Characterisation of Nematode Prolyl 4-Hydroxylase Collagen Modifying Enzymes

Alan D. Winter

Wellcome Centre for Molecular Parasitology
University of Glasgow

Submitted for the degree of Doctor of Philosophy at the University of Glasgow

December 2002

Declaration

The work presented in this thesis was performed entirely by the author except where indicated. This thesis contains unique work and will not be submitted for any other degree, diploma or qualification at any other university.

Alan Winter, December 2002

Acknowledgements

The following people are thanked for their kind provision of reagents and materials: The *C. elegans* Genetic Centre (University of Minnesota, Minnesota, USA) and *C. elegans* community labs for nematode strains; Rick Maizels and Yvonne Harcus (both ICAPB, University of Edinburgh) for *Brugia malayi* worms; Bill Gregory (ICAPB, University of Edinburgh), for *B. malayi* time course cDNA; Mark Blaxter and Jen Daub (both ICAPB, University of Edinburgh) for *B. malayi* ESTs and BAC library screening; Steven Williams (FGP, Smith College) for *B. malayi* ESTs and cDNA library; Iain Johnstone (WCMP, University of Glasgow) for *C. elegans* timecourse cDNA, dpy-7-GFP plasmid and the DPY-7 antibody; Andy Fire and co-workers (Carnegie Institution of Washington, Baltimore) for plasmid vectors from the pPD series, plasmid, pRF-4, the synthetic intron insertion protocol and *E. coli* strain HT115(DE3); Robert Horvitz (MIT, Massachusetts, USA) for p76-16B; Chuck Shoemaker (AgResearch Ltd., New Zealand) for the BA7-1 plasmid; Melanie Thein and Tony Page, (WCMP, Glasgow) for strains TP12 and TP13; Päivi Riihimaa, (Collagen Research Unit, Oulu, Finland) for the *Ce-phy-3* reporter construct; Liisa Kukkola (Collagen Research Unit, Oulu, Finland) for the *Ce-phy-4* and *Ce-phy-5* reporter constructs; Gillian McCormack (WCMP, Glasgow) for the *C. briggsae* gene cloning. Johanna Myllyharju (Collagen Research Unit, Oulu, Finland) is thanked for human αI and αII antibodies and especially for her collaborative work in insect cell expression of *Ce*-PHY-2 and *Bm*-PHY-1. Some of the work performed by Laura Gilchrist and Jimi-Carlo Bukowski-Wills for their BSc (Hons) projects at the University of Glasgow is included here.

The following people are acknowledged for the permission to quote their unpublished data; J. Novelli and J. Hodgkin (*dpy-31*); C. Thaker and A. Rose (*dpy-5*); J. Crew and J. Kramer (*bli-1* and *bli-2*); M. Sapio, M. Hilliard and P. Bazzicalupo (*cut* genes); E. Maine and A. Smardon (*dpy-17*).

The help and advice of everyone at WUMP, past and present, was appreciated, especially Iain Johnstone, Brett Roberts, Collette Britton and my project assessors Dave Barry and Tansy Hammarton. I am also grateful to Brett for proof-reading this thesis. My supervisor Tony Page is thanked particularly for the opportunity to undertake this project and for his encouragement and support.

Lastly, to my family, friends and Linda a huge thank-you for everything.

Contents

Chapter 2

Materials and methods

Chapter 3

The combined functions of *C. elegans phy-1* and *phy-2* are essential for development and morphology due to their cuticle collagen modifying activity

Chapter 4

C. elegans PDI-2 is essential for development as it is the single P4H β subunit

Chapter 5

C. elegans PHY-1, PHY-2 and PDI-2 combine in unique ways to form enzyme complexes and are the only cuticle collagen modifying P4H subunits

Chapter 6

Cloning of a hypodermally expressed independently active *Brugia malayi phy* homologue

Chapter 7

General discussion

List of figures and tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

Summary

The function of prolyl 4-hydroxylases (P4H) in the formation of the nematode cuticle was studied. The cuticle is one of the two major forms of extracellular matrix (ECM) in the nematode and performs vital roles in these animals including acting as an exoskeleton to maintain body morphology. Nematodes develop from an embryo through four larval moults to the adult stage. Each larval stage is characterised by the synthesis of a new cuticle and shedding of the existing structure. The nematode cuticle is a complex multi-layered structure formed principally from collagens that are synthesised by the underlying hypodermal tissue. Collagens are characterised by repeats of the amino acid sequence Gly-X-Y, where Gly is glycine and X and Y can be any residue but are most commonly proline and 4-hydroxyproline respectively. Three collagen monomers combine to form a triple helix, with the presence of 4-hydroxyproline residues stabilising the structure. The enzyme P4H modifies Y position proline residues in newly synthesised collagen molecules within the endoplasmic reticulum (ER) of the cell to produce 4-hydroxyproline.

P4H role in cuticular ECM formation was examined primarily using the free-living species *Caenorhabditis elegans* due to its ease of handling and culture in the laboratory, the range of genetic and transgenic techniques available, and the complete genome sequence. Characterised collagen P4H from other species are most often multi-enzyme complexes formed from catalytically active α subunits with the β subunit being the enzyme protein disulphide isomerase (PDI). The role of PDI in these complexes is not connected to its enzymatic activity but instead functions to keep the α subunits in a catalytically active form within the ER of the cell. The described vertebrate P4Hs are $\alpha_2\beta_2$ tetramer complexes. Two different α subunits have been identifed which form into separate enzyme complexes with a common PDI β subunit.

In this study the genes *Ce-phy-1*, *Ce-phy-2* and *Ce-pdi-2* were examined for their role in cuticular ECM formation in *C. elegans*. These genes were shown to be expressed throughout development in cuticle collagen synthesising hypodermal tissue in a manner reflecting the expression of their substrate, placing the encoded enzymes in the appropriate tissue for collagen modification, at times of maximal collagen synthesis.

Disruption of *Ce-phy-1* gene function produced nematodes with a mutant body shape known as dumpy (Dpy). This demonstrated that this gene is required for correct body morphology and led to the identification of the strain *dpy-18* as a *Ce-phy-1* mutant. This represented the first identification of a P4H mutant in any organism. Double disruption of *Ce-phy-1* and *Ce-phy-2* or *Ce-pdi-2* singly resulted in an embryonic lethal phenotype due to the loss of P4H activity resulting in a cuticle unable to maintain nematode body shape. Disruption of these genes was demonstrated to have an affect on the localisation of specific cuticle collagens. The forms of P4H complex present were examined which revealed the presence of a unique mixed α subunit tetramer, in addition to the already described dimer form of the enzyme. Examination of three divergent putative *Ce-phy* genes did not reveal any role for these in modification of the major ECMs in this nematode and showed that only *Ce*-PHY-1, *Ce*-PHY-2 and *Ce*-PDI-2 perform the essential modification of cuticle collagens.

P4H was also studied in the human infective filarial nematode *Brugia malayi* which is one of the causative agents of lymphatic filariasis in humans, a disease that affects over 120 million people. A *phy* gene homologue, named *Bm-phy-1*, was cloned and characterised from this organism. In contrast to the both human α subunit-encoding genes, which were shown to rescue the *C. elegans dpy-18* P4H mutant, expression of *Bm-phy-1* was not sufficient to repair the mutant form of these nematodes. Expression of this gene was demonstrated in all life cycle stages examined, with the gene promoter directing expression of a reporter gene to the hypodermal cells of *C. elegans*.

Chapter 1
General introduction

1.1. Introduction

In this study the role of the prolyl 4-hydroxylase (P4H) enzyme class in formation of nematode extracellular matrices (ECMs) was examined. ECMs in nematodes and all animals are formed predominantly from collagens, which are produced via a number of enzymatic modification steps, including hydroxylation of proline residues by P4H. The wide range of human diseases resulting from the mutation or mis-processing of collagens demonstrates their critical function in development (Myllyharju and Kivirikko, 2001). Nematodes have two forms of ECM, the exoskeleton or cuticle that almost entirely encases the body, and the basement membranes that cover most of the internal organs (Kramer, 1997). The function of P4Hs in formation of the nematode cuticle was the particular focus of this work and was studied primarily in the free-living nematode *Caenorhabditis elegans*. This organism is an excellent model for the study of ECM formation due to the array of genetic, molecular and biochemical techniques possible in this system. Additionally, completion of the entire genome sequence (Consortium, 1998) enables identification of entire gene families of putative ECM components (Johnstone, 2000) and modification enzymes. A number of intra- and extracellular modification steps are required for formation of mature collagen (Prockop and Kivirikko, 1995), which must then be assembled in to higher order structures to form complex ECM structures, such as the nematode cuticle. The components and structure of the nematode cuticle, which have been studied most comprehensively in *C. elegans*, and the enzymes required to process and modify collagens, for which much work has focused on vertebrates, will be the central topics of this chapter. *C. elegans*, in addition to being a model animal, is also being increasingly utilised as a model for the less experimentally accessible parasitic nematode species (Bürglin *et al.*, 1998), such as the filarial nematode *Brugia malayi*. Homologues of parasitic genes of interest can be studied in their free-living relative and knowledge of gene function gained in *C. elegans* then applied to the parasite species. *C. elegans* can also be used more directly to examine parasite genes by its use as a heterologous expression system (Hashmi *et al.*, 2001). Both of these approaches were taken in this work to investigate a P4H from *B. malayi*.

1.2. Structure and functions of the nematode cuticle ECM

The cuticle covers the outermost surface of nematode epithelial cells and also lines the pharynx and rectum. The cuticle is an ECM that performs a range of vital functions in nematode worms, as its correct formation is required for maintenance of body shape, motility and interaction with the environment. ECMs are a network of secreted extracellular macromolecules through which cells in tissues are in contact. These matrices hold cells and tissues together and provide an organised lattice for cells to migrate and interact (Prockop, 1998). ECM is found predominantly in connective tissues where it is more abundant than the cells it surrounds. Unlike most metazoans, no interstitial matrix is found between the cells of tissues in *C. elegans*. Nematode cuticular ECMs are composed primarily of collagens (Cox *et al.*, 1981a) which, along with other cuticular components, are secreted from the underlying hypodermal tissue. Nematodes progress through four larval stages (L1-L4) to the adult with a new cuticle being synthesised for each stage (Singh and Soulston, 1978). An alternative third-stage, the dauer larvae, is formed in conditions of low food or high population density. This highly resilient developmentally-arrested stage can re-enter the developmental cycle at L4 when food is available. Five cuticles are therefore synthesised throughout nematode development, the first is formed at the end of embryogenesis and becomes the cuticle of the first larval stage. Newly synthesised cuticles are formed beneath existing cuticles that are then shed. The new cuticle allows for growth during the following developmental stage due to its highly folded nature. The process of moulting can be subdivided into three stages (Singh and Soulston, 1978); the first, early lethargus, is characterised by a gradual decease in activity and separation of the old cuticle from the underlying hypodermis. The new cuticle is then synthesised in late lethargus during which nematodes rotate rapidly around their longitudinal axis which loosens the old cuticle. Shedding of the old cuticle during ecdysis completes the moulting process.

The structure of all nematode cuticles examined appears similar in terms of overall structure and composition, and has been analysed most extensively in *C. elegans* (Cox *et al.*, 1981a). The cuticle is a multi-layered extracellular matrix that completely encases the animal with the exception of small openings at the pharynx, anus, excretory pore and vulva. The ultrastructure of the cuticle consists of six main layers (Figure 1.1). These are from external to internal; the epicuticle, cortical, medial, two fibrous layers and a basal layer. The basal layer is an amorphous region directly above the hypodermal

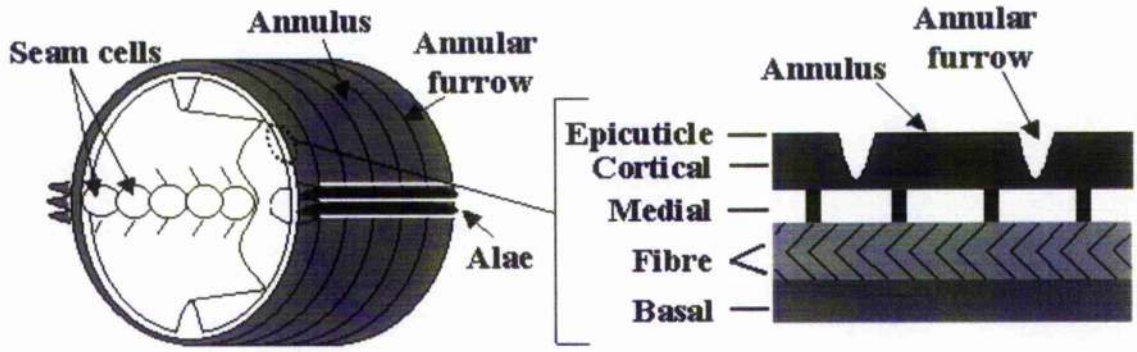


Figure 1.1. Diagrammatic representation of the *C. elegans* cuticle structure
Transverse section of the cuticle and hypodermis of an adult worm showing the annuli, annular furrows and lateral alae structures visible on the cuticle surface. A section through the cuticle is also represented showing the different layers of the cuticle; the amorphous basal layer just above the hypodermis, the striated fibrous layers, the adult specific medial layer, cortical and surface epicuticle layers. The lateral hypodermal seam cells required for formation of the alae are also represented.

tissue which is followed by the two fibrous layers, each of these consisting of tightly organised fibres that spiral along the length of the worm in opposite directions. The internal cortical layer has no visible regularity to its structure and is confined by the thin external cortical layer, which is then covered by the epicuticle. Discernible external features on the cuticle surface of all stages are the annuli that are created by narrow evenly spaced indents (or annular furrows) that run circumferentially around the animal (Figure 1.1). Also visible on the cuticle surface are the alae that are only found in the L1 (double alae), dauer (five-fold alae) and adult (three alae) stages (Cox *et al.*, 1981b; Singh and Sulston, 1978). The alae run along the lateral sides of the animal and consist of raised tread-like protrusions that form over the hypodermal seam cells. As well as these surface differences found between different developmental stages, differences exist between stages in the number of layers, the fine structure of particular layers, and, in the case of dauer larvae, the thickness of the entire cuticle in relation to body diameter (Cox *et al.*, 1981b). A defined medial layer is unique to the adult cuticle and consists of columns of material termed struts, generally positioned on either side of an annular furrow with the spaces between the struts thought to be fluid filled (Cox *et al.*, 1981b). The species of collagens present in the cuticle is also stage-specific with particular collagens being developmentally regulated such as *col-19* which is restricted to the adult stage (Liu *et al.*, 1995) and *col-2* which is expressed only in dauer larvae (Kramer *et al.*, 1985).

1.3. Components of the cuticle

The cuticle of *C. elegans,* like other nematodes, is composed predominately of collagen, with collagenous material forming approximately 80% of this ECM (Cox *et al.*, 1981a). These reducible and non-reducible cross-linked collagen proteins form the basal, fibre, struts, and internal cortical layers of the *C. elegans* cuticle. The non-collagenous fraction of the cuticular material is termed cuticlin. This is the insoluble, collagenase resistant, material that remains after treatment of cuticles with strong reducing agents and detergents. This insoluble portion of the cuticle consists of the external cortical layer and a limited amount of the internal cortical layer (Cox *et al.*, 1981a). The proteins in these regions are held together by non-reducible covalent cross-links. Although cuticlin can be isolated from all stages of the *C. elegans* life cycle it is particularly prevalent in the resistant dauer stage larvae (Cox *et al.*, 1981b), and may thus contribute to the particularly robust cuticle found in this larvae.

1.4. Collagen structure

Collagens are a family of closely related though distinct molecules that form the majority of ECMs in all animals. Collagen molecules are defined by the presence of the repeat sequence of residues Gly-X-Y, where Gly represents glycine with X and Y representing any residue. Three monomeric collagen molecules combine together to form a triple helical structure formed by winding of each molecule round a common axis. This unique tertiary structure formed from the Gly-X-Y repeat regions requires that every third residue be the smallest amino acid glycine, as a result of space constraints at the centre of this coiled-coil molecule. No such size limitation is imposed upon the residues occupying the X and Y positions of the repeat sequence. Proline and 4-hydroxyproline most often occupy positions X and Y of the repeats, as these two ring amino acids provide stability to the triple helix. 4-hydroxyproline further stabilises the triple helix as it is required for many of the hydrogen bonds and water bridges that provide additional stability to the triple helical structure (Kivirikko *et al.*, 1992).

Nematode cuticular collagens are most similar to the vertebrate FACIT (fibril-associated collagens with interrupted triple helices) type collagens (Shaw and Olsen, 1991) although in terms of size, nematode collagens are about half that of the FACIT collagens. FACITs differ from other vertebrate collagens as they posses more than one triple helical region formed from Gly-X-Y repeat regions. FACIT collagens do not undergo proteolytic processing from a larger precursor form. As their name implies these multi-domain collagens are located on the surface of fibrils (staggered arrays of fibrillar collagen molecules) and are thought to be important as molecular bridges required for the organisation and stability of extracellular matrices.

It is not currently known whether nematode cuticle collagen trimers result from homo- or hetero-trimerisation of monomeric chains. Vertebrate collagens have been demonstrated to form from the combination of three identical chains, three different chains, or two chains of one form with a single chain of a different form. Although genetic evidence exists for the interaction of more than one type of *C. elegans* cuticular collagen within a triple helical molecule (Kramer and Johnson, 1993; Levy *et al.*, 1993; Nyström *et al.*, 2002) no direct evidence has confirmed this. Rather than involvement in the same triple helix the observed interactions could reflect higher order associations between triple helices or larger complexes.

1.5. Collagen biosynthesis

Collagens have been studied most extensively in vertebrate systems (Engel and Prockop, 1991) and this provides a model on which to base what is known of nematode collagens. In vertebrates, production of a mature collagen occurs through a number of enzymatic modification steps and processing both intra- and extracellularly (Figure 1.2). As collagen molecules are synthesised a number of hydroxylation reactions occur. Hydroxylated residues in collagen occur from the modification of peptide-bound residues, not from the incorporation of hydroxylated amino acids. Y-position prolines are hydroxylated to 4-hydroxyproline (4Hyp) by prolyl 4-hydroxylase (P4H) enzymes in a step which is critical for collagen maturation. Lysine residues in the Y position of collagen repeats are also hydroxylated to 5-hydroxylysine. Hydroxylation of some X position prolines to 3-hydroxyproline has also been described; however this occurs exclusively in the sequence Gly-3Hyp-4Hyp-Gly (Kivirikko *et al.*, 1992). Glycosylation of hydroxylysine residues by galactose or galactose and glucose, with glycosylation of certain asparagine residues is also found. Three collagen monomers then associate which, in many vertebrate collagens, is directed by the structure of their non-collagen like C-terminal domain. At this stage a number of intra- and inter-chain disulphide bonds are formed by the enzyme protein disulphide isomerase (PDI). A triple helical region is then formed which propagates along the length of the molecules in a C- to N-terminal direction (Engel and Prockop, 1991). A rate-limiting step in this procedure is the *cis-trans* isomerisation of peptidyl proline bonds. The presence of the sequences Gly-Pro-Y and Gly-Pro-4Hyp predisposes unfolded chains to form *cis* peptide bonds. Peptide bound proline residues must be in the *trans* configuration in the native triple helix and thus the slow *cis-trans* isomerisation step is rate-limiting. This reaction is catalysed in collagen biosynthesis by a group of enzymes termed the peptidyl prolyl *cis-trans* isomerases (PPIase). The fibril collagens of vertebrates are synthesised as trimeric procollagen precursors that contain non-Gly-X-Y domains at both their N- and C-termini. After secretion of the procollagen, the N- and C-propeptides are each cleaved by specific peptidases. The final steps of collagen assembly involve the cross-linking of molecules which for vertebrate fibril forming collagens is performed by lysyl oxidase converting some lysine and hydroxylysine residues. As well as catalysing the formation of disulphide bonds, PDI functions in other important ways during collagen biosynthesis. As will be discussed in a latter section, PDI is the β-subunit of the enzyme

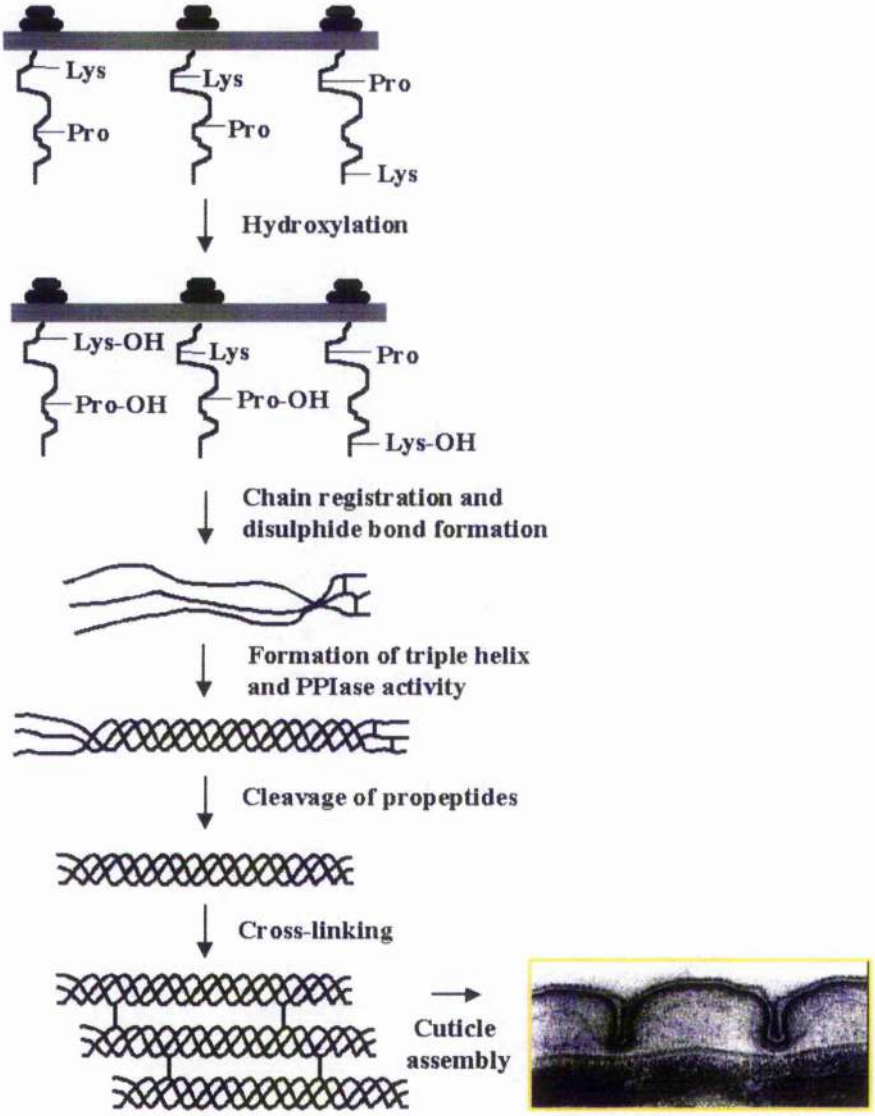


Figure 1.2. Collagen biogenesis

Some of the intra- and extra- cellular events in the formation of mature processed collagen are shown. Hydroxylation of proline by P4H residues occurs co-translationally as collagens are translated by membrane bound ribosomes. Three collagen monomers then associate and are held in registration by inter-chain disulphide bonds, catalysed by PDI. Formation of a triple helix then occurs which most likely occurs in an N- to C-terminal direction with the rate-limiting isomerisation of peptidyl-proline bonds performed by PPIase enzymes such as the CYPs and FKBs. These pro-collagen molecules are then processed by N-terminal and possibly C-terminal proteases. Cross-linking of collagens occurs both within and between trimers to form higher order structures, eventually leading to the complex multi-layered structure depicted.

complex P4H (Pihlajaniemi *et al.*, 1987) and is required for this complex to function correctly. Folding of collagens also requires the function of molecular chaperones. Chaperones involved in collagen biogenesis are the P4H complex (Walmsley *et al.*, 1999), independently functioning PDI (Wilson *et al.*, 1998) and the heat shock protein Hsp47. Hsp47 function has been shown to be essential for development in vertebrates as a mouse genetic knock out of Hsp47 is embryonic lethal (Nagai *et al.*, 2000).

1.6. Cuticle collagens of *C. elegans*

While the assembly of most collagens in vertebrates follows this basic outline, important exceptions occur that should be borne in mind for analysis of nematode collagens. Additionally, many non-fibrillar vertebrate collagens contain N- and C-regions that are non-Gly-X-Y but are not cleaved, such as the FACIT collagens mentioned previously (Shaw and Olsen, 1991). Some also lack large C-terminal domains and may associate through disulphide bonding. The triple helices of the transmembrane vertebrate collagen types XIII and XVII may fold in the N- to C-terminal direction (Myllyharju and Kivirikko, 2001). Also, importantly, hydroxylysine residues are not generally found in the *C. elegans* cuticle only being present in the dauer stage larval cuticle (Cox *et al.*, 1981b). No 3-hydroxyproline enzyme has been cloned for any species and in *C. elegans* no homologue of the chaperone Hsp47 has been identified. Additionally the cross-links in nematode cuticles appear to be of a different form from those found in vertebrates as they consist of di-, tri-, and/or isotyrosine residues (Fetterer *et al.*, 1993), which may explain the lack of hydroxylysine detected in nematode cuticles.

Completion of the *C. elegans* genome sequencing project (Consortium, 1998) has enabled identification of the entire complement of cuticular collagen genes in this organism. A total of approximately 154 genes, the majority of which are predicted gene products, have been identified as cuticular collagens (Johnstone, 2000), with the encoding genes being dispersed throughout the genome. For those genes that have not been experimentally confirmed, their classification as cuticle collagens is based on the identification of particular molecular characteristics. Nematode cuticular collagens have the relatively small size of between 26 and 35 kDa, compared to vertebrate collagens of over 150 kDa, and *C. elegans* type IV basement membrane collagens (Guo and Kramer, 1989) which are approximately 180 kDa in size. The genes encoding these proteins are

also relatively small, typically less than 2 kilobases in length with few (generally 1-3) intronic regions. The proteins are defined by two regions of Gly-X-Y repeats flanked by three clusters of highly conserved cysteine residues (Figure 1.3). The N-terminal Gly-X-Y domain is smaller containing 8-10 repeats, with the larger C-terminal triple helical region containing 40-42 repeats. The C-terminal region typically contains 1-2 small regions of non-Gly-X-Y sequence, the exact position of which varies between molecules. Collagen genes also encode non-Gly-X-Y repeat N- and C-terminal domains. The N-terminal non-repeat region is variable in size, and contains regions of homology that include the signal peptide cleavage site and a proposed subtilisin-like cleavage site (Yang and Kramer, 1994). The signal peptide is required for directing proteins to the ER for secretion while the subtilisin site is likely to be necessary for processing of the pro-collagen in to a mature collagen molecule (Yang and Kramer, 1999). The C-terminal non-collagen domain is also variable in size often being extremely short, for example 14 residues encoded by *col-19* (ZK1193.1). In contrast *bli-1* (C09G5.6) encodes a C-terminal extension of 400 amino acids, approximately ten times longer than average. However, whether the C-terminal regions of any of the *C. elegans* cuticle collagens are cleaved by a propeptidase, as seen with the vertebrate fibrillar collagens, has not yet been firmly established.

The number, position and spacing of the cysteine residue clusters flanking the Gly-X-Y domains are important when attempting to classify the collagen multi-gene family. The collagens are compared and grouped with an emphasis on their conserved cysteine clusters as the high degree of conservation in the Gly-X-Y regions of these molecules complicates analysis of inter-relatedness. These classifications are supported by the similar groupings generated when collagens are classified on either cysteine residues alone or according to homology in their N- and C-termini regions (Johnstone, 2000). This results in the ordering of the predicted collagens into six groups, 1, 1A, 2, 3, *dpy-7* and *dpy-2* (Johnstone, 2000). The size of the groups varies from group 1 with sixty-eight members to the *dpy-2* group with only two members. These groups can be viewed at (http://www.worms.gla.ac.uk/collagen/cecolgenes.htm).

1.6.1. *C. elegans* collagen mutants

Fifteen body morphology mutants have been characterised as being due to mutation in members of this collagen multi-gene family. These loci were originally identified by

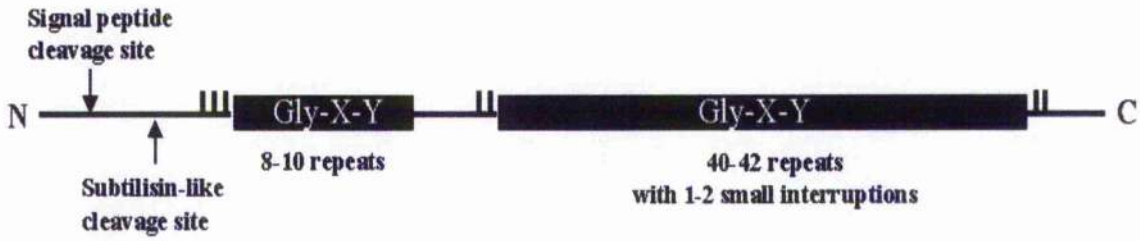


Figure 1.3. Structure of a *C. elegans* collagen monomer

Shaded boxes represent the Gly-X-Y repeat triple helix forming domains. In nematode collagens these domains are punctuated by small areas of non-repetitive sequence. The non Gly-X-Y regions present at the N- and C- termini, and separating the two major regions of repetitive sequence are represented by horizontal lines. Conserved cysteine residue clusters are depicted by vertical lines. These residues form disulphide bonds and their spacing and positioning in the monomer are important in collagen classifications. The signal peptide, that directs the collagen to the ER and the N-terminal protease cleavage, which is removed to form the mature molecule, are indicated.

random mutagenic screens and have subsequently been characterised as cuticular collagen genes. Morphologically aberrant forms resulting from mutated collagens take the form of Dpy (dumpy, short and fat), Lon (long, up to 50 % increase in body length compared to wild type), Rol (roller, helically twisted body), Bli (blister, fluid filled blistering of the cuticle) and Sqt (squat, which are a complex mixture of Dpy and Rol phenotypes). Different mutations are thought to affect the cuticle in different ways resulting in the varied phenotypes observed. The Dpy phenotype has been proposed to result from greater radial extensibility of the cuticle which, when subjected to the high internal hydrostatic pressure, produces an animal of a more spherical appearance (Kramer *et al.*, 1988). In contrast, the Lon phenotype could be due to a less radially extensible cuticle resulting in an increase in body length (Kramer *et al.*, 1988).

Genetic mutants of *C. elegans* cuticle collagens are described in Table 1.1. Genes from the group 1 collagens that have been identified by mutation are, *sqt-1* (Kramer *et al.*, 1988), *dpy-17* (E. Maine and A. Smardon, personal communication), *lon-3*, (Nyström *et al.*, 2002), *rol-6* (Kramer *et al.*, 1990), *dpy-5* (C. Thaker and A. Rose, personal communication), and *bli-1* (J. Crew and J. Kramer, personal communication); from group 2, *rol-8* (Bergmann *et al.*, 1998), *bli-2* (J. Crew and J. Kramer, personal communication) and *dpy-3* (McMahon *et al.*, 2003); and from group 3, *sqt-3* (van der Keyl *et al.*, 1994) and *dpy-13* (von Mende *et al.*, 1988). Also, from the *dpy-7* group, *dpy-7* (Johnstone *et al.*, 1992) and *dpy-8* (McMahon *et al.*, 2003) have been identified; and from the *dpy-2* group, *dpy-2* and *dpy-10* (Levy *et al.*, 1993). The only cuticle collagen group not to be represented by mutation is the group 1A, which are highly homologous and may be functionally redundant (Johnstone, 2000). The dramatic morphological changes resulting from mutation of single collagen genes illustrates that the wild type cuticle of *C. elegans*, and therefore most likely all nematodes, is critical for determination of worm body shape.

1.6.2. Allele-specific phenotypes

Multiple mutant alleles have been isolated for many of the collagen genes. Depending on the nature of the molecular lesion, these alleles can give rise to different phenotypes (Kramer and Johnson, 1993). These complex phenotypes are a result of these genes encoding components of a mechanical structure with multiple interactions possible between each component at the levels of trimer formation, inter-trimer associations, and

Gene	Cosmid	Group	Phenotypes
sqt-1	B0491.2	1	Dpy, RRol, LRol, wLon w.t.(null), Dom
dpy-17	F54D8.1	1	Dpy
lon-3	ZK836.1	1	Lon
rol-6	T01B7.7	1	RRol,wDpy, w.t. (null), Dom
dpy-5	F27C1.8	1	Dpy
bli-1	C09G5.6	1	Bli
rol-8	ZK1290.3	2	LRol
bli-2	F59E12.12	2	Bli
dpy-3	EGAP7.1	2	Dpy, DLRol
sqt-3	F23H12.4	3	Dpy, LRol, Dom
dpy-13	F30B5.1	3	Dpy
dpy-7	F46C8.6	*dpy-7*	Dpy, DLRol
dpy-8	C31H2.2	*dpy-7*	Dpy, DLRol
dpy-2	T14B4.7	*dpy-2*	Dpy, DLRol
dpy-10	T14B4.6	*dpy-2*	Dpy, DLRol, LRol, Dom

Table 1.1 *C. elegans* collagen genes defined by mutation

C. elegans collagen genes defined by genetic mutation. Cosmid gene name, collagen group and the range of phenotypes resulting from mutation in each gene are listed. Abbreviations used are: w.t., wild type; Dom, dominant mutant alleles; Bli, blister; Dpy, dumpy; DLRol, dumpy left roller; Lon, long; LRol, left roller; RRol, right roller.

interactions between cuticular layers. Phenotypes can also show stage-specific differences such as nematodes homozygous for *sqt-1(e1350)* which are wild type from L1-L2, Rol in L3 and dauer, and are Dpy in L4 and adult (Park and Kramer, 1994). Temperature specific effects have also been noted, such as the temperature-dependant *sqt-3(e2117)* phenotype which is embryonic or early larval lethal at high temperatures but almost wild type at the permissive lower temperature (Priess and Hirsh, 1986; van der Keyl *et al.*, 1994). Genetic interactions of several collagen genes have been described which may reflect interactions between their gene products. For example, a nonsense mutation of *sqt-1* suppresses the *rol-6* phenotype suggesting that the *rol-6* gene product is dependent on the presence of *sqt-1* collagen (Kramer and Johnson, 1993). The Lon phenotype of *lon-3* is also strongly suppressed by the absence of both *sqt-1* and *rol-6* gene products (Nyström *et al.*, 2002).

Null or putative null mutant alleles of *C. elegans* cuticle collagens are the most straightforward to interpret as they result in the complete absence of protein, with phenotype reflecting levels of redundancy between different collagens. Genes that produce severe phenotypes can have null alleles that are essentially wild type, such as *sqt-1(sc103)* (Kramer and Johnson, 1993; Kusch and Edgar, 1986) and *rol-6(n1178)* (Kramer and Johnson, 1993). Null alleles from other genes such as *dpy-10(cg36)* (Levy *et al.*, 1993), and *lon-3(sp23)* (Nyström *et al.*, 2002) are mutant, demonstrating that the presence of these proteins is required for normal function of the cuticle.

The most common molecular changes in collagen mutants are glycine substitutions in the Gly-X-Y repeat regions with the vast majority producing severe phenotypes. Nearly all these mutations in *C. elegans* are recessive to wild type. Replacement within the Gly-X-Y of glycine with bulkier amino acid residues would distort the formation of a triple helix as described earlier. This would produce distortions and prevent normal interchain bonding within the distorted region thereby preventing triple helix assembly and stability. The recessive nature of these mutations suggests that such distorted trimers do not become incorporated into the cuticle. If these aberrant triple helices were incorporated they could be expected to disrupt higher order structures and may thus exert a dominant effect. As dominance is not generally observed implies that the mutant trimers are not incorporated into the cuticle, and the phenotype is a result of absence of collagen rather than incorporation of mutant protein. Support for this proposition comes

from the similarity in phenotype between null and glycine substitution alleles of collagen genes such as *dpy-10* (Levy *et al.*, 1993) and *sqt-1* (Kramer and Johnson, 1993). *sqt-1* glycine substitution mutants such as *sc99* displays a much milder phenotype (abnormal hermaphrodite tail and weak long) than these lesions normally engender and as mentioned *sqt-1* nulls are essentially wild type, whereas *dpy-10* null and glycine substitutions both result in mutant phenotypes (Levy *et al.*, 1993). Additionally, a glycine substitution mutation of *dpy-7* has been shown to accumulate mutant protein in the cytoplasm with only a very small amount becoming secreted (Johnstone, 2000). It is not currently known whether cuticle collagen molecules are formed from homo- or hetero-trimers. The disruption of trimer assembly by glycine substitutions suggests that heterotrimerisation could occur. However, this association could take place only between the most similar collagens due to the differences in spacing and organisation of the repeat regions between collagens (Johnstone, 1994).

Dominant alleles of some genes exist, such as *sqt-1*(*e1350*) and *rol-6*(*su1006*), that are substitutions of conserved arginine residues in the N-terminal non-Gly-X-Y region protease recognition site (Kramer and Johnson, 1993). The protein may therefore not be processed correctly and the mutant collagen would be assembled in to the cuticular matrix causing the distortions seen (Yang and Kramer, 1999). Different phenotypes can be found dependent on homo- or heterozygosity as exemplified by the *sqt* genes. The characteristics of dominance and differing phenotypes depending on hetero- or homozygosity are also found in *dpy-10* (*cn64*). This is also a substitution mutation in the N-terminal protease recognition site that displays a left Rol phenotype when heterozygous to wild type and is Dpy when homozygous. In a few cases dominant alleles can also result from glycine substitution mutations, as demonstrated by *sqt-3* dominant alleles *sc63* and *e24*. These are due to glycine substitutions in the triple helical encoding domain of *sqt-3* (van der Keyl *et al.*, 1994). *sqt-3*(*sc63*) is a recessive Dpy and dominant left Rol while *e24* is semi-dominant, generating moderate Dpys in the heterozygote state. Another interesting observation comes from the *lon-3* gene whose null phenotype is Lon, as mentioned. However, when *lon-3* was overexpressed from a transgene, Dpy animals are produced as a result of excessive production of protein (Nyström *et al.*, 2002). These examples demonstrate the complex range of phenotypes resulting from cuticular collagen gene mutations.

1.6.3. Collagen RNAi mutants

The current ongoing genome wide RNA interference (RNAi) screens of *C. elegans* (Fraser *et al.*, 2000; Gönczy *et al.*, 2000; Maeda *et al.*, 2001) have targeted a number of cuticular collagen encoding genes (data accessible through WormBase, http://www.wormbase.org). A total of twenty-nine genes have been assessed by RNAi, twenty-five of these were collagen genes not already represented by a mutant, three of which produced identifiable RNAi phenotypes. With RNAi treatment, cuticle collagen genes D2023.7 (group 1, chromosome V), C39E9.3 (group 1, chromosome IV) and F38B6.5 (group 2, chromosome X) all display embryonic lethality, or lethality immediately post-hatch. Some escapers for F38B6.5 were reported as displaying a morphologically mutant Sma (small) phenotype which may be Dpy. The RNAi screens also targeted four genes already defined by mutation, *dpy-17*, *bli-1*, *sqt-3* and *dpy-8*. This enables a comparison between genetic mutant and RNAi induced phenotypes to be made. For *dpy-17* and *dpy-8*, RNAi gave comparable phenotypes to the genetic mutant phenotype, *sqt-3* gave Dpy early larval lethality with *bli-1* showing embryonic lethality. The remaining twenty-two collagen genes covered by these screens were scored as wild type. Therefore a mutant phenotype has resulted from a number of the relatively small proportion of cuticle collagen genes covered by these RNAi screens. Thus it may be the case that from the remaining 114 genes, not included in the genetic mutants and not assessed by RNAi, more mutants could be discovered.

There are also a number of morphological mutants described for which the molecular basis has not been determined. Some of these may prove to be a result of mutant cuticle collagens. The mapping and cloning of these genes by phenotypic rescue methods is a labour intensive process, therefore an approach to reveal if these strains, as well as displaying gross morphological defects, have specific defects in formation of the cuticular collagens would be useful for further defining candidate cuticular collagen mutants. This process may also be useful in determining more subtle phenotypes such as collagen disruptions that do not result in gross changes in the nematode body shape. Importantly, among the remaining morphological mutants are genes that may be required for the process of X-chromosome dosage compensation and are thus not collagen-related mutants (Meyer, 1997).

1.7. Parasitic nematode collagen gene families

Cuticular collagen genes have been isolated from a number of parasitic nematode species including *Brugia malayi* (Scott *et al.*, 1995), *Brugia pahangi* (Bisoffi and Betschart, 1996), *Haemonchus contortus* (Shamansky *et al.*, 1989), *Ascaris suum* (Kingston *et al.*, 1989) and *Ostertagia circumcincta* (Johnstone *et al.*, 1996). Estimates of collagen gene family size in some of these species (Selkirk *et al.*, 1989; Shamansky *et al.*, 1989) suggest all are multi-gene families although they may be smaller than the *C. elegans* family. Collagen gene cloning from these nematodes demonstrated that like *C. elegans*, the encoded collagens are small molecules and are highly conserved between species. In some cases, direct homologues of particular *C. elegans* collagens have been identified. Homologues of the tandem gene pair *col-12* and *col-13* in *C. elegans* are represented in *O. circumcincta* and are extremely similar both in terms of genomic organisation and sequence (Johnstone *et al.*, 1996). In some cases members of a particular group will have as their most similar sequence a collagen from a different organism. This is represented by the *C. elegans* collagen *dpy-7* whose most similar sequence is found in a plant parasitic nematode species *Meloidogyne javanica Mjcol-3* gene product (Johnstone, 2000; Koltai *et al.*, 1997).

