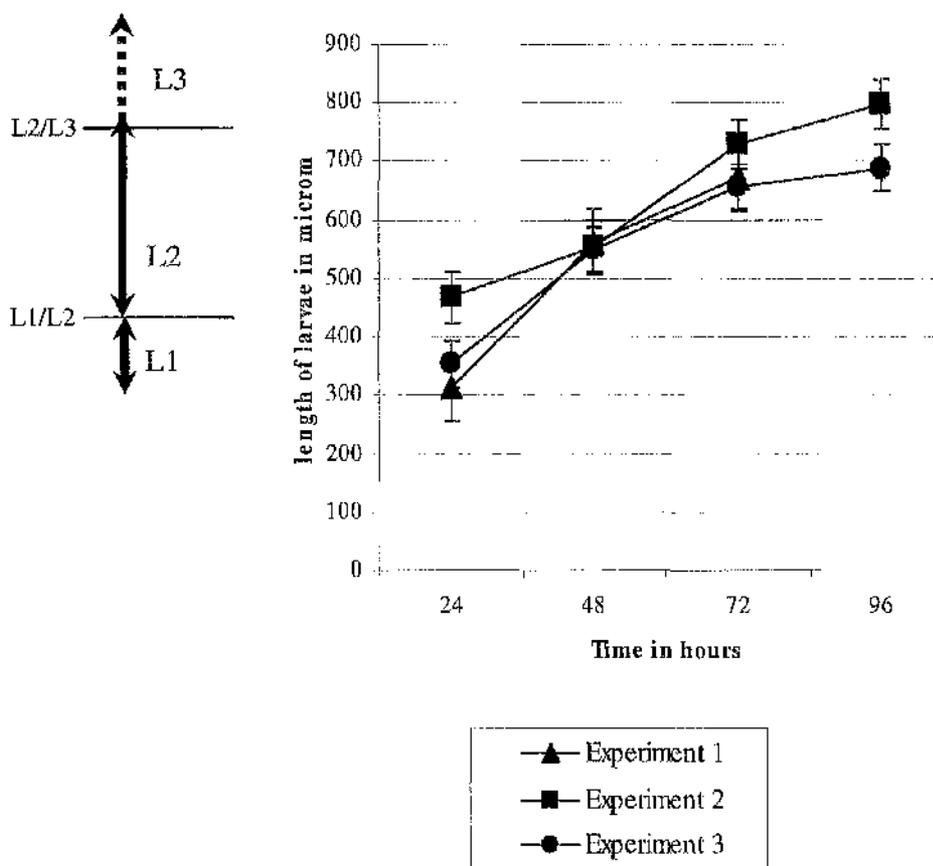
Figure 6.2: Growth of *H. contortus* larvae from eggs harvested from faeces on NGM plates seeded with HT115 bacteria, at 25 °C.

The mean length of 10 larvae measured at each time point is shown with the error bars indicating the size range. Three independent experiments were performed. The size range of each larval stage normally obtained from faecal culture is indicated on the left of the Y-axis
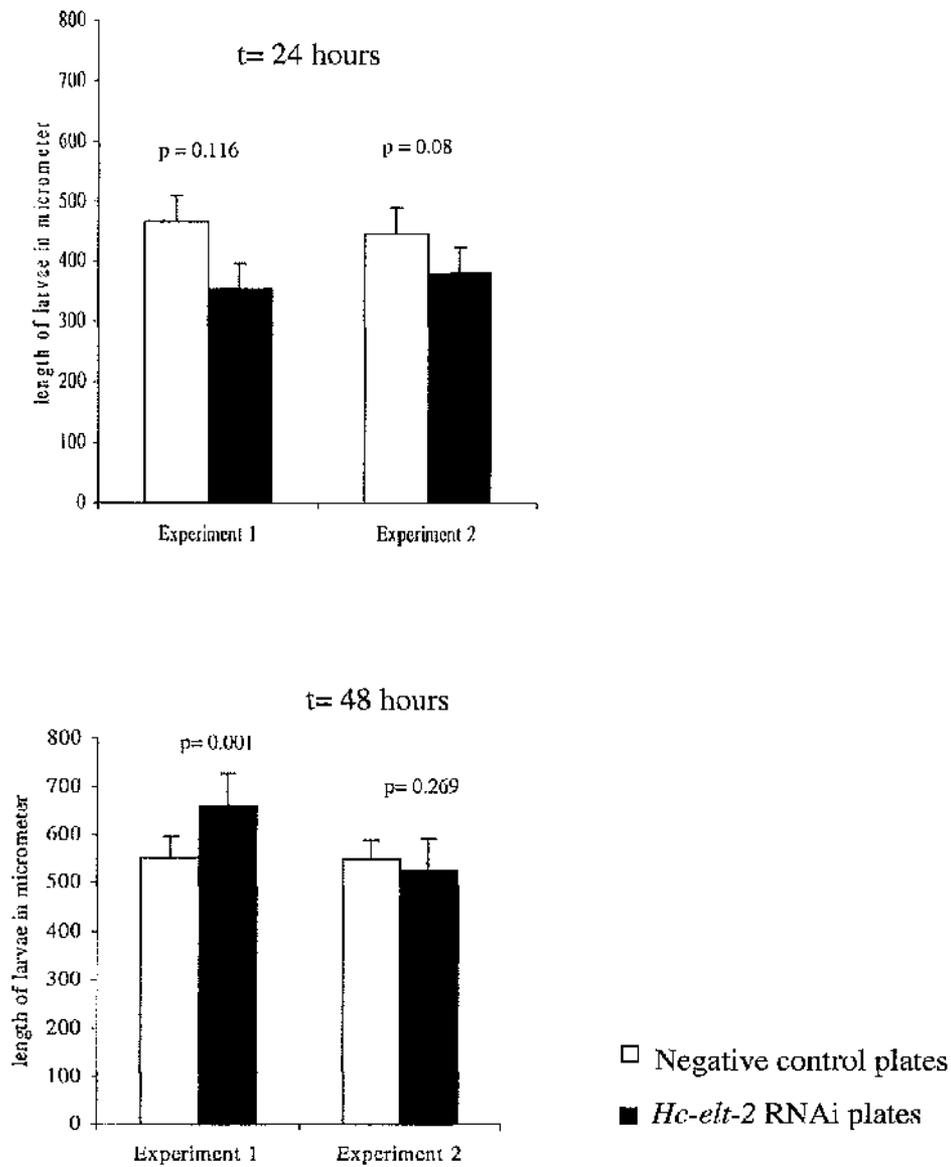
Figure 6.3: Growth of *H. contortus* larvae on NGM RNAi plates seeded with either L440 (negative control) or pAC8 (*Hc-elt-2* RNAi treatment) .

Each graph shows the mean length of larvae (n= 10 larvae) obtained at one of four time points in two independent experiments for both the negative control and the *Hc-elt-2* RNAi plates. The p-values (calculated by Fisher's exact) test for the significance of the difference between the control and RNAi plates are shown above the graphs.

t= 72 hours
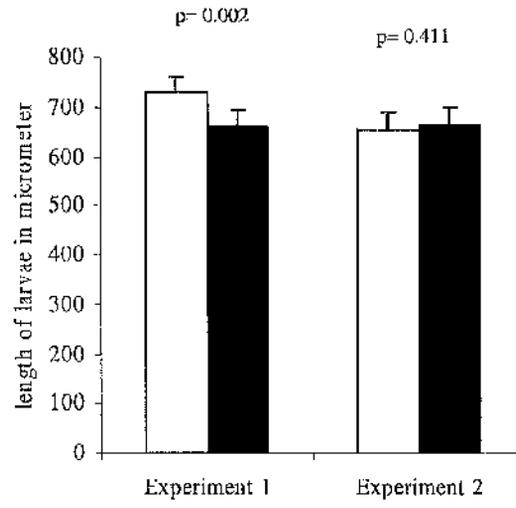
t= 96 hours
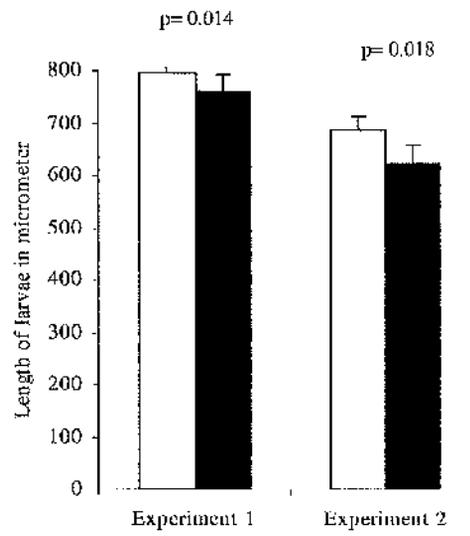
□ Negative control plates

■ *Hc-elt-2* RNAi plates

234
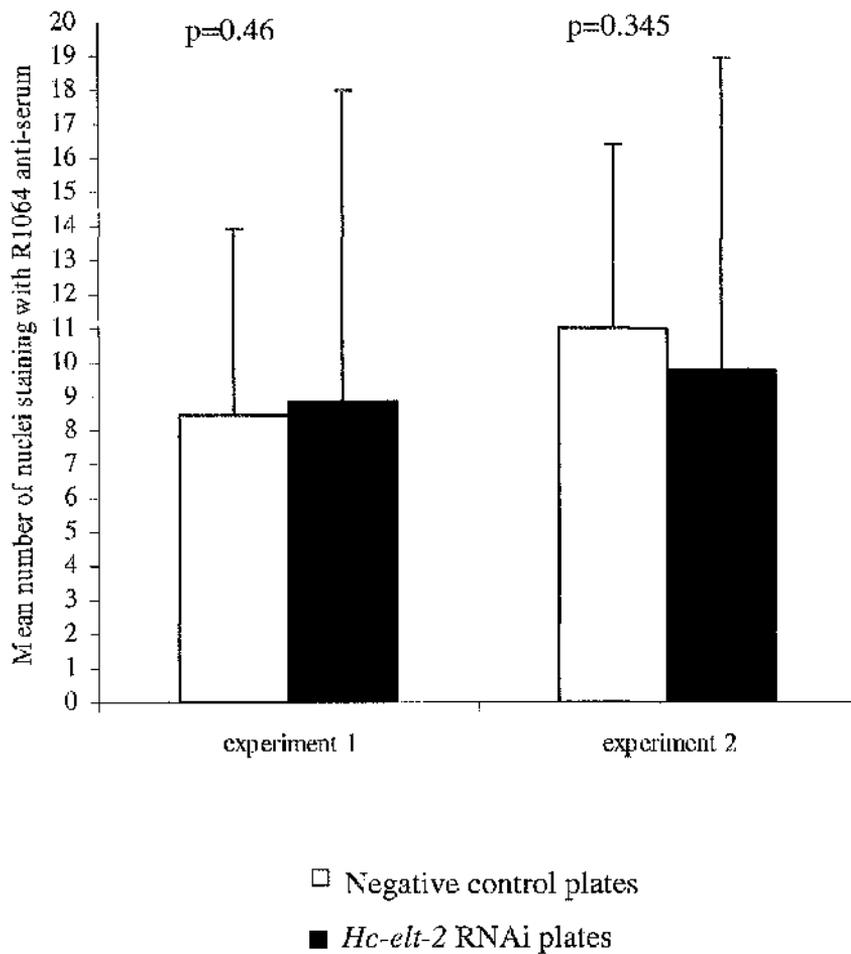
Figure 6.4: Mean number of nuclei staining with R1064 anti-serum in *H. contortus* larvae after 72 hours incubation on either negative control plates or *Hc-elt-2* RNAi plates.

The number of nuclei staining with the R1064 anti-serum was counted for 10 worms from each plate. P-values as calculated by t-test are indicated for each experiment on the graph and show that there is no significant difference in the number of gut nuclei expressing Hc-ELT-2 between negative controls and RNAi treated larvae.

235

Chapter 7: General discussion

The free-living nematode *Caenorhabditis elegans* is well established as one of the most powerful model systems in biology due to it being amenable to classical genetic analysis and the mapping of its entire cell lineage. More recently, the completion of its genome sequence along with the development of techniques such as transgenesis and RNA-mediated interference have put this organism at the forefront of biological research. In contrast, parasitic nematodes are difficult experimental subjects and techniques for functional analysis of genes are generally unavailable. Consequently there is a great deal of interest in the use of *C. elegans* as a tool for parasitic nematode research both as a source of biological information and as potential surrogate expression system for the *in vivo* study of parasitic gene function. However at present there is little information on the extent to which particular genes are functionally conserved between *C. elegans* and particular parasitic nematodes and few parasite genes have yet been functionally expressed in *C. elegans*. In *C. elegans* the endoderm is one of the most simple and best characterised tissues; the endoderm is also a tissue present in all nematodes, parasitic or free-living and is likely to have many conserved functions. The regulation of endoderm development in *C. elegans* involves a cascade of GATA transcription factors: *med-1/-2*, *end-1/-3* and *elt-2/-7*. Their role in the regulation of endoderm initiation and specification is reviewed in detail in chapter 1. The aim of this project was to identify and isolate parasitic homologues of *C. elegans* endodermal GATA transcription factors from available EST databases and genome sequencing projects and examine their functional conservation using *C. elegans* as a heterologous expression system. The project also aimed to examine the conservation of gene

regulation by examining the ability of parasite regulatory elements to direct the expression of reporter genes in transgenic.

A search of the *C. briggsae* genome database identified only seven potential GATA transcription factors and a search of the *B. malayi* genome sequence database revealed only five. Sequence analysis could clearly identify each of the *C. briggsae* genes as homologues of *C. elegans* family members. In the case of the *B. malayi* genes, the relationships were less clear but nevertheless it was possible to discern putative homologues. In both cases, those GATA factor family members which had essential functions in *C. elegans* (*elt-1*, *elt-2* and *elt-5*) appeared to have homologues in both *C. briggsae* and *B. malayi*. A search of the parasitic nematode EST databases also identified a single GATA transcription factor from each of the species *Ancylostoma ceylanicum*, *Strongyloides ratti* and *Meloidogyne arenaria*. However, the relationship of these genes to individual members of the *C. elegans* GATA transcription factor family was not clear. Hence a current hypothesis would be that the three essential GATA factors in *C. elegans* are conserved in many nematode species but some of the family members with redundant functions are not.

A GATA transcription factor was also isolated from the strongylid parasitic nematode *Haemonchus contortus*. This gene was characterised and sequence analysis suggested this was homologue of the *C. elegans elt-2* gene. Consequently this gene was named *Hc-elt-2*. Immunofluorescence experiments with Hc-ELT-2 specific antisera demonstrated that the Hc-ELT-2 was expressed in the *H. contortus* endoderm and the overall expression pattern was very similar to that of *C. elegans* ELT-2. However one difference between the *C. elegans* and *H. contortus* ELT-2 expression patterns was that additional expression was seen in two - probably neuronal - cells either side of the *H.*

contortus pharynx. The ability of the regulatory elements of the *H. contortus elt-2* gene to direct expression in neuronal cells - probably amphids and phasmids – supports the antibody expression data. Hence it is possible that the *H. contortus elt-2* gene has additional functions to those of the *C. elegans* gene. Since the GATA factor family appears to have fewer members in some nematode species than in *C. elegans,* it is possible that a particular GATA factor may have additional roles in such a nematode. However, as yet, none of the *C. elegans* GATA factors have been shown to be expressed in amphid or phasmid neurons. These Hc-ELT-2 immunolocalisation experiments also revealed that development of the endoderm lineage is extremely similar in *C. elegans* and *H. contortus.* The total number of endoderm cells at the end of embryogenesis is identical and the timing of endodermal cell divisions with respect to the stage of embryogenesis is very similar.

There is remarkably little sequence identity between the Ce-ELT-2 and Hc-ELT-2 polypeptides with the DNA binding domain being the only region of high identity. In spite of this, the function of the molecule is highly conserved since ectopic expression of Hc-ELT-2 can activate a program of gut cell differentiation in early *C. elegans* embryos almost as well as the *C. elegans* ELT-2 polypeptide itself. One explanation of this could simply be that the DNA binding domain is the only part of the polypeptide that is necessary and sufficient for the activation of endodermal differentiation. However deletion and domain swap experiments (J. Smith and J. Gilleard, personal communication) suggest that there are regions of the polypeptide outside the DNA binding domain that are also important for function *in vivo.* Similar findings have been found in GATA-1. Indeed it was shown that an N-terminal domain, although not necessary for DNA-binding or transcriptional activation, was necessary for

erythropoiesis *in vivo* (Shimizu, R. *et al.* 2001). Hence it is possible that the small regions of conserved residues between Hc-ELT-2 and Ce-ELT-2, outside the DNA-binding domain, are of functional importance. For example they could be involved in protein-protein interactions. Indeed, there is now evidence that although these interactions occur over large interfaces, only a small number of residues contribute to high affinity interactions and are critical for binding (Wells, J. 1996; Yu, C. *et al.* 2002). If this is true, comparative sequence information between the C. *elegans* and *H. contortus* ELT-2 polypeptides may provide a rapid approach to identify key functional domains and residues. The approach currently being adopted is to mutate individual residues in these regions of both the C. *elegans* and *H. contortus* polypeptides and test the ability of the altered molecule to activate endodermal differentiation in the ectopic expression assay (J. Gilleard and J. Smith, Personal communication). Another interesting use of the ectopic expression assay would be to examine the functionality and tissue specificity of those parasitic nematode GATA factors that don't have clear homologues in C. *elegans*.