1.8. Collagen associations and gene expression

Examination of the levels of specific collagen gene transcripts throughout post-embryonic development has revealed that a precise expression pattern of individual genes is repeated during the synthesis of a new cuticle (Johnstone and Barry, 1996). Collagen gene transcripts were found at specific intervals during the period between moults when the new cuticle is synthesised, and can be classified as early, middle and late expressers. The specific timing of expression within a moult for each gene is then precisely repeated during each subsequent stage. This temporal control of collagen gene expression provides a possible mechanism through which association of appropriate collagen trimerisation partners could occur. As discussed earlier, it is not currently known whether collagen trimers are composed of homo- or heterotrimers. However, if only relatively similar collagens are assumed to associate then expression of these collagens in the same temporal compartment would favour the interactions of correct partners (Johnstone and Barry, 1996). The structural class to which a collagen gene belongs should also then be reflected in the temporal class in which it is expressed. Additionally, the timing of expression of these cuticular components may also reflect

their positioning within the final matrix with early expressed collagens possibly forming the most exterior layers and late expressed components the inside layers. The multiple complex interactions between different trimers would also be simplified by this temporal control over the pool of possible interacting molecules.

1.9. The role of the hypodermis in development and cuticle synthesis/moulting

The hypodermal tissue of the nematode is the site of expression of all the enzymes and components involved in biogenesis of the nematode cuticle. The hypodermis is also central to morphogenesis of the embryo, the maintenance of this shape which then becomes one of the primary functions of the cuticle. As such the hypodermis is the central tissue of nematode body morphology having first been instrumental in defining body shape, the hypodermis then synthesises the structure that preserves it. A brief description of the form and function of this tissue is therefore required, much of which is known from the study of *C. elegans*.

The nematode epidermis is a single layer of cells historically termed the hypodermis as the extracellular cuticle covers the apical surface of these cells. During early embryonic development in *C. elegans*, rapid cell proliferation occurs with little associated change in shape. Near the end of this period, cells on the dorsal surfaces of the embryo stop dividing and differentiate as hypodermal cells which flatten, spread and become linked through adherens junctions. These hypodermal cells then spread ventrally and anteriorly to enclose the embryo in a continuous epithelial surface closing at the ventral midline. Approximately midway through development, the embryo begins to elongate along its anterior-posterior axis, during this time there is almost no migration or division of cells (Sulston *et al.*, 1983). Change of shape of the embryo to the vermiform shape of the unhatched larvae is then driven by the contractions of circumferentially orientated actin microfilament bundles present at the apical membranes of hypodermal cells (Priess and Hirsh, 1986). These contractile forces are distributed evenly by microtubules and an extracellular layer surrounding the hypodermis (the embryonic sheath). These changes in the shape of the hypodermal cells cause the body to decrease in circumference and elongate fourfold along its longitudinal axis. Complete elongation also requires the underlying muscle cells. Body wall muscle cells lie directly beneath the hypodermis separated by a layer of basement membrane. Mutant embryos defective in the function

of body wall muscles are defective in elongation. The requirement of muscle function for complete elongation is however not yet fully understood.

1.10. Control of cuticle component gene expression

The hypodermis is also the tissue where the factors controlling expression of the components of the cuticle and the enzymes of biosynthesis function.

1.10.1. Nuclear hormone receptors

Nuclear hormone receptors (*nhr*) are a class of ligand-regulated transcriptional modulators, members of which are thought to be involved in the control of the moulting process in *C. elegans*. In *Drosophila*, moulting is controlled by the steroid hormone ecdysone through the function of the *DHR3 nhr* gene (Lam *et al.*, 1997; White *et al.*, 1997). *C. elegans* possesses over 200 putative *nhr* genes, many of which have been demonstrated to be transcriptionally active (Sluder *et al.*, 1999). Most *nhr*s are referred to as orphan receptors as their specific ligands are not known. *C. elegans nhr-23* is the homologue of the ecdysone inducible *Drosophila* gene *DHR3*, with expression of *nhr-23* found in the nematode hypodermis throughout development (Kostrouchova *et al.*, 1998). Disruption of *nhr-23* function through RNAi delivery at varied developmental stages show that its action is required during all four larval moults (Kostrouchova *et al.*, 2001). RNAi disruption results in moulting and body shape defects in affected animals, specifically with an inability to shed the previous stage cuticle (Kostrouchova *et al.*, 2001). *C. elegans nhr-25* is the homologue of *Drosophila* βFTZ-F1 which is regulated by DHR3 (Lam *et al.*, 1997; White *et al.*, 1997). *nhr-25* is expressed in the hypodermis and the somatic gonad of *C. elegans* (Asahina *et al.*, 2000; Gissendanner and Sluder, 2000). The null mutant of this gene displays embryonic lethality with embryos failing in the epidermally controlled process of elongation (Asahina *et al.*, 2000). RNAi interference reveals additional post-embryonic requirements where affected animals exhibit early larval lethality with associated moulting and morphology defects (Asahina *et al.*, 2000; Gissendanner and Sluder, 2000). These moulting defects are similar to *nhr-23* and indicate that *nhr-25* may be involved in control of moulting. However, whether *nhr-25* is downstream of *nhr-23* in *C. elegans*, as their homologues in *Drosophila* are, has not been directly demonstrated (Kostrouchova *et al.*, 2001).

Analysis of the tissue-specific expression patterns and overexpression effects of other member of the large *nhr* gene class in *C. elegans* revealed additional genes with a role in moulting or the control of expression of cuticular components. Analysis of twenty-one of the *nhrs* in the genome revealed eight that are expressed in the lateral hypodermal (or seam) cells (Miyabayashi *et al.*, 1999). Overexpression of three of these genes; *nhr-77*, *nhr-81* and *nhr-82*, gave defects in body morphology and the cuticular structure (Miyabayashi *et al.*, 1999). This suggests these *nhr* genes may be involved in the pathway of expression of stage-specific genes required for the functions of the seam cells.

1.10.2. Megalin receptor

The *lrp-1* gene of *C. elegans* is also required for correct moulting in this nematode and is related to mammalian megalin. Mammalian megalin is involved in the uptake of cholesterol. Mutation of *C. elegans lrp-1* confers a moderate Dpy phenotype with defects in the shedding and degradation of old cuticles apparent during moulting from one larval stage to the next (Yochem *et al.*, 1999). This causes most larvae to arrest at the L3-L4 moult or earlier. Nematode growth is known to require the presence of cholesterol and the phenotype of this *lrp-1* mutant is similar to that resulting from the absence of cholesterol in the growth media (Yochem *et al.*, 1999). However, although the ultimate phenotypes of the mutant and sterol starvation are identical, the effect of sterol deprivation required two generations before taking effect whereas the mutant did not show this time-lag in expression of the phenotype. LRP-1 is localised to the apical surface of the cells comprising the majority of the hypodermis (Yochem *et al.*, 1999). Thus the tissue and cellular positioning of LRP-1, its resemblance to mammalian megalin, and phenocopy of mutation by cholesterol starvation suggests that LRP-1 endocytoses sterols such as cholesterol. Defects in moulting with mutant LRP-1 could therefore be due to a failure to endocytose cholesterol. In insects, moulting is controlled through ecdysteroids which are synthesised from precursors such as cholesterol. As described earlier, the *C. elegans* homologue of DHR3 when disrupted displays moulting defects and cholesterol also appears to be required for moulting. This was shown both by the moulting effects arising from mutation of LRP-1, that may endocytose cholesterol, and by the dietary requirement for cholesterol in nematode media. These findings suggest that the control of moulting in nematodes may also occur through an ecdysone-induced process. The moulting defects found from disruption or mutation of

genes such as *nhr-25*, that may potentially be downstream targets of *nhr-23*, support this contention. However, no direct evidence for ecdysteroids controlling nematode moulting is available. Additionally, synthesis of ecdysteroids from cholesterol in nematodes has not been demonstrated and *C. elegans* appears not to have homologues of the *Drosophila* ecdysone receptor complex genes.

1.10.3. Heterochronic genes

Heterochronic genes such as *lin-29* have been shown to regulate the expression of collagen genes (Liu *et al.*, 1995). The predicted product of the *lin-29* gene is a transcription factor and has been demonstrated to affect expression of stage-specific collagen genes (Liu *et al.*, 1995). In a *lin-29* loss-of-function mutant moulting is affected, with mutants failing to terminate the moulting cycle after the L4 moult and going through repeated rounds of larval cuticle synthesis. *lin-29* is a downstream member of the heterochronic regulatory pathway with its action being needed to terminate the moulting cycle. NHRs may therefore function in regulating the expression or function of genes such as *lin-29*, possibly in response to a hormonal signal.

1.11. Enzymes of collagen biosynthesis

As described earlier, many steps are required for the formation of a mature collagen molecule. This section will describe this process with a focus on the enzymes involved.

1.11.1. P4H and its PDI subunit

A critical step in collagen biogenesis is the hydroxylation of Y-position proline residues within the Gly-X-Y repeat regions. Vertebrate fibrillar collagens contain approximately 10-12% Y-position hydroxyproline, representing approximately half of the total proline residues (Kivirikko *et al.*, 1992). Hydroxylation of proline at sequences X-Pro-Gly is catalysed by the enzyme prolyl 4-hydroxylase (P4H; EC 1.14.11.2), with this modification being required for thermal stability of collagen molecules at physiological temperatures (Kivirikko and Pihlajaniemi, 1998). This enzyme has been extensively characterised in vertebrates and is an $\alpha_2\beta_2$ tetramer (Annunen *et al.*, 1997; Helaakoski *et al.*, 1995; Vuori *et al.*, 1992a). Catalytic activity resides in the α subunits (Lamberg *et al.*, 1995; Myllyharju and Kivirikko, 1997; Myllyharju and Kivirikko, 1999), with the β

subunits consisting of the enzyme protein disulphide isomerase (PDI; EC 5.3.4.1) (Pihlajaniemi *et al.*, 1987).

In vitro analysis of this enzyme was hindered for many years by its multi-subunit structure. This problem was solved by the use of an insect cell expression system using recombinant baculoviruses (Crossen and Gruenwald, 1998). Utilising this system enabled co-infections of the insect cells with different viruses to be performed, thus generating cells that synthesised multiple recombinant proteins (Vuori *et al.*, 1992a). Formation of functional complexes from subunits can then occur allowing activity, inhibition and site-directed mutagenesis studies to be performed *in vitro*. Using cells from a multi-cellular organism to express the protein assists the correct folding and post-translational modification of recombinant proteins. Attempts have also been made to crystallise complexes purified from insect cells but these have proven unsuccessful to date, probably due to the complex structure of P4IIs. Production of recombinant active P4H has also been beneficial for the *in vitro* synthesis of collagens. Insect cells have extremely low levels of endogenous P4H activity and yeast cells, which can also be used for collagen production, have no natural P4H. Expression of recombinant P4H in these systems along with the collagen chains results in production of thermally stable collagen trimers (Lamberg *et al.*, 1996a; Vuorela *et al.*, 1997). These experiments also led to the observation that co-expression of both collagens and P4H subunits as well as resulting in better expression and assembly of collagen, led to an increase in assembly of P4H complexes (Kivirikko and Myllyharju, 1998; Lamberg *et al.*, 1996a). Thus production of a stable tetramer assembly required expression of collagen and raised the possibility that this property of the enzyme could be found in all cell types. Another interesting feature of P4H is revealed by the ability to constitute active enzyme *in vitro*. Enzymatically active hybrid complexes can be formed from the α subunits of one organism and the β subunits from another. *C. elegans* PDI-2 can function in a tetramer with human α subunits (Veijola *et al.*, 1996a). The human PDI is extremely adaptable, forming active complexes with *Drosophila* (Annunen *et al.*, 1999), mouse (Helaakoski *et al.*, 1995) and *C. elegans* subunits (Veijola *et al.*, 1994).

1.11.1.1. Vertebrate α subunits

In vertebrates, two α subunits have been described in detail (Helaakoski *et al.*, 1995; Helaakoski *et al.*, 1989; Vuori *et al.*, 1992a) while at least one other isoform has been identified in humans, but as yet remains uncharacterised (Myllyharju and Kivirikko, 2001). The two human α subunits, α(I) and α(II), both form $(\alpha)_2(\beta)_2$ tetramers with the human PDI subunit. Co-expression with both forms of the α subunit together with PDI indicate that $(\alpha I)(\alpha II)(\beta)_2$ tetramers do not form (Annunen *et al.*, 1997). The processed human α subunits are 517 and 514 amino acids long, after removal of their ER signal peptides, and show an overall amino acid identity of 64%. Expression studies showed that the α(I) containing enzyme is the major form of the enzyme in most tissues but that the α(II) type enzyme predominates in some specific tissues, and may play a major role in the development of cartilage and capillaries (Annunen *et al.*, 1998). Both forms of human α subunit undergo alternative splicing (Helaakoski *et al.*, 1994; Nokelainen *et al.*, 2001) and possess two sites for asparagine-linked oligosaccharide attachment. The enzymes produced from the different splice variants have no detectable differences (Nokelainen *et al.*, 2001; Vuori *et al.*, 1992a) and mutation of the glycosylation sites of human α(I) has no effect on the assembly or catalytic activity of the tetramer (Lamberg *et al.*, 1995).

1.11.1.2. Enzymatic reaction

Enzymatic activity of P4H enzymes requires the presence of Fe^{2+}, 2-oxoglutarate, O_2, and ascorbate (Figure 1.4). One atom from the O_2 becomes incorporated in to succinate and the other into the hydroxy group formed on the proline residue. CO_2 is generated from the reaction from the decarboxylation of 2-oxoglutarate. (Kivirikko and Myllyharju, 1998). From site directed mutagenesis studies for human αI residues, H412, H483 and D414 represent the Fe^{2+}-binding pocket, with the 2-oxoglutarate binding site corresponding to residues K493 and H501 (Myllyharju and Kivirikko, 1997). The human αI peptide-binding domain has been defined as beginning at G138 and ending at approximately S244 (Myllyharju and Kivirikko, 1999). While no covalent bonding is formed between subunits, the conserved cysteines found in α chains have been implicated, by site directed mutagenesis of human αI, in intra-chain disulphide bonding (John and Bulleid, 1994; Lamberg *et al.*, 1995). The first conserved cysteine is not thought to be important for this process but the second and third cysteine residues are

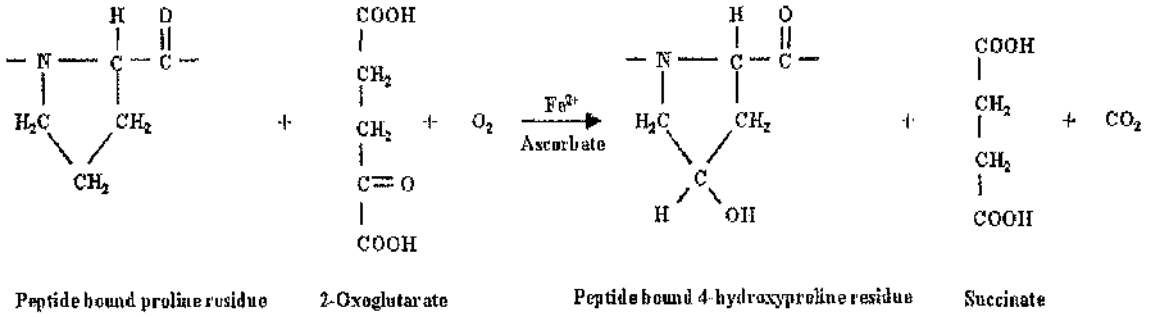


Figure 1.4. P4H reaction

Schematic representation of the reaction catalysed by P4H showing the cofactors 2-oxoglutarate, $Fe^{2+,}$ oxygen and ascorbate required for hydroxylation of Y-position proline residues in the Gly-X-Y repeats of collagen molecules. Succinate and CO_2 are produced from the reaction.

thought to form one intrachain disulphide bond, and the fourth and fifth cysteine residues a second. Both these bonds are essential for the native subunit structure.

1.11.1.3. P4H subunits from other species

As well as vertebrates, P4H subunits have been characterised from *Drosophila melanogaster* (Abrams and Andrew, 2002; Annunen *et al.*, 1999), nematodes (Merriweather *et al.*, 2001; Veijola *et al.*, 1996a; Veijola *et al.*, 1994; Wilson *et al.*, 1994) *Arabidopsis thaliana* (Hieta and Myllyharju, 2002), algae (Kaska *et al.*, 1987; Kaska *et al.*, 1988) and a virus (Eriksson *et al.*, 1999). Nineteen potential P4H may be present in the genome of *Drosophila melanogaster* (Abrams and Andrew, 2002), eight of which have been examined (Abrams and Andrew, 2002; Annunen *et al.*, 1999). These eight genes are clustered, possibly with another two predicted genes, in a 183 kb region of the *Drosophila* genome (Abrams and Andrew, 2002). Conservation of most of the residues and motifs found for other P4Hs are found in these proteins (Abrams and Andrew, 2002; Annunen *et al.*, 1999). Characterisation of one of these isoforms reveals identity of between 30-35% with the *C. elegans* and vertebrate α subunits (Annunen *et al.*, 1999). When co-expressed in an insect cell expression system, this *Drosophila* α subunit forms an active $\alpha_2\beta_2$ tetramer with its own PDI, and forms a tetramer weakly with human PDI (Annunen *et al.*, 1999). Tissue-specific embryonic expression patterns of six of the P4H genes from *Drosophila*, including the biochemically characterised form, have been demonstrated (Abrams and Andrew, 2002). Interestingly, the *Drosophila* genome only appears to encode two collagens, which are related to the type IV collagens of vertebrates, although it does encode a number of predicted proteins with collagen-like repeat regions (Abrams and Andrew, 2002). A P4H has been isolated from the parasitic filarial nematode *Onchocerca volvulus* and forms an active complex when co-expressed with its own PDI. Analysis suggests a number of other α subunit encoding genes are also present in this organism (Merriweather *et al.*, 2001). The genome of *A. thaliana* encodes six predicted P4Hs. Characterisation of one of these demonstrated that it was a low-molecular weight active monomer when expressed in an insect cell expression system (Hieta and Myllyharju, 2002). The enzyme contained the conserved catalytically important residues and could hydroxylate a range of substrates including synthetic peptides corresponding to proline-rich repeat regions in plant glycoproteins, collagen-like peptides and sequences present in hypoxia-inducible transcription factor

(Hieta and Myllyharju, 2002). Small soluble monomeric forms of P4Hs have also been identified from *Paramecium bursaria Chlorella* virus-1 (Eriksson *et al.*, 1999) and unicellular and multicellular green algae (Kaska *et al.*, 1987; Kaska *et al.*, 1988). The recombinant viral polypeptide was capable of hydroxylating synthetic peptides corresponding to proline-rich repeats coded for by the viral genome.

1.11.1.4. PDI function in P4H complexes and individually in collagen biosynthesis

PDI is an abundant ER luminal protein found in many cell types and is highly conserved between species. Usually the protein is isolated as a homodimer although monomers and homotetramers are known to occur (Ferrari and Söling, 1999). In addition to its role as a P4H subunit, the multi-functional PDI has chaperone functions and is the catalyst of the rate limiting reactions of disulphide bond formation, isomerisation and reduction within the ER. In the yeast *Saccharomyces cerevisiae*, PDI mutants are inviable (Scherens *et al.*, 1991) and its essential role in this organism appears to be a result of its disulphide bond isomerisation activity (Laboissière *et al.*, 1995). PDIs are 55 kDa modular proteins consisting of four domains; *a*, *b*, *b'*, *a'*, and an acidic C terminal domain, *c*. Domains *a* and *a'* each contain an independent active site motif CGHC, and show sequence similarity to thioredoxin. The *b* and *b'* domains have no catalytic sequence site and show no amino acid sequence similarity to thioredoxin, but are thought to have a thioredoxin-fold structure. PDI can therefore be viewed as containing two active and two inactive thioredoxin domains. Deletion of the C-terminal domain, *c*, does not appear to affect any on the enzyme, chaperone or subunit functions of PDI (Koivunen *et al.*, 1999). Coexpression of human α subunits and PDI domains demonstrated that domains *b'* and *a'* fulfilled the minimum requirement assembly for an active P4H complex (Pirneskoski *et al.*, 2001). However addition of domain *b* resulted in higher activity while the highest was achieved with all four domains (Pirneskoski *et al.*, 2001). ER resident proteins related to PDI are ERp57 (also called ERp60), ERp72 and PDIp. PDIp and ERp57 share the similar organisation of four thioredoxin-like domains (Freedman *et al.*, 2002). ERp72 also has a similar domain configuration to PDI but contains an additional thioredoxin-like domain.

PDI enzymatic activity is not required for its function as a P4H subunit as demonstrated when human PDI with either or both catalytic sequences mutated, formed fully active

P4H tetramers when co-expressed with α subunits in insect cells (Vuori *et al.*, 1992b). PDI functions as the β subunit within P4H complexes to maintain the α subunits in a non-aggregated conformation as P4H α subunits in the absence of PDI form into inactive insoluble aggregates (John *et al.*, 1993). The role of PDI within P4H complexes is however more specific than that of keeping complexes soluble as demonstrated by attempts to replace PDI with other proteins. The human α subunit had been shown to bind to the immunoglobulin-heavy-chain binding protein (BiP) which keeps the α subunits in a soluble assembly competent form (John and Bulleid, 1996). Co-expression of α subunits with BiP in insect cells confirmed this association but demonstrated that the soluble complexes had no P4H activity (Veijola *et al.*, 1996b). PDI therefore, in addition to keeping the α subunits soluble maintains them in a catalytically active conformation. The PDI-related protein ERp60 likewise could not substitute for PDI in a P4H complex, as no tetramer was formed from co-expression of α subunits with this protein (Koivunen *et al.*, 1996). PDI is also a subunit of the microsomal triglyceride transfer protein and again the role of PDI as the β subunit in this αβ dimer is not as a disulphide isomerase and serves to keep the insoluble α subunits in an active soluble complex (Lamberg *et al.*, 1996b). P4H α subunits do not contain an ER retention signal and so the P4H complexes are retained in this subcellular compartment by a retention signal contained within the C-terminus of PDI. The critical role of PDI in this process is demonstrated by deletion of this region which causes the secretion of the tetramer and free PDI polypeptide from the ER (Vuori *et al.*, 1992b).

In the cuticle of *C. elegans,* reducible bonds are formed which occur possibly between collagen trimers or at the level of higher order structure formation (Cox *et al.*, 1981a). The conserved C-terminal cysteine residues present in *C. elegans* collagens may also form disulphide bonds between monomeric collagen chains to nucleate triple helix formation. The importance of these interactions is demonstrated by the EMS *sqt-1* mutants in which replacement of one of these conserved C-terminal cysteines causes a recessive LRol phenotype (Kramer and Johnson, 1993). This mutant phenotype is distinct from the loss of function phenotype demonstrating the presence of mutant protein is inducing the morphology observed. The recessive nature of the mutant indicates that, either in the presence of wild type protein the aberrant protein fails to produce a phenotype, or, that the wild type protein is more effectively assembled than

the mutant protein. *In vitro* substitution of either of these conserved cysteines in SQT-1 or ROL-6 generates a LRol phenotype when expressed from a transgene (Yang and Kramer, 1994). Therefore disulphide bonding is required for normal collagen function but not assembly. This analysis does not however reveal at what level the disulphide bond formation is occurring. Thus PDI has multiple roles in collagen biogenesis; it catalyses disulphide bond formation possibly at both the early and late stages of collagen synthesis, is the β subunit in P4H complexes, and as described below acts as a molecular chaperone.

1.11.2. P4H and PDI as molecular chaperones in collagen biosynthesis

Molecular chaperones are required for the correct folding and assembly of proteins such as collagens within the ER. These chaperones bind to unassembled or misfolded proteins thereby preventing their aggregation or premature secretion. Proteins are then either maintained in an assembly competent form or are targeted for degradation. Human PDI mutants with double active site mutations demonstrated that PDI lacking its isomerase activity was still able to accelerate protein folding (Hayano *et al.*, 1995). As well as their enzyme related properties, P4H and PDI also perform chaperone functions during collagen biosynthesis (Walmsley *et al.*, 1999; Wilson *et al.*, 1998). Molecules that interact with retained collagen were identified using an *in vitro* translation system with semi-permeabilised cells which mimicked the initial stages in collagen assembly and modification (Bulleid, 1996). When hydroxylation was inhibited, P4H was shown to bind to unhydroxylated, non-triple helical chains. While hydroxylation was inhibited, P4H remained bound, but when activity was reactivated the enzyme became dissociated from the collagen (Wilson *et al.*, 1998). Binding of P4H to collagen was also analysed using a proline analogue that permitted hydroxylation but prevented folding. Under these conditions P4H remained bound to the non-helical, hydroxylated collagen trimers indicating that the folding status and not the extent of hydroxylation regulated the binding of P4H (Walmsley *et al.*, 1999). Thus this post-translational modifying enzyme has a role in the regulation of secretion of its substrate. The ability of P4H to interact with fully modified but non-triple helical molecules suggests a mechanism by which cells recognise misfolded mutant collagens such as those resulting from glycine substitutions. PDI plays a role in this P4H chaperone function as it is a subunit of the complex and contains the ER retention signal that retains improperly folded collagens

intracellularly. In addition to this, PDI has independent chaperone activities as it associates independently with the C-propeptide of monomeric procollagens (Wilson *et al.*, 1998). In vertebrates, the chaperone Hsp47 is essential for development as demonstrated by the embryonic lethal phenotype of the mouse knockout with associated defects in collagen biosynthesis (Nagai *et al.*, 2000). The chaperone Hsp47 interacts with the highest affinity with the triple helical form of collagen, and is hypothesised to function in stabilisation of regions of thermal instability within collagens (Tasab *et al.*, 2000). No Hsp47 homologue is identifiable in the *C. elegans* genome and this gene has not been found in any other nematode to date, highlighting the differences in collagen biogenesis existing between vertebrates and invertebrates.

1.11.3. Peptidyl prolyl *cis-trans* isomerases

Peptidyl prolyl *cis-trans* isomerases (PPIase; EC 5.2.1.8) catalyse the interconversion of prolyl imide bonds in peptide substrates. Three classes of this superfamily exist; the cyclophilins, FK506-binding proteins (FKBP) and parvulins (Göthel and Marahiel, 1999). Members of different classes are related only in terms of enzymatic activity and are unrelated in amino acid sequence. The receptors for the immonosuppresent drugs cyclosporin A (CsA) and FK506 are cyclophilin (Fischer *et al.*, 1989; Takahashi *et al.*, 1989) and FK506-binding protein (Harding *et al.*, 1989; Siekierka *et al.*, 1989) respectively. The immunosuppresent effect of the drugs is however not a consequence of suppression of their PPIase activity (Schreiber and Crabtree, 1992; Walsh *et al.*, 1992). Although isolated from many organisms, the functions of most of these enzymes are not clear with a defined role being described for only a few (Göthel and Marahiel, 1999). One proposed role for PPIases is in collagen folding. The *in vitro* re-folding of type III denatured collagen was moderately increased by the presence of partially purified PPIase (Bächinger, 1987). The folding of procollagen I in suspended chick embryo tendon fibroblasts is inhibited by the presence of cyclosporin A (Steinmann *et al.*, 1991) and in human skin fibroblasts, CsA delays helix formation of collagens I and III, resulting in overmodification and degradation (Steinmann *et al.*, 1991). These results support a role for PPIase catalysed folding of collagens both *in vitro* and *in vivo*.

Independent of its immunosuppresent effects, CsA and derivative compounds also have antiparasitic effects on a wide range of helminths, which typically results in damage to the tegument (Chappell and Wastling, 1992; Page *et al.*, 1995a). The mode of action by

which CsA compounds exert this effect is unknown but one possibility is the disruption of collagen folding. Cyclophilins (*cyp*) and FK506-binding proteins (*fkb*) have been characterised from free living and parasitic nematode species. Two *cyp* genes have been identified in *B. malayi* (Ma *et al.*, 1996; Page *et al.*, 1995b) with homologues of these being found in both *O. volvulus* and *D. immitis* (Hong *et al.*, 1998b; Ma *et al.*, 1996). A third isoform also being identified from *D. immitis* (Hong *et al.*, 1998a). CsA inhibition sensitive and insensitive isoforms of *cyps* have been described. These insensitive forms may play a part in resistance to CsA found in some nematodes. Crystal structure data from a CsA insensitive *cyp* from *B. malayi* identified alterations in residues required for CsA binding, explaining the insensitivity of this isoform to the compound (Taylor *et al.*, 1998).

Eighteen members of the *cyp* and eight members of the *fkb* families have been identified in *C. elegans* (Page *et al.*, 1996) (A. Page, personal communication). A number of these have been characterised (Dornan *et al.*, 1999; Page, 1997; Page and Winter, 1998; Page and Winter, 1999) (A. Page, personal communication) and a role in formation of the nematode cuticle proposed for one *cyp, cyp-9,* and two *fkbs. Cyps* can be classified as cytosolic, secreted or mitochondrial, *cyp-9* however does not come under any of these categories, and is grouped along with *cyp-4*, *-8*, *-10* and *-11* in a divergent *cyp* class. *cyp-9* is found in an operon with a protein disulphide isomerase gene, *pdi-1* (Page, 1997). *C. elegans* and related species are unusual among multicellular animals in having many genes arranged in operons (Blumenthal *et al.*, 2002; Page, 1999). These genes are co-transcribed from a common upstream promoter although unlike bacteria, nematode operons are not co-translated; being first processed to moncistronic units by *trans*-splicing. Examination of the tissue-specific expression pattern from the common promoter of *pdi-1* and *cyp-9* demonstrates both are expressed in the cuticle synthesising hypodermal tissue (Page, 1997). This places *cyp-9* in the appropriate tissue for modification of *C. elegans* cuticular collagens. The cycling pattern of transcript expression found between moults also points towards a cuticle synthesising function for this gene. The proposed role of the enzymes in this operon, based on biochemical function and expression, in the modification of cuticle collagens has not been conclusively demonstrated, however, as RNAi of *cyp-9* does not result in an identifiable morphological or collagen related defect (A. Page, personal communication). This may

be due to a redundancy of function where the hypodermally expressed FKBs may compensate for loss of CYP-9 activity.

1.11.4. Collagen pro-peptidases

Collagens, like many biologically important proteins, are often synthesised as precursor molecules that must be proteolytically cleaved to produce the mature functional molecule. As mentioned above, *C. elegans* cuticle collagens contain a subtilisin-like cleavage site in their N-terminal non-triple helical domain. Analysis of both mutations of the collagens at these sites (Kramer and Johnson, 1993; Levy *et al.*, 1993; Yang and Kramer, 1994; Yang and Kramer, 1999), and mutation in enzymes proposed to process the collagens at these sites (Peters *et al.*, 1991; Thacker *et al.*, 1995), provide evidence that *C. elegans* collagens are specifically cleaved at their N-terminus by proteases. Enzymes have also been identified that process both the N- (Colige *et al.*, 1999) and C-termini (Kessler *et al.*, 1996; Li *et al.*, 1996; Scott *et al.*, 1999) of vertebrate pro-collagen precursors. Subtilisin-like cleavage sites can also be identified in other parasitic nematode collagens such as *B. pahangi Bpcol-1* (Bisoffi and Betschart, 1996) (residues 14-17), *H. contortus* 3A3 cuticle collagen (Shamansky *et al.*, 1989) (at residues 76-79) *M. javanica Mjcol-3* product (Koltai *et al.*, 1997) (residues 88-91) and in the *colost-1* and *colost-2* encoded collagens from *O. circumcincta* (Johnstone *et al.*, 1996).

1.11.4.1. Cleavage of *C. elegans* N-termini

The subtilisin-like recognition motif of *C. elegans* collagens is contained in the most C-terminal region of the four blocks of homology found in the N-terminal extensions of collagens and consists of a consensus amino acid sequences of (R/K)XX(R/K) (Kramer, 1994). Mutations in the cuticle collagen genes *sqt-1*, *rol-6* (Kramer and Johnson, 1993; Yang and Kramer, 1994; Yang and Kramer, 1999), and *dpy-10* (Levy *et al.*, 1993) have been identified that alter this proposed site and are thought to block cleavage through substitution of the conserved arginines. As these initial alleles were arginine to cysteine replacements, the possibility existed that the phenotype was due to ectopic disulphide bond formation. Analysis *in vivo*, of *in vitro* mutagenised *rol-6* or *sqt-1* collagens demonstrated that loss of the protease cleavage site was the cause of the phenotype rather than addition of disulphide bonds (Yang and Kramer, 1994). Replacement of either conserved arginine of this motif with anything other than lysine (strongly charged

like arginine) resulted in a dominant right roller phenotype. Additionally, replacement of the second conserved arginine residue with cysteine or serine (structurally similar to cysteine but cannot form disulphide bonds) had the same effect. Also tested was cysteine substitution of other residues within the same homology block. No phenotype was observed in these mutants, suggesting that no ectopic disulphide bonds were formed and that functional protein was produced. Therefore mutations in this conserved N-terminal region of cuticle collagens block cleavage at the subtilisin-like site and prevent correct processing (Yang and Kramer, 1994). Analysis of two arginine substitution alleles of *sqt-1*, *e1350* and *sc1*, demonstrated that when this protease cleavage site is altered, cuticle collagen extracts showed lower levels of the SQT-1 protein, indicating that mutant collagens were not assembled into the cuticle as efficiently (Yang and Kramer, 1999). The size of mutant monomers also displayed an increase in size consistent with retention of the fragment normally cleaved by the protease (Yang and Kramer, 1999). This provides direct evidence for the cleavage of collagens of *C. elegans* at these subtilisin-like recognition sites by an endopeptidase.

The enzymes in *C. elegans*, such as *bli-4* that appear to process these collagens belong to the family of kex2/subtilisin-like proprotein convertases, due to their overall similarity to the endoprotease Kex2p in yeast and the sequence similarity of their catalytic domains to the bacterial subtilisins (Thacker and Rose, 2000). All members of this family share a similar protein domain structure, with an ER signal peptide followed by a prodomain, then subtilisin-like catalytic domain. The P domain is specific to the proprotein convertases and is followed by a unique C-terminal domain (Thacker and Rose, 2000). *C. elegans* has four members of this gene family but only one member, *bli-4* (also called *kpc-4*, for *k*ex2/subtilisin-like *p*roprotein *c*onvertase), is involved in cuticle collagen assembly. This gene is also the most complex as it produces a number of alternatively spliced products. The first twelve of the 24 exons of the *bli-4* gene are alternatively spliced to specific downstream exons. This produces nine proteins (*bli-4* A-I) with common pro-, catalytic and P domains but different C-terminal domains (Peters *et al.*, 1991; Thacker *et al.*, 1999; Thacker *et al.*, 1995). The C-terminal domains are thought to contain motifs that determine different sub- and possibly extracellular localisations of the isoforms.

The *bli-4* gene products are required for formation of a functional cuticle in *C. elegans*. The *bli-4* locus was originally identified by the recessive allele *e937* that manifests itself in the formation of fluid-filled blisters (Bli) of the adult nematode cuticle (Brenner, 1974). Thirteen additional alleles of the gene were subsequently identified which arrest during late embryonic development and are unable to complete elongation (Peters *et al.*, 1991; Thacker *et al.*, 1995). For one member of this class of lethal alleles, *e90*, approximately a third of mutant animals progress further in development and are able to hatch to produce larvae, however these are extremely Dpy in appearance (Thacker *et al.*, 1995). The failure to complete embryonic elongation, along with viable Dpy and Bli phenotypes is characteristic of cuticle collagen associated function. The cloning of this gene and demonstration of expression in the hypodermis of *C. elegans* provided an explanation of the phenotypes observed. The original allele, *bli-4*(*e937*), is a result of a deletion of exons specific for isoforms A, E, F, G and H (Thacker *et al.*, 1999; Thacker *et al.*, 1995; Thacker and Rose, 2000). The lethal alleles were due to lesions in exons 1 to 12 of *bli-4* encoding the domains shared by all forms of the BLI-4 enzyme. Thus the activity of all the BLI-4 isoforms would be affected, therefore completely removing collagen processing activity. The viable phenotype of allele *e937* suggests that the different isoforms are developmentally regulated. The Bli phenotype of *e937* is found only in the adult, this is also the only stage in which the strut structures of the cuticle are found (Cox *et al.*, 1981b). Blisters are thought to be a result of defects in the collagens forming the struts holding the fibrous and cortical layers of the cuticle together. Collagen genes *bli-1* and *bli-2* display the Bli phenotype and are missing the struts from the medial layer of the adult cuticle (J. Crew and J. Kramer, personal communication). The isoforms present in nematodes containing the *e937* allele must therefore modify all other collagens correctly during the development but are not able to correctly process the collagens making up the adult specific struts. As the rest of the adult cuticle is intact demonstrates that, as well as developmental control, BLI-4 isoforms show substrate specificity.

1.11.5. Thioredoxin

Thioredoxins catalyse various thiol-disulphide exchange reactions. PDI and other ERp proteins are members of the thioredoxin superfamily. All have at least one domain with thioredoxin homology and contain a redox-active thiol/disulphide site of the general sequence CXXC. As described in Section 1.11.1.4, cuticle collagens contain many intra-

and inter-chain disulphide bonds. Homozygous knockout mice for the single thioredoxin gene are embryonic lethal, demonstrating the essential function of this gene in vertebrate development (Matsui *et al.*, 1996). An existing genetic mutant of *C. elegans*, *dpy-11*, was shown to correspond to a novel thioredoxin, which in addition to its thioredoxin domain, which is active when expressed *in vitro*, also contained a putative signal peptide and transmembrane domain (Ko and Chow, 2002). The gene is expressed exclusively in the hypodermis from early embryogenesis and throughout larval development. Examination of the subcellular localisation of DPY-11 shows it is associated with membranous organelles, possibly the ER. Both the signal peptide and transmembrane domain are required for function of DPY-11 as constructs lacking these regions were no able to repair the mutant phenotype. No substrate was identified for *dpy-11* although its tissue and subcellular localisation, enzymatic activity and altered body morphology when mutated point strongly towards a role in modifying components of the *C. elegans* cuticle.

1.11.6. Cross-linking enzymes

1.11.6.1. Tyrosine derived cross-links

Cross-linking via di- and tri-tyrosine linkages are an important step in the synthesis of both the major fractions of the nematode cuticle, as cross-links have been found in the soluble collagenous component and the insoluble fractions of *A. suum* cuticles (Fetterer *et al.*, 1993; Fujimoto *et al.*, 1981). Synthesis of tyrosine-derived cross-links have also been identified in the sheaths of *H. contortus* larvae (Fetterer and Rhoads, 1990) and in cuticle extracts from *C. elegans* (Edens *et al.*, 2001). The cuticular collagens from *C. elegans* have a typical size of 26-35 kDa. However, when analysed, the molecular weights of over 200 kDa can be found even after treatment with reducing agents (Cox *et al.*, 1981a). Thus in addition to the reducible disulphide bridges other non-reducible bonds, such as those derived from tyrosine, are involved in the cross-linking of nematode cuticle collagens. Many *C. elegans* collagen genes encode a conserved carboxy terminal tyrosine that may be involved in the formation of these cross-links. For mutations in *C. elegans* collagen *sqt-1*, alleles *sc13* and *sc113*, the mutant protein is assembled into the cuticle, however, these mutants have different patterns of cross-linking (Yang and Kramer, 1999). The molecular change in these mutants alters an amino acid residue directly adjacent to the conserved tyrosine. This change is thus thought to affect the cross-links derived from this tyrosine imparting the roller

phenotype observed (Yang and Kramer, 1999). The lack of hydroxylysine found in nematode cuticles and the cross-links derived from them demonstrate the difference in cross-linking between nematodes and mammalian collagens.

Tyrosine derived cross-links that stabilise the *C. elegans* cuticular ECM are catalysed by two *C. elegans* Duox genes. Disruption of these genes singly or in combination produced severe body morphology and cuticle structure defects consistent with this role in modification of nematode cuticle components (Edens *et al.*, 2001). Two Duox genes (for dual oxidase, as they posses both a peroxidase domain and a gp91 *phox* domain) were identified in *C. elegans*. *Ce*-Duox1 is expressed in the hypodermis and RNAi of this gene singly, or with sequences predicted to knockdown expression of both *Ce*-Duox1 and *Ce*-Duox2, gave identical results. Over half the affected progeny from RNAi displayed a Bli phenotype and approximately a third were Dpy (Edens *et al.*, 2001). Additionally, RNAi affected animals appeared translucent and defects in movement were observed. Electron micrograph analysis of RNAi animals showed clear disruption of the cuticle with abnormal strut regions (Edens *et al.*, 2001) similar to that found in other genes with mutant Bli phenotypes (J. Crew and J. Kramer, personal communication). Analysis of wild-type *C. elegans* extracts showed that all di-, and tri-tyrosine linkages were found in the cuticle of *C. elegans*, extracts from RNAi-treated animals however showed a complete absence of these links (Edens *et al.*, 2001). The peroxidase domain of both isoforms when expressed *in vitro* demonstrated an ability to form di- and tri-tyrosine linkages using a synthetic substrate (Edens *et al.*, 2001). Homologues of the Duox genes have not yet been identified in any parasitic nematode species although EST and BAC-end sequences from *B. malayi* demonstrate a high degree of homology. Tyrosine-derived bonds have been identified in parasitic nematode cuticles (Fetterer and Rhoads, 1990; Fetterer *et al.*, 1993) and these enzymes may also be required in these species for formation of a fully functional cuticle.