The conservation of regulation between *Ce-elt-2* and *Hc-elt-2* was examined. The 3.3 kb of 5' flanking sequence immediately upstream of the *Hc-elt-2* gene two notable clusters of consensus GATA binding sites. This region was shown to be able to direct the expression of a GFP reporter gene in the endoderm of transgenic C. *elegans*. The *H. contortus* regulatory regions directed endodermal GFP expression from late embryogenesis (pretzel stage). This is different to the temporal expression pattern of the endogenous *H. contortus* ELT-2 polypeptide which is expressed from the two E cell stage of early embryogenesis. The spatial expression pattern of GFP was very conserved with that of the endogenous Hc-ELT-2 polypeptide in *H. contortus*; the GFP fusion was

expressed in the endoderm and in 4 neuronal cells adjacent to the pharynx and 2 neuronal cells in the tail region. In *H. contortus*, Hc-ELT-2 is likewise expressed in the endoderm as well as in at least two cells adjacent to the pharynx in a position consistent with being neurones. These results support the use of *C. elegans* as a heterologous expression system for studying the spatial expression and regulation of genes from closely related nematodes such as *H. contortus*. However this work, and that of others (Britton 2002), suggests that the elements regulating temporal expression pattern may have evolved more rapidly and are not highly conserved between *C. elegans* and *H. contortus*. Further experiments using larger fragments of 5' flanking sequence from *Hc-elt-2* would be useful to rule out the possibility that the discrepancies observed between the reporter gene expression in *C. elegans* and the endogenous gene in the parasite are due to a lack of important regulatory regions in the reporter gene fusion

Finally, RNAi on *H. contortus elt-2* in larval stages was attempted. It was found that *H. contortus* L1 could develop to apparently normal L3 on the standard plates used for *C. elegans* culture. However attempts to knock down the expression of the *H. contortus elt-2* gene proved unsuccessful. The ability to perform reverse genetics on parasitic nematodes such as *H. contortus* will be critical to enable functional biological research on these organisms. Hence development of RNAi should be a high priority for future work

References

Aamodt, E. J., M. A. Chung and J. D. McGhee (1991). "Spatial Control of gut-specific gene expression during *Caenorhabditis elegans* development." Science **252**(5005): 579-82.

Aboobaker, A. A. and M. L. Blaxter (2003). "Use of RNA interference to investigate gene function in the human filarial nematode parasitic *Brugia malayi*." Molecular and Biochemical Parasitology **129**: 41-51.

Akiyama, Y., N. Watkins, H. Suzuki, K. W. Jair, M. van Engeland, M. Esteller, H. Sakai, C. Y. Ren, Y. Yuasa, J. G. Herman and S. B. Baylin (2003). "GATA-4 and GATA-5 transcription factor genes and potential downstream antitumor target genes are epigenetically silenced in colorectal and gastric cancer." Molecular and Cellular Biology **23**(23): 8429-39.

Allen, J. E., J. Daub, D. Guiliano, A. McDonnell, M. Lizotte-Waniewski, D. Taylor, W. and M. Blaxter (2000). "Analysis of genes expressed at the infective larval stage validates utility of *Litomosoides sigmodontis* as a murine model for filarial vaccine development." Infection and Immunity. **68**(9): 5454-5458.

Altschul, S. F., W. Gish, W. Miller, E. W. Meyers and D. J. Lipman (1990). "Basic Local Alignment Search Tool." Journal of Molecular Biology **215**(3 SU -): 403-410.

Anderson, R. C. (1992). Nematode parasite of vertebrates, their development and transmission, CAB International.

Argiropoulos, B., J. Ho, B. Blachuta, J., I. Tayyab and A. Percival-Smith (2003). "Low-level ectopic expression of Fushi tarazu in *Drosophila melanogaster*

results in ftzUa1/Rp1-like phenotypes and rescues ftz phenotypes."

Mechanisms of Development **120**(12): 1443-1453.

Ash, C. (1999). "Exploring parasites genome." Trends in Microbiology **7**(1): 10-12.

Ashton, F. T., J. Li and G. A. Schad (1999). "Chemo -and thermosensory neurons: structure and function in animal parasitic nematodes." Veterinary Parasitology **84**: 297-316.

Ashton, F. T. and G. A. Schad (1996). "Amphids in *Strongyloides stercoralis* and other parasitic nematodes." Parasitology Today **12**(5): 187-194.

Bird, D. M., C. H. Opperman, S. J. Jones and D. L. Baillie (1999). "THE CAENORHABDITIS ELEGANS GENOME: A Guide in The Post Genomics Age." Annual Review of Phytopathology. **37**(1): 247-265.

Birnboim, H. C. and J. Doly (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Research **7(6):1513-23**(6): 1513-23.

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

Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman, A. Vierstraete, J. R. Vanfleteren, L. Y. Mackey, M. Dorris, L. M. Frisse, J. T. Vida and W. K. Thomas (1998). "A molecular evolutionary framework for the phylum Nematoda." Nature **392**: 71-75.

Blouin, M. S., C. A. Yowell, C. H. Courtney and J. B. Dame (1995). "Host mouvement and the gentic structure of populations of parasitic nematodes." Genetics **141**: 1007-1014.

Blumenthal, T. and K. Steward (1997). RNA processing and gene structure. _C. elegans_ II. D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press: 117-45.

Boag, P. R., P. Renb, S. E. Newton and R. B. Gasser (2003). "Molecular characterisation of a male-specific serine/threonine phosphatase from _Oesophagostomum dentatum_ (Nematoda: Strongylida), and functional analysis of homologues in _Caenorhabditis elegans_." International Journal for Parasitology **33**,(3): 313-325.

Bosher, J. M. and Labouesse M. (2000). "RNA interference: genetic wand and genetic watchdog." Nature Cell Biology **2**(2): E31-E36.

Bossinger, O., T. Fukushige, M. Claeys, G. Borgonie and J. D. McGhee (2004). "The apical disposition of the _C. elegans_ intestinal terminal web is maintained by LET-143." Developmental Biology **268**(2): 448-56.

Bowerman, B., B. A. Eaton and J. R. Priess (1992). "_skn-1_, a maternally expressed gene required to specify the fate of ventral blastomeres in the early _C. elegans_ embryo." Cell **68**: 1061-1075.

Bowerman, B., M. K. Ingram and C. P. Hunter (1997). "The maternal par genes and the segregation of cell fate specification activities in early _Caenorhabditis elegans_ embryos." Development **124**: 3815-3826.

Breathnach, R. and P. Chambon (1981). "Organization and expression of eucaryotic split genes coding for proteins." Annual Review of Biochemistry **50**: 349-83.

Brenner, S. (1974). "The genetics of Caenorhabditis elegans." Genetics **77**(1): 71-94.

Brinster, R. L., J. M. Allen, R. R. Behringer, R. E. Gelinas and R. D. Palmiter (1988). "Introns increase transcriptional efficiency in transgenic mice." Proceedings of the National Academy of Science, U S A. **85**(3): 836-840.

Britton, C. and L. Murray (2002). "A cathepsin L protease essential for

   Caenorhabditis elegans embryogenesis is functionally conserved in parasitic

   nematodes." Molecular and Biochemical Parasitology **122**(1): 21-33.

Britton, C., D. L. Redmond, D. P. Knox, J. H. McKerrow and J. D. Barry (1999).

   "Identification of promoter elements of parasite nematode genes in transgenic

   Caenorhabditis elegans." Molecular and Biochemical Parasitology **103**: 171-

   181.

Brooks, D. R. and R. E. Isaac (2002). "Functional genomics of parasitic worms: The

   dawn of a new era." Parasitology International **51**(4): 319-325.

Buchman, A. R. and P. Berg (1988). "Comparison of intron-dependent and intron-

   independent gene expression." Molecular and Cellular Biology **8**(10): 4395-

   4405.

Bujard, H., R. Gentz, M. Lanzer, D. Stuber, M. Muller, I. Ibrahimi, M. T. Hauptle

   and B. Dobberstein (1987). "A T5 promoter based transcription-translation

   system for the analysis of proteins *in vivo* and *in vitro*." Methods in

   Enzymology **155**(416-433).

Burglin, T. R., E. Lobos and M. L. Blaxter (1998). "*Caenorhabditis elegans* as a

   model for parasitic nematodes." International Journal for Parasitology **28**:

   395-411.

Charron, F. and M. Nemer (1999). "GATA transcription factors and cardiac

   development." Cell and developmental biology **10**: 85-91.