Tyrosine-derived cross-links are also found in the insoluble portions of nematode cuticles and these may also be modified by Duox enzymes. The collagenease-resistant insoluble residue remaining after treatment of nematode cuticles with ionic detergents and reducing agents is termed cuticlin. This fraction in *C. elegans* is found almost entirely in the external cortical layer and has a biochemical composition similar to but distinct from the soluble fraction, with decreased levels of glycine and hydroxyproline

(Cox *et al.*, 1981a). The cuticles from dauer stage larvae contain a higher quantity of cuticlin those from other stages (Cox *et al.*, 1981b). This demonstrates the importance of this material in the resilience of the cuticle that is especially pronounced in dauer larvae.

Two genes, *cut-1* (Sebastiano *et al.*, 1991) and *cut-2* (Lassandro *et al.*, 1994), have been identified in *C. elegans* that encode components of cuticlin. CUT-1 is a component of the dauer stage cuticle and localises underneath the lateral alae of this stage (Ristoratore *et al.*, 1994; Sebastiano *et al.*, 1991). Immunolocalisation and transcriptional analysis of the second cuticlin gene, *cut-2*, indicates that its gene product is present in the cuticles of all stages, with expression peaking preceding the moults (Lassandro *et al.*, 1994). RNAi of *cut-1* results in Dpy-like dauer larvae that lack the alae (M. Sapio, M. Hilliard and P. Bazzicalupo, personal communication). These worms appear not to undergo radial shrinking with the enlarged regions corresponding to the normal positioning of the alae. The sequencing of the entire *C. elegans* genome reveals that a family of approximately thirty *cut* genes are present. Two in particular, *cut*-3 and *cut-5,* have close homology to the other cuticlins. *cut-3* is expressed in late embryogenesis and RNAi gives Dpy L1 larvae that lack alae. RNAi of *cut-5*, which is expressed in late embryogenesis and during dauer formation, affects the alae of both dauer and L1s (M. Sapio, M. Hilliard and P. Bazzicalupo, personal communication). No effect on adult body shape or alae formation was observed for any of these genes.

Recombinant CUT-2 is a substrate for cross-linking *in vitro* by horseradish peroxidase, which, in the presence of H_2O_2, polymerises CUT-2 to give high molecular weight complexes containing di-tyrosine cross-links (Lassandro *et al.*, 1994). Tyrosine residues found in many of the amino acid repeat motifs found in CUT-2 may be participating in the cross-linking of these proteins. Work on the CUT-2 protein from *C. elegans* involved the use of a heterologous cross-linking enzyme. Although candidate enzymes have been proposed for this activity in other nematodes, only recently in *C. elegans* have Duox enzymes been characterised with peroxidase activity and cuticular specific defects upon disruption of activity. It would therefore be interesting to assess the ability of the recombinant Duox enzymes (Edens *et al.*, 2001) to cross-link *C. elegans* cuticlins *in vitro* to determine if these may be one of their natural substrates. Cuticlin genes have also been found from *Brugia* species (Lewis *et al.*, 1999). The transcript of *Bpcut-1*

from *B. pahangi* cycles in a manner consistent with a component of the cuticle and the protein is localised to the median layer of the *Brugia* cuticle using a heterologous antibody (Lewis *et al.*, 1999).

1.11.6.2. Transglutaminases

The presence of another cross-linking activity contributing to the insolubility of nematode cuticles was suggested when transglutaminase (TGase; EC 2.3.2.13) activity and enzymatic products were identified in *B. malayi* (Mehta *et al.*, 1990; Mehta *et al.*, 1992). Chemical inhibitors of TGases were noted to affect the development and survival of filarial nematodes including *B. malayi* (Mehta *et al.*, 1992; Rao *et al.*, 1991). TGase inhibitors also inhibit the moulting of *O. volvulus* from L3-L4 with the products of TGase-catalysed reaction found localised to the larval cuticle (Lustigman *et al.*, 1995). TGases post-translationally modify proteins through the introduction of a covalent bond between internal glutamine and peptide bound lysine to form an ε-(γ-glutamyl)lysine isopeptide bond. This link is resistant to proteolysis and is destroyed only after complete degradation of both peptide chains. TGases are widely distributed enzymes and occur in both intra- and extracellular forms.

Isolation and characterisation of active 56 kDa TGases from *B. malayi* (Singh and Mehta, 1994) and from the dog filarial parasite *Dirofilaria immitis* (Singh *et al.*, 1995) suggested that these nematode enzymes were similar to each other but differed from the mammalian forms of the enzyme. Subsequent cloning of the enzyme from *D. immitis* revealed that this enzyme displayed no sequence homology to any other known TGase enzymes, but instead was similar to PDI and PDI-like endoplasmic reticulum proteins ERp60s (Chandrashekar *et al.*, 1998). The recombinant protein displayed both TGase cross-linking activity and PDI activity (Chandrashekar *et al.*, 1998) and inhibition of the enzyme showed effects on moulting of *D. immitis* (Chandrashekar *et al.*, 2002) similar to that found for inhibition of *O. volvulus* TGases. In *C. elegans*, a gene encoding a 57 kDa protein has also been identified that displays both PDI and TGase activities (Eschenlauer and Page, 2003; Natsuka *et al.*, 2001).

1.12. Conservation of collagen biosynthesis in *C. briggsae*

In addition to the genes encoding collagen and processing enzymes described above for parasitic nematode species, conservation of collagen biogenesis has been demonstrated between *C. elegans* and *C. briggsae*. Like its relative, *C. briggsae* is a free-living nematode, with the two species having diverged around 40 million years ago. Although almost indistinguishable in terms of morphology and development, the genomes of these two species have diverged. Analysis in *C. briggsae* is assisted by a genome sequencing project (completed September 2002) thus enabling easier identification of putative homologues in this organism. Comparison of sequences demonstrates that conservation of DNA sequence is confined almost entirely to protein coding regions and short flanking sequences. Analysis of ECM components and enzymes in both *Caenorhabditis* species allows comparisons between gene coding and control regions as well as allowing genome organisation and gene function to be examined. The *bli-4* cuticular collagen N-terminal proteinase locus (Thacker *et al.*, 1999), a *cyp/pdi* collagen specific operon (Page, 1999) and the homologue of the *dpy-7* cuticular collagen encoding gene (Gilleard *et al.*, 1997) have all been characterised in *C. briggsae*. These comparative studies have illustrated conservation of function between species, and identified important coding sequences and promoter elements.

1.12.1. *C. briggsae bli-4*

Cloning of the complex *bli-4* locus from *C. briggsae* demonstrated the conservation of an essential collagen-modifying gene between nematodes, in terms of sequence, genomic organisation and function. *cb-bli-4* was able to repair both the lethal and Bli phenotypes of *C. elegans bli-4* mutants (Thacker *et al.*, 1999) indicating that the heterologous enzymes could process the large number of *C. elegans* collagens correctly from all stages. Comparison of upstream regions additionally enabled identification of control elements that direct expression of this complex locus through conserved sequence elements. Comparison of the two genomic coding sequences revealed that four regions of homology existed. These had not been previously predicted as exons by computer analysis and were not represented in either species EST data. These regions were confirmed as transcriptionally active, thus identifying four additional isoforms encoded by this locus (Thacker *et al.*, 1999). The comparative analysis in a related organism therefore revealed the complexity of this gene locus and demonstrates the utility of this process to provide a more complete understanding of gene function.

1.12.2. *C. briggsae pdi-1/cyp-9* operon

Both genes of the *pdi-1/cyp-9* operon described in *C. elegans* (Page, 1997) have been identified in *C. briggsae* with the proteins from both species demonstrating between 86-90% identity (Page, 1999). Like *C. elegans*, the *C. briggsae* genes are arranged in a genomic organisation suggestive of an operon. This is further confirmed by the operon specific mode of *trans*-splicing. Analysis of the upstream region of the *C. briggsae* operon reveals sequence homology to the region in *C. elegans* that was shown to drive expression of both genes (Page, 1997; Page, 1999). Therefore these species show conservation of a protein folding operon which is likely to be involved in collagen synthesis, demonstrating the importance of the functionally related proteins encoded by this operon.

1.12.3. *C. briggsae* DPY-7

A component of the *C. elegans* cuticle DPY-7 is required for formation of a fully functional cuticle with mutants displaying a Dpy phenotype (Johnstone *et al.*, 1992). The homologous gene in *C. briggsae* was capable of functioning in the *C. elegans* cuticle as assayed by phenotypic rescue by interspecies transformation of *C. elegans* mutants with the *C. briggsae* gene (Gilleard *et al.*, 1997). This identified *Cb*-DPY-7 as the true homologue and not only a structurally related collagen. Additionally, as expression was directed by the *C. briggsae* promoter elements this indicated that these sequences were sufficient to promote expression in the correct tissue and development time, and to provide sufficient levels of expression (Gilleard *et al.*, 1997). Thus the sequence identity found between the two promoter regions is likely to be functionally important. For functional *C. briggsae* DPY-7 to be incorporated correctly into the *C. elegans* cuticle it must also be processed and modified by all the enzymes of biosynthesis in *C. elegans*, then associate with the correct molecules to form the appropriate structure within the ECM. Also significant is the high degree of conservation of non-Gly-X-Y domains of the protein and of the X and Y position amino acids with the triple helical domain highlighting the importance of these regions and residues for function.

1.13. Basement membrane (type IV) collagens

As mentioned previously, in addition to the cuticle, a second type of ECM is found in *C. elegans*, the basement membrane. Basement membranes are thin sheets of specialised

molecules that form an extracellular matrix surrounding tissues in all animals. In *C. elegans* basement membranes underlie the hypodermis and surround the pharynx, intestine, gonad and body wall muscles (Kramer, 1997). Type IV collagens are the major constituent of basement membranes. The structures of type IV collagens are highly conserved, with the majority of the molecules consisting of a Gly-X-Y repeat region containing numerous small interruptions. Two type IV collagens have been identified in *C. elegans*. These are encoded by the genes *emb-9* and *let-2* which are homologues of the human α1(IV) and α2(IV) genes respectively (Guo *et al.*, 1991; Guo and Kramer, 1989; Sibley *et al.*, 1993). Mutation in both *C. elegans emb-9* and *let-2* results in embryonic arrest in the developing nematode (Guo *et al.*, 1991; Gupta *et al.*, 1997; Sibley *et al.*, 1994; Sibley *et al.*, 1993). Type IV collagen encoding genes have also been isolated from parasitic nematode species *B. malayi* (Caulagi and Rajan, 1995) and *A. suum* (Pettitt and Kingston, 1991).

Type IV collagens in vertebrates have been demonstrated to contain 4-hydroxproline and hydroxylysine in Y positions of Gly-X-Y repeats (Kivirikko *et al.*, 1992) and must therefore be modified by the hydroxylating enzymes. Examination of the gene *let-268*, the only lysyl hydroxylase enzyme in the *C. elegans* genome, shows this is specific for the modification of the basement membrane collagens (Norman and Moerman, 2000). This is in agreement with the finding that cuticle collagens of most developmental stages of *C. elegans* do not contain hydroxylysine (Cox *et al.*, 1981b). Expression of *let-268* transgenes were detected in the body wall muscle cells and glial-like cells (Norman and Moerman, 2000), an expression pattern similar to that of the type IV collagens (Graham *et al.*, 1997). A putative null allele of this gene shows embryonic lethality at the two-fold stage of development with the type IV collagens, EMB-9 and LET-2, being intracellularly retained (Norman and Moerman, 2000).

1.14. *C. elegans* as a model organism and a heterologous expression system

C. elegans has a number of advantages as an organism for studying many biological processes including ECM formation. In addition to its ease of culture, small size and rapid life cycle, *C. elegans* is a self-fertilising hermaphrodite that can, if desired, be crossed with males, therefore making this organism extremely amenable for genetic manipulation (Brenner, 1974). *C. elegans* is also transparent making analysis by light

microscopy possible. Extensive analysis has provided a complete embryonic and post-embryonic cell lineage map (Sulston and Horvitz, 1977; Sulston *et al.*, 1983) and detailed information on development.

A large number of mutagenic screens have been performed on this organism, since the first screens nearly 30 years ago, which give a range of visible phenotypes (Brenner, 1974). A bank of mutant strains has been constructed which are freely accessible through a central resource facility. Crosses can be performed with most mutant strains allowing mapping of mutations to specific chromosomal locations and construction of a genetic map. Many loci are represented by multiple alleles, and for those for which the corresponding the gene has been cloned, this allows a more extensive analysis of gene function to be undertaken.

C. elegans was the first multicellular organism for which the entire genome sequence was completed (Consortium, 1998). Availability of the data from the 97-megabase genomic sequence and the gene predictions of the >19000 genes abolishes the time consuming process of gene cloning based on heterologous sequences. Complete genomic sequence is also required for analysis of gene families, as the full complement of potential family members can be identified on the basis of sequence homology as illustrated by the large collagen (Johnstone, 2000) and *nhr* (Sluder *et al.*, 1999) gene families. EST projects have also provided useful data for confirming and refining computer based gene predictions, giving estimates of expression levels and providing data on alternative splicing of gene transcripts. This physical map of the genome can be compared to the genetic map using the *C. elegans* database ACeDB and can be used to identify candidate genes for particular mutant loci. The ability to rapidly genetically transform this nematode via microinjection (Mello and Fire, 1995; Mello *et al.*, 1991), and more recently by ballistic transformation (Wilm *et al.*, 1999), underpins much of the research performed on this organism and allows analysis of molecular function and expression *in vivo*. As the entire genome is available in the form of cosmids and YACS, mutants, which have been mapped genetically to a region of the genome, can be cloned by phenotypic rescue using transformation of pools of cosmids covering this area of the genome. Injection of sub-population of cosmids followed by injection of individual genes can be used to clone genes responsible for mutant phenotypes.

The advent of double-stranded RNAi techniques developed from previous sense and antisense RNA based gene interference techniques (Guo and Kemphues, 1995). dsRNAi was first performed in *C. elegans* (Fire *et al.*, 1998) and is now used in many organisms. Again this can be performed rapidly by microinjection, and due to bacterial feeding based RNAi (Kamath *et al.*, 2000; Timmons *et al.*, 2001; Timmons and Fire, 1998) and RNAi soaking techniques (Tabara *et al.*, 1998) has been extended to a genome-wide RNAi project (Fraser *et al.*, 2000; Gönczy *et al.*, 2000; Maeda *et al.*, 2001). The temporary nature of the RNAi effect, and the difficulty in disrupting late acting and neuronally expressed genes has been addressed by transforming with DNA encoding sequences of target gene in an inverted repeat conformation under the control of an inducible promoter (Tavernarakis *et al.*, 2000). Information on gene sequence, structure, genetic mutants, RNAi and expression pattern data are available through AceDB. *C. elegans* closest relative *C. briggsae* is also the subject of a now completed genome sequencing project thus revealing conserved functional and control features.

Many fundamental processes of biology such as development of body plan, and programmed cell death pathways are found in *C. elegans*. Likewise many of the components of the cuticular ECM and enzymes involved in its generation in *C. elegans* are conserved both between other nematodes and vertebrates. This nematode is therefore ideal for studying ECM formation due to the genetic and molecular biological techniques available. Additionally the nematode cuticle is a visible and examinable ECM with mutations in its formation having dramatic and readily identifiable outcomes. Manipulation and examination of these outcomes using the full complement of molecular genetic and biochemical techniques provides an invaluable system in which to study this process.

In addition to its function as a model organism to study gene function, *C. elegans* can be used to study functions of heterologous genes more directly. Phenotypic rescue of *C. elegans* genes by homologous genes from other organisms has been demonstrated for human (Levitan *et al.*, 1996; Stern *et al.*, 1993), *Drosophila* (Stern *et al.*, 1993) and parasitic nematode genes (Britton and Murray, 2002; Kwa *et al.*, 1995). The correlation of rescue of a *C. elegans* mutant by an *H. contortus* gene, based on drug resistance, was in one study shown to corresponded directly to whether the allele injected came from a drug sensitive or resistant strain (Kwa *et al.*, 1995). This enabled identification of amino

acids critical for sensitivity. While this process relies on an existing mutant, a recent study described a new RNAi approach to study heterologous rescue. Wild type *C. elegans* was transformed with a gene encoding a cathepsin L cysteine protease gene from *H. contortus* that produced no deleterious effects. Transformed *C. elegans* carrying the transgene were then subjected to RNAi of their endogenous homologue of the cysteine protease. In non-transformed worms RNAi is embryonic lethal but in lines expressing the *H. contortus* protein selective removal of the *C. elegans* protein is not lethal, demonstrating that the heterologous gene is capable of functionally compensating for loss of the endogenous activity (Britton and Murray, 2002). This technique is rapid, does not rely on a stable mutant nematode strain, and can, in theory, be performed for any gene for which a homologue is identifiable in *C. elegans*. Conservation of control elements has also been demonstrated for a number of genes between parasitic nematode species and *C. elegans* (Britton *et al.*, 1999; Qin *et al.*, 1998) thus identifying conserved elements directing tissue specific control and also indicating patterns of expression for genes for which function may not be assigned. Upstream sequences from parasitic nematodes were fused to reporter genes and transformed in *C. elegans* and expression of the transgene assayed. A spatial expression pattern in defined cells and tissues, and any temporal specificity of expression observed, may assist in defining gene function.

1.15. Inhibition of collagen biosynthetic enzymes

Due to P4Hs central role in the biosynthesis of collagens it is an attractive target for the inhibition of collagen formation. Inhibition of this process is desirable in two respects. Firstly, loss of P4H function could in parasitic nematodes interfere with formation of the cuticle, and as such be used as a basis for drug control of these species. Inhibitors of P4H almost all bind to the 2-oxoglutarate binding site. Competitive analogue inhibitors of 2-oxoglutarate have anti-nematode effects, and in *C. elegans* cause embryonic and larval lethality and moulting defects (Myllyharju *et al.*, 2002). In *B. malayi* inhibition of P4H results in cuticle-associated defects with a corresponding reduction in 4-hydroxproline residues (Merriweather *et al.*, 2001). Secondly fibrotic diseases of humans involve the excessive accumulation of collagen fibres which often occurs during wound healing. While it is an essential process, the excessive build up of collagen can result in the replacement of cellular tissue with fibrous tissue with accompanying impairment of tissue function.

1.16. Project aims

The aim of this study was the definition and characterisation of P4H subunit-encoding genes involved in the formation of nematode ECMs, particularly that of the cuticle, from free-living and parasitic species. Firstly, the expression patterns, RNAi/mutant phenotypes and subunit associations were to be defined for P4Hs from *C. elegans*. Thereafter homologues of interesting genes were to be identified from the filarial parasitic nematode *B. malayi*. Modification by P4H had been demonstrated *in vitro* to be essential for the thermal stability of vertebrate collagens (Berg and Prockop, 1973) and, as such, P4H activity was hypothesised to be central to nematode development and formation of the collagen-rich cuticle. Understanding the enzymes involved in formation of the cuticle, which is essential for many nematode functions, could lead to identification of specific enzymes as targets for development of anti-nematode compounds and, in addition, provide greater insight into ECM formation as it may occur in all species. Potential P4H subunits identified from the completed *C. elegans* genome sequence, and two P4H subunit-encoding genes identified previously were examined, as well as a P4H cloned and characterised from the filarial parasite *B. malayi*.

Chapter 2

Materials and methods

2.1. Standard reagents and media

Ampicillin:	100 mg/ml ampicillin (Sigma) in sterile distilled H_20 Filter sterilised and stored at -20°C
BSA:	bovine serum albumin 10 mg/ml (NEB). Stored at –20°C.
Chloramphenicol:	12.5 mg/ml chloramphenicol (Sigma) in 100% ethanol. Stored at –20°C.
DEPC H_2O:	0.1% (v/v) diethylpyrocarbonate (Sigma) in sterile distilled H_2O mixed overnight and autoclaved. Stored at room temperature.
DTT:	dithiothreitol (Sigma) in sterile distilled H_2O. 1M stock stored at -20°C.
EDTA:	ethylenediaminetetra-acetic acid in sterile distilled H_2O. Stock solution of 0.5 M, pH 8.0. Autoclaved and stored at room temperature.
Ethidium bromide:	8 mg/ml in sterile distilled H_2O. Stored at 4°C.
IPTG:	isopropyl-β-D-thiogalactoside (Promega) in sterile distilled H_2O. Stock concentration of 1 M filter sterilised and stored at –20°C
L-broth:	1% bacto tryptone (Difco), 0.5% yeast extract (Difco), 0.5% NaCl in sterile distilled H_2O, pH 7.0 NaOH. Autoclaved and stored at room temperature.
LB-agar:	L-broth + 15 g/L bacto-agar (Difco). Autoclaved and stored at room temperature.
M9 buffer	3% KH_2PO_4, 6% Na_2HPO_4, 5% NaCl, 10 mM $MgSO_4$. 10X stock autoclaved and stored at room temperature.
2-mercaptoethanol:	14.3 M 2-hydroxethylmercaptan (Sigma).
NGM-agar:	0.3% NaCl, 1.7% agar (Difco), 0.25% peptone (Difco), 0.0003% cholesterol (1 ml/L of 5 mg/ml stock in ethanol), in sterile distilled H_2O. Autoclaved and 1 ml/L 1 M

	$CaCl_2$, 1 ml/L 1M $MgSO_4$ and 25 ml/L KPO_4 (pH 6.0) added.
PBS:	phosphate buffered saline: 7.31 g NaCl, 2.36 g Na_2HPO_4, 1.31 g NaH_2PO_4 2H20 in 1 L, pH 7.2. Autoclaved and stored at room temperature.
Proteinase K:	20 mg/ml proteinase K (Roche) in sterile distilled H_2O. Stored at -20°C.
SDS:	sodium dodecyl sulphate (Sigma) in H_2O. Stock solution 10% stored at room temperature.
TBE:	10X stock, 0.9 M Tris-HCl, 0.9 M Boric acid (Sigma), 25 mM Na_2EDTA pH 8.0, in sterile distilled H_20. Stored at room temperature.
TE buffer:	10 mM Tris, 1mM EDTA, pH 8.0).
Tetracycline:	12.5 mg/ml tetracycline hydrochloride (Sigma) dissolved in 50% ethanol. Stored at -4°C
Tris-HCl:	2-amino-2-(hydroxymethyl)-1,3-propanediol-hydrochloride.
Tween-20:	Polyoxyethylenesorbitan monolaurate (Sigma).
X-gal:	5-bromo-4-chloro-3-indoyl-β-D-galactoside (Promega) dissolved in N,N'-dimethyl-formamide and stored at –20°C out of light. Stock concentration of 2% (w/v) X-gal.

Preparation of any standard reagents or media not detailed can be found in Sambrook *et al.*, Molecular cloning: a laboratory manual, 1989.

2.2. Agarose gel electrophoresis

Nucleic acids were separated on 0.7-1.5% (w/v) agarose gels. Agarose (Gibco BRL Life Technologies) was dissolved in 1X TBE by heating until in solution, ethidium bromide was added to 0.08 μg/ml and gels cast. Gels were run in 1X TBE using electrophoresis equipment from Gibco BRL and 1 kb ladder (Gibco BRL). Nucleic acids were visualised and images captured using a UV transilluminator and imaging systems (Biorad).

2.3. Nematode strains and culture

All nematode strains were received from the *C. elegans* Genetics Centre (CGC) (University of Minnesota, Minnesota, USA) or the relevant laboratories. Nematodes were cultured as described previously (Brenner, 1974) on NGM-agar inoculated with *E. coli* strain OP50 as a food source. Freezing and cleaning of *C. elegans* strains was performed as described in The Nematode *Caenorhabditis elegans* (Cold Spring Harbour Laboratory Press, Cold Spring Harbour, N.Y.).

C. elegans strains

N2:	wild type Bristol strain (Brenner, 1974)
CB364:	*dpy-18*(*e364*) III (Brenner, 1974)
CB2590:	*tra-1*(*e1099*)/*dpy-18*(*e1096*) III (Hodgkin and Brenner, 1977); heterozygotes are wild type and segregate wild type, fertile wild-type males (*tra-1* homozygotes) and Dpy (*dpy-18* homozygotes). Dpys were selected, these worms have the genotype, *dpy-18*(*e1096*) III.
EM76:	*dpy-18*(*bx26*) III; *him-5*(*e1490*) V (Baird and Emmons, 1990)
JK2757:	*phy-2*(*ok177*) IV (Friedman *et al.*, 2000)
TP7:	*phy-3*(*ok199*) V (Riihimaa *et al.*, 2002)
**phy-4* Δ:	*phy-4*(*tm360*) X
GR1029:	*let-44*(*mg21*) *lon-2*(*e678*) X; mnDp31 (X;f)
DR96:	*unc-76*(*e911*) V (Brenner, 1974)
TP12:	*kaIs*12[*col-19*:GFP]
TP13:	*dpy-18*(*e364*) III; *kaIs*12[*col-19*:GFP]

* non-backcrossed strain produced by the *C. elegans* knockout project with no designated strain name, received from S. Mitani, Tokyo Women's College, Tokyo, Japan.

C. briggsae strain

AF16 (pka G16):	wild type strain (Fodor *et al.*, 1983)

2.4. *E. coli* strains and culture

Liquid *E. coli* cultures were grown overnight in L-broth at 37°C in an orbital shaker. LB-agar was used for growth of bacteria on plates with overnight incubation at 37°C. Antibiotic selection was used where appropriate, generally 100 µg/ml ampicillin, unless otherwise stated. Permanent bacterial stocks were made by addition of 0.8 mls of overnight cultures to 1 ml of 2% peptone/ 40% glycerol and storage at -80°C. Full genotypes of the *E. coli* strains used can be found from supplier or in reference given.

OP50:	strain used was a variant of the uracil requiring OP50 strain (Brenner, 1974) with a tetracycline selectable marker. Strain received from CGC.
XL-10 Gold:	ultracompetent cells, suitable for blue-white colour screening of recombinant plasmids. Multiple methylated DNA restriction systems removed. *recA* recombination and *endA1* endonuclease deficient giving greater stability to transformed DNA and better quality of purified plasmid. Purchased from Stratagene.
HT115(DE3):	tetracycline resistant, IPTG inducible T7 polymerase, RNase III minus (Timmons *et al.*, 2001). Strain received from CGC.
LE392:	used for infection and propagation of bacteriophages, no antibiotic resistance used for growth of bacterial culture. Purchased from New England Biolabs (NEB).
GM2163:	chloramphenicol resistant, *dam* and *dcm* minus strain, endogenous adenine methylation at GATC and cytosine methylation at CC(A/T)GG abolished. Purchased from NEB.

2.5. Purification and synthesis of DNA and RNA

2.5.1 Phenol/chloroform extraction and ethanol precipitation of DNA

An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma) was added to a solution of DNA, mixture vortexed for 1 minute then spun at 12k in a bench top centrifuge for 2 minutes. Top layer was removed and an equal volume of

chloroform:isoamyl alcohol (24:1) (Sigma) added, vortexed, spun and separated as above. Aqueous layer was precipitated with 1/25 volumes of 5 M NaCl and 2 volumes of 100% ethanol. Precipitate was stored for at least 1 hour at –20°C then pelleted in a bench top centrifuge at full speed for 15 minutes. Pellet was washed in 75% ethanol, air-dried and the DNA suspended in TE buffer.

2.5.2 Genomic DNA isolation

C. elegans strains for genomic DNA extraction were grown on NGM OP50 plates and washed off plates with H_20. Worms were concentrated by centrifugation and stored at –80°C. 200-500 µl of concentrated worm pellet was added to 6 volumes of 1 X worm lysis buffer. For *B. malayi* genomic extraction, 10 adults were added to 6 volumes of 1X worm lysis buffer. Worm suspensions were disrupted in a glass hand held homogeniser followed by incubation at 65°C for 4 hours. Debris was removed by centrifugation and DNA purified by repeated phenol:chloroform and chloroform extractions. DNA was treated with RNAse A (DNase free) (Sigma) at a final concentration of 100 µg/ml for 1 hour at 37°C then phenol/chloroform extracted, chloroform extracted and ethanol precipitated. 100-500 ng of genomic DNA was used in 50 µl PCR volumes.

5X Worm lysis buffer:	0.25 M Tris-HCl (pH 8.0), 0.50 M NaCl, 0.25 M EDTA. pH to 8.0, solution autoclaved and stored at 4°C
1X Worm lysis buffer:	50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 50 mM, EDTA 1% SDS, 30 mM 2-mercaptoethanol, 100 µg/ml Proteinase K (Roche). 1X lysis buffer was made by from 5X stock with freshly added SDS, 2-mercaptoethanol and proteinase K.

2.5.3. Total RNA isolation

4 X volume of Trizol Reagent (Gibco BRL Life Technologies) was added to 100-1000 µl of concentrated *C. elegans* pellet. For *B. malayi*, 800 µl of Trizol was added to three adult females. Suspensions were vortexed to solubilise and disrupt worms and left at room temperature for 20 minutes with occasional mixing. Insoluble material was removed by centrifuging 1 ml aliquots at 4°C at top speed. To the supernatants 160 µl of chloroform isoamyl alcohol was added. These were vortexed for 15 seconds,

incubated at room temperature for 2 minutes, then centrifuged for 15 minutes at 4°C. 400 µl isopropanol was added, solutions mixed and precipitated at room temperature for 10 minutes. RNA was pelleted by centrifugation for 10 minutes at 4°C. RNA was washed with 75% ethanol (25% DEPC treated H_2O), spun at 7,500 k for 10 minutes and resuspended in DEPC treated H_2O.

2.5.4. Purification of mRNA

mRNA purification was accomplished using a Poly (A) Quik mRNA isolation kit (Stratagene). Total RNA, prepared as described above, was heated to 65°C then rapidly cooled on ice to remove any secondary structure, then sample buffer added to 1X concentration. An oligo (dT) cellulose column was equilibrated with high-salt buffer and the sample applied to the column with a syringe, collected, then reapplied. Column was washed twice with high-salt buffer then three times with low-salt buffer. mRNA bound to the oligo (dT) cellulose was eluted four times in elution buffer preheated to 65°C. RNA was then precipitated with 1/10 volume sample buffer and 2.5 volumes 100% ethanol, washed in 75% ethanol (25% DEPC H_2O) and resuspended in DEPC H_2O.

10X Sample buffer:	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 5 M NaCl.
High-salt buffer:	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.5 M NaCl.
Low-salt buffer:	10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.1 M NaCl.
Elution buffer:	10 mM Tris-HCl (pH 7.5), 1 mM EDTA.

2.5.5. Synthesis of first strand cDNA

First strand cDNA was synthesised using a RT-PCR kit (Stratagene). 100 ng of mRNA, prepared as described above, was annealed to 7.5 ng/µl of oligo d(T) primer by heating to 65°C for 5 minutes then slowly cooling to room temperature. First strand buffer was added to 1X concentration, dNTP's to 4 mM and 50 units of Moloney murine leukaemia virus reverse transcriptase (MMLV-RT) added. Reaction was incubated at 37°C for 1 hour, heat inactivated and 2 µl used per 50 µl PCR reaction.

2.6. PCR based cloning procedure

2.6.1. PCR conditions and polymerases

All PCR reactions were performed using a Robocycler Gradient 96 PCR machine (Stratagene) in a 50 µl volume unless otherwise stated. Routine PCR conditions used were; denaturation at 92°C for 1 minute, primer annealing at 56°C for 2 minutes, and extension at 72°C for 1-2 minutes per kb of target sequence. Final concentrations of 0.4-0.6 µM of each primer and 250 µM of each dNTP were used. Oligonucleotide primers were purchased from Sigma Genosys or Gibco BRL, the sequences of all primes used are given in Table 2.2. *Taq* polymerase (Applied Biosystems) was used in 1X buffer IV (Applied Biosystems), supplemented with 2.5 mM final concentration $MgCl_2$, using 5 units of enzyme per reaction. PCR with *Pfu* and *Pfu* Turbo DNA polymerases (Stratagene) was performed in 1X Cloned *Pfu* DNA polymerase buffer (Stratagene) using 2.5 to 5 units of enzyme per reaction. 2-4 units of Vent DNA polymerase (NEB) was used per reaction in 1X Thermopol buffer (NEB). Proof-reading polymerases *Pfu*, *Pfu* Turbo and Vent contain 3' to 5'-exonuclease activity that enable them to proof-read for nucleotide mis-incorporations and were used where appropriate to minimise errors in the amplified sequences.

10X Buffer IV:	200 mM $(NH_4)_2SO_4$, 750 mM Tris-HCl pH 9.0 (at 25°C), 0.1% (w/v) Tween-20.
10X Cloned *Pfu* buffer:	200 mM Tris-HCl (pH 8.8), 20 mM $MgSO_4$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 1% Triton X-100, 1 mg/ml nuclease-free BSA.
10X Thermopol buffer:	100 mM KCl, 200 mM Tris-HCl (pH 8.8 at 25°C), 100 mM $(NH4)_2SO_4$, 20 mM $MgSO_4$, 1% Triton X-100.

2.6.2. Purification of PCR products

QIAquick Gel extraction and QIAquick PCR purification kits from QIAGEN were used to purify DNA from agarose gels or directly from PCR reactions, respectively. Both kits use a column with a silica gel membrane that binds DNA in the presence of high salt concentrations and pH ≤7.5. Gel extraction of DNA was accomplished by separating DNA on a 1% agarose gel, excising the desired band(s) and solubilising at 50°C in 3 volumes (w/v) of QX1 buffer. 1 volume of isopropanol was also added if DNA

fragment to be purified was <0.5 kb or >4 kb. DNA was then bound to the column by spinning the solution through the column. Column was washed with QX1 buffer followed by PE buffer to remove agarose and salt impurities. DNA was eluted with H_2O (preheated to 50°C) by spinning 4 X 50 µl volumes through the column. Ethanol precipitation was performed as described in Section 2.5.1. For QIAquick PCR purification 5 volumes of PB buffer was added to the PCR reaction, mixture spun through the column and washed with PE buffer. Elution and precipitation of DNA carried out as described with the addition of 10 µg of glycogen as a carrier where required.

2.6.3. PCR product cloning using pCRScript

The pCRScript cloning kit (Stratagene) was used for cloning blunt ended PCR fragments. These fragments were produced either by PCR with a proofreading DNA polymerase or were made blunt ended. The pCRScript vector is supplied pre-linearised with *Srf* I, a rare cutting restriction enzyme that produces blunt end termini. Inclusion of *Srf* I in the ligation reaction reduces re-ligation of the vector, as, if this occurs, the enzyme recognition site is reformed and the plasmid cleaved. Ligation of an insert destroys the enzyme recognition site and so recombinant plasmids are positively selected for. PCR product purified by one of the above methods was either ligated directly with pCRScript vector or made blunt ended in 1X polishing buffer (Stratagene) with 0.25 mM each dNTP and 0.5 units of *Pfu* DNA polymerase. Reaction was incubated for 30 minutes at 72°C then 1/3 of reaction used for ligation. A 10 µl ligation was performed by adding approximately 40X molar excess of PCR product to 10 ng of linearised pCRScript vector, in 1X pCR Script reaction buffer with 1 mM rATP, 5 units *Srf* I restriction enzyme and 4 units T4 DNA ligase. Reaction was heat inactivated after 1 hour at room temperature.

2.6.4. T-overhang PCR cloning vectors

Vectors such as pGEM-T (Promega) and pTAg (Invitrogen) are supplied linearised with 3' terminal thymidine at both ends. These serve to prevent re-cicularisation of the vector and provide compatible bases for ligation with PCR product produced with *Taq* DNA polymerase. Amplification with *Taq*, and some other non-proof reading polymerases, causes the addition of single deoxyadenosine to the 3' ends of amplified DNA. *Taq*

amplified product was purified using one of the above methods then ligated in a 10 µl reaction with 50 ng of vector in 1X T4 DNA ligase buffer (NEB) with 1 µl T4 DNA ligase (NEB). A molar ratio of approximately 3:1 insert:vector was used in the ligation and reaction incubated overnight at 4°C. Blunt ended PCR product to be cloned in to a T-overhang vector had 3' A-tails added using *Taq*. Purified PCR product was added to 1 X Buffer IV (AB), 0.2 mM final concentration dATP, and 5 units *of Taq* (AB) in a 10 µl reaction. After incubation at 72°C for 30 minutes the reaction was heat inactivated. A proportion of this reaction was then used in a ligation with a T-overhang vector.

1 X T4 DNA ligase: 50 mM Tris HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP, 25 µg/ml BSA.

2.6.5. TOPO T-A cloning

The pCR 2.1 TOPO kit (Invitrogen) was also used for cloning A-tailed PCR fragments. Purified PCR products were added with 1 µl TOPO vector and 1 µl Salt solution (1.2 M NaCl, 0.06 M $MgCl_2$) in a 6 µl reaction at room temperature for 5 minutes. The vector is supplied linearised with 3' overhanging-T residues and covalently bound Topoisomerase I at vector ends. The usual function of this enzyme is nicking supercoiled DNA then re-joining ends. This activity is used in TOPO cloning to enable ligase independent joining of fragments. After T-A base recognition between the vector and insert DNA, the vector bound Topoisomerase cleaves the phospodiester backbone of one strand at its recognition site. Formation of a covalent bond between the 3' phosphate from the cleaved strand of the backbone and a residue from topoisomerase enables joining of DNA the ends and release of enzyme when the 5'-hydroxyl from the original cleaved strand attacks this bond.

2.6.6. Transformation of *E. coli*

Transformation of *E. coli* was performed using 40 µl of XL10-Gold ultracompetent cells (Stratagene) with 2 µl of a ligation reaction from a cloning kit, followed by incubation on ice for 30 minutes, heat shock at 42°C for 30 seconds then ice for 2 minutes. 400 µl of SOC media was added and transformation incubated at 37°C in an orbital shaker for 1 hour. Cell suspension was spread on LB + antibiotic plates, dried and grown overnight at 37°C. For blue-white colour selection of transformants 100 µl

0.1 M IPTG and 100 µl 2% X-gal was added to the surface of a 9 cm plate prior to plating of cells.

2.6.7. Identification of bacterial transformants

2.6.7.1. Blue/white colour selection

When vector permitted, blue/white screening of colonies was performed. Vectors that can be used for colour screening contain the regulatory sequences and N-terminal sequences for *E. coli* protein β-galactosidase, coded for by *lacZ* gene. Embedded within the *lacZ* coding sequence is a multiple cloning site that does not affect the reading frame of the peptide but results in the addition of a few amino acids. Plasmids encoding the N-terminal of β-galactosidase are used in conjunction with host *E. coli* cells that code for the C-terminal portion of protein. Neither peptide alone is active but when both are present they can associate and form an active protein. Cells expressing both peptides turn blue when grown on the artificial substrate analogue X-gal. Insertion of a fragment into the cloning site of the vector inactivates the N-terminal peptide resulting in no β-galactosidase activity and white coloured colonies. White colonies were screened by PCR.

2.6.7.2. PCR screening

Colonies selected for screening (either white when blue/white colour selection or random when no colour selection was applied) were picked and touched on to a LB+antibiotic master plate. The bacteria remaining was added to 100 µl H_2O and boiled for 5 minutes and cell debris pelleted by centrifugation. 2 µl of colony boils were then added to a 20 µl PCR. PCRs were performed in 1X PCR buffer (45 mM Tris-HCl pH 8.8, 11 mM Ammonium sulphate, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 µM EDTA pH 8.0, 1 mM each dNTP, 113 µg/ml BSA) with 100 ng of each primer and *Taq* polymerase. Primers used were complementary to the vector backbone flanking the cloning site. Colonies positive for presence of the insert were then identified on the master plate.

Table 2.1. Primer combinations for screening bacterial colonies transformed with standard vectors

Vector	Primer combinations (sequences in Table 2.2)
pCRScript/pBluescript SKM (Stratagene)	1224 and M13 reverse
pGEM (Promega)	pGEM forward and pGEM reverse
pTAg (Invitrogen)	M13 Reverse and 1224.
PCR 2.1TOPO (Invitrogen)	M13 reverse and pGEM forward
pPD95.03*	M13 reverse and NLS
pPD96.04*	M13 reverse and 96.04 reverse
pPD129.36* (Timmons *et al.*, 2001)	L4440 F and L4440 R
pVL1392 (Pharmingen)	BV F and BV R

*pPD series of vectors were a gift from A. Fire (Carnegie Institution of Washington, Baltimore).

2.6.8. Plasmid DNA preparation

10 ml (small scale or mini plasmid preparation) or 50 ml (large scale or midi plasmid preparation) overnight bacterial cultures of LB-antibiotic were grown at 37°C in an orbital shaker. QIAGEN Mini or Midi plasmid kits or QIAprep spin columns were used to prepare DNA using solutions and procedure supplied by the manufacturer. The procedure is based on an alkaline lysis process that utilises the selective renaturation of plasmid DNA following denaturation. Cells were lysed with NaOH/SDS in the presence of RNase A. NaOH denatured the cellular proteins, chromosomal and plasmid DNA, the lysate was then neutralised with acidic potassium acetate. The high salt concentration causes SDS precipitation of denatured proteins, membrane bound chromosomal DNA and cell debris. Plasmid DNA renatures under these conditions and remains in solution. Columns containing an anion-exchange resin whose positively charged DEAE groups bind the phosphates of the DNA backbone but not degraded RNA, cellular proteins or metabolites, were used to purify plasmids. The salt conditions and pH of buffers determined whether DNA was bound or eluted from the column. Column-bound

plasmid DNA was washed, eluted and isopropanol precipitated. DNA was examined by electrophoresis.