Christophe, D., C. Christophe-Hobertus and B. Pichon (2000). "Nuclear targeting of

   proteins: how many different signals?" Cellular Signalling **12**(5 SU -): 337-

   341.

Coles, G. C. (1998). "Ivermectin-resistant *Cooperia* species from calves on a farm in Sommerset." Veterinary Record **142**: 255-256.

Daub, J., A. Loukas, D. I. Pritchard and M. Blaxter (2000). "A survey of genes expressed in adults of the human hookworm, *Necator americanus.*" Parasitology **120**(Pt 2): 171-84.

Davis, R. E., A. Parra, P. T. LoVerde, E. Ribeiro, G. Glorioso and S. Hodgson (1999). "Transient expression of DNA and RNA in parasitic helminths by using particle bombardment." Proceedings of the National Academy of Sciences, USA **96**(15): 8687-8692.

Eagleson, J. S. (1992). "Benzimidazole resistance in *Trichostrongylus axei* in Australia." Veterinary Record **131**: 317-318.

Eagleson, J. S. and J. Y. Bowie (1986). "Oxfendazole resistance in *Trichonstrongylus axei* in cattle in Australia." Veterinary Research **119**: 604.

Finney, M. and G. Ruvkun (1990). "The *unc-86* gene product couples cell lineage and cell identity in *C. elegans.*" Cell **63**: 895-905.

Fire, A. (1995). "Fire lab vector kit."

Fire, A. (1999). "RNA-triggered gene silencing." Trends in Genetics **15**(9): 358-363.

Fire, A., D. Albertson, S. White Harrison and D. G. Moerman (1991). "Production of antisense RNA leads to effective and specific inhibition of gene expression in *C. elegans* muscle." Development **113**: 503-14.

Fire, A., S. White Harrison and D. K. Dixon (1990). "A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans.*" Gene: 189-198.

Fjose, A., S. Ellingsen, A. Wargelius and H. C. Seo (2001). "RNA interference: mechanisms and applications." Biotechnology Annual Review **7**: 31-57.

Fraser, A. G., R. S. Kamath, P. Zipperlen, M. Martinez-Campos, M. Sohrmann and J. Ahringer (2000). "Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference." Nature **408**: 325-30.

Fukushige, T., B. Goszczynski, H. Tian and J. D. McGhee (2003). "The evolutionary duplication and probable demise of an endodermal GATA factor in *Caenorhabditis elegans*." Genetics **165**: 575-588.

Fukushige, T., M. G. Hawkins and J. D. McGhee (1998). "The GATA factor *elt-2* is essential for formation of the *Caenorhabditis elegans* intestine." Developmental Biology **198**: 286-302.

Gasser, R. B. and S. E. Newton (2000). "Genomic and genetic research on bursate nematodes: significiance, implications and prospects." International Journal for Parasitology **30**: 509-534.

Geary, T. G. and D. P. Thompson (2001). "*Caenorhabditis elegans*: how good a model for veterinary parasites?" Veterinary Parasitology **101**: 371-386.

Geerts, S. (1987). "Suspected resistance of Ostertagia ostertagi to levamisole." Veterinary Parasitology **23**: 77-82.

Georgi, J. R., M. E. Georgi and V. J. Theodorides (1990). Parasitology for veterinarians. Philadelphia, London, Toronto, Montreal, Syndney, Tokyo, W. B. Saunders Company.

Gilleard, J. S. and J. D. McGhee (2001). "Activation of hypodermal differentiation in the *Caenorhabditis elegans* embryo by GATA transcription factors ELT-1 and ELT-3." Molecular and Cellular Biology **21**(7): 2533-2544.

Gilleard, J. S., Y. Shafi, J. D. Barry and J. D. McGhee (1999). "*Elt-3*: a *Caenorhabditis elegans* GATA factor expressed in the embryonic epidermis during morphogenesis." Developmental Biology **208**: 265-280.

Goldstein, B. (1993). "Establishment of gut fate in the E-lineage of *C. elegans*: the roles of lineage-dependant mechanisms and cell interactions." Development **118**: 1267-1277.

Gomez-Escobar, N., W. F. Gregory, C. Britton, L. Murray, C. Corton, N. Hall, J. Daub, M. L. Blaxter and R. M. Maizels (2002). "Abundant larval transcript-1 and -2 genes from *Brugia malayi*: diversity of genomic environments but conservation of 5' promoter sequences functional in *Caenorhabditis elegans*." Molecular and Biochemical Parasitology **125**(1-2): 59-71.

Gopal, R. M. (1999). "Resistance of fiel isolates of *Trichostrongylus colubriformis* and *Ostertagia circumcincta* to ivermectin." International Journal of Parasitology **29**: 781-786.

Grishok, A. and M. C. C. (2002). "RNAi (Nematodes: *Caenorhabditis elgans*)." Advances in Genetics **46**: 339-60.

Gross, J. M., R. E. Peterson, S.-Y. Wu and D. R. McClay (2003). "LvTbx2/3: a T-box family transcription factor involved in formation of the oral/aboral axis of the sea urchin embryo." Development **130**(9): 1989-1999.

Hashmi, S., G. Hashmi and R. Gaugler (1995). "Genetic Transformation of an Entomopathogenic Nematode by Microinjection." Journal of Invertebrate Pathology **66**(3): 293-296.

Hashmi, S., W. Tawe and S. Lustigman (2001). "*Caenorhabditis elegans* and the study of gene function in parasites." Trends in parasitology **17**(8): 387-393.

Haun, C., J. Alexander, D. Stainier, Y. and P. Okkema, G. (1998). "Rescue of *Caenorhabditis elegans* pharyngeal development by a vertebrate heart specification gene." Proceedings of the National Academy of Sciences. USA **95**(9): 5072-5075.

Hawkins, M. G. and J. D. McGhee (1995). "*elt-2*, a second GATA factor from the

nematode *Caenorhabditis elegans*." The Journal of Biological Chemistry

270(24): 14666-14671.

Hekimi, S., B. Lakowski, T. M. Barnes and J. J. Ewbank (1998). "Molecular genetcis

of life span in *C. elegans:* how much does it teach us?" Trends in Genetics

14: 14-19.

Higazi, T. B., A. Merriweather, L. Shu, R. Davis and T. R. Unnasch (2002). "*Brugia*

malayi: transient transfection by microinjection and particle bombardment."

Experimental Parasitology **100**(2): 95-102.

Hobert, O. (2002). "PCR fusion-based approach to create reporter gene constructs for

expression analysis in transgenic *C. elegans*." BioTechniques **32**(4): 728-730.

Hoekstra, R., A. Visser, M. Otsen, J. Tibben, J. A. Lenstra and M. H. Roos (2000).

"EST sequencing of the parasitic nematode *Haemonchus contortus* suggests a

shift in gene expression during transition to the parasitic stages." Molecular

and Biochemical Parasitology **110**: 53-68.

Hope, I. A. (1999). Background on *Caenorhabditis elegans*. C. elegans, a practical

approach. I. A. Hope. Oxford, Oxford University Press: 1-16.

Houthoofd, W., K. Jacobsen, C. Mertens, S. Vangestel, A. Coomans and G. Borgonie

(2003). "Embryonic cell lineage of the marine nematode *Pellioditis marina*."

Developmental Biology **258**: 57-69.

Hunter, C. P. (1999). "Genetics: a touch of elegance with RNAi." Current Biology

9(12): R440-2.

Hussein, A. S., K. Kichenin and M. E. Selkirk (2002). "Suppression of secreted

acetylcholinesterase expression in *Nippostrongylus brasiliensis* by RNA

interference." Molecular and Biochemical Parasitology **122**(1): 91-94.

Hutter, H. and R. Schnabel (1994). "*glp-1* and inductions establishing embryonic axes in *C. elegans*." Development **120**(7): 2051-2064.

Jackson, F. and R. L. Coop (2000). "The development of anthelminthic resistance in sheep nematodes." Parasitology **120 (Suppl)**: S95-S107.

Jackstadt, P., T. P. Wilm, H. Zahner and G. Hobom (1999). "Transformation of nematodes via ballistic DNA transfer." Molecular and Biochemical Parasitology **103**(2): 261-266.

Jagannathan, S., D. L. Laughton, C. L. Critten, T. M. Skinner, L. Horoszok and A. J. Wolstenholme (1999). "Ligand-gated chloride channel subunits encoded by the *Haemonchus contortus* and *Ascaris suum* orthologues of the *Caenorhabditis elegans gbr-2 (avr-14)* gene." Molecular and Biochemical Parasitology **103**: 129-40.

Jungblut, B. and R. J. Sommer (1998). "The *Pristionchus pacificus mab-5* gene is involved in the regulation of ventral epidermal cell fate." Current Biology **8**: 775-778.