2.6.9. Analytical restriction digests

100-200 ng of plasmids were digested with 10 units of restriction enzymes in a 20 µl digest in 1X restriction digest buffers (NEB) for 2 hours at 37°C or other recommended temperature. For single digests the buffer used was as recommended and supplied by NEB. Double digests reactions were performed in a buffer compatible for both enzymes as recommended by NEB. 5-10 µl of digests were analysed by electrophoresis.

2.7. Subcloning from plasmids and bacteriophage

2.7.1. Restriction digests for subcloning

Restriction digests were performed using 5-10 µg of insert containing plasmid or bacteriophage and target plasmid vector digested with 20 units of restriction enzyme in a 40 µl volume, at the appropriate temperature. For directional cloning digests were then phenol extracted, precipitated with ethanol (see Section 2.5.1), resuspended in H_2O and digested with a second enzyme in the same manner. For non-directional cloning the recipient plasmid vector was treated with calf intestinal alkaline phosphatase (CIAP) (NEB) for 1 hour at 37°C, directly after treatment with the first restriction enzyme. Vector and insert fragments to be ligated were purified from agarose gels and ethanol precipitated as described in Section 2.6.2 and analysed for purity and quantity by agarose gel electrophoresis.

2.7.2. Ligations

Ligation of insert and vector DNA was performed in a 20 µl reaction with 100 ng vector DNA and a 3:1 molar ratio of insert to vector in 1X T4 DNA ligase buffer. Ligations were left at 16°C overnight. 5-10 ng of DNA was then used to transform XL10-Gold ultracompetent cells. All subsequent manipulations including bacterial transformation, screening of bacterial colonies, plasmid preparations and restriction digests were performed as described earlier in Sections 2.6.6 to 2.6.9.

2.8 Sequencing analysis

DNA for sequence analysis was prepared in either precycled or uncycled form. Pre-cycled reactions were prepared using the ABI PRISM Dye Terminator Sequencing Ready Reaction kit. 8 µl Terminator Ready Reaction mix was added to 200 ng plasmid DNA with 3.2 pmoles of primer in a 20 µl reaction. Oligonucleotide primers were purchased from Sigma Genosys or Gibco BRL with sequences given in Table 2.2. Cycling was performed in a GeneAmp PCR System 2400 (Perkin-Elmer) with parameters; 96°C - 10 seconds, 50°C - 5 seconds, 60°C - 4 minutes for 25 cycles. Extension products were purified by ethanol precipitation with 1/10 volumes 3 M NaAcetate, pH 4.6, and 2.5 volumes 95% ethanol. Reactions were precipitated on ice for 10 minutes, centrifuged for 30 minutes, pellet washed in 250 µl 70% ethanol, then air-dried. These were taken to the Molecular Biology Support Unit (MBSU) Sequencing Service at Glasgow University for analysis. Uncycled DNA was supplied as plasmid along with primer(s), which were then processed by the MBSU Sequencing Service Glasgow University.

2.8.1. Sequencing analysis software and computer analysis of DNA and protein sequences

DNA sequence data was analysed, assembled and compared using; the ABI Prism Sequencing analysis software (PE Applied Biosystems), Vector NTi (Informax) and its associated programs Contig Express and Align X, and the Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, 1990) from the National Centre for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/). Searches for *B. malayi* EST and GSS sequences were performed using BLAST at the EMBL-EBI Parasite-genome server (http://www.ebi.ac.uk/parasites/parasite-genome.html). Data on sequenced *C. elegans* genes was accessed using WormBase (http://www.wormbase.org/) and Intronerator (Kent and Zahler, 2000) (http://www.soe.ucsc.edu/~kent/intronerator/) sites. The prediction of ER signal peptide cleavage sites of proteins was performed using the Signal P program (Nielsen *et al.*, 1997) through the ExPASy molecular biology server (http://www.expasy.ch/). Glycosylation predictions were performed using NetNGlyc 1.0 Prediction Server (Gupta *et al.*, 2002) from the Centre for Biological Sequence Analysis (http://www.cbs.dtu.dk/services/NetNGlyc/). Amino acids sequences were aligned for

comparison using ClutalW (http://www.ebi.ac.uk/clustalw/) and processed using Boxshade (http://www.ch.embnet.org:80/software/BOX_form.html).

2.9. Protein techniques

2.9.1. Peptide antisera

Peptides for immunisation were made by N-terminal fusion of synthetic peptides for proteins *Ce*-PHY-1, *Ce*-PHY-2, *Ce*-PDI-2 and *Bm*-PHY-1 to keyhole limpet heamocyanin. (Sigma Genosys) via a cysteine residue. Polyclonal antisera were raised in two rabbits for each peptide (Sigma Genosys). C-terminal residues used for immunisations are shown below, C represents added cysteine residues, which are not present in the native protein sequences.

Ce-PHY-1:	CEPRNAPNVSPNLAKDVWETL
Ce-PHY-2:	CLEEEVQENFIGDLSPYANDP
Ce-PDI-2:	CGASEEEKAEEEADEEGHTEL
Bm-PHY-1:	CRRPCGLSRSVEEQFVGDLSA

2.9.2. SDS PAGE

Samples were boiled in 5% final v/v 1M DTT and 1X NuPAGE SDS sample buffer (Invitrogen). Extracts were examined on a NuPAGE 4-12% Bis-Tris polyacrylamine gel (Invitrogen) in 1X NuPAGE MOPS SDS buffer using NuPAGE antioxidant (Invitrogen) and run at room temperature for 45-60 minutes.

4X SDS buffer:	1.17 M sucrose, 563 mM Tris base, 423 mM Tris HCl, 278 mM SDS, 2.05 mM EDTA, 0.88 mM Sera Blue G250, 0.7 mM Phenol red.
20X MOPS SDS buffer:	1 M MOPS, 1 M Tris base, 69.3 mM SDS, 20.5 mM EDTA.

2.9.3. Western blotting

Proteins were transferred using NuPAGE (Invitrogen) Western blotting apparatus and reagents. Blots were run for 1 hour at room temperature in 1X Transfer buffer. Hybond-P PVDF membrane (Amersham Pharmacia Biotech) was used after activation by

soaking in methanol. The membrane was blocked by soaking overnight in a solution of 5% (w/v) Membrane blocking agent (Amersham Life Sciences) made in PBST (PBS with 0.1 v/v Tween-20). Primary antibody dilutions were incubated at room temperature for 2 hours and membranes washed three times in PBST for 5 minutes at room temperature. Secondary antibody, a 1/15,000 dilution of goat anti-rabbit (or anti-mouse) IgG whole molecule alkaline phosphate conjugate (Sigma) was added for 2 hours at room temperature followed by washes. Antibodies were detected using Sigma Fast Tablets diluted in H_2O.

Transfer buffer 20X: 500mM bicine, 500mM Bis-Tris, 20.5 mM EDTA
10-20% methanol added to 1X transfer buffer.

2.10. Transformation of *C. elegans*

2.10.1. Microinjection procedure

C. elegans was transformed by microinjection (Mello *et al.*, 1991), detailed methods for which can be found in C. Mello and A. Fire, 1995 (Mello and Fire, 1995). DNA for microinjection was prepared using the standard QIAGEN method. Plasmids were co-injected with a selectable marker to identify transformed progeny. Mixes for injection were made up to a final concentration of 120-200 µg/ml plasmid DNA in sterile distilled H_2O and centrifuged before use to pellet debris. Pads for mounting worms for injection were made from a small volume of molten 2% agarose solution added to a large coverslip. This was covered with another coverslip to spread solution into thin layer. The top coverslip was removed and the agarose pad air-dried. Needles were pulled on a computer controlled electrode puller model 773 (Campden Instruments) using 1.2 mm O.D X 0.69 mm I.D. borosillicate glass capillaries with standard wall and inner filament (Clark Electromedical Instruments). Nematodes were added to agarose pads under liquid paraffin (BDH) and injected using an Axiovert-100 inverted microscope (Zeiss) equipped with a flat, free-sliding glide stage with centred rotation with DIC/Normaski optics. DNA was injected in to the hermaphrodite gonad using a micromanipulator guided needle and pressurised nitrogen. Recovery buffer (2% glucose, 1X M9) was added to injected worms on pads, after which they were transferred to fresh NGM OP50 plates in a pool of recovery buffer. Transformed progeny from injected hermaphrodites (F1 generation) were identified by the phenotype

conferred by the co-injected marker plasmid. These are selected either clonally or in small groups of 4-10, and the next generation (F2) surveyed for transformants.

2.10.2. Behaviour of injected DNA

Oocyte nuclei in the adult gonad share a common core of cytoplasm. As they mature, individual nuclei are incorporated into plasma membranes along with a portion of the core cytoplasm. DNA injected in to this syncitium can also become incorporated into the oocyte. In a fraction of the F1 progeny (1-10%), recombination reactions occur between injected sequences that result in the formation of long tandem arrays. Recombination between co-injected plasmids occurs due to regions of homology within their vector backbones. Arrays that attain a large size can become heritable as extrachromosomal elements. Once assembled these elements are no longer targets for further intra-array recombination. The arrays are then transmitted at a frequency of between 5-95% and lines can be maintained indefinitely by selection of the phenotype conferred by the marker DNA. Both marker and construct DNA should be contained within these arrays thus nematodes expressing the marker phenotype should contain the construct. It is possible for arrays to form that not contain any copies of the construct, especially if this is injected at a low concentration. To ensure construct DNA was examined and that a result found in one line is not due to the particular rearrangement of DNA within a particular array, multiple lines (generally more than three) were normally examined for each set of injections. Only one F2 line was selected from any positive F1 plate of transformants, thus ensuring different lines were the product of separate rearrangement events.

2.10.3. Selectable markers of transformation

Plasmid pRF-4 (Mello *et al.*, 1991) (a gift from A. Fire, Carnegie Institution of Washington, Baltimore) was used as a marker for transformation by co-injecting it at 100 µg/ml. Plasmid pRF-4 contains a semi-dominant allele of the collagen gene *rol-6(su1006)* which is expressed from L3 onwards and in dauers. Transformants were identified by the right-hand roller phenotype conferred by this gene. Plasmid p76-16B (a gift from R. Horvitz, MIT, Massachusetts, USA) contains wild-type *unc-76*. UNC-76 protein is found in all neuronal cells throughout development (Bloom and Horvitz, 1997). The strain DR96 contains the mutant allele *unc-76* (*e911*) and is uncoordinated,

injection of DR96 with p76-16B rescues this phenotype (Bloom and Horvitz, 1997). *unc-76* rescue was used as a marker of transformation by co-injecting p76-16B, along with a construct of interest, into the strain DR96 and identifying transformants by restoration of wild type movement. Expression of GFP (Chalfie *et al.*, 1994; Heim *et al.*, 1995), from the construct dpy-7-GFP (a gift from I. Johnstone, WCMP, Glasgow), which contains the *dpy-7* cuticle collagen promoter fused to GFP, was also used as a selectable marker. This construct gives non-nuclear localised GFP expression in the hypodermal cells of transformed worms, with strong GFP expression detectable from embryo to L4, in dauers, and less strongly in adults. dpy-7-GFP was co-injected at a concentration of 5-10 μg/ml and transformants were selected by detection of GFP expression under UV light using a Stemi SV-6 dissecting microscope (Zeiss).

2.11. Microscopy of live nematodes

Nematodes were washed in 1X M9 in a watch glass and then transferred to microscope slides (with 2% agarose/0.065% sodium azide pads). A small volume of 1X M9 was kept on the pad to prevent drying of samples and the coverslip added and sealed with white soft paraffin BP. Nematodes were viewed as live specimens using an Axioskop 2 microscope with a Hamamatsu digital camera. Images were processed and pseudocoloured (where appropriate) using Improvision Openlab software.

2.12. SL-PCR

To determine SL *trans*-splicing for *C. elegans phy-1*, *phy-2* and *pdi-2*, RT-PCR was performed with control primers phy-1 F (pMal) and phy-1 R (pMal), phy-2 F and phy-2 R (pMal), and pdi-2 F and pdi-2 R (pMal) on *C. elegans* N2 mixed stage cDNA with *Taq* for 30 cycles. The reverse primer from each primer pair was also used individually for RT-PCR with SL1 (consensus nematode splice leader sequence 1, sense primer) and SL2 (consensus nematode splice leader sequence 2, sense primer). Products from these reactions determined if primary transcripts were *trans*-spliced by either the SL1 or SL2 *trans*-spliced leader RNA sequences. SL RT-PCR was also performed for *Ce-phy-4* and *Ce-phy-5* using both SL1 and SL2 primers in combination with; phy-4 R1, R2 and R3, for *Ce-phy-4*; and phy-5 cDNA R1, R2 and R3, for *Ce-phy-5*. For *Ce-phy-4* 40 cycles of PCR were used and for *Ce-phy-5* 30 cycles were performed.

2.13. Promoter analyses

2.13.1 Promoter/reporter gene constructs

The *Ce-phy-1* promoter was PCR amplified from N2 genomic DNA for 27 cycles with primers phy-1 PF (*Pst* I)/phy-1 PR (*Bam* HI) and *Taq* polymerase to generate a fragment representing sequences from –2755 to +5 relative to the translational start ATG. The insert was cloned in to pCRScript, excised with *Bam* HI and *Pst* I and ligated with the *lacZ* reporter gene vector pPD95.03. The *Ce-phy-2* promoter was similarly amplified with primers phy-2 PF 2 (*Pst* I) and phy-2 PR 2 (*Bam* HI). Sequences from – 1715 to + 11, relative to the translational start site, were cloned into the pTAg vector and subcloned into pPD95.03 with *Pst* I and *Bam* HI. 1.5 kb of putative *Ce-phy-5* promoter was PCR generated as above using primers phy-5 F (*Hind* III) and phy-5 R (*Xba* I) and cloned into pGEM-T by Laura Gilchrist as part of a BSc (Hons) project. A fragment, representing sequences from – 1402 to + 125 relative to the translational start site, was subcloned in to pPD95.03 with *Hind* III and *Xba* I. N2 genomic DNA was used for production of *Ce-pdi-2* promoter sequences from –2620 to +5 using primers pdi-2 PF (*Sph* I) and PR (*Bam* HI), *Taq* and 27 cycles of PCR. The 2.6 kb product was cloned into pTAg vector then pPD95.03 with *Sph* I and *Bam* HI. Final promoter/reporter gene constructs were sequenced to confirm identity of insert and ensure an in-frame fusion of insert sequences with the *lacZ* reporter gene. Transformation of *C. elegans* was performed by microinjection into the syncitial gonad of N2 or DR96 hermaphrodites. N2s were injected with constructs at 20 µg/ml and plasmid marker pRF-4 at 100 µg/ml. DR96 was injected with marker plasmid p76-16B at 100 µg/ml and reporter construct at 20 µg/ml. At least three independent lines were examined for expression of β-galactosidase for each reporter construct.

2.13.2. Staining for β-galactosidase activity

The method used is a variation on the procedure described by Andrew Fire, 1992. Transgenic *C. elegans* containing reporter constructs were washed from mixed population NGM plates in 1X M9, concentrated by centrifugation and washed twice. The volume was reduced to 100 µl and an equal volume of 2.5% gluteraldehyde (Sigma) added. Suspensions were incubated for 20 minutes at room temperature with occasional gentle mixing, followed by two washes with 1X M9. The volume was reduced to 100 µl, worms pipetted onto microscope slides, vacuum dried, immersed in

acetone at –20°C for 5 minutes and air-dried. 50-100 µl of staining mixture was added per slide, a microscope coverslip (BDH) added and sealed with nail varnish. Worms were checked every hour for staining or left overnight.

Staining mix (1 ml): 500 µl 0.4 M $NaPO_4$ (pH7.5), 1 µl $MgCl_2$, 100 µl Redox buffer (50 mM potassium ferricyanide, 50 mM potassium ferrocyanide), 4 µl 1% SDS, 400 µl H_2O. For the standard staining procedure 12 µl of 2% (w/v) X-gal was used in the mix; for sensitive staining the mixture was pre-heated to 65°C then 12 µl freshly made 20% (w/v) X-gal added.

2.14. Semi-quantitative (sq-) RT-PCR

2.14.1. RT-PCR of *Ce-phy-1*, *Ce-phy-2* and *Ce-pdi-2* from staged samples

PCR was performed on staged cDNA samples representing two hourly intervals throughout post-embryonic development (Johnstone and Barry, 1996). PCR amplification was performed using two sets of primers pairs, one corresponding to the test gene, and a second set, ama-1 F and ama-1 R, for the control gene *ama-1* (Bird and Riddle, 1989), which encodes the largest subunit of RNA polymerase II. *Ce-phy-1* primers used were phy-1 F (pMal) and phy-1 R (pMal), *Ce-phy-2* primers were phy-2 F (pMal) and phy-2 R (pMal), and for *Ce-pdi-2* primer pair pdi-2 F (pMal) and pdi-2 R (pMal) was used. 15 pmol of each primer and 1.5 µl of staged cDNA samples were used in each 50 µl reaction. *Taq* DNA polymerase was added to 1X PCR buffer (45 mM Tris-HCl pH 8.8, 11 mM Ammonium sulphate, 4.5 mM $MgCl_2$, 6.7 mM 2-mercaptoethanol, 4.4 µM EDTA pH 8.0, 1 mM each dNTP, 113 µg/ml BSA). Reactions were cycled in a Robocycler Gradient 96 PCR machine (Stratagene) (92°C - 5 minutes for 1 cycle followed by, 92°C - 1 minute, 56°C - 1 minute, 72°C - 2 minutes for 35 cycles) and samples separated on 1.25% agarose gels

2.14.2. Southern blotting

Agarose gels were treated with 0.25 M HCl acid for 20 minutes, rinsed with neutralisation buffer, soaked for 20 minutes in denaturation buffer, followed by rinsing in neutralisation buffer. Gels were then treated for a further 20 minutes with neutralisation buffer. DNA was transferred using a standard southern blot method in

10X SSC with wicks cut from 3 MM paper and Hybond-N (Amersham Life Sciences) nylon membrane. Membranes were washed in 3X SSC, air-dried and cross-linked in a UV Stratalinker 2400 (Stratagene).

20 X SSC:	3 M NaCl, 0.3 M NaCitrate, pH to7.5 with conc. HCl.
Denaturation buffer:	0.4 M NaOH, 0.8 M NaCl.
Neutralisation buffer:	0.5 M Tris, 1.5 M NaCl, pH 7.6.

2.14.3. Radiolabelling oligonucleotide primers

For each gene tested, the four primers used for PCR were radioactively end-labelled with γ^{32}-P using T4 polynucleotide kinase, in a single reaction. This enzyme transfers γ^{32}-P from rATP to the 5' end of oligonucleotides. 25 µl reactions were performed in 1X T4 polynucleotide kinase buffer (NEB) with 10.5 pmol/µl of each primer. To this 35 µCi of [γ^{32}-P]ATP and 10 units of T4 polynucleotide kinase (NEB) were added and reactions incubated at 37°C for 30 minutes. Reactions were terminated at 95°C for 5 minutes.

2.14.4. Probing blots

Herring sperm DNA (10 mg/ml) was denatured at 100°C for 5 minutes and 50 µl added to 50 mls Church and Gilbert solution (preheated to 56°C) in a hybridisation bottle and incubated at 56°C for 1-2 hrs. Radiolabelled probe (half of the 25 µl reaction) was added along with the corresponding membrane and incubated at 56°C overnight in a rotating hybridisation oven (Stuart Scientific). Membranes were washed twice for 5 minutes with preheated wash buffer at 56°C in hybridisation bottles in the hybridisation oven. Blots were then washed twice in a container with wash buffer in a rocking 56°C water bath, before a final wash in 3X SSC. Filters were dried on Watmann 3 MM paper and exposed to X-ray film (Kodak) for 1-16 hours.

Church & Gilbert solution:	100 mls 1M $NaPO_4$ pH 7.2 (71 g anhydrous Na_2HPO4 with 4 mls of 85% orthophosphoric acid in 1 litre of H2O), 70 mls 20% SDS, 0.4 mls 0.5 M EDTA (pH 8), 29.6 mls H2O.
Wash buffer:	2X SSC, 0.1% SDS.

2.14.5. Quantification of radioactive bands

Radioactive bands corresponding to the test gene and *ama-1* for each PCR were located on the nylon membrane by comparison to the autoradiograph and then excised. Excised bands were added to 2 mls scintillation fluid (National Diagnostics) and counted in a liquid scintillation counter (LKB Wallace). The relative abundance of the test gene transcript was calculated from the ratio of test gene signal to *ama-1* signal for every point in the lifecycle. Results were plotted in arbitrary units.

2.15. RNA interference (RNAi)

2.15.1. RNAi by injection of *in vitro* synthesised double-stranded (ds) RNA

This procedure was based on the method originally described by Fire *et al.*, 1998.

2.15.1.1. Construction of plasmids for *in vitro* RNA synthesis

For *Ce-phy-1* the construct phy-1cDNA,pCRScript was made by PCR on N2 mixed stage cDNA with primers phy-1 F (pMal) and phy-1 R (pMal) with *Taq* polymerase using 25 amplification rounds. The 1.6 kb product was cloned into pCRScript. Clone was separately digested with *Not* I and *Sma* I for the T7 and T3 *in vitro* transcription reactions respectively (see below). A *Ce-phy-2* RNAi construct was similarly made using primers phy-2 F (pMal) and phy-2 R (pMal). The 1.6 kb product was cloned into pCRScript and separately digested with *Not* I and *Sma* I for the T7 and T3 reactions respectively. *Ce-phy-3* RNAi construct phy-3,pPD129.36 (see Section 2.15.2.1) was used to produce RNA using two T7 promoters. Sense and antisense RNA was produced separately after linearisation with *Xba* I and *Xho* I. *Ce-phy-4* RNAi construct phy-4,pGEM (see Section 2.15.2.1) was used to produce RNA using T7 and Sp6 promoters. Sense and antisense RNA was produced separately after linearisation with *Not* I (T7) and *Nco* I (Sp6). *Ce-phy-5* RNAi construct phy-5,pPD129.36 clone (see Section 2.15.2.1) was used to produce RNA using two T7 promoters. Sense and antisense RNA produced separately after linearisation with *Not* I and *Nco* I. *Ce-pdi-2* RNAi construct pdi-2 cDNA,pCRScript was made by PCR on N2 mixed stage cDNA with primers pdi-2 F(pMal) and pdi-2 R (pMal) with *Taq* polymerase using 25 amplification rounds. The 1.4 kb product was cloned into pCRScript and digested separately with *Not* I and *Sma* I for T7 and T3 reactions respectively

2.15.1.2. *In vitro* transcription of RNA

10 µg of plasmid DNA was digested for each strand to be transcribed using the restriction enzymes indicated for the appropriate reaction. Digests were purified by phenol/chloroform extraction and 4 µg of linearised template DNA used in a 100 µl transcription reaction with either the T7, T3 or Sp6 Ribomax large scale RNA production kits (Promega). Reactions were performed in the appropriate 1X reaction buffer, with 1X enzyme mix and 7.5 mM final concentration of each dNTP. Reactions were allowed to proceed for 4 hours after which 4 units of RQ1 DNase (RNase-free) was added and incubated at 37°C for 15 minutes. After extraction with phenol (pH 4.5) and chloroform, RNA was precipitated with isopropanol and sodium acetate. Samples were resuspended in DEPC treated H_2O. 10-20 µg of each single-stranded RNA was then mixed in a volume of 40 µl in 1X injection buffer (20 mM KPO_4, 3 mM K Citrate, 2% polyethylene glycol 6000, pH 7.5) and incubated at 37°C for 30 minutes. The presence of dsRNA was verified by comparison of annealed and single-stranded samples separated on an agarose gel.

2.15.1.3 dsRNA injections

dsRNA was microinjected in to the syncitial gonad of *C. elegans* using the procedure described above for DNA transformation (Section 2.10.1). A concentration of 0.5–1.0 mg/ml dsRNA was injected. After overnight recovery of injected animals to allow clearing of embryos, worms were plated singly on to NGM OP50 plates and transferred daily.

2.15.2. Bacterially-mediated RNAi

2.15.2.1. Constructs for RNAi feeding

25 cycles of PCR on N2 mixed stage cDNA with *Taq* and primers pairs phy-1 HSC F/phy-1 HSC R, and phy-2 HSC F/phy-2 HSC R were used to produce full-length Ce-*phy-1* and Ce-*phy-2* cDNA fragments respectively. Products were cloned in to the pTAg vector and subcloned with *Eco* RV into CIAP treated pPD129.36 (Timmons *et al.*, 2001). 900 bp of *Ce-phy-3* sequence was PCR amplified for 30 cycles with *Taq* from N2 cDNA with primers T20B3.7 cLF and T20B3.7 NR. This was cloned in to pGEM-T and subcloned into pPD129.36 with *Xba* I and *Xho* I. 40 cycles of PCR were used for the amplification of *Ce-phy-4*, using *Taq* on N2 cDNA with primers phy-4 F1 and phy-4

R3. The 484 bp product was cloned in to pGEM-T and subcloned in to pPD129.36 with *Not* I and *Nco* I. For *Ce-phy-5* primers phy-5 cDNA F2 and phy-5 cDNA R3 were used for 30 cycles with *Pfu* on N2 cDNA to give a 560 bp product which was cloned in to pGEM-T, this clone was made by Laura Gilchrist as part of a BSc (Hons) project. *Ce-phy-5* was subcloned from this plasmid into pPD129.36 with *Not* I and *Nco* I. The construct that was used to produce *in vitro* RNA for *Ce-pdi-2* (see Section 2.15.1.1) was used to subclone *Ce-pdi-2* sequences into the feeding vector using *Xmn* I and *Pst* I to excise the gene and ligated with *Sma* I and *Pst* I digested pPD129.36. *C. briggsae phy-1* sequences were PCR amplified with *Taq* from *C. briggsae* G16 mixed stage cDNA using primers CB phy-1 F and CB phy-1 R. The product was cloned into pGEM-T then in to pPD129.36 using *Sal* I and *Nco* I. *Cb-phy-2* was cloned as for *Cb-phy-1* using primers CB phy-2 F and CB phy-2 R. *Cb-pdi-2* was cloned as above but using the primers CB pdi-2 F and CB pdi-2 R and subcloned using *Sal* I and *Apa* I. Cloning of the *C. briggsae* gene fragments was performed by Jimi-Carlo Bukowski-Wills as part of a BSc (Hons) project. All pPD129.36 clones were sequenced with primers L4440 F and L4440 R to ensure integrity of double T7 sites in vector backbone and the identity of inserts.

2.15.2.2. Transformation of *E. coli* HT115(DE3)

E. coli HT115(DE3) cells (Timmons *et al.*, 2001) were made competent by inoculating an overnight culture of LB+ 12.5 µg/ml tetracycline. LB (+tet) was then inoculated with a 1:100 dilution of the overnight culture and cells grown to mid-log phase (OD_{600}=0.4). Cells were pelleted for 10 minutes at 3000 rpm at 4°C and gently resuspended in 0.5X original volume of cold, sterile 50 mM $CaCl_2$. After ice incubation for 30 minutes, cells were spun as before and resuspended with $CaCl_2$ at 0.1X original volume. Cells were frozen at –80°C in 10% glycerol in single use aliquots. Transformation with pPD129.36 constructs was performed using 1-100 ng/µl plasmid DNA with 100 µl of competent cells. Cells with DNA added were incubated on ice for 30 minutes then 37°C for 1 minutes and ice for 2 minutes. Cells were then put in an orbital shaker for 1 hour with 400 µl SOC. Transformed cells were selected on LB (+amp+tet). Transformants were PCR tested to confirm the presence of plasmid.

2.15.2.3. RNAi feeding

The feeding method used was based on the optimised procedure described by Kamath *et al.*, 2000 (Kamath *et al.*, 2000). Cultures were inoculated overnight with Amp (no Tet) and used to seed NGM plates containing 100 µg/ml Amp and 1 mM IPTG. These were left overnight at room temperature to induce T7 RNA polymerase. Nematodes were added and incubated at 25°C for 2 days.

2.16. Rescue of *C. elegans* mutant *dpy-18* with *Ce-phy-1*

Ce-phy-1 control and coding sequences were amplified from N2 genomic DNA using 27 cycles of PCR with primers Phy-1 (Not I) F and Phy-1 (Sal I) R and a mixture of *Taq* and *Pfu* polymerases in a unit ratio of 10:1. The 7 kb product was ligated with pGEM-T vector and sequenced over the cloning junctions to confirm identity. The clone contained sequences from –2006, relative to ATG, to position 227 after TAA, therefore including the *Ce-phy-1* promoter, coding sequence and 3' UTR (containing the polyadenylation signal sequence and transfer site). This clone was injected at a concentration of 5 µg/ml in to strains carrying *dpy-18* alleles *e364* and *e1096*. As a marker of transformation, dpy-7-GFP was injected at 10 µg/ml and the final DNA concentration increased with pBluescript SKM at 100 µg/ml. Control injections were performed with dpy-7-GFP and pBluescript plasmids alone.

2.17. Cloning of *phy*-1 alleles from *dpy-18* strains

cDNA and genomic DNA were produced for *dpy-18* alleles *e364*, *e1096* and *bx26* using the methods described in Section 2.5.

2.17.1. *Ce-phy-1* cloning from *dpy-18(e364)*

Full-length coding sequence clones were amplified from *dpy-18(e364)* with primers phy-1 HSC F and phy-1 HSC R following 27 cycles of PCR with *Taq*. Two pGEM-T clones from separate PCR reactions were analysed, one was fully double strand sequenced with primers pGEM forward, pGEM reverse, phy-1 IS IF, phy-1 IS 2F, phy-1 IS 3F, phy-1 IS 4F, phy-1 IS 5F, phy-1 IS 1R, phy-1 IS 2R, phy-1 IS 3R, phy-1 IS 4R, phy-1 IS 5R, phy-1 F (pMal) and phy-1 R (pMal). The second clone was double-strand sequenced only over potential mutation regions. Full-length *phy-1* genomic coding sequence was produced from *dpy-18(e364)* with primers phy-1 HSC F and phy-1 HSC

R following 28 cycles of PCR using a 10:1 *Taq/Pfu* mix. The pGEM-T clone was double strand sequenced over the region of mutation using primers phy-1 IS 4R and phy-1 (new) GS F.

2.17.2. *Ce-phy-1* cloning from *dpy-18(e1096)*

PCR on cDNA prepared from *dpy-18(e1096)* with full-length *Ce-phy-1* primers, phy-1 HSC F and phy-1 HSC R, gave no product. Control PCRs were performed with the *Ce-phy-1* primer combination on N2 cDNA and primers specific to *Ce-phy-2* on *e1096* cDNA. PCR on genomic *e1096* DNA with primers phy-1 (*Not* I) F and phy-1 (*Sal* I) R with a 10:1 *Taq/Pfu* mix and 28 cycles produced a band approximately 800 bp smaller than N2. This was cloned into pGEM-T. The area of deletion was further defined by PCR using, phy-1 IS 1F with Phy-1 (*Sal* I) R, and Phy-1 (*Not* I) F with phy-1 IS 5R. The genomic *e1096* clone was then sequenced with primers pGEM forward, pGEM reverse, phy-1 PS 1F, phy-1 PS 2F and phy-1 IS 5R to define the deletion. Transcription of *Ce-phy-1* sequences from *e1096* mRNA was mapped by PCR using the full length antisense primer phy-1 HSC R in combination with a range of sense primers phy-1 F (pMal), phy-1 IS 1F, phy-1 IS 2F, phy-1 IS 3F, phy-1 IS 4F and phy-1 IS 5F.

2.17.3. *Ce-phy-1* cloning from *dpy-18(bx26)*

The *dpy-18(bx26)* allele was analysed by PCR cloning full-length coding sequence from cDNA prepared from this mutant strain with primers phy-1 BV F and phy-1 BV R using 27 cycles and *Pfu*. This product was A-tailed, cloned in to pGEM-T and sequenced as above. The genomic sequence of *Ce-phy-1* from *bx26* genomic DNA was cloned over the proposed mutation region (performed by L. Murray, WCMP, Glasgow). PCR with primers phy-1 IS 1F and phy-1 IS 2R with 30 cycles using *Pfu* gave a 1.6 kb product. This was cloned in to pCRScript and sequenced with phy-1 IS 2R.

2.18. Developmental time-course analysis of RNAi embryos

RNAi was performed by injection, worms left to recover for 10 hours then transferred to fresh plates. Embryos being produced after this time should be RNAi affected. Pre-gastrulation embryos were collected and washed in M9 and mounted on slides prepared with dried 2% agarose, covered with a coverslip and sealed with White soft paraffin (Pinewood Laboratories). Embryos were then monitored throughout development and

images taken at intervals of 30 minutes using a Hamamatsu digital camera with a Zeiss Axioskop 2 microscope and Improvision Openlab software.

2.19. Sample preparation for scanning electron microscopy

RNAi feeding was performed on wild type N2 strain with *Ce-pdi-2,* and *dpy-18(e364)* strain with *Ce-phy-2* on ten 9 cm plates with ten L4s added per plate. Plates were grown for 11 days at 25°C, then nematodes were washed with 10 mls PBS, allowed to settle on ice or spun very gently. The volume was taken down to approximately 150 µl and 500 µl 2.5% gluteraldehde (25% stock grade 1, EM grade, diluted in PBS) was added and left for 90 minutes at room temperature. Nematodes were washed twice in PBS for 5 minutes and volume reduced to 100 µl. Samples were taken to the Electron Microscopy Centre at the University of Glasgow. Samples were processed there by fixing in 1% osmium and dehydration in acetone, these were then processed in a critical point drier. Mounted samples were viewed using a Philips 500 scanning electron microscope.

2.20. Injection of *col-19::gfp* construct

Plasmid BA7-1, a gift from C. Shoemaker (AgResearch, New Zealand), contained the *C. elegans* collagen *col-19* (ZK119.3) genomic sequence and 5' control sequence fused to *gfp*. Wild type N2 nematodes were injected with BA7-1 at 20 ug/ml along with pRF-4 as a marker for transformation. After confirming GFP fluorescence in transformed lines the BA7-1 plasmid was injected alone with 100 ug/ml pBluescript SKM into N2, *dpy-18(e364)* and *phy-2(ok177)* strains, and transgenic lines generated. Strains TP12 and TP13 were generated by integration of transgenic arrays in N2 and *dpy-18(e364)* backgrounds respectively. Generation of integrated strains was performed by M. Thein and A. Page (both WCMP, Glasgow).

2.21. Antibody staining of nematodes

Nematodes were washed of plates in 1X M9 buffer, washed and volume reduced to 100 µl. Poly-L-lysine microscope slides were prepared by adding a drop of a 0.1% w/v solution of poly-L-lysine (Sigma) to slides and baking at 80°C for 30 minutes. Nematodes were added to slides, covered with a coverslip and immediately frozen on a cold block in dry ice for 15 minutes (or overnight). After removal of the coverslip slides were immersed in –20°C methanol for 10 minutes, followed by –20°C acetone for 10

minutes. Slides were air dried, washed with 1X PBST (PBS with 0.1 v/v Tween-20), blocked with 1% w/v dried milk in PBST and washed again. The anti-DPY-7 monoclonal antibody (a gift from I. Johnstone, WCMP, University of Glasgow) was raised in mice and is specifically active against the unique C-terminal domain of the *C. elegans* collagen DPY-7 (McMahon *et al.*, 2003). This was used at a concentration of 1/50 dilution of anti-DPY-7(5b) in PBST, with 100 µl of diluted antibody used per slide. This was left for 1 hour at room temperature (or overnight at 4°C), followed by washing in PBST. 100 µl of a 1/100 dilution of secondary antibody, Alexa Flour 488 goat anti-mouse IgG (Molecular probes), was added and incubated at room temperature for 1 hour out of the light. Slides were washed, 14 µl mount solution (50% glycerol, 0.5X PBS, 2.5% DABCO (Sigma) added, and coverslips sealed with nail varnish. Nematodes were viewed as described.

2.22. Rescue experiments with *C. elegans dpy-18* strains by expression of alternative α subunits

2.22.1. Construction of vector pAW1

Vector pAW1 was constructed for the expression of alternative α subunit coding sequences with the 5' and 3' sequences from *Ce- phy-1*. The *Ce-phy-1* promoter from plasmid phy-1,pPD95.03 (see Section 2.13.1) was excised with *Pst* I and *Bam* HI and the 2.8 kb fragment cloned into pBluescript SKM. *Ce-phy-1* 3' UTR sequences were generated by PCR from N2 genomic DNA using primers phy-1 3'UTR F (*Sac* I) and phy-1 3'UTR R (*Sac* I) with 27 cycles and Vent polymerase. This product was cloned into pCRScript, digested with *Sac* I, cloned into plasmid phy-1 prom,pBluescript SKM and selected for the correct orientation of the 3' UTR insert. The resulting construct was sequenced at the junctions of inserts with the primers 1224 and M13 reverse and named pAW1.

2.22.2. Human α subunit rescue constructs

Adult human lung cDNA was produced from adult human lung total RNA (Stratagene). Full-length cDNA sequences from human prolyl 4-hydroxylase α subunits I and II were amplified from this template by PCR. α subunit I was produced with the primers hP4H αI F (*Bam* HI) and hP4H αI R (*Not* I) using *Pfu* Turbo polymerase (Stratagene) with 30 PCR cycles. Subunit αII was amplified similarly using the primers hP4H αII F (*Bam*

HI) and hP4H αII F R (*Not* I). Both products were cloned into pCRScript then subcloned into vector pAW1 with *Bam* HI and *Not* I. Clones were sequenced at their junctions using primers M13 Reverse and pAW Seq F.

2.22.3. Human α subunit rescue constructs with synthetic intron sequences

The synthetic intron (SI) insertion protocol and intronic sequence used were based upon a protocol received from A. Fire (Carnegie Institution of Washington, Baltimore). Two complementary oligos, IVS sense and IVS antisense, were designed to represent a typical *C. elegans* intron, containing the required splice donor and acceptor sequences to be recognised by the *C. elegans* splicing machinery. As the double-stranded oligo is ligated into a blunt site the reading frame is unaltered after its excision. Equal volumes of these primers were mixed, heated to 65°C for 5 minutes then allowed to cool slowly to room temperature, to allow annealing of oligos to form a double-stranded blunt ended section of DNA. The construct hP4H αI cDNA, pAW 1 contains a single *Bsa* BI recognition site within the coding region of αI. However the enzyme is methylation sensitive and DNA from αI clone prepared with standard strains of *E. coli* is not cleavable with *Bsa* BI. Plasmid hP4H αI cDNA, pAW1 was therefore transformed into the *E. coli* strain GM2163 which is *dcm* and *dam* minus. *Bsa BI* digestible plasmid DNA was purified from this strain. Approximately 10 ng of linearised plasmid was then ligated with a large molar excess of the annealed SI (approximately 4000 times) by adding 250 ng SI in a 10 µl reaction. *Bsa BI* was included in the ligation reaction to reduce the background resulting from self-ligation of the non-CIAP treated plasmid. Positive clones for hP4H αI, SI, pAW1 were identified by PCR, using a gene-specific primer in combination with a single stranded IVS primer. These were then sequenced with hP4H αI primer to confirm the presence of the insert. hP4H αII cDNA, pAW1 was cut with restriction endonuclease *Stu I* and the synthetic intron ligated using a similar process to that described above. The construct was sequenced with hP4H αII IS 1F.

2.22.4 *Ce-phy-2 dpy-18* rescue construct

C. elegans phy-2 cDNA was cloned by PCR from mixed stage cDNA using primers phy-2 FL F (*Xba* I) and phy-2 FL R (*Sac* II). This product was cloned into pCRScript, subcloned into pAW1 by digestion with *Xba* I and *Sac* II and sequenced with M13 Rev.

2.22.5. Microinjection of *C. elegans dpy-18* with alternative α subunits

Human α cDNA and SI constructs, and *Ce-phy-2* rescue clones were injected into *dpy-18(e364)* at concentrations of 100 μg/ml with the plasmid dpy-7-GFP at 5 μg/ml as a marker and injection mixes made up to a final concentration of 150 μg/ml with pBluescript SKM. The *Ce-phy-2* rescue clone was also injected at the lower concentration of 10 μg/ml.

2.22.6. Single worm RT-PCR of human α subunit rescue transgenic lines

Single worms (transformed or non-transformed) were picked onto unseeded plates and washed in a pool of 1X M9. These were picked into a 0.2 ml PCR tube in 2 μl DEPC treated H_2O and frozen at –80°C. Lysates were prepared by adding 1 μl of single worm lysis buffer (SWLB) plus proteinase K (final concentration 0.5 μg/μl) with an overlay of mineral oil. These were incubated at 65°C for 1 hour followed by 95°C for 15 minutes. 1 μl of Promega DNAse I (RNAse free RQ1 1unit/μl) was added and reaction incubated at 37°C for 30 minutes and heat inactivated for 15 minutes at 95°C. The QIAGEN One Step RT-PCR procedure was used to attempt RT-PCR amplification from human α subunit mRNA. This procedure allows the reverse transcription and PCR to be performed as a one-step procedure by use of an enzyme mix containing Omniscript (QIAGEN) and Sensiscript (QIAGEN) reverse transcriptases along with HotStart DNA polymerase (QIAGEN). This mix of reverse transcriptase enzymes enables transcription from a range of RNA concentrations, thus allowing for production from picogram concentrations of starting RNA. The DNA polymerase included is inactive during the reverse transcription step and is activated after a heating step at 95°C. Therefore the polymerase should not interfere with the reverse transcription step and the production of non-specific and primer-dimer products in the first cycle is reduced. PCR conditions used were as suggested in the QIAGEN protocol with 1X Q solution being included, a final concentration of 6 mM was used for each primer. 21 μl of mix was added to 4 μl of worm lysate. Cycling conditions were:-

50°C -30 minutes} 1 cycle

95°C -15 minutes} 1 cycle

92°C - 1 minutes, 62°C - 2 minutes, 72°C - 2 minutes} 40 cycles

72°C - 10 minutes} 1 cycle.