Kamath, R. S., A. G. Fraser, Y. Dong, G. Poulin, R. Durbin, M. Gotta, A. Kanapin, N. Le Bot, S. Moreno, M. Sohrmann, D. P. Welchman, P. Zipperlen and J. Ahringer (2003). "Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi." Nature **421**(6920): 231-7.

Kamath, R. S., M. Martinez-Campos, P. Zipperlen, A. G. Fraser and J. Ahringer (2001). "Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*." Genome Biology **2**(1).

Kampkotter, A., T. E. Volkmann, S. Hegi de Castro, B. Leiers, L.-O. Klotz, T. E. Johnson, C. D. Link and K. Henkle-Duhrsen (2003). "Functionnal analysis of

the glutathione S-transferase 3 from *Onchocerca volvulus* (*Ov-GST-3*): a

parasite GST confers increased resistance to oxidative stress in

Caenorhabditis elegans." Journal of Molecular Biology **325**: 25-37.

Kennedy, B. P., E. J. Aamodt, F. L. Allen, M. A. Chung, M. F. Heschl and J. D.

McGhee (1993). "The gut esterase gene (*ges-1*) from the nematodes

Caenorhabditis elegans and *Caenorhabdtis briggsae*." Journal of Molecular

Biology **229**(4): 890-908.

Kenny, A. P., D. W. Oleksyn, L. A. Newman, R. C. Angerer and L. M. Angerer

(2003). "Tight regulation of SpSoxB factors is required for patterning and

morphogenesis in sea urchin embryos." Developmental Biology **261**(2): 412-

425.

Kim, R., K. Tanabe, M. Emi, Y. Uchida, H. Inoue and T. Toge (2003). "Inducing

cancer cell death by targeting transcription factors." Anticancer Drugs **14**(1):

3-11.

Ko, L. J. and J. D. Engel (1993). "DNA-binding specificities of the GATA

transcription factor family." Molecular and Cellular Biology **13**(7): 4011-

4022.

Koh, K. and J. H. Rothman (2001). "ELT-5 and ELT-6 are required continuously to

regulate epidermal seam cell differentiation and cell fusion in *C. elegans*."

Development **128**: 2867-2880.

Kosman, D. and S. Small (1997). "Concentration-dependent patterning by an ectopic

expression domain of the Drosophila gap gene knirps." Development **124**(7):

1343-1354.

Kramer, J. M., R. P. French, E. C. Park and J. J. Johnson (1990). "The

Caenorhabditis elegans rol-6 gene, which interacts with the *sqt-1* collagen

gene to determine organismal morphology, encodes a collagen." Molecular and Cellular Biology **10**(5): 2081-9.

Krause, S., A. Sommer, P. Fisher, P. M. Brophy, R. D. Walter and E. Liebau (2001). "Gene structure of the extracellular glutathione S-transferase from *Onchocerca volvulus* and its overexpression and promoter analysis in transgenic *Caenorhabditis elegans*." Molecular and Biochemical Parasitology **117**: 145-154.

Kwa, M. S. G., J. G. Veenstra and M. H. Roos (1993). "Molecular characterisation of [beta]-tubulin genes present in benzimidazole-resistant populations of *Haemonchus contortus*." Molecular and Biochemical Parasitology **60**(1): 133-143.

Kwa, M. S. G., J. G. Veenstra, M. Van Dijk and M. H. Roos (1995). "Beta-tubulin genes from the parasitic nematode *Haemonchus contortus* modulate drug resistance in *Caenorhabditis elegans*." Journal of Molecular Biology **246**: 500-510.

Laufer, J. S., P. Bazzicalupo and W. B. Wood (1980). "Segregation of developmental potential in early embryos of *Caenorhabditis elegans*." Cell. : **19**(3): 569-77.

Laverierre, A. C., C. MacNeill, C. Mueller, R. E. Poelmann, J. B. Burch and T. Evans (1994). "GATA-4/5/6, a subfamily of three transcription factors transcribed in developing heart and gut." Journal of Biological Chemistry **269**(37): 23177-23184.

LaVoie, H. A. (2003). "The role of GATA in mammalian reproduction." Exp Biol Med **228**(11): 1282-90.

Legname, G., G. Fossati, G. Gromo, i. N. Monzin, F. Marcucci and D. Modena (1995). "Expression in *Escherichia coli*, purification and functional activity

of recombinant human chaperonin 10." Federation of Biochemical Societies Letters **361**(2-3): 211-4.

Lewis, J. A. and J. T. Fleming (1995). Basic culture methods. *Caenorhabditis elegans*: Modern biological analysis of an organism. H. F. Epstein and D. C. Shakes. San Diego, Academy Press.

Li, J., F. T. Ashton, H. R. Gamble and G. A. Schad (2000). "Sensory neuroanatomy of a passively ingested nematode parasite, *Haemonchus contortus*: amphidial neurons of the first stage larva." The Journal of Comparative Neurology **417**: 299-314.

Li, J., X. Zhu, F. T. Ashton, H. R. Gamble and G. A. Schad (2001). "Sensory neuroanatomy of a passively ingested nematode parasite, *Haemonchus contortus*: amphidial neurons of the third stage larva." Journal of Parasitology **87**(1): 65-72.

Li, J., X. Zhu, R. Boston, F. T. Ashton, H. R. Gamble and G. A. Schad (2000). "Thermotaxis and theromsensory neurons in infective larvae of *Haemonchus contortus*, a passively ingested nematode parasite." The Journal of Comparative Neurology **424**: 58-73.

Li, X., Y. Fang, X. Zhao, X. Jiang, T. Duong and S. R. Kain (1999). "Characterization of NFkappa B Activation by Detection of Green Fluorescent Protein-tagged Ikappa B Degradation in Living Cells." Journal of Biological Chemistry. **274**(30): 21244-21250.

Lindqvist, A., Ljungsteom B. L., O. Nilsson and P. J. Walker (2001). "The dynamics, prevalence and impact of nematode infection of organically raised sheep in Sweden." Acta Vet Scandinavia **42**(3): 377-89.

Linton, K. J. and C. F. Higgins (2002). "P-glycoprotein misfolds in Escherichia coli: evidence against alternating-topology models of the transport cycle." Molecular and Membrane Biology 19(1): 51-8.

Lizotte-Waniewski, M., W. Tawe, D. B. Guiliano, W. Lu, J. Liu, S. A. Williams and S. Lustigman (2000). "Identification of Potential Vaccine and Drug Target Candidates by Expressed Sequence Tag Analysis and Immunoscreening of *Onchocerca volvulus* Larval cDNA Libraries." Infection and Immunity. 68(6): 3491-3501.

Lok, J. B. and J. Massey, H. C. (2002). "Transgene expression in *Strongyloides stercoralis* following gonadal microinjection of DNA constructs." Molecular and Biochemical Parasitology 119(2): 279-284.

Lowry, J. A. and W. R. Atchley (2000). "Molecular evolution of the GATA family of the transcription factors: conservation within the DNA-binding domain." Journal of Molecular Evolution 50: 103-115.

Lyons, G. E. (1996). "Vertebrate heart development." Current Opinion in Genetic and Development 6(4): :454-60.

Maduro, F. M. and J. H. Rothman (2002). "Making worm guts: the gene regulatory network of the *Caenorhabditis elegans* endoderm." Development Biology 246: 68-85.

Maduro, M. F., M. D. Meneghini, B. Bowerman, G. Broitman-Maduro and J. H. Rothman (2001). "Restriction of mesendoderm to a single blastomere by the combined action of SKN-1 and a GSK-3 beta homolog is mediated by MED-1 and -2 in *C. elegans*." Molecular cell 7: 475-485.

Maeda, I., Y. Kohara, M. Yamamoto and A. Sugimoto (2001). "Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi." Current Biology **11**(3): 171-176.

Maizels, R. M., K. K. A. Tetteh and A. Loukas (2000). *"Toxocara canis*: genes expressed by the arrested infective larval stage of a parasitic nematode." International Journal for Parasitology **30**(4 SU -): 495-508.

Manfield, I. W., L. A. Reynolds, J. Gittins and G. G. Kneale (2000). "The DNA-binding domain of the gene regulatory protein AreA extends beyond the minmal zinc-finger region conserved between GATA proteins." Biochemica et Biophysica Acta **1493**: 325-332.

Mango, S. E., C. J. Thorpe, P. R. Martin, S. H. Chamberlain and B. Bowerman (1994). "Two maternal genes, *apx-1* and *pie-1*, are required to distinguish the fates of equivalent blastomeres in the early *Caenorhabditis elegans* embryo." Development **120**(8): 2305-2315.

Marin, I. and B. S. Baker (1998). "The evolutionary dynamics of sex determination." Science **281**: 1990-4.

Marshall, S. D. G. and J. D. McGhee (2001). "Coordination of *ges-1* expression between the *Caenorhabditis* pharynx and intestine." Developmental Biology **239**: 350-363.

Mello, C. C. and A. Fire (1995). DNA transformation. *Caenorhabditis elegans*: modern biological analysis of an organism. H. F. Epstein and D. C. Shakes. San Diego, Academy Press: 451-482.

Mello, C. C., J. M. Kramer, D. Stinchcomb and V. Ambros (1992). "Efficient gene transfer in *C.elegans*: extrachromosomal maintenance and integration of

transforming sequences." <u>European Molecular Biology Organisation Journal</u> **10**(12): 3959-70.

Merika, M. and S. H. Orkin (1993). "DNA-binding specificity of GATA family transcription factors." <u>Molecular and Cellular Biology</u> **13**(7): 3999-4010.

Metzstein, M. M., G. M. Stanfield and H. R. Horvitz (1998). "Genetics of rpgrammed cell death in *C. elegans*: past, present and future." <u>Trends in Genetics</u> **14**(10): 410-6.

Miller, D. M. and D. C. Shakes (1995). Immunofluorescence microscopy. <u>*Caenorhabditis elegans*: modern biological analysis of an organism</u>. H. F. Epstein and D. C. Shakes. San Diego, Academic Press: 365-394.

Miriami, E., J. Sperling and R. Sperling (1994). "Heat shock affects 5' splice site selection, cleavage and ligation of CAD pre-mRNA in hamster cells, but not its packaging in InRNP particles." <u>Nucleic Acids Research</u> **22**(15): 3084-91.

Molkentin, J. D. (2000). "The zinc finger-containing transcription factors GATA-4, -5 and -6." <u>Journal of Biological Chemistry</u> **275**(50): 38949-38952.

Morrisey, E. E., H. S. Ip, M. M. Lu and M. S. Parmacek (1996). "GATA-6, a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm." <u>Developmental Biology</u> **177**: 309-22.

Morrisey, E. E., H. S. Ip, Z. Tang and M. S. Parmacek (1997). "GATA-4 activates transcription via two novel domains that are conserved with the GATA-4/5/6 subfamily." <u>Journal of Biological Chemistry</u> **272**(13): 8515-8524.

Moskowitz, I. P., S. B. Gendreau and J. H. Rothman (1994). "Combinatorial specification of blastomere identity by *glp-1*-dependent cellular interactions in the nematode *Caenorhabditis elegans*." <u>Development</u> **120**(11): 3325-3338.

Muhich, M. L., M. P. Hsu and J. C. Boothroyd (1989). "Heat-shock disruption of

    trans-splicing in trypanosomes: effect on Hsp70, Hsp85 and tubulin mRNA

    synthesis." Gene **82**(1): 169-75.

Muro-Pastor, M. I., R. Gonzalez, J. Strauss, F. Narendja and C. Scazzocchio (1999).

    "The GATA factor *AreA* is essential for chromatin remodelling in a

    eukaryotic bidirectional promoter." European Molecular Biology

    Organisation Journal **18**(6): 1584-1597.

Newton, A., J. Mackay and M. Crossley (2001). "The N-terminal Zinc finger of the

    erythroid transcription factor GATA-1 binds GATC motifs in DNA." The

    Journal of Biological Chemistry **276**(38): 35794-35801.

Nikolaou, S., D. Hartman, P. J. A. Presidente, S. E. Newton and R. B. Gasser (2002).

    "HcSTK, a *Caenorhabditis elegans* PAR-1 homologue from the parasitic

    nematode, *Haemonchus contortus**1." International Journal for Parasitology

    32(6): 749-758.

Ohneda, K. and M. Yamamoto (2002). "Roles of hematopoietic transcription factors

    GATA-1 and GATA-2 in the development of red blood cell lineage." Acta

    Haematologia **108**(4): 237-45.

Okkema, P. G., S. W. Harrison, V. Plunger, A. Aryana and A. Fire (1993).

    "Sequence requirements for myosin gene expression and regulation in

    Caenorhabditis elegans." Genetics **135**(2): 385-404.

Omichinski, J. G., C. D. Trainor, T. Evans, A. Gronemborn, M., and G. M. Clore

    (1993). "A small single-"finger" peptide from the erythroid transcription

    factor GATA-1 binds specifically to DNA as a zinc or iron complex."

    Proceedings of the National Academy of Sciences, U S A **90**: 1676-1680.

Orkin, S. H., R. A. Shivdasani, Y. Fujiwara and M. A. McDevitt (1998).

"Transcription factor GATA-1 in megakaryocyte development." Stem Cells.
16(Suppl 2): 79-83.

Orlowski, R. Z. and J. Baldwin, A. S. (2002). "NF-[kappa]B as a therapeutic target in
cancer." Trends in Molecular Medicine 8(8): 385-389.

Page, B. D., W. Zhang, K. Steward, T. Blumenthal and J. R. Priess (1997). "*elt-1*, a
GATA-like transcription factor, is required for epidermal cell fates in
Caenorhabditis elegans." Genes and Development 11: 1651-1661.

Pal, S., M. Lo, D. Schmidt, I. Pelczer, S. Thurber and S. Walker (1997). "Skn-1:
Evidence for a bipartite recognition helix in DNA binding." Proceedings of
the National Academy of Sciences, USA 94(11): 5556-5561.

Patient, R. K. and J. D. McGhee (2002). "The GATA family (vertebrates and
invertebrates)." Current Opinion in Genetics & Development 12(4): 416-422.

Prichard, R. K. (2001). "Genetic variability following selection of Haemonchus
contortus with anthelmintics." Trends in Parasitology 17(9): 445-453.

Ravagnani, A., L. Gorfinkiel, T. Langdon, G. Diallinas, E. Adjadj, S. Demais, D.
Gorton, H. N. Arst, Jr., and C. Scazzocchio (1997). "Subtle hydrophobic
interactions between the seventh residue of the zinc finger loop and the first
bas of an HGATAR sequence determine promoter-specific recognition by the
Aspergillus nidulans GATA factor AreA." European Molecular Biology
Organisation journal 16(13): 3974-3986.

Rehman, A. and D. P. Jasmer (1998). "A tissue specific approach for analysis of
membrane and secreted protein antigens from *Haemonchus contortus* gut and
its application to diverse nematode species." Molecular and Biochemical
Parasitology. 97(1-2): 55-68.

Rocheleau, C. E., W. D. Downs, R. Lin, C. Wittmann, Y. Bei, Y.-H. Cha, M. Ali, J.
R. Priess and C. C. Mello (1997). "*Wnt* signaling and an APC-related gene
specify endoderm in early *C. elegans* embryos." Cell **90**.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory
manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Sambrook, J., E. F. Fritsch and T. Maniatis (1989). Molecular cloning: a laboratory
manual. Cold Spring Harbor, Cold Spring Harbor Laboratory Press.

Samson, S. I., O. Richard, M. Tavian, T. Ranson, C. A. Vosshenrich, F. Colucci, J.
Buer, F. Grosveld, I. Godin and J. P. Di Santo (2003). "GATA-3 promotes
maturation, IFN-gamma production, and liver-specific homing of NK cells."
Immunity **19**(5): 701-11.

Sanger, F., S. Nicklen and A. R. Coulson (1977). "DNA sequencing with chain-
terminating inhibitors." Proceedings of the National Academy of Sciences, U
S A **74**(12): 5463-7.

Sangster, N. C. and J. Gill (1999). "Pharmacology of anthelminthic resistance."
Parasitology Today **15**: 141-146.

Schierenberg, E. (1987). "Reversal of cellular polarity and early cell-cell interaction
in the embryo of *Caenorhabditis elegans.*" Developmental Biology **122**: 452-
463.

Schlesinger, A., C. A. Shelton, J. N. Maloof, M. D. Meneghini and B. Bowerman
(1999). "Wnt pathway components orient a mitotic spindle in the early
Caenorhabditis elegans embryo without requiring gene transcription in the
responding cell." Genes and Development **13**: 2028-2038.

Shimizu, R., S. Takahashi, K. Ohneda, J. D. Engel and M. Yamamoto (2001). "*In
vivo* requirements for the GATA-1 functional domains during primitive and

definitive erythropoiesis." European Molecular Biology Organisation journal **20**(18): 5250-5260.

Shompole, S. and D. P. Jasmer (2003). "*Haemonchus contortus* UNC-18 binds to *Caenorhabditis elegans* neuronal syntaxin." Molecular and Biochemical Parasitology **130**(1): 55-59.