Primers used for hP4HαI, SI, pAW1 lines were hP4HαI cDNA F and hP4HαI SW RT R or hP4HαI SW RT F and hP4HαI SW RT R. Two lines were tested one with each set of primers. One line of hP4HαII, SI, pAW1 was tested with hP4HαII cDNA F and hP4HαII SW RT R only. Primers tested on plasmids to confirm that SW RT R primers which span the synthetic intron region were positive on cDNA clones and negative on SI containing clones.

SWLB: 10 mM Tris (pH 8.0), 50 mM KCl, 2.5 mM $MgCl_2$, 0.45% Tween-20, 0.05% gelatin. Made without Tween-20, autoclaved, Tween-20 added, stored at – 20°C.

2.22.7. Westerns for human α subunit transgene expression

Details of buffers, SDS PAGE and western blotting can be found in Sections 2.9 and 2.25. GFP positive and negative worms (20-30 animals) from human α SI rescue experiments were collected in 50 µl of 1X M9. Whole plates were also washed along with a plate of *dpy-18* (*e364*) for comparison. Worms were pelleted and washed with 1X M9 and pellets frozen at –80°C. To the whole plate extracts P4H buffer was added to 1X concentration, samples homogenised and spun for 30 minutes at 4°C. Supernatants were removed and boiled in 1X NuPAGE sample buffer and 5% DTT. Pellets were resuspended and boiled in 1X P4H buffer, 1X SDS sample buffer and DTT. GFP positive and negative worm collections were boiled in 1X SDS sample buffer and 5% DTT. Samples were run on SDS PAGE gels and western blotted. One line for each human α-subunit was tested using antibodies K17 (anti human αI polyclonal antibody raised in rabbits) at 1/3000 dilution and a 1/1000 dilution of 95K4B (anti human αII monoclonal antibody raised in mouse). Both these antibodies were a gift from J. Myllyharju, Collagen Research Unit, Oulu, Finland.

2.23. Examination of *C. elegans* mutant *let-44* as a candidate *Ce-pdi-2* mutant

2.23.1. *let-44* strain GR1029

Strain GR1029 has the genotype *let-44*(*mg41*) *lon-2*(*e678*)X:mnDp31(X;f). The *let-44* and *lon-2* mutations are balanced by a free duplication that contains a wild-type copy of both loci. Approximately a third of animals from this strain loose the free duplication and arrest as dead embryos due to the *let-44* mutation. Animals that retain the free

duplication are phenotypically wild type and segregate wild type and dead embryos. Strain GR1029 was investigated to determine if the locus *let-44* was mutant for *Ce-pdi-2*.

2.23.2. *let-44* developmental timecourse analysis

let-44 embryos was examined at regular time intervals throughout development to determine if their embryonic arrest phenotype was similar to that of *Ce-pdi-2* RNAi treated nematodes. The procedure used here is described in the developmental timecourse analysis of RNAi treated embryos (Section 2.18). As approximately 33% of GR1029 embryos observed were *let-44* mutants any embryos that went on to hatch were excluded from analysis.

2.23.3. *Ce-pdi-2* genomic sequence from *let-44* mutants

To generate genomic *Ce-pdi-2* sequence from *let-44* mutants, GR1029 hermaphrodites were allowed to egg lay for 8 hours. These embryos were then left to develop for 16 hours. After this period any unhatched embryos were considered dead, the vast majority of which should be *let-44* mutants. Single *let-44* embryos were transferred using a drawn out glass capillary to a watch glass and washed in 1X M9 buffer. Embryos were then singly transferred using a capillary containing Chitanase (20 mg/ml) in salts and put in thin walled PCR tubes. These were frozen at –80°C and 4 µl single worm lysis buffer (SWLB) with proteinase K added along with a mineral oil overlay, and the reaction incubated at 65°C for 1 hour followed by 100°C for 15 minutes. This treatment lysed the embryos releasing their DNA, which was used as a template for PCR using *Taq* polymerase. Primers pdi-2 HSC F and pdi-2 HSC R resulted in the wild type sized 2 kb product. PCRs were repeated with proofreading polymerases, however no products were obtained. *Taq* products were isolated and cloned into vector pCR2.1 TOPO. Clones were sequenced with M13 reverse, pdi-2 IS 1F, pdi-2 IS 2F, pdi-2 IS 1R and pdi-2 IS 2R and pGEM forward. This was repeated for three individual embryos.

2.23.4. *Ce-pdi-2* cDNA sequence from *let-44* mutants

Ce-pdi-2 sequence was also generated from *let-44* mRNA by RT-PCR from pools of embryos. Attempts were initially made to clone *Ce-pdi-2* sequences from *let-44* mutants using 1-3 embryos, generated as described above, lysed in SWLB with proteinase K

followed by DNase treatment. These were then used as a template for reverse transcription and PCR using the QIAGEN One Step RT-PCR procedure. This approach was however unsuccessful in generating consistent results. Instead a pool of approximately thirty *let-44* embryos was collected, washed twice with 200 µl PBS, resuspended in 180 µl lysis buffer (0.5% SDS, 10 mM EDTA, 10 mM Tris pH 7.5), 10 µl 2-mercaptoethanol and 10 µl proteinase K (10 mg/ml) added, and stored at –80°C. A modified version of the Total RNA Isolation Reagent (TRIR) (ABgene) protocol was followed, which allows the isolation of RNA using a formulation of guanidine salts, urea and phenol. Embryos were lysed by incubation at 55°C for 1 hour then cooled to 4°C. Total RNA was isolated using the TRIR, with the lysis step above replacing the suggested homogenisation procedure, and with 2.5X volume of reagent used rather than the suggested 5X volume. RNA was resuspended in DEPC treated water and 1/100 volume RNasin (Promega 30 units/µl). RNA was divided between six separate RT-PCRs and the QIAGEN One Step RT-PCR procedure used as described previously using the cycling conditions below and the primers pdi-2 HSC F and pdi-2 HSC R. The 1.4 kb products were cloned into vectors pGEM-T or pCR2.1 TOPO and sequenced with vector specific sequencing primers and pdi-2 IS 1F, pdi-2 IS 2F, pdi-2 IS 1R and pdi-2 IS 2R.

PCR conditions:-

50°C -30 minutes} 1 cycle

95°C -15 minutes} 1 cycle

92°C - 1 minutes, 56°C - 2 minutes, 72°C - 2 minutes} 40 cycles

72°C - 10 minutes} 1 cycle.

2.23.5. *let-44* rescue experiments with *Ce-pdi-2*

Repair of *let-44* phenotype by microinjection of the wild-type copy of the mutated gene would result in the production of viable worms that have lost the free duplication but contain the introduced free array. Progeny from injections in which this has occurred would be identifiable due to the presence of the *lon-2* genetic locus that would result in worms with a phenotypically long appearance. *lon-2* worms are typically 50% longer than wild type worms at all stages.

2.23.5.1. Cosmid rescue of *let-44*

To investigate the potential of *Ce-pdi-2* to rescue the *let-44* phenotype the strain GR1029 was injected with wild type *Ce-pdi-2* and progeny scored for presence of Lon. Cosmid C07A12 contains the entire coding region, promoter and 3' UTR regions for *Ce-pdi-2* at position C07A12.4. Cosmid DNA was prepared by the same procedure used for plasmid DNA preparations. The cosmid was checked for the presence of *Ce-pdi-2* sequences by PCR using primers pdi-2 PF and pdi-2 Resc R and microinjected into the strain GR1029 at a concentration of 1 μg/ml or 5 μg/ml, with 5 μg/ml dpy-7-GFP marker and 120 μg/ml pBluescript SKM.

2.23.5.2. Plasmid rescue of *let-44*

A *Ce-pdi-2* plasmid designed to rescue *let-44* phenotype was constructed. Primers pdi-2 PF and pdi-2 Resc R were used with 28 cycles of PCR using 10:1 *Taq:Pfu* mix. This fragment corresponded to position –2628 relative to translational initiation, includes all of coding sequence and 1112 bases after TAA (included predicted poly-adenylation signal sequence). The product was cloned into pGEM-T vector and sequenced with pGEM forward. pGEM reverse and pdi-2 IS 1F. Plasmid DNA was injected at 5 μg/ml with 10 μg/ml dpy-7-GFP and pBluescript SKM at 100 μg/ml. Transgenic worms were not taken to F2.

2.24. Baculovirus expression

The baculovirus expression vector system (Crossen and Gruenwald, 1998) was used for the expression of proteins. This results in high expression levels, due to host cell gene expression shutoff and expression from a strong late viral promoter, and post-translational modification of the desired protein. In addition, this system enables co-expression of multiple proteins by co-infection of cells with more than one recombinant virus. The Pharmingen system used here utilises the *Autographa californica* nuclear polyhedrous virus (AcNPV). In tissue culture, baculovirus genes, such as polyhedrin, are non-essential and can be replaced by a heterologous gene. As insertion directly into viral genome is not easy, a transfer vector is used where genes can be cloned into the vector, mixed with AcNPV DNA and the heterologous gene recombined into the viral genome by virtue of homologous flanking regions. Modification of AcNPV DNA by deleting the non-essential polyhedrin gene increases the efficiency of the process, a

lethal deletion downstream from this locus is also included. Co-transfection of insect cells (*Spodoptera frugiperda*) with a transfer vector, containing the gene to be expressed, flanked by sequences surrounding the polyhedron locus, results in recovery of the lethal phenotype and insertion of the gene into AcNPV DNA. This vector contains a polyhedrin promoter, lacks the polyhedrin coding sequence, but contains flanking sequence allowing homologous recombination into AcNPV DNA and rescue of the lethal mutation. This results in high level expression of the heterologous gene from the polyhedrin promoter, and expression of a non-fused protein.

2.24.1. Insect cell expression construct for *Ce-phy-2*

Ce-phy-2 was cloned from N2 mixed stage cDNA by PCR with primers phy-2 BV F (*Not I*) and phy-2 BV R (new) (*Xba I*) for 30 cycles with *Pfu*. The product was cloned into pCRScript and subcloned with *Xba* I and *Not* I in to the vector pVL1392. Sequencing was carried out with primers phy-2 BV F and phy-2 BV R, phy-2 IS 1F, 2F, 1R and 2R. The Collagen Research Unit, Oulu, Finland, performed the expression of *Ce*-PHY-2 in insect cells.

2.25. Native extract analysis of *C. elegans* and *B. malayi* worms and insect cell samples

Protein extracts were made from a 500 µl volume of concentrated *C. elegans*, and from approximately 100 adult female *B. malayi* parasites, collected from the peritoneal cavity of jirds (a gift from R. Maizels and Y. Harcus ICAPB, University of Edinburgh). Worms were disrupted using a glass hand held homogeniser in 1X P4H buffer with DTT and protease inhibitors added fresh prior to use. Samples were left on ice for 2 hours, spun and supernatants and Native sample buffer added to 1X concentration. Insect cell extracts were made from 10 ml cultures with cells washed in PBS. Cell pellets were then resuspended in 500 µl Homogenisation buffer, homogenised, spun and supernatant analysed in 1X Native sample buffer. Samples were separated on 4-12% Tris Glycine polyacrylamide gels (Invitrogen) in 1X Native running buffer (Invitrogen), gels run overnight at 4°C with 20 volts. Proteins were transferred to PVDF membranes by Western blotting, performed as described in Section 2.9.3. For *C. elegans* PHY-1, PHY-2 and PDI-2 detection the corresponding peptide antisera were used at 1/5000

dilution in PBST. For *Bm*-PHY-1, antisera was used at 1/4000 dilution in PBST (mix of two rabbits final bleeds).

4 X P4H buffer:	0.4 M NaCl, 0.4 M Glycine, 40 mM Tris pH 8.0, 0.4% Triton-X100, make up in H_2O, pH to 7.8 (at 4°C). DTT added to a final concentration of 40 µM and protease inhibitors PMSF (to 4mM), EDTA (to 4mM), EGTA (to 4mM), E64 (to 8µM) Pepstatin (to 0.4 µM) added.
Homogenisation buffer:	67mM Tris, 267 mM NaCl, 0.2% Triton-X 100.
4X Native sample buffer:	2mls of 1.25 M Tris, 4 mls glycerol, 0.96 mg bromophenol blue, make up to 10 mls with H_2O).
10X Tris-glycine native running buffer:	29g Tris Base, 144g glycine, make to 1 litre with H_2O.

2.26. PCR mapping of *Ce-phy-3*, *Ce-phy-4* and *Ce-phy-5*

For *Ce-phy-3* sense primers T20B3.7 cF, T20B3.7 cSF and T20B3.7 cLF were used for PCR in combination with the antisense primer T20B3.7 NR, firstly on genomic DNA then on cDNA to map the transcribed region of the gene. *Ce-phy-4* sense primers phy-4 F1 and phy-4 F2 were used in PCR with phy-4 R1, phy-4 R2 and phy-4 R3 in all sense and antisense combinations on genomic and cDNA. *Ce-phy-5* sense primers phy-5 cDNA F1 and phy-5 cDNA F2 were used for PCR with antisense primers phy-5 cDNA R1, phy-5 cDNA R2 and phy-5 cDNA R3 in all possible combinations with genomic and cDNA templates. Attempts were also made to map *Ce-phy-4* and *Ce-phy-5* by SL PCR as described in Section 2.12.

2.27. Cloning of *Bm-phy-1* cDNA

2.27.1. *B. malayi* ESTs

SW3D9CA480SK was supplied as a dried *Taq* PCR product from the amplification of bacteriophage clones with the primers T7 and T3. This 500 bp fragment was resuspended in 10 µl H_20 and cloned into pCRScript. Multiple attempts were made to sequence this clone but were unsuccessful. MBAFCX8G05T3, MBAFCZ7H09T3 and SWAMCAC30B11SK (*Bm*-PDI EST) were supplied as bacteriophage. 5 µl of phage

was amplified by PCR using primers T7 and T3 at 100 ng/µl in a 50 µl reaction using *Taq* polymerase with the following conditions:-

95°C for 3 minutes} 1 cycle
94°C for 15 seconds, 55°C - for 20 seconds, 72°C - 3minutes} 35 cycles
72°C for 10 minutes} 1 cycle.

PCR products of 1.5 kb and 1.4 kb respectively were cloned into pGEM. Clone X8G5, pGEM was sequenced with primers pGEM forward, pGEM reverse, X8G5 F1, X8G5 R1, X8G5 F2, X8G5 R2. Z7H9, pGEM was sequenced with pGEM forward and pGEM reverse primers.

2.27.2. Phage cDNA library screens

2.27.2.1. Preparation of plating cells

A single colony of LE392 from a freshly streaked LB agar plate was added to 20 mls of LB with 200 µl 20% maltose and grown for 16 hours. Cells were pelleted by centrifugation for 5 minutes at 6000g and resuspended in 10 -15 mls filter sterile 10mM $MgSO_4$ to an OD_{600} of 2 (approximately 1.5 X 10^9 cells per ml).

2.27.2.2. Library titres

B. malayi adult male cDNA library (SAW94NLBmAm) was titred by taking serial 1/10 dilutions of phage library stock in phage buffer (20 mM Tris-HCl pH7.4, 100 mM NaCl, 10 mM $MgSO_4$). 20 µl of 1/100 to 1/100,000 dilutions were added to 0.1 ml LE392 plating cells in a 15 ml falcon tube, mixed gently and left at room temperature for 15 minutes. Top agarose was melted and kept at 42°C. 3 mls of top agarose was added to falcon tube, mixed gently and plated on 100 mm prewarmed LB plates. After setting, plates were incubated at 37°C overnight. 20 µl of 1/10,000 dilution gave 100 plaque forming units (pfu) therefore library titre is 5 X 10^7 pfu/ml.

2.27.2.3. Radiolabelled probe

A 312 bp PCR product was generated with *Taq* polymerase and the primers X8G5 F1 and BM N Phy Reverse covering the predicted 3' coding region of *Bm-phy-1*. This was cloned into pGEM and sequenced. Clone Bm-phy 3' frag, pGEM was digested with *Sac*

II and *Spe* I to liberate insert. A Prime It II Random primer-labelling kit (Stratagene) was used to label probe. Random 9-mer primers anneal to multiple sites along the length of DNA. The Klenow fragment of DNA polymerase I incorporates nucleotides, one of which is radioactive, into the primer template complexes. Using 3' exonuclease deficient Klenow results in rapid incorporation of nucleotides. 100 ng of gel purified insert from clone Bm-phy 3' frag,pGEM was labelled using 10 µl random oligonucleotide primers in a reaction volume of 34 µl. This was heated to 100°C for 5 minutes then 5 µl of 5 X dCTP buffer, 5 µl of labelled α^{32}P dCTP (10mCi/ml) and 1 µl of Exo$^-$ Klenow enzyme added. The reaction was incubated at 37°C for 30 minutes and stopped by addition of 2 µl stop solution. The probe was purified using NucTrap probe purification column (Stratagene) with Push column Beta shield Device (Stratagene) that removes unincorporated nucleotides from the probe. The column was equilibrated with 70 µl 1X STE (100 mM NaCl, 20 mM Tris pH 7.4, 10 mM EDTA). The probe volume was made up to 70 µl with 1X STE and applied to the column and collected along with two washes with 70 µl 1X STE. The radiolabelled probe was denatured at 100°C for 5 minutes in a screw top tube then rapidly cooled on ice immediately before use. 50 µl of probe was used per hybridisation bottle.

2.27.2.4. Library screen

The bacteriophage library was plated on two 25 cm x 25 cm square LB plates. For each plate 200 µl of 1/1000 dilution of phage stock with 1 ml of plating cells in a 50 ml falcon tube were processed as above using 40 mls of top agarose. The total number of plaques screened was approximately 2 x 10^4. Plates were left at 4°C for a few hours to harden top agarose. Hybond-N nylon membranes (Amersham) were cut to size and labelled to identify plate and contact side. Dry filters were placed carefully onto plates, labelled with inked syringe pricks and left for 1 minute then removed. Duplicate plaque lifts were performed from each plate with the second filter left in contact with the plate for 2 minutes. Filters were processed by immersion in the following solutions; 1.5 M NaCl, 0.5 M NaOH for 2 minutes, 1.5 M NaCl/0.5 M Tris pH 8 for 1 minute, 1.5 M NaCl/0.5 M Tris pH 8 (second container) for 1minute, 20X SSC for 1 minute, 1 M NH_4Ac, 1X SSC for 1 minute. Filters were air-dried on Watmann 3MM paper for 1 hour then placed DNA side up in UV cross-linker. Pre-hybridisation and hybridisations were performed at 60°C as described in Section 2.14.4.

2.27.2.5. Phage plaque elution

Positive plaques were identified by first orientating the autoradiographs on the filters using luminous markers (Stratagene), autoradiographs where then marked where ink marks appeared on the filter, this then enabled the autoradiograph to be orientated on plates and positive plaques located. The position of positive plaques was checked with the duplicate filters. Only plaques positive on both filters were used. Eleven positive plaques were selected for further screening. These were picked with a sterile Pasteur pipette by plugging the plate and adding this plug to 5 mls of phage dilution buffer with a drop of chloroform. Plugs were left for at least 4 hours at 4°C in the dark to allow elution of the phage from the agarose.

2.27.2.6. Subsequent library screening rounds

All positive phage clones were titred as before then plated on 9 cm plates with 0.1 mls of LE392 cells and 3 mls top agar and left overnight at 37°C. Rounds of plaque lifts, hybridisations and isolations were performed as before until approximately 100% of plaques were positive. This required nine clones to be screened to the tertiary stage and two to the quaternary stage.

2.27.3. Phage DNA preparation

Liquid phage preparations and DNA isolations were performed following the LambaSorb Phage Adsorbant (Promega) protocol. This utilises a reagent consisting of a conjugate of *Staphylococcus aureas* cells and rabbit polyclonal antibody directed against lambda bacteriophage particles which allows purification of phage particles from crude lysates. Phage were grown on LB agar plates and single well isolated phage plaques picked by removing a plug of agar into 100 mls of phage buffer, which was then left at 4°C. LE392 cells were freshly grown in 0.2% maltose and 100 mM $MgSO_4$. 500 µl of LE392 culture was incubated with the agar phage plug at 37°C for 20 minutes then added to 100mls of LB + 100 mM $MgSO_4$ in 37°C orbital shaker until bacterial lysis occurred. Phage particles were isolated with LambaSorb Phage Adsorbant (Promega) and heated to release phage DNA. DNA was phenol:chloroform extracted, precipitated with ammonium acetate and ethanol and RNase A treated as described in the protocol. Eleven clones were subcloned into the plasmid vector pBluescript SKM using *Eco* RI and *Xho* I. Two clones of around 1.6 kb were selected and sequenced

completely on both strands. Primers used were 1224, M13 reverse, BM *phy* 1.1 IS 1F, 2F, 3F and BM *phy* 1.1 IS 1R, 2R, 3R(B). Nine clones were sequenced with 1224 and M13 Reverse.

2.27.4. *Bm-phy-1* 5' RACE

The amplification of the 5' end of *Bm-phy-1* was performed using the Rapid Amplification of cDNA ends kit (Gibco BRL). This method involves conversion of total RNA to cDNA using a gene specific antisense oligonucleotide and an RNAase H⁻ derivative of Moloney monkey leukemia virus reverse transcriptase. Terminal deoxynucleotidyl transferase (TdT) is used to transfer a homopolymeric tail to the 3' end of this cDNA (which represents the 5' end of mRNA). Nested PCR using a further two gene specific primers in combination with primers complementary to the added homopolymeric tails, provided in the kit, are used to amplify the 5' end of the cDNA sequence.

Total RNA was extracted from three *B. malayi* adult females using Trizol reagent (Gibco BRL Life Technologies). The solution was drawn through a narrow gauge needle several times then purified as described in Section 2.5.3. 350 ng of total RNA was used in the reverse transcription reaction using 2.5 pmols primer BM *phy* RACE 1. RNA and primer were incubated at 70°C for 10 minutes then cooled before addition of other components. The reaction was performed in 1X PCR buffer (Gibco) with $MgCl_2$ and dNTP's (20 mM Tris-HCl pH8.4, 50 mM KCl, of 2.5 mM $MgCl_2$ and 400 µM each dNTP), incubated at 42°C for 1 minute, then 50 minutes at 42°C using 200 units of Superscript II reverse transcriptase. Products were purified using the QIAGEN PCR purification procedure. TdT tailing was performed using all purified reverse transcribed DNA in 1X Tailing buffer with dCTP (10 mM Tris-HCl pH8.4, 25 mM KCl, 1.5 mM $MgCl_2$, 200 µM dCTP). The reaction was incubated at 94°C for 3 minutes, then 37°C for 10 minutes with Terminal deoxynucleotidyl transferase (Gibco). A tenth of this reaction was used as a template for PCR, with *Taq* using primers 5' RACE Abridged anchor primer (Gibco) and BM *phy* RACE 2. 1 µl of a 1:10 dilution of this PCR product was used in a subsequent reaction using nested primers 5' RACE Universal amplification primer (Gibco) and BM *phy* RACE 3 with *Taq*. The 454 bp 5' RACE product was cloned into pCR 2.1 TOPO and sequenced.

2.28. Genomic sequencing of *Bm-phy-1*

Three clones were generated by PCR using primers BM *phy* Res F (*Bam* HI) and BM *phy* Res R (*Not* I) on *B. malayi* genomic DNA (made from 10 adult females received from Rick Maizels in ICAPB Edinburgh) with *Pfu* or *Taq* polymerases. These were made by cloning PCR products into pCRScript (then subcloning in to pAW1) or pCR2.1 TOPO. These clones were then sequenced with M13 reverse, pGEM forward, pAW1 Seq F, BM *phy* 1.1 IS 1F, 2F, 3F, 1R, 2R and 3R(B); and BM phy G seq 1F, 2F, 3F, 4F, 1R, 2R, 3R.

2.29. Construct for baculovirus expression of *Bm-phy-1*

Full length *Bm-phy-1* cDNA was cloned by PCR from *B. malayi* cDNA (cDNA from A. Page, WCMP, Glasgow) using *Pfu* polymerase and primers BM *phy-1* BV F and BM *phy-1* BV R. The product was cloned into pCRScript cloning vector then subcloned into vector pVL1392. Clone Bmphy-1, pVL1392 was sequenced with primers BM *phy* 1.1 IS 1F, 2F, 3F, 1R, 2R and 3R(B). Details of vector and expression system are given in Section 2.24. The Collagen Research Unit, Oulu, Finland, performed the expression of *Bm*-PHY-1 in insect cells

2.30. Glycosidase treatment of *Bm*-PHY-1

Carbohydrates were removed from protein samples by treatment with PNGase F (NEB). 1X denaturation buffer (NEB) was added and incubating at 100°C for 10 minutes then Buffer G7 (NEB) was added to 1X concentration and NP-40 (NEB) to 1%. 2 µl of enzyme was added per reaction. For each sample a minus enzyme control was included, with samples being otherwise treated identically. Samples were incubated at 37°C overnight, separated by SDS PAGE and Western blotted (performed as described in Section 2.9.3). Anti-*Bm*-PHY-1 antibody was a 1/4000 dilution of a mixture of rabbits 2624 and 2625 final bleeds in PBST.

10 X G7 buffer:	0.5 M Sodium phosphate (pH 7.5)
10 X Denaturing buffer:	5% SDS, 10% 2-mercaptoethanol

2.31. *C. elegans dpy-18* rescue experiments with *Bm-phy-1*

2.31.1. *Bm-phy-1* rescue constructs

Bm-phy-1 coding sequence was cloned from both cDNA (cDNA from A. Page, WCMP, Glasgow) and genomic DNA (made from 10 adult females received from Rick Maizels, ICAPB, Edinburgh) by PCR with the primers BM *phy* Res F (*Bam* HI) and R (*Not* I) using *Pfu* and *Pfu* Turbo polymerase respectively. Products were cloned into pCRScript then subcloned with *Bam* HI and *Not* I into the vector pAW1 (see Section 2.22.1). The clones were sequenced over the coding sequence insertion site with the primers pAW Seq F and M13 Reverse. A synthetic intron was inserted into clone BM phy-1cDNA, pAW1 using the process described (Section 2.22.3). The clone was digested with *Stu* I which cuts once within *Bm-phy-1* coding sequence and not at any other site in the plasmid. Linearised plasmid was ligated with the double-stranded synthetic *C. elegans* intron and the construct was sequenced to check for correct insertion of the intronic region.

2.31.2. Microinjection of *Bm-phy-1* rescue constructs

The *Bm-phy-1* cDNA rescue clone was injected into *dpy-18(e364)* at concentrations of 10 µg/ml and 100 µg/ml with the plasmid dpy-7-GFP at 5 µg/ml as a marker; injection mixes were made up to a final concentration of 150 µg/ml with pBluescript SKM. Transformants were selected by GFP fluorescence and over 5 semi-stable transmitting lines were examined for each concentration. The *Bm*-phy-*1* genomic rescue clone was injected into *dpy-18(e364)* at concentrations of 10 µg/ml, 30 µg/ml and 100 µg/ml as above. The construct Bmphy-1 SI, pAW1 was injected in to *dpy-18(e364)* at 10 µg/ml and 100 µg/ml as above and at least five lines scored for rescue for each set of conditions.

2.31.3. *Bm-phy-1* single worm RT-PCR from transformed lines

Single worm RT-PCR (see Section 2.22.6) was performed with primers BM phy Res F and BM phy-1 SWRT R on two different BM phy-1 SI, pAW1 lines (injected at 100 µg/ml) and on one *Bm*-phy-*1* genomic rescue line (injected at 10 µg/ml).

2.31.4. Western blotting for transgene expression

Details of buffers, antibodies, SDS PAGE and Western blotting can be found in Sections 2.9 and 2.25. Two different synthetic intron containing cDNA lines, (100 µg/ml) and one *Bm-phy-1* genomic rescue line (10 µg/ml) were tested by collecting GFP positive and negative worms. One synthetic intron containing cDNA (100 µg/ml) was tested by extracting protein from a mixed population of transformed and non-transformed worms from a whole plate and comparing to *dpy-18(e364)*. Protein were run on SDS PAGE gels, Western blotted and probed with anti-*Bm*-PHY-1 antibody.

2.32. *B. malayi* developmental timecourse RT-PCR

A panel of cDNAs made from daily extracts of infected jirds (up to day 14 post infection, after which 2-4 day extracts were taken) (Gregory *et al.*, 2000) (a gift from Bill Gregory, ICAPB, University of Edinburgh) was examined. PCR conditions used were; 1 µl of staged cDNA per 20 µl reaction and 30 cycles of PCR with *Taq* polymerase. Two pairs of primers were used simultaneously; BM phy 1.1 IS 1F and BM phy 1.1 2R for *Bm-phy-1*, and BM Tub A and BM Tub B specific to the *B. malayi* tubulin gene. Products were separated by electrophoresis, blotted, probed and excised radioactive bands quantified by scintillation counting as described previously (see Section 2.14.2.-2.14.5).

2.33. *Bm-phy-1* promoter cloning

2.33.1. TOPO Walking technique

A region of the *Bm-phy-1* promoter was cloned using a linker PCR technique. The TOPO walker (Invitrogen) technique uses the ligation activity of Topoisomerase I from Vaccinia virus to enable the amplification of unknown genomic sequences. Genomic DNA was digested with a range of restriction enzymes that leave 3' overhangs. This was then dephosphorylated and primer extension performed, using a primer from a region of known sequence, with *Taq* polymerase which adds 5' A-tails. This template is then combined with a 58 bp double-stranded section of DNA that has topoisomerase I covalently attached to a 3' T-overhang at one end. Dephosphorylated 3' A-overhangs provide the acceptor sites for the TOPO linker. Addition of this linker allows nested PCR to be performed using primers designed against sequence at the known end of the DNA, and primers designed to the TOPO linker sequences at the end of the DNA where

the sequence is not known. 1.5 µg of genomic DNA was digested per reaction; this was then dephosphorylated with calf intestinal alkaline phosphatase, extracted with phenol:chloroform and ethanol precipitated. Half of this was then used as a template for primer extension with BM *phy* GW 1R, an antisense primer designed against the 5' end of the coding sequence, and cycled once for; 94°C - 4 minutes, 56°C - 2 minutes, 72°C - 20 minutes. A tenth of this reaction was then used in the TOPO linking reaction, of which a third was used in a 50 µl PCR with primers LinkAmp 1 and BM *phy* GW 2R. 1 µl of a 1/10 dilution of this PCR was used as a template in a nested PCR with primers LinkAmp primer 2 and BM *phy* GW 3R. A 700 bp product from a *Pst* I digested product was cloned into pGEM, from which three clones were sequenced with pGEM forward and pGEM reverse primers. Further attempts to "walk" further 5' in this region were unsuccessful.

2.33.2. Isolation of BAC clones

Isolation of *B. malayi* BAC clones containing 5' genomic sequence from *Bm-phy-1* was done with the help of J. Daub in the laboratory of M. Blaxter (both ICAPB, University of Edinburgh). A BAC library was screened by a combination of two methods, probe hybridisation to a gridded BAC library filter and PCR.

2.33.2.1. Hybridisation screening of BAC library

A biotin labelled probe was generated with a biotin labelled T7 primer (NEB) along with a gene specific primer BM *phy* 1.1 IS 3R(B) using a plasmid clone containing 5' genomic sequences from *Bm-phy-1* and a T7 site as a template. The 1.7 kb PCR purified product was hybridised to BAC filters using the NEB NEBlot phototope and Phototope-star detection kit protocols. For prehybrisation and overnight hybridisation a temperature of 55°C was used. Hybridisation was followed by two washes in 2X SSC/0.1% SDS followed by two 55°C washes in 0.1% SSC/0.1% SDS. Filter was then blocked with blocking solution (5% SDS, 125 mM NaCl, 25 mM sodium phosphate, pH 7.2) for 10 minute at room temperature. 50 µl of Avidex (Streptavidin-Alkaline phosphatase conjugate) solution, diluted 1/2000 with block solution, was added for 5 minutes. Four washes were carried out in 10 mM Tris-HCl, 10 mM NaCl, 1 mM $MgCl_2$, pH 9.5. 20 mls CDP-Star reagent was diluted 1/250 with 1X CDP Star dilutent and the filter incubated at room temperature for 5 minutes with gentle shaking. The filter was

air-dried and sealed in hybridisation bags, exposed to X-ray film and the identity of clones giving a positive signal determined by using the co-ordinates of the positive signal on the grid to determine clone number.

2.33.2.2. PCR screening of BAC library

PCR screening was performed on pre-made BAC PCR pools. These were made by replicating 12 of the 364 well plate format that the library is in (full library is contained in 48 364 well plates therefore PCR pools only represent 25% of library) onto 48 96 well plates. From each of these plates a pool from each row was made, row pools were then combined to give a plate pool. PCR was performed using the primers BM phy BV F and BM phy GW 1R to give a product of approximately 100 bp from the 5' end of *Bm-phy-1* coding sequence. 20 µl PCR using 2 µl of pools as template were cycled in a Hybaid PCR Express as follows:-

94°C - 3 minutes} 1 cycle
94°C - 15 seconds, 55°C - 20 seconds, 72°C - 3 minutes} 35 cycles
72°C - 10 minutes} 1 cycle

Plate pools were screened to identify plates containing positive clones. From these positive plates the row pools were then screened. Six clones were identified by hybridisation that had a supporting PCR result (i.e. a positive result from a row pool PCR containing a clone identified by hybridisation). From three of these clones, 3a17, 4b8 and 4o8, BAC DNA was isolated by growing 400 ml cultures in 12.5 µg/ml chloramphenicol and using QIAprep spin columns (QIAGEN) to purify BAC DNA. BACs were then sequenced with primers BM phy GW 1R, BM phy GW 4R, BM phy BAC seq 1R, 2R and 3R.

2.34. Reporter gene expression in *C. elegans* from *Bm-phy-1* promoter

2.2 kb of *Bm-phy-1* 5' putative promoter region was amplified from *B. malayi* genomic DNA using *Pfu* polymerase and the primers BM phy-1 PF (*Sph* I) and BM phy-1 PR (*Bam* HI). This region represents position –2189 to +8, relative to the translational start site. The product was cloned into pCRScript, subcloned into the *C. elegans* reporter gene vector pPD96.04 with *Sph* I and *Bam* HI and the construct sequenced with primers

M13 Reverse and 96.04 reverse. The construct was then injected along with the marker plasmid p7616B into the *C. elegans* strain DR96, with both plasmids at 100 µg/ml. Six lines were examined for reporter gene expression by viewing GFP expression of live worm and using the sensitive X-gal staining procedure, described in Section 2.13.2, for β-galactosidase activity.

Table 2.2. Oligonucleotide primer sequences

Primer sequences are given in 5' to 3' orientation with engineered sequences given in lower case and restriction sites in lower case and underlined.