Simmer, F., C. Moorman, A. M. van der Linden, E. Kuijk, P. V. E. van den Berghe, R. S. Kamath, A. G. Fraser, J. Ahringer and R. H. A. Plasterk (2003). "Genome-Wide RNAi of *C. elegans* Using the Hypersensitive *rrf-3* Strain Reveals Novel Gene Functions." Public Library of Science Biology **1**(1): e12.

Sommer, R. J. (2000). "Evolution of nematode development." Current opinion in genetics and development **10**: 443-8.

Sommer, R. J., A. Eizinger, K.-Z. Lee, B. Jungblut, A. Bubeck and I. Schlak (1998). "The *Pristionchus* HOX gene *Ppa-lin-39* inhibits programmed cell death to specify the vulva equivalence group and is not required during vulval induction." Development **125**: 3865-3873.

Sommer, R. J. and P. W. Sternberg (1995). "Evolution of cell lineage and pattern formation in the vulval equivalence of Rhabditid nematodes." Developmental Biology **167**: 61-74.

Sommer, R. J. and P. W. Sternberg (1996). "Evolution of nematode vulval fate patterning." Developmental Biology **173**: 396-407.

Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." Journal of Molecular Biology **98**(3): 503-17.

Spieth, J., Y. H. Shim, K. Lea, R. Conrad and T. Blumenthal (1991). "*elt-1*, an embryonically expressed *Caenorhabditis elegans* gene homologous to the

GATA transcription factor family." <u>Molecular and Cellular Biology</u> **11**(9): 4651-9.

Stein, L. D., Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent, N. Chen, A. Chinwalla, L. Clarke, C. Clee, A. Coghlan, A. Coulson, P. D'Eustachio, D. H. A. Fitch, L. A. Fulton, R. E. Fulton, S. Griffiths-Jones, T. W. Harris, L. W. Hillier, R. Kamath, P. E. Kuwabara, E. R. Mardis, M. A. Marra, T. L. Miner, P. Minx, J. C. Mullikin, R. W. Plumb, J. Rogers, J. E. Schein, M. Sohrmann, J. Spieth, J. E. Stajich, C. Wei, D. Willey, R. K. Wilson, R. Durbin and R. H. Waterston (2003). "The Genome Sequence of *Caenorhabditis briggsae*: A Platform for Comparative Genomics." <u>Public Library of Science Biology</u> **1**(2): e45.

Stiernagle, T. (1997). Maintenance of *C. elegans*. <u>*C. elegans*, a practical approach</u>. I. A. Hope. Oxford, Oxford University Press: 51-67.

Stinchcomb, D. T., J. E. Shaw, S. H. Carr and D. Hirsh (1985). "Extrachromosomal DNA transformation of *Caenorhabditis elegans*." <u>Molecular and Cellular Biology</u> **5**(12): 3484-96.

Stringham, E. G., D. K. Dixon, D. Jones and E. P. Candido (1992). "Temporal and spatial expression patterns of the small heat-shock (hsp-16) genes in transgenic *Caenorhabditis elegans*." <u>Molecular Biology of the Cell</u> **3**: 221-233.

Stroeher, V. L., B. P. Kennedy, K. J. Millen, D. F. Schroeder, M. G. Hawkins, B. Goszczynski and J. D. McGhee (1994). "DNA-protein interactions in the *Caenorhabditis elegans* embryo: oocyte and embryonic factors that bind to the promoter of the gut specific *ges-1* gene." <u>Developmental Biology</u> **163**: 367-380.

Stuber, D., H. Matile and G. Garotta (1990). System for high-level production in
    Escherichia coli and rapid purification of recombinant proteins: application to
    epitope mapping, preparation of antibodies, and structure-function analysis.
    Immunological Methods. I. Lefkovits and B. Pernis. New York, Academic
    Press. **IV:** 121-152.

Sudhaus, W. (1976). "Vergleichende Untersuchungen zur Phylogenie,
    Systematik,Okologie,Biologie und Ethologie der Rhabditidae (Nematoda)."
    Zoologica **43**: 1-230.

Sulston, J. E. and H. R. Horvitz (1977). "Post-embryonic lineage of the nematode
    Caenorhabditis elegans." Developmental Biology **56**(1): 110-56.

Sulston, J. E., H. R. Horvitz and J. Kimble (1988). Cell lineage. The nematode
    Caenorhabditis elegans. W. B. Wood. Cold Spring Harbor, NY, Cold spring
    harbor laboratory.

Sulston, J. E., E. Schierenberg, J. G. White and J. N. Thomson (1983). "The
    embryonic cell lineage of the nematode *Caenorhabditis elegans*."
    Development biology **100**: 64-119.

Sutcliffe, J. G. (1979). "Complete nucleotide sequence of the *E. coli* plasmid
    pBR322." Cold Spring Harbor Quant. Biol. **43**: 77-90.

Tabara, H., A. Grishok and C. C. Mello (1998). "REVERSE GENETICS:RNAi in *C.
    elegans*: Soaking in the Genome Sequence." Science **282**(5388): 430-431.

Tabish, M., Z. Kidwai Siddiqui, K. Nishikawa and S. S. Siddiqui (1995). "Exclusive
    Expression of *C. elegans osm-3* Kinesin Gene in Chemosnsory Neurons Open
    to the External Environment." Journal of Molecular Biology **247**(3): 377-389.

Thompson, F. J., A. C. Cockroft, I. Wheatley, C. Britton and E. Devaney (2001). "Heat shock and developmental expression of hsp83 in the filarial nematode Brugia pahangi." European Journal of Biochemistry. **268**(22): 5808-15.

Thorpe, C. J., A. Schlesinger and B. Bowerman (2000). "Wnt signalling in *Caenorhabditis elegans:* regulating repressors and polarizing the cytoskeleton." Cell Biology **10**: 10-17.

Timmons, L., D. L. Court and A. Fire (2001). "Ingestion of bacterially expressed dsRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans.*" Gene **263**(1-2): 103-112.

Timmons, L. and A. Fire (1998). "Specific interference by ingested dsRNA." Nature **395**(6705): 854.

Trainor, C. D., R. Ghirlando and M. A. Simpson (2000). "GATA zinc finger interactions modulate DNA binding and transactivation." The Journal of Biological Chemistry **275**(36): 28157-28166.

Tremblay. J. J. and R. S. Viger (2003). "Novel roles for GATA transcription factors in the regulation of steroidogenesis." Journal of Steroid Biochemestry and Molecular Biology **85.**(2-5): 291-8.

Tsang, A. P., J. E. Visvader, C. A. Turner, Y. Fujiwara, C. Yu, M. J. Weiss, M. Crossley and S. H. Orkin (1997). "FOG, a multitype zinc finger protein, acts as a cofactor for transcription factor GATA-1 in erythroid and megakaryocytic differentiation." Cell **90**: 109-119.

Uchida, O., H. Nakano, M. Koga and Y. Ohshima (2003). "The *C. elegans che-1* gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons." Development. **130**(7): 1215-24.

Urquhart, G. M., J. Armour, J. L. Duncan, A. M. Dunn and F. W. Jennings, . (1987). Ovine haemonchosis. Veterinary parasitology. G. M. Urquhart, Blackwell Science: 19-22.

Urwin, P. E., C. J. Lilley and H. J. Atkinson (2002). "Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference." Molecular Plant Microbe Interactions 15(8): 747-52.

Vermunt, J. J. (1995). "Multiple resistance to ivermectin and oxfendazole in Cooperia species of cattle in New-Zealand." Veterinary Record 137: 43-44.

Walker, G. R., K. D. Feather, P. D. Davis and K. K. Hines (1995). "Supersignal™ CL-HRP: A new enhanced chemiluminescent substrate for the development of the horseradish peroxide label in Western blotting applications." Journal of National Institute of Health Research 7: 76.

Waller, P. (1997). "Anthelmintic resistance." Veterinary Parasitology 72: 391-412.

Wells, J. (1996). "Binding in the growth hormone receptor complex." Proceedings of the National Academy of Sciences, U S A 93: 1-6.

White, J. (1988). The anatomy. The nematode Caenorhabditis elegans. W. B. Wood. Cold Spring Harbor, NY, Cold spring harbor laboratory,.

Wiegner, O. and E. Schierenberg (1998). "Specification of gut cell fate differs significantly between the nematodes *Acrobeloides nanus* and *Caenorhabditis elegans*." Developmental Biology 204: 3-14.

Wiegner, O. and E. Schierenberg (1999). "Regulative development in a nematode embryo: a hierachy of cell fate transformations." Developmental Biology 215: 1-12.

Williams, S. A. (1999). "Deep within the Filarial Genome: Progress of the Filarial Genome Project." Parasitology Today 15(6 SU -): 219-224.