Oligonucleotide primer	Sequence (5'–3')
5' RACE Abridged anchor primer (Gibco kit)	ggccacgcgtcgactagtacGGGIIGGGIIGGGIIG (lower case= various restriction sites, I=deoxinosine)
5' RACE Universal amplification primer (Gibco kit)	cuacuacuacuaGGCCACGCGTCGACTAGTAC (u=dUMP, for UDG cloning which was not used here)
1224	CGCCAGGGTTTTCCCAGTCACGAC
96.04 reverse	TCTGAGCTCGGTACCCTCCAAGGG
ama-1 F	TTCCAAGCGCCGCTGCGCATTGTCTC
ama-1 R	CAGAATTTCCAGCACTCGAGGAGCG
BM N Phy Reverse	CCAATAGTATTTAAGCAC
BM *phy* 1.1 IS 1F	GCTTCTGGTGTTCAACCG
BM *phy* 1.1 IS 1R	GTTAACACAGGACACGCAG
BM *phy* 1.1 IS 2F	GGAAACAGCGGAGCATG
BM *phy* 1.1 IS 2R	GGTATGATGCTGTTTCAAG
BM *phy* 1.1 IS 3F	GGTGGTCATTATGATCC
BM *phy* 1.1 IS 3R(B)	GCGTGGATGATTTGGATC
BM phy BAC seq 1R	TCCAGGCACTTGACGATTG
BM phy BAC seq 2 R	GAACAATTTGAGGCTTATTG
BM phy BAC seq 3R	CGCTTTCTTCTAGCCACCATC
BM phy-1 BV F (*Not* I)	gagcggccgcATGATAGCTACCGTGGTGTTC
BM phy-1 BV R (*Xba* I)	gctctagaTTAAGCACTTAGATCGCCCAC

Oligonucleotide primer	Sequence (5'–3')
BM phy G seq 1 F	CCAGTCAGTGCTTATCTGC
BM phy G seq 1 R	GCAGAATGAATTGAATACTC
BM phy G seq 2 F	CACCCCAAATCTATCTAAG
BM phy G seq 2 R	GTTAGTTAGCTTTTGACAACAG
BM phy G seq 3 F	GTAAGGATTTTCGAAAGCGA
BM phy G seq 3 R	GATCAAACTCCTAGACCC
BM phy G seq 4 F	CTGACTTTTAGACTAATAATC
BM phy GW 1 R	GCTTCCAGAAGTAACTCC
BM phy GW 2 R	GTGTAGTACTCCGCAGTGC
BM phy GW 3 R	GAACACCACGGTAGCTATC
BM phy-1 PF (*Sph* I)	ggcgcatgcGAATGAGACAATTGCACAAG
BM phy-1 PR (*Bam* HI)	ggcggatccGCTATCATCACTGGCTCTGGA
BM phy RACE 1	CATCAGCTATCTCCTTAAC
BM phy RACE 2	GTAAGGCTATTGCAGCTC
BM phy RACE 3	CCAAATCTTCCGTTGTAG
BM phy Res F (*Bam* HI)	gcggatccgATGATAGCTACCGTGGTGTTC
BM phy Res R (*Not* I)	gagcggccgcTTAAGCACTTAGATCGCCCAC
BM phy-1 SW RT R	GATCGGCTTAAGCCACAAGGCCTTCG
BMTub A	AATATGTGCCACGAGCAGTC
BMTub B	CGGATACTCCTCACGAATTT
BV F	ATCACAAACTGGAAATGTCTATC
BV R	GGATTTCCTTGAAGAGAGTGAG
CB pdi-2 F (*Kpn* I)	cgggtaccGTCATTTGAAGAAGAAGAGA
CB pdi-2 R (*Sac* I)	CGGAGCTCAACTTAGACGAGACTTTCTC
CB phy-1 F (*Bam* HI)	cgggatccTCTACGCTCCAATTAAAGTC
CB phy-1 R (*Sac* I)	gcgagctcTGTTGACTCTCTCGACGACT
CB phy-2 F (*Kpn* I)	gcggtaccAGCAGAAAAAGATGTGACGAC
CB phy-2 R (*Sac* I)	cggagctcCAGTGATATTCTTGAGGAAAG
hP4H αI F (*Bam* HI)	gcggatccgATGATCTGGTATATATTAATTATAGG
hP4H αI R (*Not* I)	gcgcggccgcTCATTCCAATTCTGACAACG
hP4H αII IS 1F	CAACCAAGTCACAGGTGCTG

Oligonucleotide primer	Sequence (5'–3')
hP4HαI SW RT F	GGCAGAAGAGGACAAGTTAG
hP4HαI SW RT R	TTGGCTGCCCCAACCTGATCTTCATC
hP4HαII F (*Bam* HI)	gcggatccgATGAAACTCTGGGTGTCTGC
hP4HαII R (*Not* I)	gcgcggccgcTCAGTCAACTTCTGTTGATC
hP4HαII SW RT R	CAGGCAGGTAGTCCACAGGCCTCTC
IVS sense	GTAAGTTTAAACTATTCGTTACTAACTAACTTTA AACATTTAAATTTTCAG
IVS antisense	CTGAAAATTTAAATGTTTAAAGTTAGTTAGTAA CGAATAGTTTAAACTTAC
L4440 F	GAGTGAGCTGATACCGC
L4440 R	GTGCTGCAAGGCGATTAAG
M13 reverse	AACAGCTATGACCATGATTA
NLS	CACCCACCGGTACCTTACGC
pAW Seq F	CATCTTGCGGGTCTTGCTCAG
pdi-2 F (pMal) (*Xmn* I)	ggcgaaggatttcGCCGTCATTGAAGAAGAAGAG
pdi-2 R (pMal) (*Pst* I)	ggcctgcagTTAGAGCTCGGTGTGTCCCT
pdi-2 HSC F (*Kpn* I)	gccggtaccATGTTCCGGCTCGTCGGTCTG
pdi-2 HSC R (*Kpn* I)	gccggtaccTTAGAGCTCGGTGTGTGTCCCTC
pdi-2 IS 1 F	CTCAAGCTTTTCCGCAAC
pdi-2 IS 1 R	CGAGCTTGTCCCAGGTTGG
pdi-2 IS 2 F	GTAAAGTCCGAGATTGAG
pdi-2 IS 2 R	CGAAGTCTGGCTTGAACTTG
pdi-2 P F (*Sph* I)	GATGGAGAGCATGCATGTTTTG
pdi-2 P R (*Bam* HI)	cgcgggatccAACATCACGATGAATAGCGAATGG
pdi-2 Resc R	CACTGCTCAATCGGATTCG
pGEM forward	GTTTCCCAGTCACGACGTTG
pGEM reverse	CAGGAAACAGCTATGACCA
phy-1 (*Not* I) F	gcggcggccgcTTGGCTCTCCTAAGTTTCAGC
phy-1 (*Sal* I) R	gcgtcgacGGCTTGCAGCCATCACTTCACAGG
phy-1 3'UTR F (*Sac* I)	gcggagctcCTCTAAGCATTGGTTTTCATTG
phy-1 3'UTR R (*Sac* I)	gcggagctcACTAGGGAATTGTCGGCTGC

Oligonucleotide primer	Sequence (5'–3')
phy-1 F (pMal) (*Eco* RI)	ggcgaattcGATCTGTTCACCTCGATTGC
phy-1 R (pMal) (*Pst* I)	ggcctgcagTTAGAGGGTCTCCCAGACGT
phy-1 HSC F (*Eco* RV)	gcggatatcATGCGCCTGGCACTCCTTGTAC
phy-1 HSC R (*Eco* RV)	gcggatatcTTAGAGGGTCTCCCAGACGTC
phy-1 IS 1 F	CACAGATGACAGTTATGG
phy-1 IS 1 R	CTTGGCCTCAGTGAAGACG
phy-1 IS 2 F	GTGGAGGTGGAAGATATTC
phy-1 IS 2 R	GGCTTCGCGAGCTCCTG
phy-1 IS 3 F	GCCGTCCCGACTCTGTCC
phy-1 IS 3 R	GTCCTCGTACCATTTGAC
phy-1 IS 4 F	CTCTAAAAGCGCGTGGC
phy-1 IS 4 R	CCATCGGCAAGATCCTTG
phy-1 IS 5 F	CTGCCCACCAAGAATGATG
phy-1 IS 5 R	GAGTACTCCTCCGACAGC
phy-1 (new) GS F	GGCTGCTCCCAACTATTG
phy-1 P F (*Pst* I)	gcgctgcagGGTCTGCTGGCCGTTTCGTCAG
phy-1 P R (*Bam* HI)	gcaggatccCGCATTCTGAAAAATTGAGAG
phy-1 PS 1 F	CTATGTGTAATGCGAAAC
phy-1 PS 2 F	CGGTTGATGCTTCTAAAACG
phy-2 BV F (*Not I*)	gacgcggccgcATGAGAGCAGTTTTGCTAGTC
phy-2 BV R (new) (*Xba I*)	gcgtctagaCTATGGATCATTGGCATATGGGGAC
phy-2 F (pMal) (*Xba I*)	ggctctagaGATTTGTTCACTGCAATTGC
phy-2 R (pMal) (*Pst* I)	ggcctgcagCTATGGATCATTGGCATATG
phy-2 FL F (*Xba I*)	gctctagagATGAGAGCAGTTTTGCTAGTC
phy-2 FL R (*Sac II*)	gaccgcggCTATGGATCATTGGCATATG
phy-2 HSC F (*Eco* RV)	gcggatatcATGAGAGCAGTTTTGCTAGTC
phy-2 HSC R (*Eco* RV)	gcggatatcCTATGGATCATTGGCATATGG
phy-2 IS 1 F	CACTGATAACCGAGTTCG
phy-2 IS 1 R	CTCCGCGTTCTGGTTGACTC
phy-2 IS 2 F	CCAACAATTGAGGAATGG
phy-2 IS 2 R	GCTCGCTTCAGCTTTGG

Oligonucleotide primer	Sequence (5'–3')
phy-2 P F 2 (*Pst* I)	ggcgctgcagAGACTATAGTCTATAGCTGAAAACG
phy-2 P R 2 (*Bam* HI)	gcgggatccACTGCTCTCATTCTGAAAGACAAATC
phy-4 F1	ATGAACTATTTTCAAACCCTC
phy-4 R1	CTACGGAAGACACGTCGGCATG
phy-4 F2	GCATCCGGTACGTTCAACAG
phy-4 R2	CGTGGCAATGATCTTCTCTCCAG
phy-4 R3	CTGGCAGCAACTGAGTAG
phy-5 P F (*Hind* III)	GCCCAGGCCCAAAGCTTCTAGCGC
phy-5 P R (*Xba* I)	gcctctagaCCACGGAGCTCCTTCCCACACTTGTC
phy-5cDNA F1 (*Nde* I)	gcgcatatgTTCAATTTTCTCACACCGTTCACCG
phy-5cDNA R1 (*Bam* HI)	gcggatccTTAAAATTCTCCCGGAAAAAGTAGAG
phy-5cDNA F2	TCCGAGCCCTTCATCCTACAG
phy-5cDNA R2	GGCTGATCCCTCATGCGGAGC
phy-5cDNA R3	TCGTGGACCCATAAAGTCGCACC
SL1	GGTTTAATTACCCAAGTTTGAG
SL2	GGTTTTAACCCAGTTACTCAAG
T3	AATTAACCCTCACTAAAGGG
T7	GTAATACGACTCACTATAGGGC
T20B3.7 cF (*Xba* I)	gcgtctagaATGATGGATTCCATCTGCATC
T20B3.7 cSF (*Xba* I)	gctctagaATGCTCCCGGTCGACATGG
T20B3.7 cLF (*Xba* I)	gctctagaATGATTTCTGTCACTTTCCG
T20B3.7 NR (*Xho* I)	gcgctcgagCGGTATAAGCCACGAAGCATG
X8G5 F1	CAGTCGCTCAACACCGG
X8G5 R1	GCCAAACACATTTTGCG
X8G5 F2	CTGTTACTGCACGTTGG
X8G5 R2	GTACCCTTCGGTTTATG

Chapter 3

The combined functions of *C. elegans phy-1* and *phy-2* are essential for development and morphology due to their cuticle collagen modifying activity

3.1. Introduction

The cuticular collagens of the nematode *C. elegans* are hydroxylated, with approximately equal levels of hydroxyproline (12%) and proline (11%) being found (Cox *et al.*, 1981a), indicating the presence of P4H activity in these nematodes. One P4H α subunit-encoding gene from *C. elegans* had been previously cloned and biochemically characterised (Veijola *et al.*, 1994). This gene, *Ce-phy-1*, was subsequently sequenced by the Genome sequencing consortium and assigned to cosmid/YAC T28D6.1/Y47D3B.10 on chromosome III. The cosmid sequence differed slightly from the published sequence and it is the consortium sequence referred to in this work. When co-expressed in an insect cell expression system, *Ce*-PHY-1 formed an active complex with the human PDI subunit (Veijola *et al.*, 1994). This complex differed from enzymes described previously as it formed a dimer. The catalytic properties of this hybrid dimer were however, similar to those of the human tetramer. Subsequent cloning of the authentic *C. elegans* PDI partner demonstrated that this also formed a dimer with *Ce*-PHY-1 in an insect cell expression system (Veijola *et al.*, 1996a). Unusually, *Ce*-PHY-1 formed a less stable association with the *C. elegans* PDI, *Ce*-PDI-2, than with human PDI, and the complex formed was much less active that the hybrid. As the *C. elegans* PDI subunit was able to form an active tetramer with human α subunits (Veijola *et al.*, 1996a), dimer formation was evidently a property of *Ce*-PHY-1 and not *Ce*-PDI-2. *Ce*-PHY-1 contains a 32 amino acid C-terminal extension which is absent in the human α subunits. The function of this extension was investigated to examine if it was the determining factor in dimer or tetramer formation. Deletion of this region removed the ability of this subunit to form either a dimer or tetramer (Veijola *et al.*, 1996a). Therefore, although sequences critical for dimer formation are present in this extension its removal was not sufficient to produce tetramers.

Sequencing of the *C. elegans* genome revealed the presence of another conserved putative P4H α subunit encoding gene. This gene is encoded by cosmid F35G2.4 from chromosome IV and was named *Ce-phy-2*. The aim of this chapter was to define the

temporal and tissue specific expression of *Ce-phy-1* and *Ce-phy-2* and assess their role in the biogenesis of *C. elegans* ECMs. The tissue-specific expression of both genes was examined by transgenic expression of reporter genes from *Ce-phy* gene promoters. Expression of gene transcript levels was examined at time points throughout post-embryonic development by a semi-quantitative RT-PCR approach. The effect of disruption/removal of gene function, using RNAi/mutant analysis, was determined on nematode development and on the expression and localisation of specific cuticle collagens. The genetic mutant strain *dpy-18* was identified as a *Ce-phy-1* mutant and the potential of inter-species phenotypic repair assessed by expression of human P4H α subunits in this genetic background.

3.2. Results

3.2.1. P4H subunit-encoding gene identification

A P4H α subunit-like encoding gene from *C. elegans, Ce-phy-1*, had been cloned and described as an active P4H which formed dimers with partner PDIs (Veijola *et al.*, 1994). By searching the *C. elegans* database a second gene, *Ce-phy-2*, was identified which was predicted to encode a protein 57% identical to *Ce*-PHY-1. Both proteins contained the conserved active site histidine, aspartic acid and lysine residues (Lamberg *et al.*, 1995; Myllyharju and Kivirikko, 1997) and intra-molecular disulphide bonding cysteine residues (John and Bulleid, 1994; Lamberg *et al.*, 1995) (amino acid sequence and conserved residues are shown in Figure 6.2 in chapter 6). Both genes were represented by a number of expressed sequence tag (EST) sequences from embryonic and post-embryonic stages (data accessible through WormBase and Intronerator databases, addresses in Section 2.8.1), although no evidence of alternative splicing was found from this data for either *Ce-phy-1* or *-2*. Additionally no indication of alternative splicing of coding sequence was found during the sequencing of these genes in the course of the study described here. This is in contrast to the human α subunit-encoding genes which are both alternatively spliced (Helaakoski *et al.*, 1994; Nokelainen *et al.*, 2001).

3.2.2. *Trans*-splicing analysis by SL RT-PCR

Trans-splicing is the addition of a short leader sequence, the spliced leader (SL), which is not associated with the gene itself, onto the 5' end of the mRNA. An estimated 70% of *C. elegans* genes are SL *trans*-spliced by the two major forms of SL, SL1 and 2

(Zorio *et al.*, 1994). The SL2 sequence is normally specific to the downstream gene(s) of an operon and is found on approximately 15% of genes (Blumenthal *et al.*, 2002; Zorio *et al.*, 1994). Although the genomic organisation of these genes did not suggest organisation of either in an operon, the *trans*-splicing pattern was analysed by SL PCR. Both genes were analysed with a control pair of gene-specific primers, Figure 3.1, lanes 1 and 4, and the gene-specific antisense primer for each transcript in combination with sense primers designed against the consensus sequences of SL1 and SL2 *trans*-splice leaders, lanes 2, 3, 5 and 6. *Ce-phy-1* and -2 primers were positive for the controls and the SL1 specific primer sets and negative for SL2. A non-coding first exon for *Ce-phy-1* had already been described which was not *trans*-spliced (Veijola *et al.*, 1994). An SL1 positive result from mixed-stage cDNA for *Ce-phy-1* shown here demonstrated that *trans*-splicing of this gene also occurred and suggested that either the original clones examined were chimeras or that stage-specific regulation of spicing may be occurring. The *Ce-phy-2* gene was also *trans*-spliced by the SL1 leader sequence. The similarity in size of both SL1 products to their gene specific products indicated that the predicted 5' ends of both genes was correct.

3.2.3. Determination of spatial expression patterns

3.2.3.1. Reporter gene vectors

To determine the tissue-specific expression of *Ce-phy-1* and *Ce-phy-2* an approach using transgenic reporter genes was taken. Other techniques are available in *C. elegans* for the study of gene expression patterns. Immuno-cytochemistry reveals protein localisation patterns, while *in situ* hybridisations shows RNA distribution, however these were not used as the former requires a specific antibody (which was not available at that time), and *in situ* hybridisations is a technically difficult approach with variable success in this organism. We chose the examination of gene expression in *C. elegans* using the *cis*-acting sequences from the genes under examination to direct expression of a reporter gene. The *E. coli lacZ* gene encodes the enzyme β-galactosidase (βGal). This protein has been used as a reporter molecule in various organisms and can be expressed in many *C. elegans* tissues.

A specific series of vectors has been generated and distributed by the laboratory of A. Fire for this purpose (Fire *et al.*, 1990). A diagram of vectors from this series used in this study is shown in Figure 3.2. The backbone for these vectors was originally derived

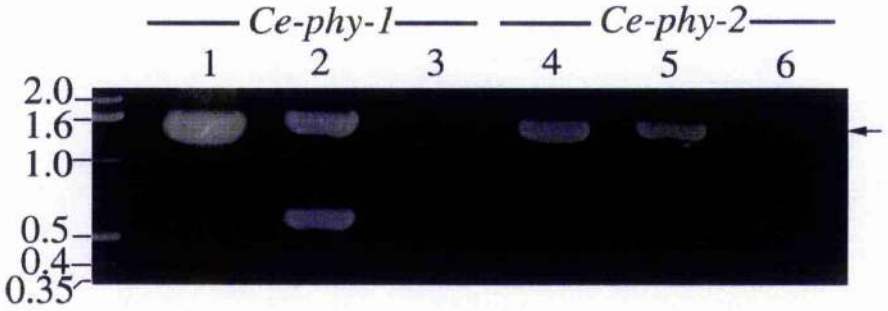


Figure 3.1. *Ce-phy-1* and *Ce-phy-2* SL RT-PCR

Agarose gel of products resulting from SL RT-PCR of *C. elegans phy* genes *-1* and *-2*. Lanes 1-3 are an examination of *Ce-phy-1*. Lane 1 shows the products amplified using gene specific primers for *Ce-phy-1*, lane 2 shows the product amplified using SL1 specific sense primer and an antisense *Ce-phy-1* gene primer, lane 3 is from PCR with SL2 specific and an antisense *Ce-phy-1* gene primer. Lanes 4-6 are an examination of SL splicing for *Ce-phy-2* with gene specific primer combination (lane 4) followed by an SL1 gene antisense primer combination (lane 5) and a SL2 gene antisense primer combination (lane 6). The arrow indicates the position of the SL1 product for both genes. The positions of molecular size standards are indicated with sizes given in kilobases.

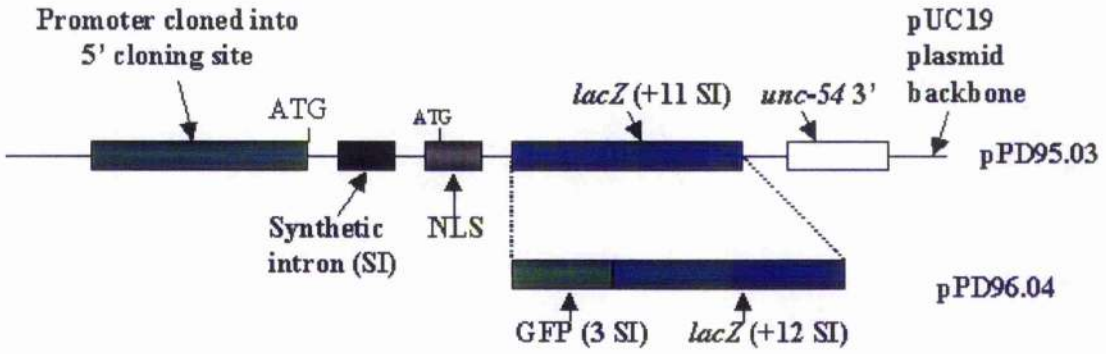


Figure 3.2. Reporter gene vectors

Diagram of the reporter gene vectors used. Promoter sequences are inserted into the 5' multiple cloning site upstream of multi-intron containing reporter genes *lacZ* and GFP. Reporter genes were either a multi-intron containing *lacZ* (pPD95.03, used in Chapters 3 and 4) or a multi-intron containing *lacZ/gfp* fusion (pPD96.04 used in Chapters 5 and 6). The positions of an additional synthetic intron and nuclear localisation signal (NLS) are indicated. The 3' region from the *C. elegans* gene *unc-54* was included in the vectors used in this study.

from pUC19 with the *lacZα* region deleted to avoid duplication from insertion of the complete copy of the *lacZ* gene. The last 51 bp of *lacZ* have also been replaced by a synthetic oligonucleotide to allow the addition of a 3' cloning site just after translational termination of *lacZ*. Additionally this sequence has been modified to introduce silent changes to eliminate codons used rarely in *C. elegans*. Into this 3' multiple cloning site 3' flanking sequences were inserted, which for the vectors used in this study came from the *C. elegans unc-54* gene. This sequence is thought to be capable of functioning in all somatic tissues. If the potential of a particular 3' region to affect gene expression was to be investigated, the 3' region present in the vector can be replaced with this potential control region. A 45 bp ATG and nuclear localisation signal (NLS) containing cassette is located upstream from the *lacZ* coding sequence. The NLS, from the SV40 T antigen, is an eight amino acid sequence that can target a normally cytoplasmic protein, such as βGal, to the nucleus when placed at either end of the protein or at a variety of internal sites. The nuclear localisation of the reporter gene aids in the identification of cells expressing reporter proteins. For transcriptional fusion constructs, translational initiation occurs at the ATG provided by this cassette with the NLS therefore being at the N-terminus of βGal. In this study translational fusions were constructed with the ATG and first few amino acids of the protein included with the promoter, therefore translation is initiated at the gene ATG. This positions the NLS at the junction between the fused proteins. A 42 bp synthetic intron segment is inserted upstream of the NLS and after the 5' cloning site, as expression has been demonstrated to be more efficient from spliced than unspliced transcripts. The vectors used in these studies were derivatives of the original vectors with multiple introns inserted into the *lacZ* coding sequence. Decoy sequences are inserted upstream of the 5' multiple cloning site of the vectors used. These consist of a short intron followed by an open reading frame that terminates just before the multiple cloning site. Inclusion of these sequences reduces the background staining in these more sensitive vectors. The vector used in this chapter was pPD95.03, which consisted of a 5' decoy sequence, upstream synthetic intron, NLS, *lacZ* coding sequence containing eleven additional synthetic introns, and the 3' region from *C. elegans unc-54* gene.

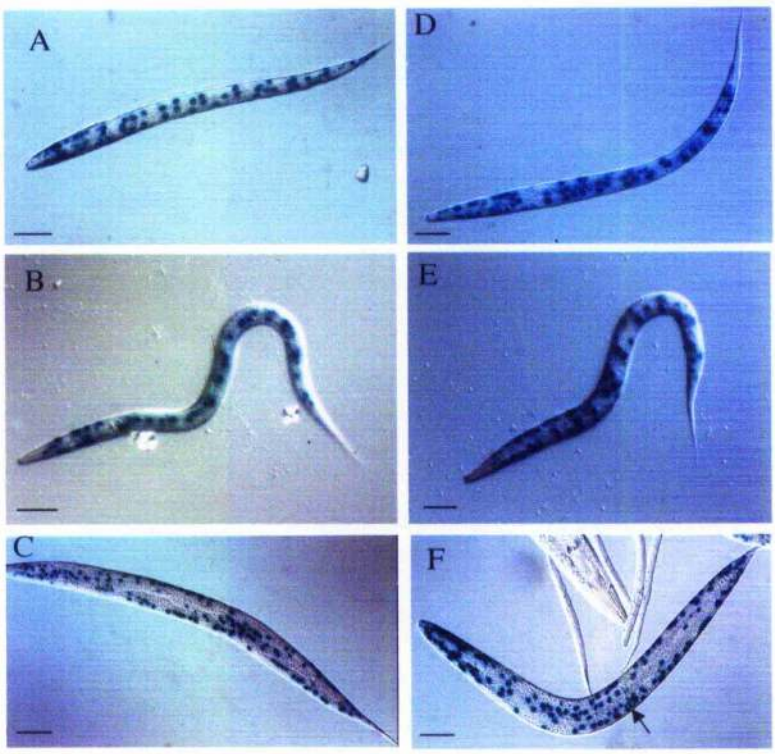
3.2.3.2 *Ce-phy-1 and Ce- phy-2 lacZ* reporter gene constructs

For *Ce-phy-1* and *Ce-phy-2*, putative promoter regions were amplified from N2 genomic DNA. Approximately 2.8 kb of *Ce-phy-1* upstream sequence, up to and

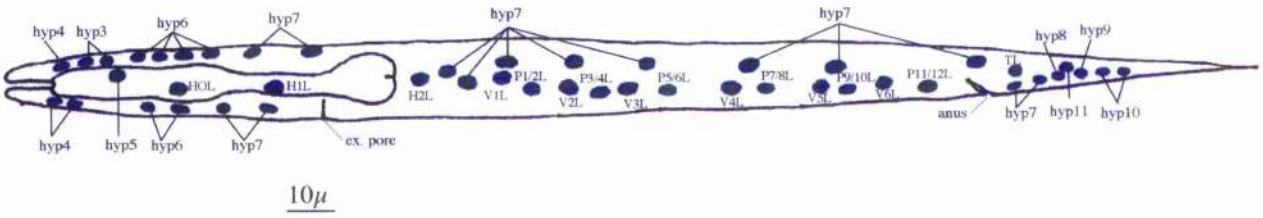
including the ATG and two bases of the second codon, was fused in-frame with the *lacZ* gene in the vector pPD95.03. 1.7 kb of *Ce-phy-2* sequence was used, incorporating the first eleven base pairs of coding sequence. Background expression using reporter gene vectors with increased sensitivity with no promoter is possible although non-specific expression is generally limited to gut and pharyngeal cell types. Constructs were transformed into the *C. elegans* germ line via microinjection and transformants selected by the expression of phenotypic markers, *rol-6* or *unc-76* rescue, then fixed and stained for βGal activity. Individual nematodes do however display mosaicism, and therefore at least three independent lines were selected for each construct and marker combination. Many individual nematodes encompassing all the different life cycle stages were examined to establish which individual nuclei were reproducibly expressing the reporter gene constructs. Both hypodermally and non-hypodermally expressed transformation markers were employed, thereby excluding the possibility of reporter expression being driven by the transcriptional regulatory units of the marker plasmid. No differences were noted in the expression patterns between transformants generated using either marker, and no significant differences were observed between the independent lines generated in this study.

For both genes, all stages from embryo to adult consistently expressed the reporter gene constructs in at least the hypodermal cell nuclei. Expression of both genes was examined in detail in the first larval stage, as the position of the hypodermal cell nuclei can be most accurately determined in this stage. For *Ce-phy-1* and *Ce-phy-2*, expression was detected in the L1 hypodermal cells; including the anterior H0L, H1L, hyp3, 4, 5, 6 and 7; posterior TL, hyp7, 8, 9, 10, 11; and the mid body hyp7 and lateral P, V, H2R and H2L cells (Figure 3.3, panels A, B, D, and E). Mosaicism in expression was evident, especially in the posterior hyp and anterior hyp3 and hyp4 cells. In accordance with the increase in numbers of hypodermal cells in the late larval and adult stages, the expression pattern becomes increasingly more complex (Figure 3.3. panels C and F). For both genes most of the identifiable stained nuclei are of hypodermal origin, additional nuclei are however apparent, particularly for *Ce-phy-2*. The hypodermal pattern for *Ce-phy-1* and *Ce-phy-2* includes the vulval cell nuclei (indicated in Figure 3.3. panel F). In addition, *Ce-phy-2* expression was occasionally detected in the body wall muscle cells, which become increasing evident when sensitive staining methods were applied (data not shown). Expression predominately in the hypodermal cells was indicative of a function for both

Ce-phy-1::*lacZ* *Ce-phy-2*::*lacZ*



G



10μ

Figure 3.3. Tissue specific expression of *Ce-phy-1* and *Ce-phy-2*

lacZ expression from *Ce-phy* gene promoters. Panels A-C show results for *Ce-phy-1*. Panel A, *lacZ* expression in an L1 co-expressing the *rol-6* marker plasmid. Panel B, *lacZ* expression in an L1 co-expressing the *unc-76* marker plasmid. Panel C, *lacZ* expression in an adult co-expressing the *rol-6* marker. Panels D-F show results for *Ce-phy-2*. Panel D *lacZ* expression in an L1 co-expressing the *rol-6* marker plasmid. Panel E *lacZ* expression in an L1 co-expressing the *unc-76* marker plasmid. (F) *lacZ* expression in an adult co-expressing the *rol-6* marker, vulval cell staining is indicated by an arrow. Bar on A, B, D and E equals 10 μm, bar for C and F equals 100 μm. (G) Diagrammatic representation of hypodermal cell nuclei in an L1 larvae.

these genes in modification of the nematode cuticular ECM. The nematode epidermis is known as the hypodermis and is the site of synthesis, modification and secretion of the components of the cuticle. Similar expression patterns to those found have been described for *C. elegans* collagens (Gilleard *et al.*, 1997; Nyström *et al.*, 2002) and collagen modifying enzymes (Thacker *et al.*, 1995). The weaker expression of *Ce-phy-2* in the body wall muscles raised the possibility that this gene could play a role in modification of the basement membrane collagens surrounding and secreted by this tissue.

3.2.4. Temporal expression of *Ce-phy-1* and *-2*

The temporal expression pattern of the transcripts from *Ce-phy-1* and *-2* was examined by applying a semi-quantitative reverse transcriptase PCR (sq-RT-PCR) approach (Johnstone and Barry, 1996). This permitted the abundance of the individual genes to be quantified via mRNA isolated from synchronised populations of *C. elegans* sampled at two hourly intervals throughout post-embryonic development. Synchronous post embryonic cultures were prepared by collecting embryos from gravid adults, these were then hatched in the absence of food which causes developmental arrest in the early L1 stage. Synchronised populations were then used to prepare RNA samples (Johnstone and Barry, 1996). RT-PCR was then performed simultaneously with two sets of primers, both of which span introns to enable signals from genomic DNA and cDNA to be distinguished. RT-PCR reactions separated on agarose gels for *Ce-phy-1* and *Ce-phy-2* are shown in Figure 3.4. A and B respectively. These gels were blotted and the radioactive bands quantified. The abundance of the test transcript was then expressed as the ratio of the amount of its amplified product to control transcript product. The control gene used was the constitutively expressed gene *ama-1* (the RNA polymerase II large subunit gene) (Bird and Riddle, 1989).

In agreement with the reporter gene assays, which indicated expression in all developmental stages, RT-PCR analysis showed the presence of transcripts from both genes at all the larval and adult time points examined. Both genes presented very similar transcript profiles (Figure 3.5), displaying an overall increase throughout larval development, with distinct peaks of abundance corresponding to the mid-larval stages. Expression is highest in the L4 larvae and is followed by a dramatic drop in the adult stage. Comparable oscillating expression patterns have been described for a number of individual cuticle collagen genes (Johnstone and Barry, 1996; McMahon *et al.*, 2003) and the genes

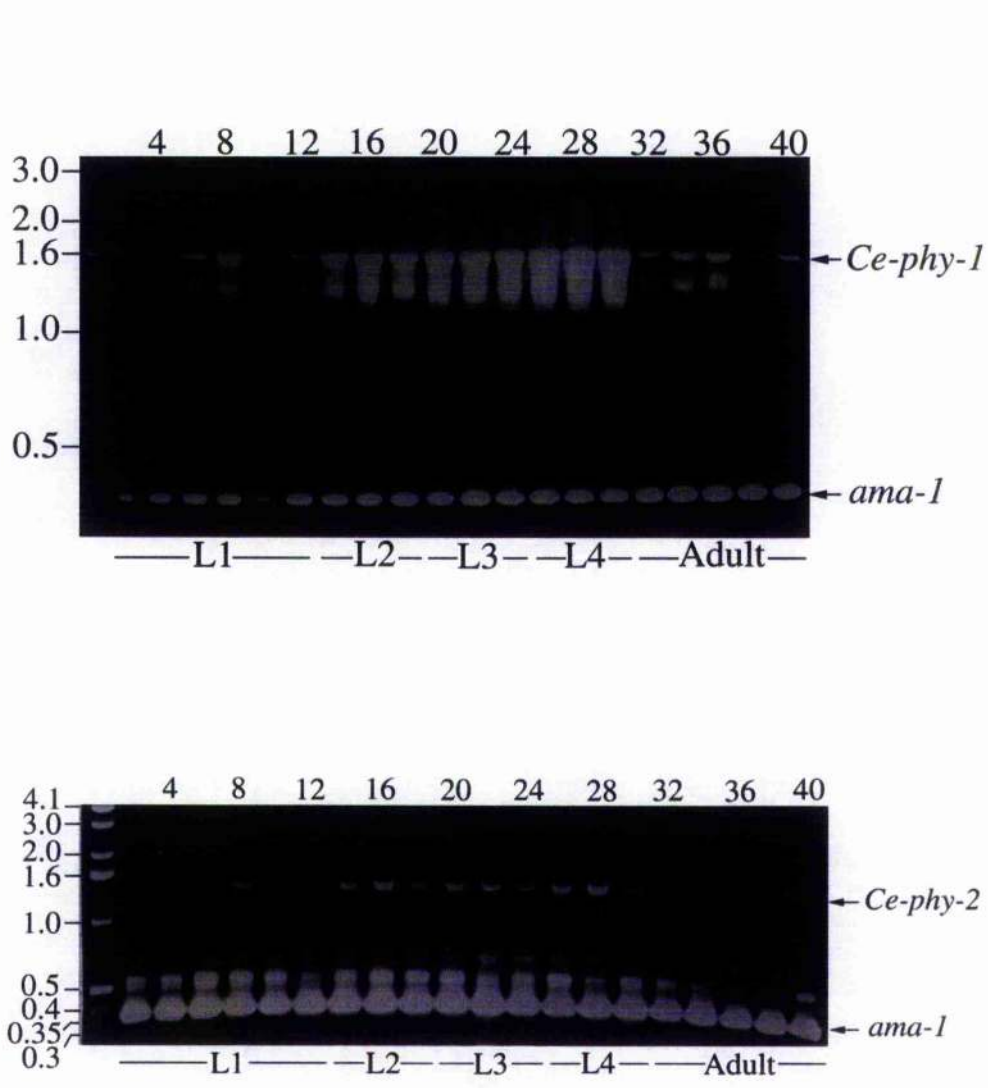


Figure 3.4. Timecourse RT-PCR of *Ce-phy-1* and *Ce-phy-2*

(A) Agarose gel showing products from RT-PCR of *Ce-phy-1* from staged cDNA with simultaneous amplification of the control gene *ama-1*. (B) Agarose gel showing products from RT-PCR of *Ce-phy-2* with *ama-1*. The positions of molecular size standards are indicated with sizes given in kilobases. Along the top of the gel picture the time in hours is given which represents hours post-hatch. The developmental stages from L1 to adult that the samples represent are indicated below the gel.

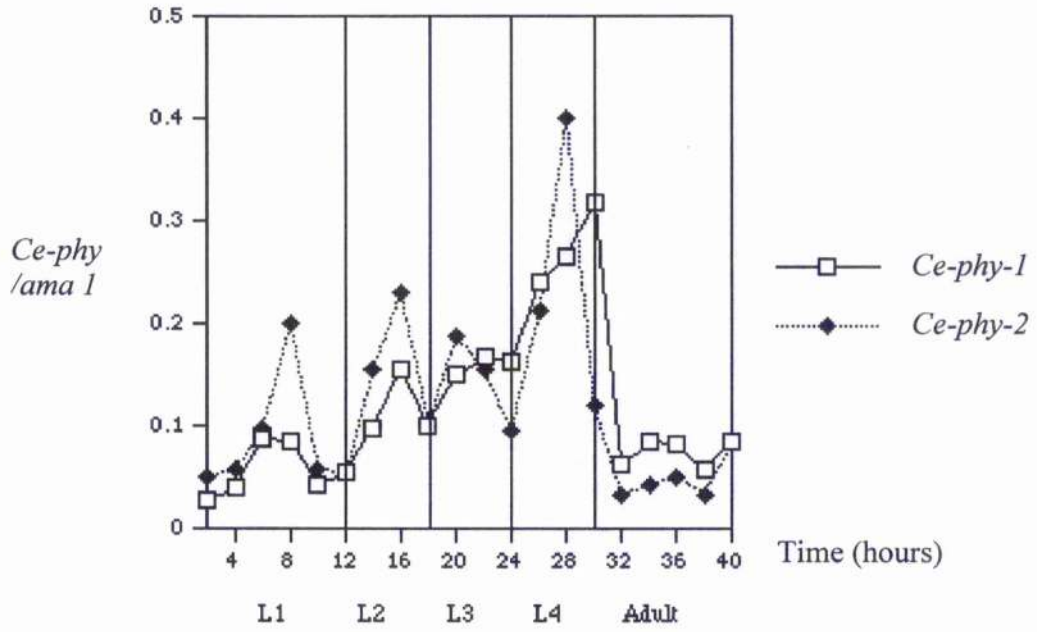


Figure 3.5. Temporal expression of *Ce-phy-1* and *Ce-phy-2*

Graph showing expression levels of transcripts, in relation to *ama-1*, throughout post-embryonic development. Levels are expressed in arbritary units along the y-axis. The x-axis shows the time in hours post-hatch. The vertical bars represent the timing of the moults with the corresponding developmental stage indicated. A single peak of expression is found for each gene within each larval stage.

encoding two potential collagen-folding enzymes, CYP-9 and PDI-1 (Page, 1997). The four larval stages are characterised by the shedding and re-synthesis of the cuticle, a structure in which more than 80% of the proteins are collagenous (Cox *et al.*, 1981a). As the exoskeleton progressively increases in size, greater pulses of collagen folding enzymes will be required to assemble this complex extracellular matrix. The high degree of similarity between the transcript abundance profiles for these two *Ce-phy* genes indicated that they may have shared or common roles, a point supported by their almost identical spatial expression pattern data. These inter-moult waves of expression in the cuticle synthesising hypodermal cells shows these genes are strongly expressed in the appropriate tissue and with the correct developmental timing for the encoded enzymes to modify the collagens of the nematode cuticle. A role in modification of the type IV basement membrane collagens possibly of the body wall muscles could not however be excluded from these results.

3.2.5. Embryonic expression

A combination of spatial and temporal expression analysis was performed to determine expression of *Ce-phy-1* and *Ce-phy-2* in the developing embryo. Expression of both genes was examined by RT-PCR, PCRs from embryo cDNA are shown in Figure 3.6A, lanes 1 and 7. The remaining lanes are pools of the samples from time points within each stage (lane 2 and 8 represent L1, lanes 3 and 9 L2, lanes 4 and 10 L3, lanes 5 and 11 L4, and lanes 6 and 12 adult) are shown for comparison. Expression of *Ce-phy-1* in the embryo is just detectable in lane 1; for *Ce-phy-2* no band is distinguishable from the gel. Expression of both genes in these samples was confirmed on quantification of these bands (data not shown), and from reporter gene data showing expression of *Ce-phy-1*, Figure 3.6B, and *Ce-phy-2*, Figure 3.6C, in embryos.

3.2.6. Interference of gene function by injection of double-stranded RNA

RNA-mediated interference or RNAi is a technique that uses exogenous RNA to interfere with the function of an endogenous gene. The use of double-stranded (ds) RNA in this technique was first described in *C. elegans* by A. Fire and co-workers with phenocopies of null gene phenotypes produced by introducing dsRNA corresponding to a target gene by microinjection (Fire *et al.*, 1998). Comparison of RNAi induced phenotypes with the characterised mutant phenotypes of a number of genes validated this procedure as identical phenotypes were produced. The RNAi effect is not normally

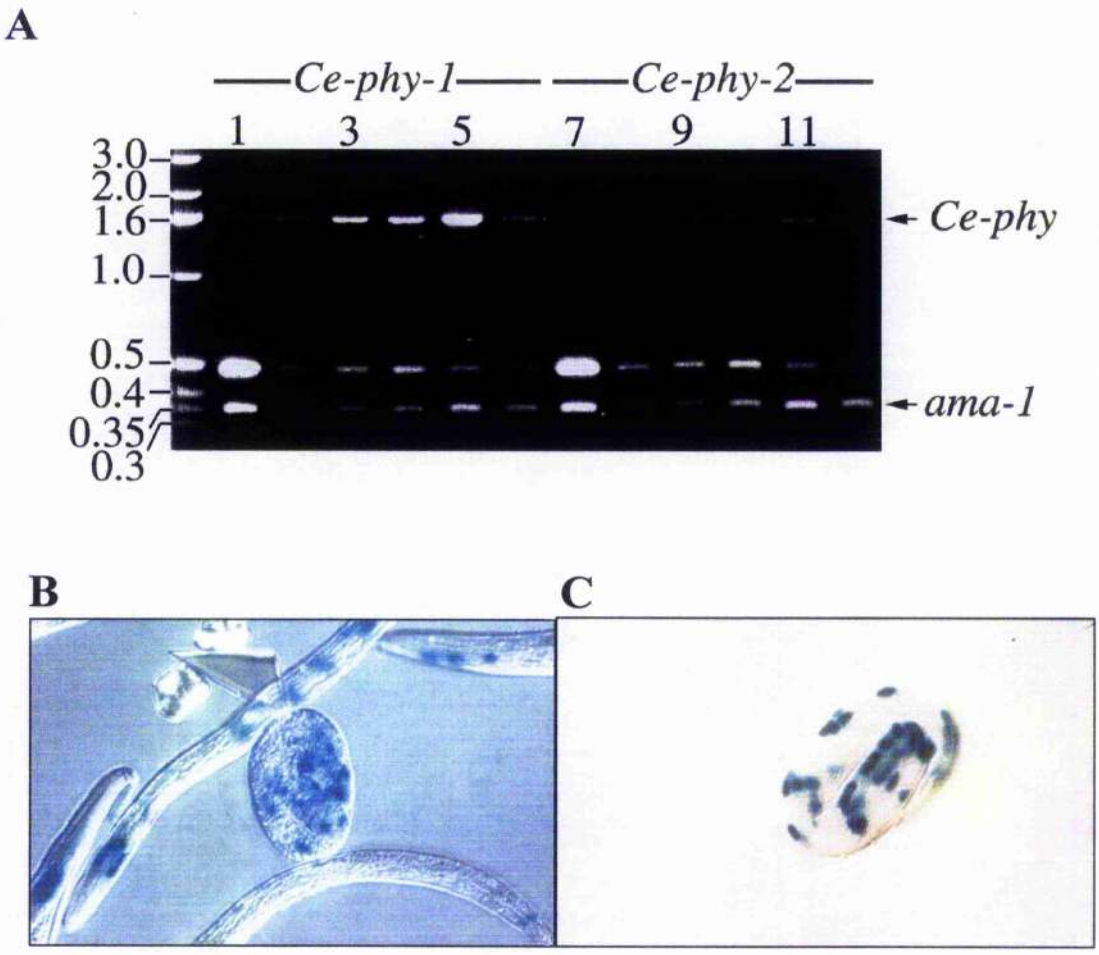


Figure 3.6. Embryonic expression of *Ce-phy-1* and *Ce-phy-2*

(A) Agarose gel showing a comparison of embryonic expression of the *Ce-phy* genes with samples representing developmental stages L1 to adult. Lanes 1 and 7 are embryo samples, lanes 2 and 8 are L1 samples, lanes 3 and 9 are L2 samples, lanes 4 and 10 are L3, lanes 5 and 11 are L4 and lanes 6 and 12 are adult. Lanes 1-6 represent expression from embryo to adult of *Ce-phy-1* along with the control gene *ama-1*, lanes 7-12 represent embryo to adult expression of *Ce-phy-2* in comparison to *ama-1*. (B) *Ce-phy-1* promoter driven reporter gene expression in the developing 3-fold embryo in the hypodermal cell nuclei. (C) *Ce-phy-2* promoter driven reporter gene expression in the developing 3-fold embryo in the hypodermal cell nuclei.

heritable and is found only in the progeny of injected animals. This procedure was developed from interference using sense or antisense RNA individually. This showed that both forms of single stranded RNA were capable of eliciting an interference effect independently (Guo and Kemphues, 1995). Interference using double-stranded RNA is however several orders of magnitude more efficient at producing an effect than the single strand procedures (Fire *et al.*, 1998). The phenotypes seen previously with single stranded RNA approaches may have in fact been due to the presence of small amount of double-stranded RNA in these preparations due to priming of reverse strand from 3' overhangs. No precise size limit has been defined for the type of fragment that is capable of producing an effect although routinely RNA of between 0.5 kb and 2 kb is used. Cross reactivity with non-target sequences has been noted (Fire *et al.*, 1998; Tabara *et al.*, 1998). Although the exact level of similarities at the nucleotide level with which cross reactivity occurs is not known, an estimate of approximately 80% similarity causing cross-interference has been suggested (Tabara *et al.*, 1998). If cross-reactivity is a potential problem a region that is present only in the target gene and not in potential cross-reactive sequences can be used. Alternatively, if necessary, untranslated regions, particularly at the 3' end, can be used. Unless absence of gene activity is verified at the transcriptional or translational level complete removal of gene activity by RNAi cannot be assumed and the resulting effect is often referred to as transcript disruption or knockdown, rather than knockout. dsRNA can be introduced into many positions in the body with the interference effect spreading throughout the tissues of the worm (Fire *et al.*, 1998), however in our experience microinjection into the syncitial gonad of the adult produced the best results. The amount of material injected could be more easily assessed within the confines of the gonad, with a similar volume being introduced to that used for DNA injections. Additionally the injected animals appeared much healthier after gonadal injections thus maximising survival rate and progeny size. The effectiveness of RNAi can also depend on a number of variables including the target tissue and the timing of expression of gene to be disrupted. Neuronally expressed genes can be resistant to RNAi (Tavernarakis *et al.*, 2000) and genes expressed late in larval development can be resistant to RNAi carried out in the parent (Fire *et al.*, 1998). Also RNAi is concentration dependent so the amount of RNA delivered may affect the results (Fire *et al.*, 1998). It is also important to note that the amount of interference required for an effect to be seen would be predicted to be gene dependent. The fraction

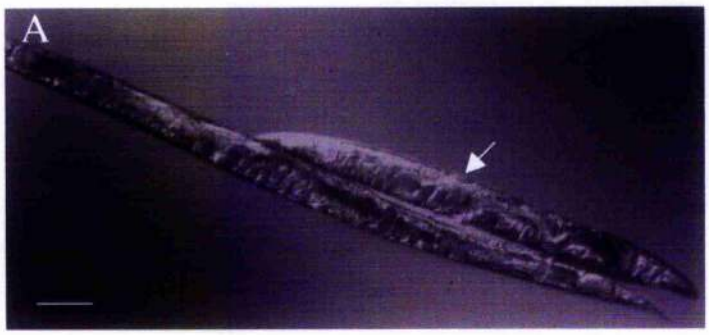
of cells that need to be affected, and the fraction of activity that must be eliminated, for an effect to be achieved will be different for individual genes.

3.2.6.1 *Ce-phy* gene RNAi

RNAi using double-stranded RNA was used to define a functional role for both *Ce-phy* genes. As the *Ce-phy-1* and *Ce-phy-2* genes were both expressed predominately in a tissue for which no resistance to RNAi had been reported, the hypodermis, and from an early point in development, none of the potential problems mentioned were predicted to affect their disruption. Similarity at the nucleotide level between *Ce-phy-1* and *Ce-phy-2* is 57%, which was considered low enough to assume no cross-interference. Additionally, both genes have each other as the most similar nucleotide sequence in the *C. elegans* genome, which made the probability of cross-reaction with another sequence low. dsRNA was produced *in vitro* for both *Ce-phy-1* and *Ce-phy-2* by PCR cloning almost full length coding sequence (both approximately 1.6 kb) in a pBluescript-like cloning vector. This vector contains the promoter and initiation sites for the bacteriophage RNA polymerases from T7 and T3 at opposite ends of the cloning site. Single stranded RNA was produced separately from each site in different reactions. Linearisation of the plasmid with an enzyme that cleaves at the opposite end of the insert from the promoter to be used for transcription gives a suitable template for the *in vitro* transcription of single stranded RNA using the appropriate enzyme. The single strands were then annealed together in a salt buffer to produce dsRNA corresponding to the cloned gene. The presence of dsRNA was verified by gel electrophoresis. dsRNA was injected into the gonad of young adult N2 hermaphrodites. After a period of recovery, from 4 hours to overnight, nematodes were transferred singly to fresh plates. This period allows for partially developed embryos to be cleared and progeny scored are those that have been exposed to dsRNA for their entire development. Injected worms were then transferred daily to fresh plates and their progeny scored.

3.2.6.2. *Ce-phy-1* RNAi injection

Injection of dsRNA at a concentration of 0.5 mg/ml into wild type N2 *C. elegans* for *Ce-phy-1* gave a highly penetrant dumpy (Dpy, short and fat body shape) phenotype in approximately 90% of progeny from affected plates. A RNAi affected individual (arrowed) is shown in comparison to a N2 worm in Figure 3.7A. The injected animals themselves appeared phenotypically wild type. RNAi affected animals were



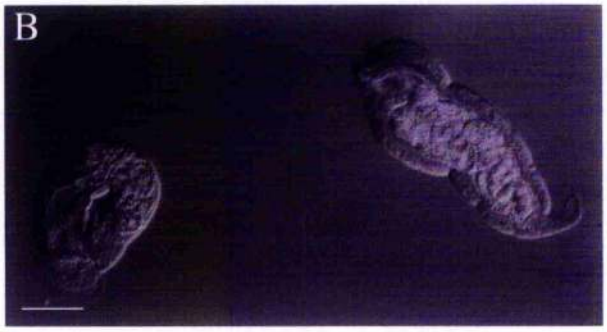


Figure 3.7. RNA interference of *Ce-phy-1* and *Ce-phy-2*

A) Typical progeny from wild type N2s following *Ce-phy-1* RNAi injection (arrowed) displaying a Dpy phenotype, shown in comparison to untreated an untreated wild type N2. Bar 100 μm.

B) Representation of the severe Dpy phenotypes resulting from injection of wild type N2 nematodes with combined *Ce-phy-1* and *Ce-phy-2* dsRNA at a overall concentration of 1 mg/ml. Bar 40 μm.

predominantly medium Dpy, approximately 2/3 the length of wild type nematodes with a very small number showing more extreme body shape defects. The degree of disruption of *Ce-phy-1* function achieved by this process was not assessed and so it was not possible to say whether the medium Dpy phenotype resulted from partial or complete disruption of gene function. The RNAi phenotype from interference of *Ce-phy-1* was found not to be heritable, as progeny from affected animals appeared wild type all respects. The medium Dpy phenotype produced by RNAi of *Ce-phy-1* suggested a role in the development or maintenance of nematode body shape for this gene through modification of the cuticle. Similar body shape defects have been found for a number of mutant collagen genes (Johnstone *et al.*, 1992; Levy *et al.*, 1993; von Mende *et al.*, 1988). The lack of phenotype in the injected adult is in agreement with this role, as we have shown this is when expression of *Ce-phy-1* is low due to the cuticle already being formed. No other phenotype was observed in these RNAi worms other than their altered body morphology. The small percentage of RNAi progeny displaying more extreme body defects may suggest that the degree of nucleotide similarity which could cause low level cross-reactivity is lower than suggested previously, a fact supported by a recent study (Hussein *et al.*, 2002), and that these extreme phenotypes were caused by disruption of *Ce-phy-1* and also cross-interference of gene(s) with a similar sequence.

3.2.6.3. *Ce-phy-2* RNAi injection

dsRNA corresponding to the *Ce-phy-2* gene was injected at a concentration of 0.5 mg/ml into N2 *C. elegans*. No effect was found in either the injected animals or their progeny after injection with *Ce-phy-2* dsRNA. Worms were assessed as being wild type on the basis of body shape, movement, internal morphology egglaying and fecundity. The level of gene disruption was not assessed and it was therefore not possible to determine if *Ce-phy-2* gene function had been successfully knocked-down. Lack of a mutant phenotype observable under laboratory conditions for *Ce-phy-2* suggested that, assuming effective interference, the function of this gene was not essential for formation of the nematode cuticle. The lack of observable phenotype was even more surprising given the additional staining of *Ce-phy-2* in the body wall muscle cells. Due to the importance of the muscle cells for producing type IV collagen and the severe embryonic lethal phenotypes resulting from mutation in these collagens (Guo *et al.*, 1991; Gupta *et al.*, 1997; Sibley *et al.*, 1994; Sibley *et al.*, 1993) and processing enzymes (Norman and

Moerman, 2000), an effect from lack of hydroxylation of proline could be predicted to have a strong phenotype. The lack of mutant phenotype by RNAi of this gene strongly suggested that *Ce-phy-2* was not involved in processing of type IV collagens. It was also possible however, that either the RNAi of this gene was not effective, or that in particular muscle tissue expression was not being disrupted.

3.2.6.4. Combined *Ce-phy-1* and *Ce-phy-2* RNAi injection

Combined RNAi using single stranded RNAs from two different genes had been reported to have equivalent effects to that produced from double genetic mutants of the corresponding genes (Rocheleau *et al.*, 1997). The effect of combined *Ce-phy-1* and *Ce-phy-2* interference in an N2 background was examined by mixing equal volumes of the preparations of dsRNA used previously, producing a final concentration of each dsRNA of 0.25 mg/ml. Injection at this concentration produced identical results as found for single disruption of *Ce-phy-1,* as most of the progeny were medium Dpy in appearance. A higher concentration of dsRNA at 0.5 mg/ml for each gene was made, injection at that concentration produced much more severe effects, as shown in Figure 3.7.B. Extreme body shape defects were evident with affected progeny approximately a quarter of the length of wild type. Many of these remained in a coiled position throughout development. The severe dumpy phenotype was evident in 97% (1,189 out of 1208 hatched progeny) of progeny. The more extreme phenotypes produced by interference of both genes showed that the *Ce-phy-2* dsRNA used was effective in knocking down *Ce-phy-2* function at the concentration used previously for single disruption. Therefore the wild type appearance of *Ce-phy-2* singly disrupted nematodes was reliable and was not a result of lack of interference. Secondly it indicated that the medium Dpy phenotype from *Ce-phy-1* disruption is due to knock-down of this gene alone and not a result of interference of *Ce-phy-2* as well. Additionally, it demonstrated that these genes are required in combination for development of wild type body shape and suggested a central role for these genes in formation of the *C. elegans* cuticle which serves to maintain the post-embryonic elongated form of the nematodes. Disruption of both genes caused production of an even weaker cuticle that was barely capable of maintaining any normal worm body shape.

3.2.7. Characterisation of *dpy-18* as a *phy-1* mutant

3.2.7.1. Identification of *dpy-18* as a candidate *phy-1* mutant

Examination of the genetic map of the regions surrounding the *Ce-phy-1* physical map position suggested the recessive medium Dpy mutant, *dpy-18,* as a candidate *Ce-phy-1* mutant. The *Ce-phy-1* RNAi phenotype was extremely similar to the appearance of this mutant and no other Dpy or body shape mutants were found in this area. *dpy-18* was first identified almost 30 years ago in one of the first genetic screens in *C. elegans* (Brenner, 1974) but had remained uncharacterised at the molecular level. Multiple alleles of *dpy-18* had been identified and the genetic position of this locus mapped on chromosome III. As illustrated in Figure 3.8. the *dpy-18* map position was 8.62 while the interpolated map position of *phy-1* was 7.8. A study of the associated male tail ray defect of three *dpy-18* alleles, *e364, e1096* and *bx26,* demonstrated that while this aspect of the phenotype was affected by temperature (not evident at 16°C) the overall body morphology phenotype was not temperature sensitive (Baird and Emmons, 1990).

3.2.7.2. Rescue of the *dpy-18* phenotype with wild type transgenic *Ce-phy-1*

To determine if *dpy-18* was mutant in *Ce-phy-1,* phenotypic rescue was attempted with wild-type *Ce-phy-1* coding and flanking sequence. Genomic sequence was chosen to promote maximum expression of the transgene. No cosmid containing complete genomic sequence for this gene was available as the coding sequence for *Ce-phy-1* is divided between a YAC, containing the promoter and first two exons, and a cosmid, containing the remainder of the genomic sequence and 3' UTR, as illustrated in Figure 3.8. *Ce-phy-1* sequences were therefore cloned from wild type genomic DNA by PCR. The region cloned included approximately 2 kb of upstream sequence, the entire genomic coding sequence and approximately 300 bp of 3' UTR. 2.8 kb of promoter had previously been shown to direct strong hypodermal expression of reporter genes but the full region was not included in this construct due to PCR size constraints. The 3' UTR included the poly(A) signal sequence and transfer site. A mixture of two polymerases, *Taq* and *Pfu,* was required to efficiently amplify this 7 kb product which was cloned into a plasmid vector. Full-length sequencing was not performed on this construct but the ends of the fragment were sequenced to confirm the identity of the insert. The construct was injected into *dpy-18* mutant strains carrying alleles *e364* and *e1096* along with a GFP expressing marker, dyp-7-GFP. Control injections in *dpy-18* were performed with DNA mixes lacking the *Ce-phy-1* sequences. Repair of the *dpy-18* body

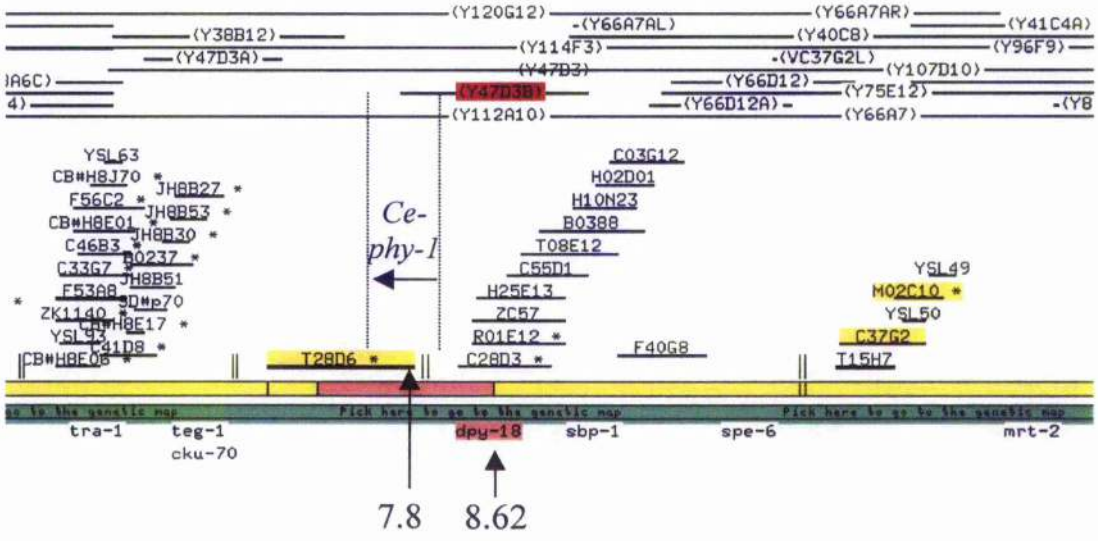


Figure 3.8. Comparison of physical and genetic maps in the region of *Ce-phy-1*

Positions of *dpy-18* locus at 8.62 on chromosome III on the genetic map and the *Ce-phy-1* encoding cosmid/YAC. The interpolated genetic map co-ordinate of *Ce-phy-1* at position 7.8 is indicated. The *Ce-phy-1* gene is encoded by the YAC Y47D3.B (shown in red) and the cosmid T28D6 (shown in yellow). The direction of gene from ATG to TAA is shown with an arrow. The figure is adapted from the data available from WormBase.

morphology defect with *Ce-phy-1* was evident in transformed F1s identified by UV microscopy. Multiple semi-stable transformed lines from both *e364* and *e1096* containing strains were identified and mutant repair found for all transformed animals. A comparison of transformed and un-transformed *dpy-18* nematodes is shown in Figure 3.9, with clear reversion to wild-type body shape seen in the transgene expressing nematode. No repair was found for injection mixes lacking wild type *Ce-phy-1* sequences. The similarities between *Ce-phy-1* RNAi and *dpy-18* phenotypes, their comparative positions on their respective maps and the phenotypic rescue of *dpy-18* mutants with wild type *Ce-phy-1*, indicated that *dpy-18* was a *Ce-phy-1* mutant.

3.2.7.3. Analysis of *Ce-phy-1* molecular changes in *dpy-18* mutants

3.2.7.3.1. Amber stop codon mutation in *dpy-18(e364)*

To confirm the relationship between gene and mutant, the *Ce-phy-1* gene sequence was analysed from three *dpy-18* mutant alleles. cDNA and genomic DNA were isolated from strains carrying *dpy-18* alleles *e364, e1096* and *bx26*. Strain CB364 carrying *dpy-18* allele *e364* had already been defined as containing an amber stop codon as it is wild type when crossed with amber mutant suppresser strains (Waterston and Brenner, 1978). Therefore any lesion found in *Ce-phy-1* from this strain should correspond to this type of mutation. Full-length coding sequence, from translational start to stop, was cloned from this strain and double-strand sequenced over the length of the insert. Two clones from independent PCR reactions were analysed in this way to discriminate between PCR induced errors, produced by the non-proof-reading DNA polymerase used for amplification of these sequences, and true changes in the gene sequence. Comparison of the two clones revealed only one consistent base change, this mutation G275-A275 was predicted to cause the alteration of a TGG, tryptophan$_{76}$ encoding, to a TAG amber stop codon. *dpy-18(e364)* was produced by ethyl methanesulphonate (EMS) mutagenesis (Brenner, 1974) and the point mutation found for *Ce-phy-1* is consistent with the G/C-A/T transitions most often produced by EMS mutagenesis. The production of an amber stop codon was also identified as predicted, which suggested this molecular change was the true *e364* molecular lesion. Cloning of genomic *Ce-phy-1* sequences from *e364* and double strand sequencing over the point mutation region confirmed this mutation as the molecular alteration responsible for the *dpy-18(e364)* phenotype. A stop codon at the start of the second *Ce-phy-1* exon is predicted to result in a severely truncated protein containing only 75 of the 559 amino acids of this protein

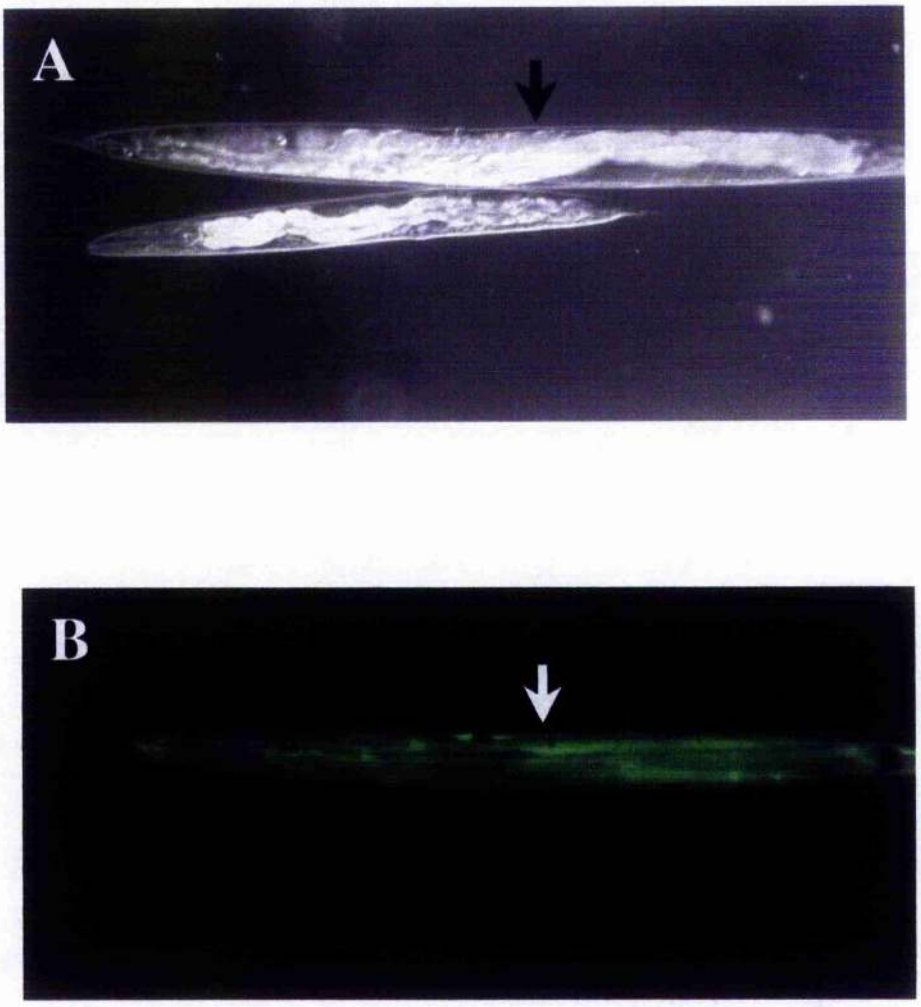


Figure 3.9. *dpy-18* rescue with the wild type *Ce-phy-1* gene

Rescue of the medium Dpy phenotype of a *dpy-18*(*e364*) showing the repair to wild type body shape of the transformed nematode (arrowed) compared to a *dpy-18* mutant. Identical results found were found for rescue of *dpy-18*(*e1096*) mutants (data not shown). The DIC image is shown in panel A, with the UV image shown in panel B demonstrating expression of the marker plasmid dpy-7-GFP.

and missing all defined functional regions and conserved residues. This strain is therefore predicted to be a functional null. Therefore the rescue of *dpy-18* with wild type *Ce-phy-1* and identification of a mutation in this gene in a *dpy-18* strain demonstrates conclusively that *dpy-18* is a *Ce-phy-1* mutant.

3.2.7.3.2. A deletion in *dpy-18(e1096)*

Allele *e1096* of *dpy-18* was also investigated to determine the nature of the molecular lesion in this strain. A similar approach to the analysis of the *e364* mutant was initially adopted, however, PCR with full-length coding sequence primers from cDNA extracted from *e1096* nematodes gave no product, see Figure 3.10. Control PCRs were performed to confirm the ability of the *Ce-phy-1* primers to amplify using these conditions and the quality of mutant template cDNA. PCRs with *Ce-phy-1* specific primers using N2 template DNA, lane 1, and primers specific to *Ce-phy-2* on mutant template cDNA, lane 3, both produced the expected amplification products. Additionally no product was amplified by PCR using mutant genomic template with the same *Ce-phy-1* primers used on cDNA (data not shown). Using primers flanking the coding region a PCR product was generated from mutant genomic DNA approximately 800 bp smaller than that amplified from wild type template, suggesting a deletion of gene sequence, Figure 3.11. A. This region was cloned and the extent of the putative deletion further defined by PCR analysis. Each primer used to clone this region was used in combination with another primer from within the *Ce-phy-1* coding sequence, Figure 3.11. B. PCR with the sense primer phy-1 IS IF in combination with phy-1 (*Sal* I) R, lane 1, did not show evidence of the deletion, however, phy-1 (*Not* I) F in combination with the antisense primer phy-1 IS 5R, lane 2, produced a product 800 bp smaller than predicted. The mutation therefore lay to the 5' end of the gene and included the ATG sequence but did not extend as far as primer phy-1 IS 5R (which lies approximately 180 bp downstream of ATG) and was an estimated 800 bp in size. Primers phy-1PS 1F and phy-1PS 2F were therefore designed within the *Ce-phy-1* promoter region approximately 800 bp 5' of the ATG and the deletion clone sequenced with these and primer phy-1 IS 5R. The extent of the mutated sequence in *dpy-18(e1096)* is shown in Figure 3.12. 776 bases of sequence were deleted, from –688 to +88, removing a significant region of the promoter and 5' sequences coding for 30 N-terminal amino acids of *Ce*-PHY-1, including the predicted signal peptide residues. No other genes are predicted in the area covered by the deletion. Lack of a signal peptide suggests that any protein produced from this gene

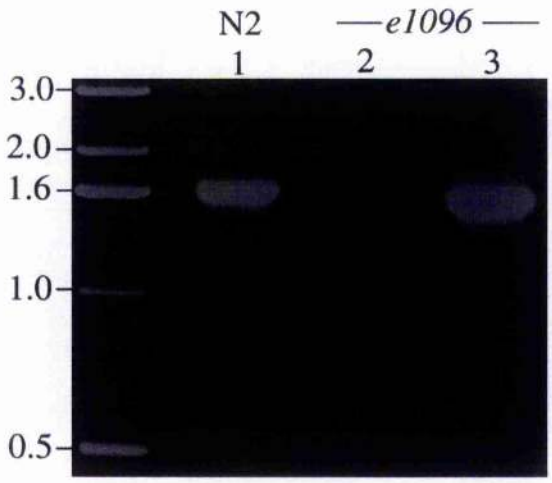


Figure 3.10. *Ce-phy-1* PCR on *dpy-18*(*e1096*) cDNA

Agarose gel of products resulting from PCR on N2 cDNA with full-length coding sequence primers for *Ce-phy-1* (lane 1) in comparison with *e1096* cDNA (lane 2), which generates no product. Full-length coding sequence primers specific for *Ce-phy-2* were applied to the *e1096* cDNA (lane 3) as a control for the ability to amplify other sequences from this template DNA. The positions of molecular size standards are indicated with sizes given in kilobases.

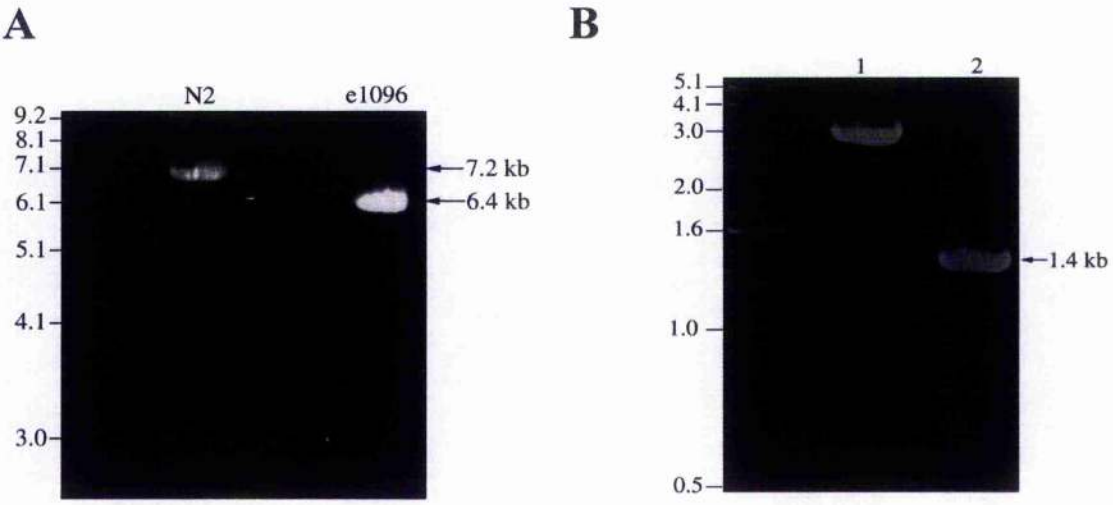


Figure 3.11. *Ce-phy-1* PCR on *dpy-18(e1096)* genomic DNA

(A) Agarose gel of products resulting from PCR of N2 and *dpy18(e1096)* genomic DNA with *Ce-phy-1* primers lying outwith the region covered in the mutant cDNA analysis. The 6.4 kb product from the mutant genomic DNA is smaller than the 7.2 kb product amplified from wild type template, indicating a deletion of *Ce-phy-1* sequences in *dpy-18(e1096)*. (B) Mapping of deletion to the 5' or 3' end of *Ce-phy-1* was performed. Primers at the 3' end gave correct sized 3 kb band (lane 1) and primers specific to the 5' end of the gene (lanes 2) showing the presence of a 800 bp deletion. The size in kilobases of this mutant band is indicated. The positions of molecular size standards are indicated with sizes given in kilobases.

```
-706  ttatatttct ctaaccctaa acaccaaaga acctatatat ctcaoaottt oactatcatc
-646  aagtcaacgt cagttttact catttctgcg tctcgttttc ttatttcttt cgcatctaat
-586  tgttgtaagt ctacgtggag aaaccagcct tatatagctg aaccaacttc ctcttttgca
-526  atttttttct ttctttgtgt tttatgccag aactttgaaa tcacaaaagt tattgttaat
-466  ttaatattta agctacctga aatttcctat atttggtcaa ttaacaattt tttagcagct
-406  agttcaagaa gtaagttaaa gcagacgaca tctagcgggt cagaattttg gaagctgtga
-346  gatgtcggcc gctatatgct tcttgaggaa aactggtgtg gtagccgaca ccttgcggga
-286  tttgatcccc ggggtttccg gatctgtaag atgtaggctg ccacaacttt tttaaactga
-226  caaagttaag aaatcaaaat ttcctttttt ccggtaagca ctgtcttttg gctattttag
-166  gcagccgaca tcttgcgggt cttgctcagt tttgattttt ttttttcgaa cccatgagat
-106  gtccgccgct ataatttttt ctggcgtttg tgtcaaatct ttattgtaaa ctttatccgc
                                                        M  R  L  A  L
 -46  ttcccgtttt gcaggttttt aattccctaa ctctcaattt ttcagaATGC GCCTGGCACT
                                                      Phy-1 HSC F ➔

       L  V  L  A  T  I  G  Y  A  V  A /D  L  F  T  S  I  A  D  M
  15   CCTTGTACTA GCCACAATCG GCTATGCAGT TGCCGATCTG TTCACCTCGA TTGCCGACAT
       Q  N  L  L  E  T  E  R  N  I  P  K  I  L  D  K  Y  I  H  D
  75   GCAAAACCTT CTGGAAACTG AAACAAATAT TCCGAAAATC CTTGACAAAT ACATTCACGA
       E  E  E  R  L  V  Q  L  K  K  L  S  E  E  Y  S  K  K  N  E
 135   CGAGGAAGAG CGACTGGTTC AGCTGAAGAA GCTGTCGGAG GAGTACTCGA AGAAAAATGA
                                       ➔ Phy-1 IS 5R
       I  S  I  E  N  G  L  K  D  I  T  N  P  I  N  A  F  L  L  I
 195   GATTTCGATT GAAAATGGGC TCAAGGATAT TACGAATCCG ATCAATGCGT TTTTATTGAT
       K  R  K
 255   CAAGAGAAAG gtaatttttgt ttgcaatagg ttttaagtta ttgcagccga catctcactg
```

Figure 3.12. *Ce-phy-1* sequences deleted in *dpy-18(e1096)*

Deleted sequences are shown in bold, uppercase represents exonic sequences, lower case represents promoter and intron sequence. The translation of exon sequences is given above and the position of relevant primers in this region indicated with underlining and arrows. The position of the signal peptide cleavage site in the N-terminal amino acid sequence is shown with a forward slash.

would not be targeted to the ER and would therefore be non-functional. Alternatively, lacking proximal promoter sequence it might be predicted that no gene transcription would occur. Interestingly, despite lacking a large region of the promoter region transcription of the gene can still be detected. RT-PCR products were produced using the full-length reverse primer in combination with a number of upstream genes from the area of the gene not deleted, Figure 3.13. Therefore the gene is a predicted null due to the lack of encoded signal peptide causing loss of targeting to the ER where it is required to function in the modification of secreted proteins.

3.2.7.3.3. Allele *bx26* is caused by an ochre stop codon

Strain EM76, carrying allele *dpy-18*(*bx26*), was also examined at the molecular level. This was analysed in a manner identical to *e364* although only one *Ce-phy-1* cDNA clone was sequenced as PCR products were made with a proof-reading polymerase, and the mutation was confirmed by analysis of genomic DNA. A stop codon was found resulting from a C-T transition, causing a CAA-TAA (glutamine$_{253}$-Ochre stop) change. EM76 was also produced from an EMS screen (Baird and Emmons, 1990) with this transition being a common result of treatment with this mutagen. The point mutation is at the start of exon five and again is predicted to produce a severely truncated and non-functional protein. The position of this mutation and those of the other two identified alleles are shown in Figure 3.14.

3.2.8. RNAi of *Ce-phy-2* in a *dpy-18* genetic background

Identification of *dpy-18* as a *Ce-phy-1* mutant allowed for a more detailed examination of the combined effects of removal of both *Ce-phy-1* and *Ce-phy-2* function. Combined RNAi of both these genes produced severely Dpy animals (Figure 3.9. B), although the degree of removal of gene function of either gene was not determined, and the effectiveness of RNAi knock-down of multiple genes has not been extensively examined. A putative null allele of *dpy-18* was used which was assumed to have no active *phy-1* present. Thus the effect of knock-down of one gene with complete absence of *Ce-phy-1* activity could be determined. Strain CB364 [*dpy-18* (*e364*)] was injected with dsRNA corresponding to *Ce-phy-2* at a concentration of 0.5 mg/ml. In a wild type genetic background this produced no effect, however, in the *dpy-18* mutant background progeny from injected animals were embryonic lethal. Animals from affected plates were being determined as dead when they remained unhatched 16 hours after the

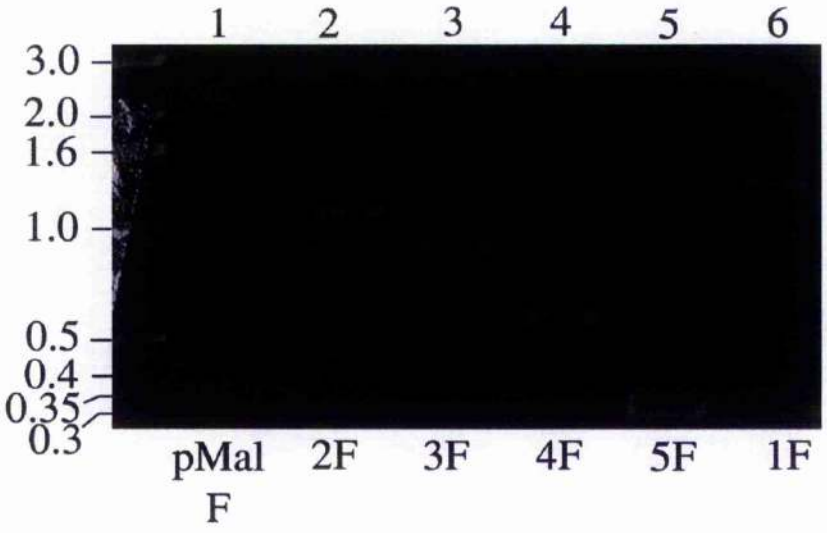


Figure 3.13. Mapping of the transcribed region of *dpy-18*(*e1096*)

Agarose gel of products resulting from PCR with the primers indicated in combination with a common antisense primer on *dpy-18*(*e1096*) mutant cDNA. The antisense primer covered the translational stop region and was used with a primer within the deletion region (pMalF) and primers lying outwith the deleted region (1F-5F). Wild type products were amplified from the areas not deleted, demonstrating the production of mutant *Ce-phy-1* mRNA missing its 5' coding sequence. The positions of molecular size standards are indicated with sizes given in kilobases.

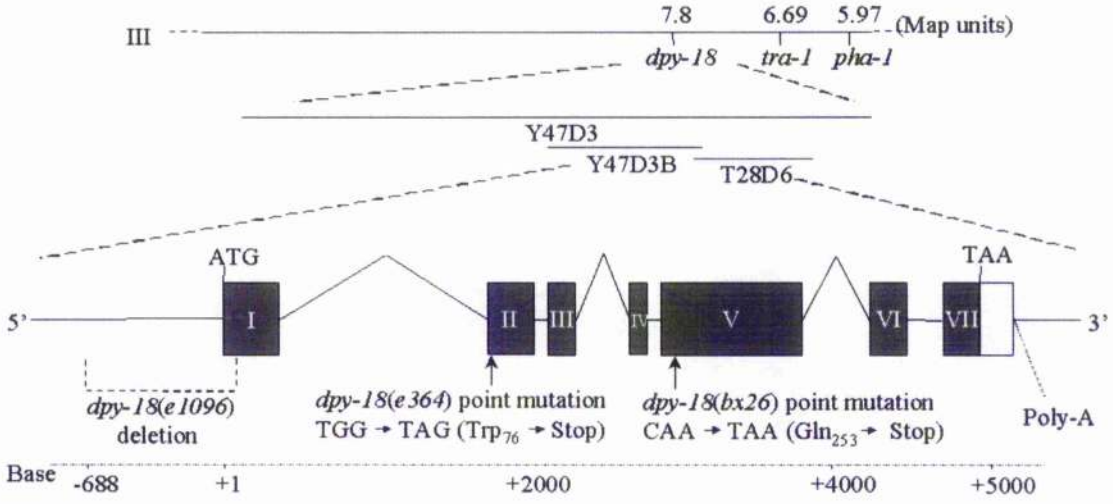


Figure 3.14. *dpy-18* alleles

The positions of molecular mutations found within the *Ce-phy-1* gene in the *dpy-18* strains investigated are shown. The gene structure of *Ce-phy-1* is also depicted as is the *Ce-phy-1* encoding cosmid/YAC in relation to the genetic map.

removal of the adult. These were counted and embryonic lethality was observed on affected plates at a penetrance of 89% (1118 dead embryos out of 1250 laid). Those animals that hatched were Dpy or severely Dpy in appearance. This identified an essential role in nematode development for the combined *Ce-phy* genes, as without their combined function nematodes could not complete embryonic development.

3.2.8.1. Developmental timecourse of *Ce-phy-2* RNAi/*dpy-18* embryos

The embryonic lethal phenotype of *Ce-phy-2* RNAi injected *dpy-18* nematodes was examined in more detail by monitoring their appearance throughout embryonic development. This was accomplished by injecting animals as described above then, after a period to ensure RNAi was effective, pre-comma stage embryos were collected and photographed at 30 minute intervals as they developed. Specific time points during this process are shown in Figure 3.15. Comma stage embryos developed normally to the three-fold stage where elongation has been completed, transforming a ball of cells in to a worm shape. Embryos at this stage were active, with movement and twitching indicating correct formation of muscular structures and attachments. After the three-fold stage the mutant phenotype becomes apparent as embryos become less active and gradually lose their elongated shape. Examination of the timing of loss of worm body morphology reveals that this corresponds exactly with the time when the cuticle becomes the support to maintain the nematodes body form (Sulston *et al.*, 1983). The embryonic development of wild type *C. elegans* is depicted in Figure 3.16, showing elongation of the embryo occurring followed by the timing of cuticle synthesis. The RNAi cuticles are weakened due to improper modification and are therefore unable to fulfil their role of supporting the nematodes vermiform shape. RNAi treated embryos continue to collapse back over time to a more disorganised state. The terminal phenotype is variable but was typically like the two-fold stage. These animals did not hatch, gradually ceased movement, vacuolated and died. These results show that both genes are essential for development and formation of a functional cuticle capable of maintaining morphology. The movement of developing embryos and attainment of a fully elongated form distinguishes them form the mutant phenotypes of genes involved in the formation of the basement membranes, typically two-fold paralysis, and further supports a non-type IV collagen modification role for *Ce-phy-1* and *Ce-phy-2*.

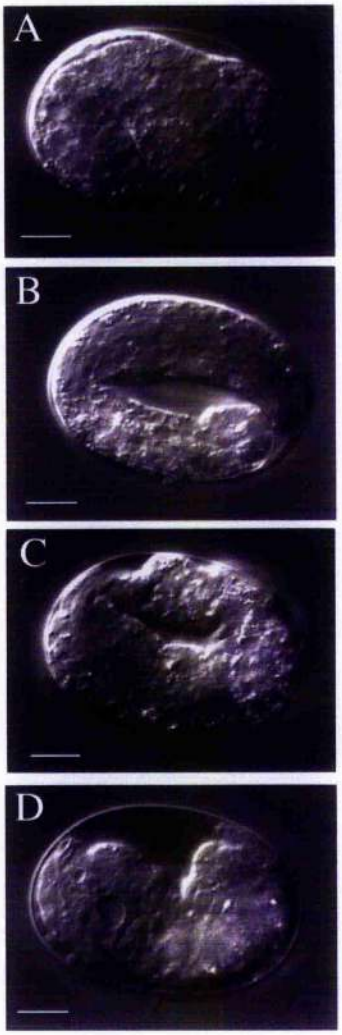


Figure 3.15. RNAi of *Ce-phy-2* in a *dpy-18* mutant

DIC images showing the developmental timecourse of a *dpy-18* mutant injected with dsRNA for *Ce-phy-2*. Bar represents 10 μm. (A) 1.5 fold embryo (440 minutes). (B) 3-fold elongated embryo (570 minutes) with the head and tail out of focal plane. (C) Retracting embryo (710 minutes). (D) Terminal phenotype (approximately 1,800 minutes), a fully retracted dying embryo. Bars equal 10 μm.

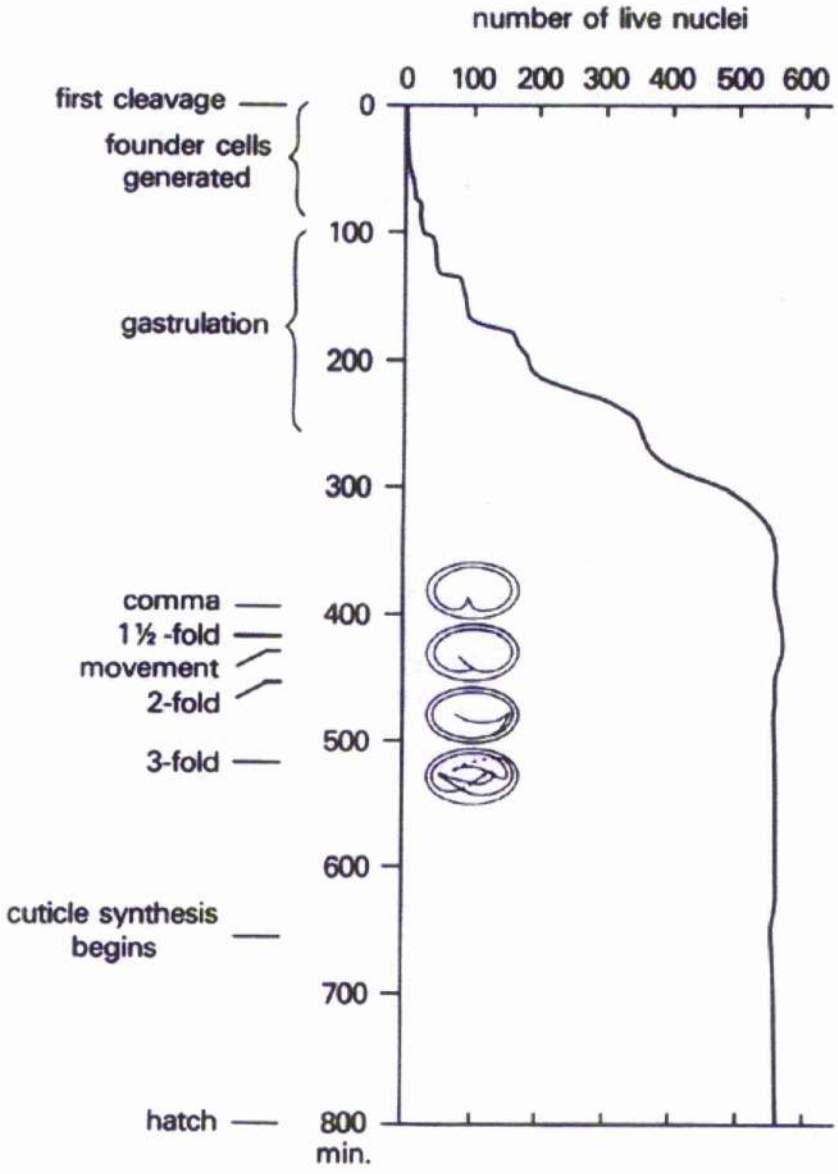


Figure 3.16. Embryonic developmental timepoints

Timing of the main events occurring during *C. elegans* embryonic development. Adapted from Sulston *et al.*, (1983). Note the timing of the beginning of cuticle synthesis.