Winter, A. D., J. Myllyharjus and A. P. Page (2003). "A hypodermally expressed prolyl 4-hydroxylase from the filirial nematode *B. malayi* is soluble and active in the absence of protein disulfide isomerase." The Journal of Biological Chemistry **278**(4): 2554-2562.

Woods, T. (1998). "Anthelmintic resistance on horse farms in north central Florida." Equine Practice **20**: 14-17.

Yost, H. J. and S. Lindquist (1986). "RNA splicing is interrupted by heat shock and is rescued by heat shock protein synthesis." Cell **45**(2): 185-93.

Yu, C., A. B. Cantor, H. Yang, C. Browne, R. A. Wells, Y. Fujiwara and S. H. Orkin (2002). "Targeted deletion of a high-affinity GATA-binding site in the GATA-1 promoter leads to selective loss of the eosinophil lineage in vivo." J Exp Med **195**(11): 1387-95.

Zhang, H., T. Toyofuku, J. Kamei and M. Hori (2003). "GATA-4 regulates cardiac morphogenesis through transactivation of the N-cadherin gene." Biochemical and Biophysical Research Communications **312**(4): 1033-1038.

Zhou, M. and W. Ouyang (2003). "The function role of GATA-3 in Th1 and Th2 differentiation." Immunol Res. **28**(1): 25-37.

Zhu, J., T. Fukushige, J. D. McGhee and J. H. Rothman (1998). "Reprogramming of early embryonic blastomeres into endodermal progenitors by a *Caenorhabditis elegans* GATA factor." <u>Genes and Development</u> **12**: 3809-3814.

Zhu, J., R. J. Hill, P. J. Heid, M. Fukuyama, A. Sugimoto, J. R. Priess and J. H. Rothman (1997). "*end-1* encodes an apparent GATA factor that specifies the endoderm precursor in *Caenorhabditis elegans* embryos." <u>Genes and Development</u> **11**: 2883-2896.

Appendix 1: media and solutions

Media for bacterial growth

LB broth

Tryptone	10 g
Yeast Extract	5 g
NaCl	10 g
NaOH 5M	0.2 ml

Complete to 1l with dH_2O. Autoclave.

LB agar

Tryptone	10 g
Yeast Extract	5 g
NaCl	10 g
Agar	15 g
NaOH 5M	0.2 ml

Complete to 1l with dH_2O. Autoclave.

SOC medium (0.5 l)

Bactotryptone	10 g
Yeast extract	2.5 g
NaCl	0.25 g
0.25 M KCl	5 ml

Autoclave and allow to cool.

Add: 2M MgCl2   2.5 ml

       1M Glucose   10 ml

Freezing solution

20 % Glycerol in LB-broth

Autoclave

Media for worm culture

NGM agar

		When medium has cool down add:	
NaCl	3g		
Agar	17 g	1M CaCl$_2$	1 ml
Peptone	2.5 g	1M MgSO$_4$	1 ml
dH$_2$O	975 ml	1M KPO$_4$	25 ml
Autoclave		5 mg/ml cholesterol	1 ml

Mix and pour plates in sterile conditions. Leave plates to dry at room temperature for 2-3 days and seed with 0.05 ml *E. coli* OP50 fresh liquid culture per plate. Allow the bacterial lawn to grow overnight at room temperature. Seeded plates can be kept for 2-3 weeks.

Solutions

Injection buffer

KPO4	20 mM
K citrate	3 mM
polyethylene glycol (PEG) 6000	2%

Adjust pH to 7.5.

M9 buffer

KH$_2$PO$_4$	3 g
Na$_2$HPO$_4$	50 mM
NaCl	85 mM
MgSO$_4$	1 mM

11.1 X buffer (for PCR)

Tris HCl, pH 8.8	45 mM
$(NH_4)_2SO_4$	11 mM
$MgCl_2$	4. 5 mM
2-mercaptoethanol	6.7 mM
EDTA, pH 8.0	4.4 mM
dATP	1 mM
dCTP	1 mM
dGTP	1 mM
dTTP	1 mM
BSA	113 µg/ml

50 X TAE

Tris base	2 M
Glacial acetic acid	5.71 %(v/v)
EDTA pH 8.0	50 mM

Adjust to 1 l with dH2O, store at room temperature.

Miniprep solutions

Resuspension solution

50 mM Tris pH 7.5

10 mM EDTA

Autoclave, and add 100 µg/ml RNase A

Lysis solution

0.2 M NaOH

1% SDS

Neutralisation solution

1.32 M potassium acetate pH4.8

Southern Blot solutions

Denaturation solution

NaOH 0.5 M

NaCl   1.5 M

Neutralisation solution II

Tris HCl, pH 7.5 0.5 M

NaCl   3.0 M

20 X SSC

Tris HCl pH 7.5 0.5 M

NaCl   3.0 M

Library screening solutions

Neutralisation solution (I)

Tris-HCl   1 M, pH 7.5

NaCl   1.5 M

Chemoluminescent detection of DIG labelled DNA probe

Maleic acid buffer

Maleic acid   0.1 M

NaCl   1.5 M, pH 7.5

Washing buffer

Maleic acid buffer + 0.3 % Tween 20

Detection buffer

Tris-HCl, pH 9.5   100mM

NaCl   100mM

$MgCl_2$ 50 mM

Immunocytochemistry solutions

10 X PBS

NaCl   136   mM

KCl   2.5 mM

Na_2HPO_4 12   mM

KH_2PO_4 1.8 mM

Adjust pH at 7.4 with HCl.

Autoclave.

DABCO

50 % glycerol in dH_2O with 2.5 % DABCO

Adjust pH at 8.0

DAPI

1 mg/ml stock in dH_2O

Store at -20°C

Ruvkun buffer

KCl	0.16 M
NaCl	0.04 M
Na_2EGTA	0.02 M
spermidine-3 HCl	0.01 M
NaPipes in 50 % methanol	0.03 M

AbA buffer

PBS	1X
BSA (fraction V)	1%
Triton-X-100	0.5%
Na Azide	0.05 %
EDTA	1mM

AbB buffer

PBS	1X
BSA (fraction V)	0.1%
Triton-X-100	0.5%
Na Azide	0.05 %
EDTA	1mM

20 X BO3 buffer

1M H_3BO_3

0.5 M NaOH

Tris-Triton buffer

0.1M Tris-HCl, pH 7.4

1% Triton-X-100

0.001M EDTA

Solutions for bacteriophage λ

SM solution

NaCl	5.8 g
MgSO$_4$	2.0 g
1M Tris. Cl (pH 7.5)	50 ml
2% gelatin solution	5 ml
Complete to 1l with dH$_2$O	

Top agarose

Agarose	0.7 %
MgSO$_4$	0.01 M

SDS-PAGE gel solutions

Stacking buffer

Tris-HCl   0.5 M, pH 6.8

Separating buffer

Tris-HCl   1.5 M, pH 8.8

SDS loading buffer

dH$_2$O	4.0 ml
Tris-HCl, pH 6.8	1.0 ml
Glycerol	0.8 ml
SDS, 10 %	1.6 ml
β-mercaptoethanol	0.4 ml
Bromophenol blue, 0.05 %	0.2 ml

Electrophoresis buffer (5X)

Tris	0.5 M,
Glycine	1.9 M
SDS	0.5 %

Western blot solution

Western transfer buffer

5X electrophoresis buffer	200 ml
Methanol	200 ml
dH2O	600 ml

Tris buffered saline (TBS) 10 X

| Tris-HCl | 0.5 M |
| NaCl | 1.5 M |

TBST

TBS 1X + 0.05 % Tween 20

Appendix 2: Primer sequences

Helt2/7:

TCT ACT GGT CAA ACA TGT GGT G

Helt9:

ATG TTG TAT TTG TTC CAC CGC

Helt8a:

CGG AGG CAA GGC TTG GTT TGT TC

Helt9a:

TGC GGT GGA AGG AAT ACA ACG

Ann1:

ATC GGA GCT CAC CCT TCA ACT TCC CAC TCA GCC TG

Ann2:

ATC GAA GCT TAA CCA AGC CTT GCC TCC GTT GCG

Ann5:

ATC GCC CGG G AT GGA GAA TTC TCA CTG ATA GCC A

Ann6:

ATC GGG TAC CAT CAA TAA AAT CCA GGC ACA AAG A

Ann7:

ATC GCG CTA GCA TGG AGA ATT CTC ACT GAT AGC CA

Ann8:

GTG GCG GCT TGC GTC TCA TCA TCG

Ann14:

AAT CCA GGC ACA AAC AGG TAT TCG

Ceelt2III:

ACT GGC TAG CCT ACC AAC TCA GAA CAT GGA T

Ceelt2IV:

ACT GCT GCA GCA CAA GTC CCT GCC GAC GGG AC