3.2.9. RNAi by bacterial feeding

Bacterially-mediated RNAi was developed by utilising the unusual ability of the RNAi effect to spread throughout the tissues of the body from the delivery site. Engineering the *E. coli* food source of *C. elegans* to produce dsRNA was shown to cause specific RNAi effects (Timmons and Fire, 1998). This procedure uses two convergent, inducible T7 RNA polymerases flanking a fragment of the gene to be disrupted. Using a bacterial strain with an inducible *lac* promoter enables production of RNA from both strands by treatment of the bacteria with IPTG, these strands then anneal within the cell to form dsRNA. As the initial experiments showed weaker disruption compared to injections, the procedure was refined further by using a viable mutant bacterial strain deficient in RNase III, the enzyme specific for double-stranded RNA (Timmons *et al.*, 2001). The RNAi feeding procedure was then optimised by the discovery that although uninduced bacteria produced no phenotypes, strong induction (in liquid culture) produced a lower penetrance in phenotypes than the presumed weaker induction on plates using moderate concentrations of IPTG (Kamath *et al.*, 2000). Severity and penetrance of the RNAi effect was also found to be temperature dependent, being more effective at higher temperatures (Kamath *et al.*, 2000). A particular advantage of the feeding procedure is that the RNAi effect can be titrated to reveal secondary phenotypes that could be masked by a strong primary phenotype such as embryonic lethality. This could be accomplished by reducing (1000-fold) the induction concentration of IPTG, or by simple dilution of induced bacteria with uninduced bacteria. Thus the possibility exists to generate a series of RNAi phenotypes analogous to a series of genetic mutants.

3.2.9.1. RNAi feeding of *Ce-phy-1* and *Ce-phy-2*

The full-length gene coding sequences virtually identical to the sequences used for RNAi injections were cloned for both genes into the double T7 bacterial-feeding vector. This was then transformed into the HT115(DE3) strain of *E. coli* and the production of RNA was induced. Nematodes were then added to these plates and grown at 25°C. The effects of feeding of dsRNA via bacteria for *Ce-phy-1* in an N2 background were not as severe as for injections, with nematodes displaying a very slight dumpy phenotype (data not shown). Likewise the effect of feeding of *Ce-phy-2* RNAi bacteria to *dpy-18* nematodes was not as severe as injection (Figure 3.17. panels A and B), with many of the progeny showing similar defects to those seen in *Ce-phy-1* and *Ce-phy-2* simultaneous RNAi injections. During the course of this work a phenotypically wild

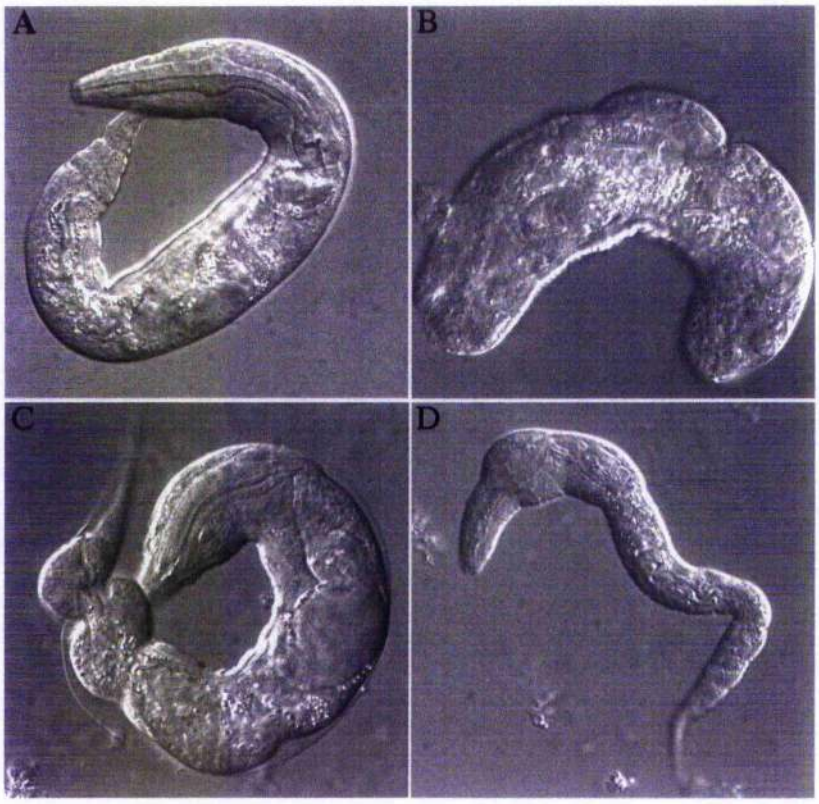


Figure 3.17. *Ce-phy-1* and *Ce-phy-2* RNAi feeding phenotypes

Body morphology mutants resulting from *Ce-phy-2* RNAi expressing bacteria fed to *dpy-18* mutants (panels A and B), and *Ce-phy-1* fed to *phy-2* deletion mutants (panels C and D). Panels A and C represent the typical results from these combined experiments with panels B and D representing the range of phenotypes observed. Unshed cuticle and cuticle constrictions are evident in panel C. All images were taken at the same magnification.

type *Ce-phy-2* deletion strain, JK2757 [*phy-2(ok177)*], had been identified. This strain contained a 1335 bp deletion removing most of *Ce-phy-2* exons 4-7, followed by a premature stop codon directly after the deletion end point (Friedman *et al.*, 2000). No other predicted coding sequences are removed by this deletion. The phenotype of this strain corroborated the *Ce-phy-2* RNAi phenotype and indicated that lack of a basement membrane specific phenotype upon *Ce-phy-2* RNAi was not due to lack of penetrance of the RNAi effect in muscle cells. Using this putative null *Ce-phy-2* deletion strain and feeding *Ce-phy-1* produced very similar results (Figure 3.17. panels C and D) to the reverse combination. The majority of animals from these experiments displayed the appearance of the nematodes shown in panels A and C, while a smaller percentage were severely malformed with a less uniform appearance, the range of these phenotypes is shown in panels B and D. These extreme body morphology mutants confirmed that the effect of double disruption of these genes is to severely disrupt cuticle formation and that this defect is the cause of embryonic lethality in these nematodes. Also it provided a means to produce large quantities of viable progeny for transgene, antibody and electron microscopy analysis. Although severely malformed, these animals have formed a cuticle and analysis using these methods enabled a more detailed look at the effects of loss of P4H activity on collagen and formation of the nematode cuticle.

3.2.10. Further characterisation of *dpy-18* and *Ce-phy* RNAi mutants

3.2.10.1. COL-19::GFP protein fusion

Due to the nature of the structural collagen molecules and the large number of processing steps required to generate mature collagen, producing a collagen GFP fusion is more problematic than a conventional protein fusion. The N-terminal regions of collagens are cleaved, therefore an N-terminal fusion at the extreme N-terminus would not be feasible as the GFP protein would be cleaved by the N-terminal specific proteases. A possible alternative to this would be an N-terminal fusion with the GFP inserted at the C-terminal site of the protease recognition site. It is possible though that the presence of the GFP protein in this region could affect cleavage. Due to the triple helical structure of most of the mature collagen regions and the severe effects of a single amino acid change in these regions, insertion of GFP at this location would not be feasible. Processing at the C-terminal ends of nematode collagens is not so well understood, and whether any or all are cleaved by a specific C-terminal protease is not currently known. Fusion of GFP to collagens at the C-terminus could thus provide a

means to tag these molecules. This approach was not successful with *C. elegans* collagens C39E9.9 and T01B7.7 (*rol-6*) which produced transgenic animals but no GFP expression (Thein *et al.*, 2003). C-terminal fusion of GFP to collagen ZK1193.1 (COL-19) was successfully performed by C. Shoemaker (AgResearch Ltd. New Zealand) from whom the plasmid BA7-1 was received. Plasmid BA7-1 contains the genomic *col-19* promoter and coding sequence fused in frame to GFP-c3 (Crameri *et al.*, 1996) in the vector pSP65. The vector also contains the 3' UTR from the *Schistosoma mansoni* triose phosphate isomerase gene. This 3' UTR region may or may not be utilised in this construct. This collagen::*gfp* fusion plasmid is depicted in Figure 3.18. Promoter/reporter gene studies of *col-19* suggested an adult specific expression pattern for this gene (Liu *et al.*, 1995), and analysis of its C-terminal non-Gly-X-Y region demonstrated that this was one of the shortest domains found for the *C. elegans* collagens, being only 9 amino acids long. Plasmid BA7-1 was injected into *C. elegans* wild type N2 strain at a concentration of 20 µg/ml with pRF-4 at 100 µg/ml. Once the GFP fluorescence of transformants was established, the plasmid was re-injected at the same concentration with pBluescript SKM at 100 µg/ml and transformants were selected on the basis of fluorescence in the adult stage from the COL-19::GFP fusion protein. Semi-stable lines were generated and adults examined under high power magnification to determine the expression pattern. Discernible external features on the cuticle surface of all stages are the annuli that are created by narrow evenly spaced indents (or annular furrows) that run circumferentially around the animal (see Figure 1.1). Also visible on the cuticle surface are the alae that are only found only in the L1, dauer, and adult stages (Cox *et al.*, 1981b; Singh and Soulston, 1978). The alae run along the lateral sides of the animal and consist of raised tread-like protrusions that form over the hypodermal seam cells. COL-19::GFP expression was found in the adult annuli and alae (data shown for integrated lines, see below, in Figure 3.19).

Transcript disruption of *Ce-phy-1* by bacterial feeding in these N2 COL-19::GFP transgenic lines produced a disrupted appearance of the collagen (data not shown, see below). *dpy-18(e364)* and *phy-2(ok177)* strains were then injected with the construct. The N2 and *dpy-18* COL-19::GFP transgenic lines were integrated (by M. Thein and A. Page, WCMP, Glasgow) to generate strains TP12 and TP13 respectively. The expression pattern of the integrated lines and the *phy-2(ok177)* transgenic non-integrated lines are shown in Figure 3.19. The COL-19::GFP pattern in TP12 is

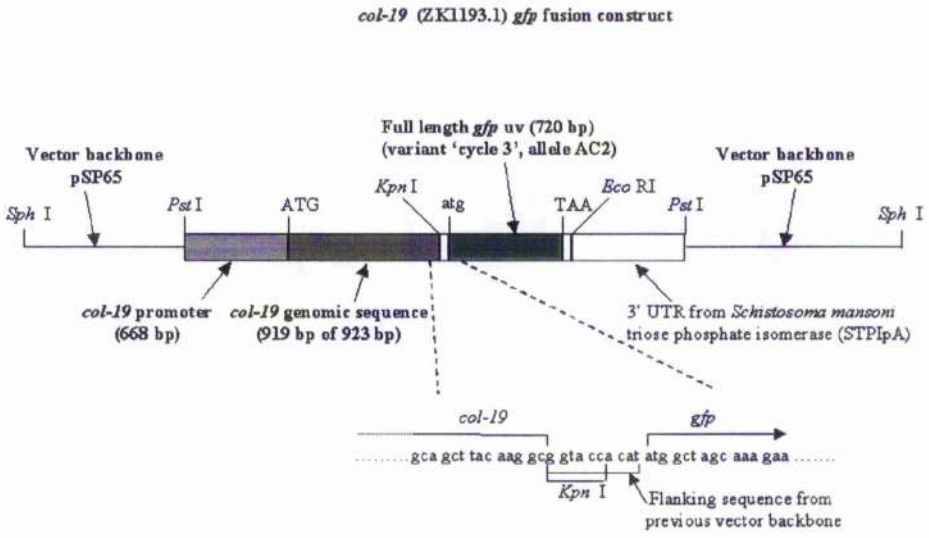


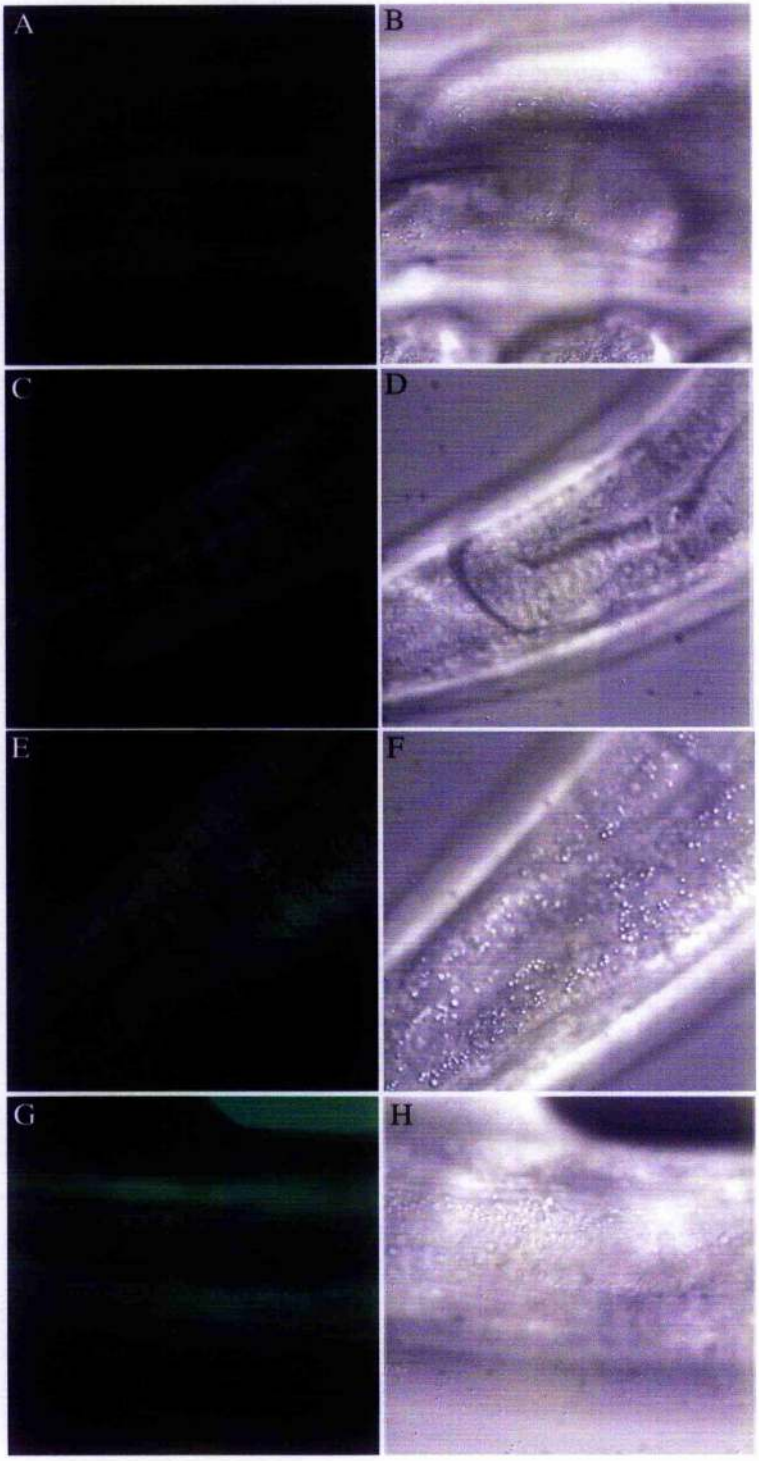
Figure 3.18. *col-19::gfp* fusion construct

The *col-19::gfp* fusion construct of plasmid BA7-1 is illustrated, with genomic *col-19* fused at its C-terminal encoding region to *gfp* (in green). An expansion of the fusion region is shown to demonstrate how the sequences were adjusted to maintain the correct reading frame.

Figure 3.19. COL-19::GFP expression patterns

The COL-19::GFP expression pattern in the annuli of the adult cuticle from strain TP12 (panels A and B); TP12 was produced form the integration of a *col-19::gfp* fusion plasmid in to the wild type N2 strain. Disrupted COL-19::GFP expression pattern in strain TP13 (panels C-F); TP13 was produced form the integration of a *col-19::gfp* fusion plasmid strain in to strain CB364 [*dpy-18(e364)*]. Wild type appearing COL-19::GFP expression pattern in a *phy-2* deletion strain (JK2757) expressing *col-19::gfp* from a free array (panels G and H) . All images were taken at the same magnification.

Facing page 130


A
B
TP12
C
D
TP13
E
F
G
H
phy-2(ok177)

identical to that produced from the transgene. COL-19::GFP localisation in TP12, panel A and B, can be seen in the adult annuli and alae (alae out of focal plane). In panels C-F, the COL-19-GFP distribution pattern in strain TP13 is shown. The *dpy-18* mutant background of this integrated strain results in the abnormal collagen expression pattern, identical to that found from *Ce-phy-1* RNAi of the N2 COL-19::GFP transgenic lines. Phenotypes ranged from that shown in panel C, where collagen is distributed in much broader and less uniform stripes than the wild type, to the more severe, where in addition to mislocalisation, the collagen has a fragmented appearance. These results demonstrate that in addition to the body morphology defect indicative of a cuticle collagen defect, loss of *Ce-phy-1* results in a specific cuticle collagen phenotype. In *phy-2(ok177),* panels G and H, the distribution pattern of COL-19::GFP from the transgene was identical to that of the wild type, consistent with the wild type appearance of this strain.

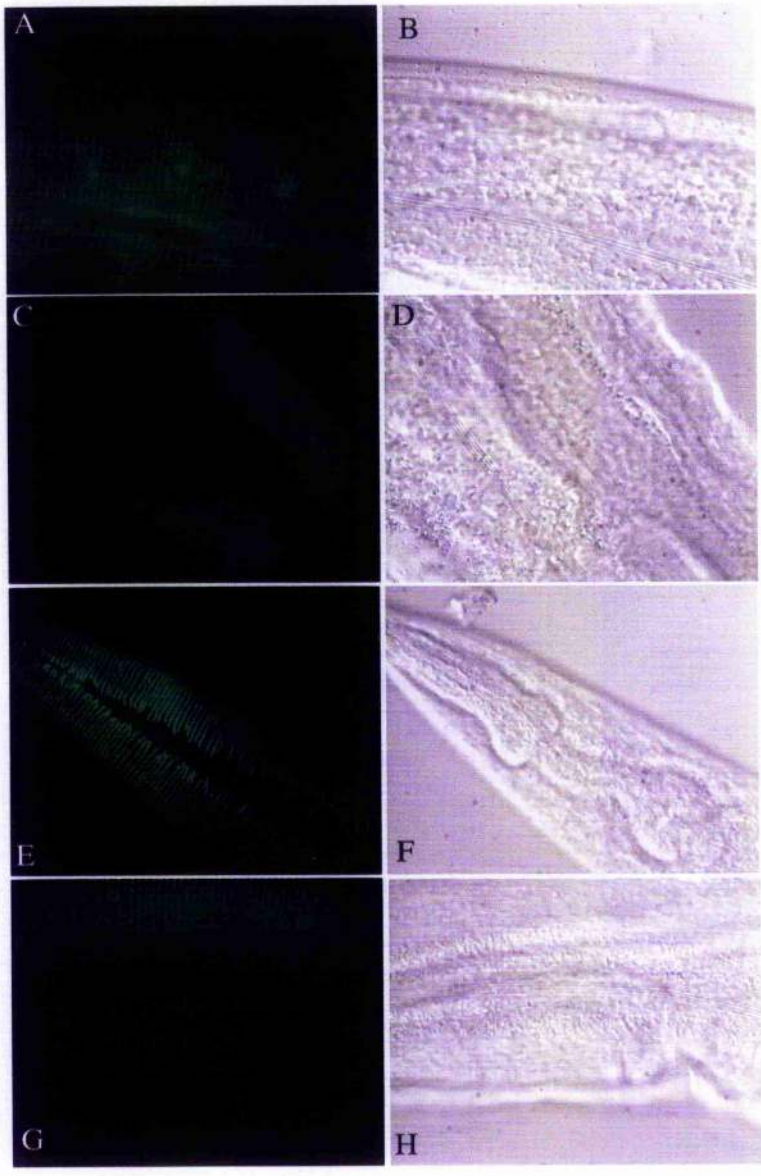
3.2.10.2. DPY-7 antibody analysis

The *C. elegans* cuticle collagen DPY-7 has an extended C-terminal domain not shared by other members of this family. A monoclonal antibody raised against this region of the protein specifically recognises the DPY-7 protein (McMahon *et al.*, 2003). The DPY-7 antibody is found to locate to the circumferential annular furrow regions of the cuticle, but not to the alae. This antibody was used as a marker for the effects of single and combined removal/reduction of *Ce*-PHY subunits. The localisation of DPY-7 was first examined in wild type, *dpy-18(e364)* and *phy-2(ok177)* strains, Figure 3.20. Panels A and B show a wild type adult with DPY-7 localising to the annular furrows with regular and evenly spaced bands, which almost contact the lateral alae. Panels C and D show localisation in the adult of the *Ce-phy-1* mutant strain *dpy-18*. The pattern of DPY-7 is clearly disrupted no longer being restricted to the defined narrow bands but localising in a more disjointed manner with larger less regular bands, and in some regions being completely absent. In panels E and F a late larval stage *dpy-18* mutant is shown with a more regular appearance of DPY-7. In wild type larval stages which lack the lateral alae structures, DPY-7 bands form over the seam cells which secrete the alae and partly interdigitate (McMahon *et al.*, 2003). This clearly does not occur for *dpy-18* mutants and while localisation to discrete bands was found, these bands do not meet and interdigitate. A correlation between the increased severity of the Dpy appearance of mutants as they get older and the more Dpy body shape towards the mid-point of the

Figure 3.20. DPY-7 immunolocalisations of *Ce-phy* mutants

DPY-7 immunolocalisations in wild type and mutant strains. Wild type N2 strain showing DPY-7 collagen localisation specifically in the annular furrows of the adult cuticle (panels A and B). Disrupted DPY-7 immunolocalisation in the adult cuticle of mutant *dpy-18(e364)* (panels C-D) and in the late larval stage of this strain (panels E-F). The wild type appearance of DPY-7 immunolocalisation in an adult of the *phy-2* deletion strain is shown in panels G and H. All images were taken at the same magnification.

Facing page 132



N2

dpy-18(*e364*)

phy-2(*ok177*)

body, was reflected in the DPY-7 collagen phenotypes of *dpy-18* mutants. DPY-7 protein mis-localisation was more severe in the adults, and in the larvae was more pronounced further from the head and tail regions. The localisation of DPY-7 in the *phy-2* deletion mutant is shown in panels G and H where localisation appeared wild type, the stage depicted is an adult, for comparison with panels A and B. These finding for the single *Ce-phy* mutant strains were in agreement with their morphological characteristics, with wild type collagen distribution for the wild type appearing *phy-2* deletion strain and collagen mislocalised or absent for the abnormal body shape *dpy-18* strain. These findings with a characterised monoclonal antibody support the findings of collagen specific disruption found using a GFP tagged collagen, and suggests that the disruption found for these two very different collagens would be reflected in all the cuticle collagens.

The studies on single mutants of *Ce-phy* genes with a collagen marker confirmed an important role for at least *Ce-phy-1* in collagen modification. As both genes were found to be essential for embryonic development and body morphology, the effect of double removal/disruption on collagen distribution was assessed by analysis of the pattern of DPY-7 expression in RNAi fed mutants. In Figure 3.21 the distribution of collagen in *Ce-phy-2* RNAi fed *dpy-18* mutants (panels A and B) and *Ce-phy-1* fed *phy-2(ok177)* mutants (panels C-F) is shown. These show that upon removal/interference of both *Ce-phy* genes that DPY-7 collagen deposition is severely decreased, with some intra-cellular retention visible in panels C and E. Also the increase in severity of body morphology upon disruption of both genes function is shown to be accompanied by an increased severity in cuticle collagen disruption.

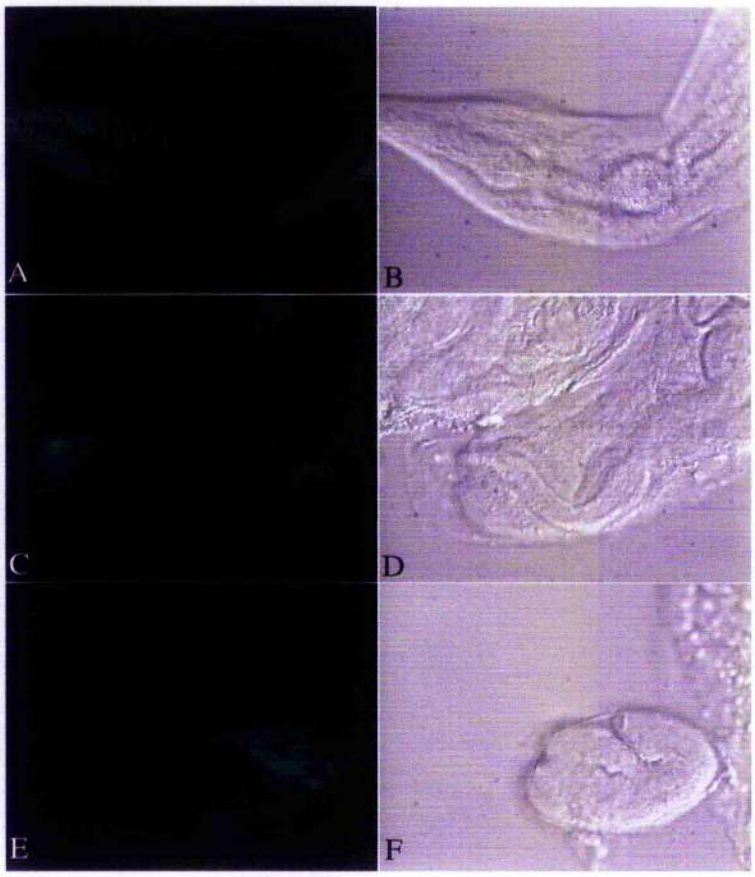
3.2.10.3. Scanning electron micrographs

The surface detail of *dpy-18* mutants and *dpy-18* mutants fed with *E. coli* expressing dsRNA for *Ce-phy-2* was examined by scanning electron microscopy. In Figure 3.22 panels A and B show the overall shape and detailed appearance of visible cuticular structures of a *dpy-18(e364)* adult. The lateral alae, that in wild type run straight along the length of the nematode, are in this mutant disordered, fragmented and branched; indicating loss of *Ce-phy-1* function affects formation of these structures. The evenly spaced annulae that run around the wild type animal appear to be formed to some degree but are irregular and disjointed due to the pitted appearance of the cuticle. Panels

Figure 3.21. DPY-7 immunolocalisation of RNAi fed mutant strains

DPY-7 immunolocalisations in *dpy-18* mutants fed with *Ce-phy-2* RNAi expressing bacteria (panels A and B), and in *phy-2* deletion mutants fed *Ce-phy-1* expressing bacteria (panels C-F). Some intracellular retention of collagen is evident in larva and embryo in panels C and E respectively. All images were taken at the same magnification.

Facing page 134



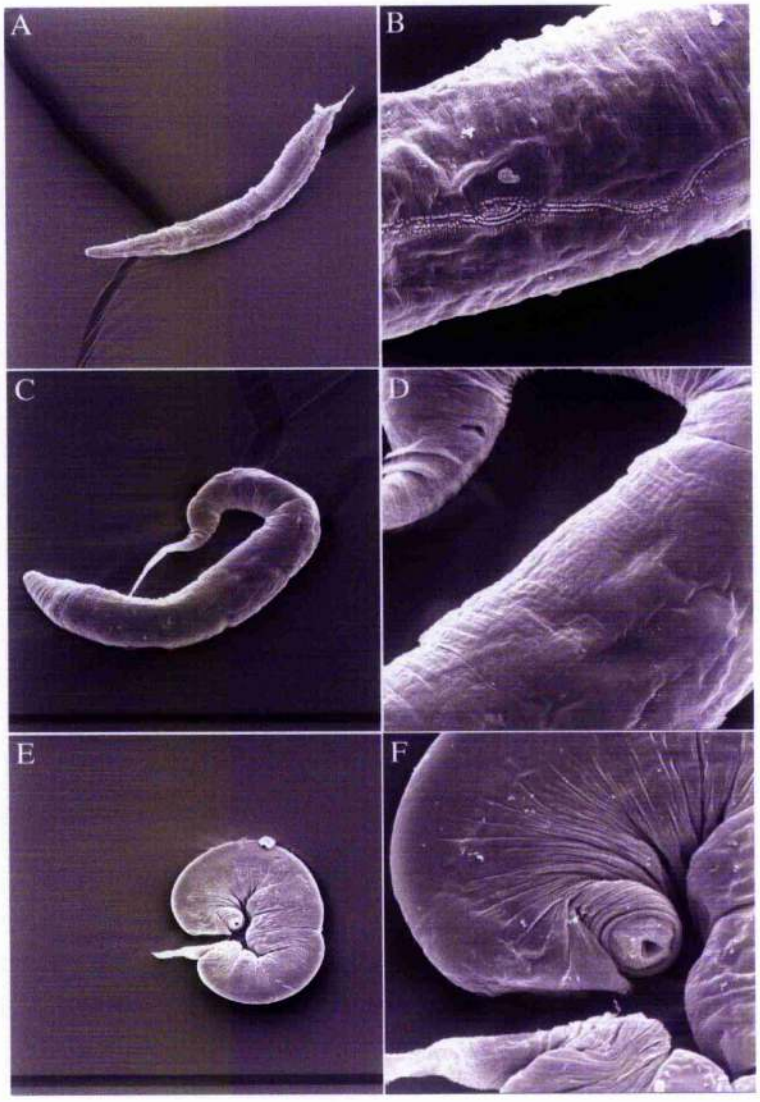
Ce-phy-2 RNAi
dpy-18(*e364*)

Ce-phy-1 RNAi
phy-2(*ok177*)

Figure 3.22. Scanning electron micrographs of mutants

Scanning electron micrographs of *dpy-18*(*e364*) (panels A, x50; and B, x400) *dpy-18* mutants fed *Ce-phy-2* RNAi expressing bacteria (panels C,x200; D, x800; E, x200; and F,x800).

Facing page 135


A
B
C
D
E
F

C-E show the effect of RNAi of *Ce-phy-2* in the *Ce-phy-1* mutant background. The highly distorted body shape is evident at lower magnifications (panels C and E) and the lack of surface cuticle structures is shown in panels D and F. Neither individual shows visible annulae or alae and the surface particularly in F has a creased appearance. Therefore removal/disruption of the *Ce-phy-1/Cephy-2* gene functions, in addition to disrupting overall body shape and collagen expression, affects the appearance of the cuticle surface and surface structures.

3.2.11. *dpy-18* rescue with alternative α subunits

3.2.11.1. Human P4H α subunit rescue constructs

Humans have two P4H α subunits (Annunen *et al.*, 1997; Helaakoski *et al.*, 1989) similar to the two conserved isoforms in *C. elegans*. The human α subunits both occur naturally in a tetramer of the form $\alpha_2\beta_2$ (Annunen *et al.*, 1997; Vuori *et al.*, 1992a) with the β subunit being human PDI (Pihlajaniemi *et al.*, 1987). Co-infection experiments suggest that a human mixed tetramer consisting of $(\alpha I)(\alpha II)(\beta)_2$ does not exist (Annunen *et al.*, 1997). The β subunit from humans can function to form an active *C. elegans* P4H enzyme in an insect cell expression system (Veijola *et al.*, 1994), and the β subunit from *C. elegans*, *Ce*-PDI-2, forms an active tetramer when coexpressed with the human αI subunit (Veijola *et al.*, 1996a). However this tetramer from human αI and *Ce*-PDI-2 is formed much less frequently than seen with the human α with human PDI, resulting in extracts with approximately 33% the activity of human tetramers (Veijola *et al.*, 1996a). Phenotypic rescue of the *C. elegans* P4H mutant, *dpy-18*, with human αI and αII subunit encoding genes was attempted in order to assess inter-species conservation of gene function. The ability of these genes to rescue could also determine if the association found for αI in insect cells can also occur in the nematode, and if the αII subunit could also form with *Ce-pdi-2*

Rescue experiments with the *phy-1* null mutant, *dpy-18(e364)*, were performed using the α subunit coding sequence with *C. elegans* upstream and downstream flanking sequences. A vector was constructed which contained the *C. elegans phy-1* promoter and 3' UTR (vector pAW1), Figure 3.23. Similar regions of *Ce-phy-1* were used in the rescue construct used to define *dpy-18* as a *Ce-phy*-1 mutant. Additionally the promoter region used had been defined as directing strong hypodermal reporter gene expression

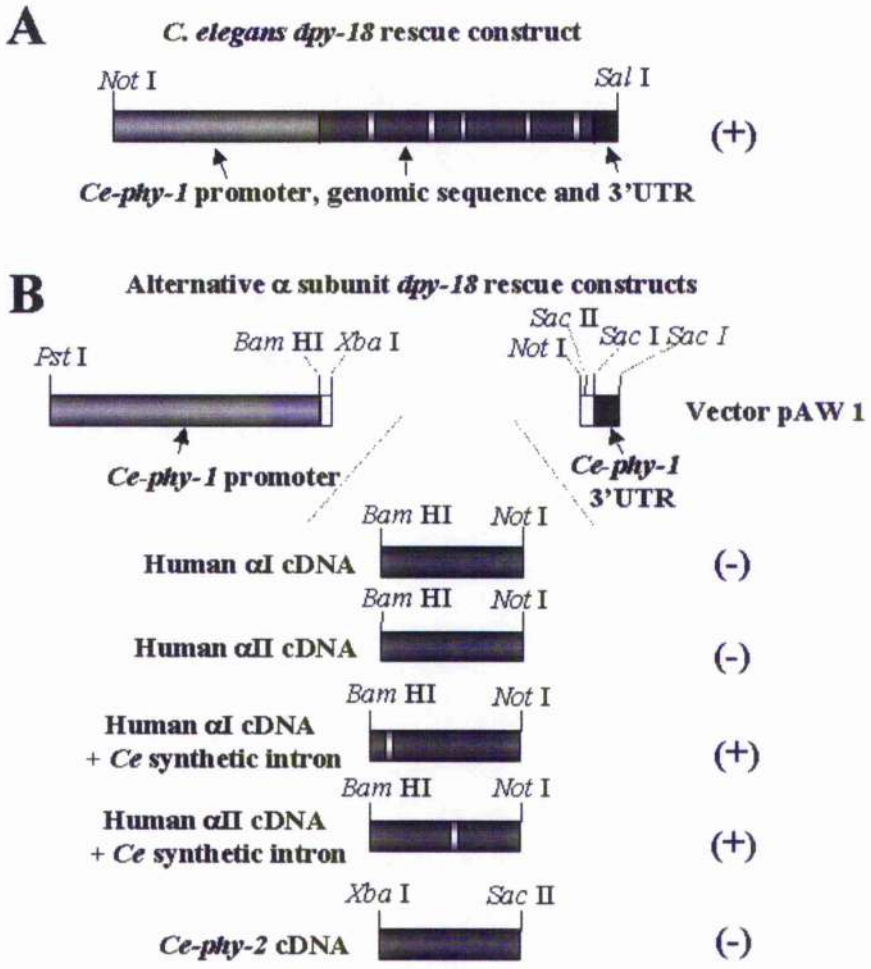


Figure 3.23. *dpy-18* mutant rescue with alternative α subunits

(A) Depiction of the *C. elegans phy-1* gene construct (containing promoter, genomic and 3' UTR sequences) used to successfully rescue *dpy-18* mutants, indicated with, (+). (B) Representation of vector pAW1 with inserted sequences. Vector pAW1 consists of the *Ce-phy-1* promoter and 3' UTR. Between these sequences human P4H α subunits cDNAs, α subunits cDNAs with a single *C. elegans* synthetic intron and *Ce-phy-2* cDNA were inserted. Sizes of fragments are not drawn to scale. Constructs were injected into the *C. elegans* P4H mutant *dpy-18(e364)*. Rescue with transformed sequences is indicated with, +; Lack of rescue is indicated, (-)

in *C. elegans*. The α subunits should therefore be expressed at high levels with the appropriate developmental timing and tissue specificity. The 3' UTR region incorporates the polyadenylation signal sequence and poly(A) transfer site, and is almost identical to the region used for rescue of *dpy-18* with *Ce-phy-1*. The human α genomic sequences were though to be prohibitively large for a PCR based cloning procedure at a size of 69 kb for αI (Helaakoski *et al.*, 1994) and 35 kb for αII (Nokelainen *et al.*, 2001). Therefore full-length cDNA coding sequence for human P4H subunits αI and αII were produced by PCR cloned from human lung cDNA. Expression from transgenes of human cDNAs has been demonstrated to produce functional products in *C. elegans* (Levitan *et al.*, 1996; Sugimoto *et al.*, 1995). Although different splice variants exist for both the α subunit encoding genes (Helaakoski *et al.*, 1989; Nokelainen *et al.*, 2001), no attempt was made to identify the specific splice variants cloned. However the αI splice from (utilising exon 9), is the more abundant form in lung tissue (Helaakoski *et al.*, 1994) from which the cDNA was derived. Both forms of αII splice variants appear to be expressed in equal proportions in lung tissue and both produce equally active enzymes in insect cells (Nokelainen *et al.*, 2001). Likewise recombinant αI produced from either splice variant can assemble into tetramers with identical properties (Vuori *et al.*, 1992a). Clones were sequenced at their junctions using primers M13 Reverse and pAW seq F. The human α subunit proteins produced from these chimeric constructs should both have four additional amino acid at their N-terminus. This is due to the reading frame set up by the CE *phy-1* promoter which includes the first five bases of the *Ce-phy-1* coding sequence. Using the Signal P signal peptide prediction site the extra amino acids were predicted not to influence the native signal peptide cleavage sites of either protein, and were therefore assumed not to interfere with signal peptide function.

Human αI and αII cDNA rescue clones, Figure 3.23, were injected separately into *dpy-18(e364)* at concentrations of 100 ng/μl together with the plasmid dpy-7-GFP at 5 ng/μl as a marker, and injection mixes made up to a final concentration of 150 ng/μl with pBluescript SKM. Transformation with pBluescript SKM and dpy-7-GFP alone had already been assessed and found to have no effect on the phenotype of *dpy-18* mutant nematodes. Transformants were selected by GFP fluorescence and over five semi-stable transmitting lines examined for each construct. Neither construct was capable of rescuing the *dpy-18* phenotype. No change in body morphology of transformed worms

was evident when compared to both *dpy-18* and non-transformed worms from injected lines.

3.2.11.2. Human α rescue with *C. elegans* synthetic introns

Although cDNA clones appear to be expressed for some genes, lack of introns in these constructs could result in transgene expression being extremely low or absent. Comparison of expression levels of intron containing and intronless transgenes in transgenic animals and tissue culture demonstrated increased expression levels of around 100-fold from spliced transcripts (Brinster *et al.*, 1988; Buchman and Berg, 1988). This is thought to be a consequence of the coupling of intron splicing with export of mRNA from the nucleus (Luo and Reed, 1999). Lack of intronic sequences could result in the message remaining untranslated as it would be unable to exit the nucleus. Insertion of synthetic introns into the coding sequences of the *E. coli lacZ* gene at various positions has been demonstrated to stimulate expression of the transgene in *C. elegans* (Fire *et al.*, 1990).

An artificial intron based on typical *C. elegans* introns (Blumenthal and Steward, 1997) was synthesised and inserted into blunt-ended restriction sites of human αI and αII. For the human αI rescue construct, *Bsa* BI was the only enzyme site available which gave blunt ended termini, cuts only once within the αI coding region, and nowhere else within the plasmid. The enzyme was however sensitive to methylation of its recognition site which blocked digestion of the plasmid. *Bsa* BI has a recognition site of GATNNNNATC, however if the sequence is GATCNNNATC or, as it is in αI, GATNNNGATC, the underlined A is methylated. This occurs in standard strains of *E. coli* that contain the methylase encoded by the *dam* gene. The αI rescue plasmid was therefore transformed in to the *E. coli* strain GM2163 which is *dcm* and *dam* minus and digestable plasmid DNA was purified. The double-stranded oligonucleotide was ligated into the *Bsa* BI site. The plasmid was then sequenced to confirm the presence of the synthetic intron in the correct orientation and to ensure no flanking sequences had been lost. Therefore a *C. elegans* synthetic intron sequence was inserted 374 bp into the coding sequence of αI. Methylase-deficient strains can show increased rates of recombination, therefore, the plasmid was digested with *Ssp* I which cuts a number of times within the inserted sequences of the plasmid. Digestion with the αI rescue construct with synthetic intron plasmid with this enzyme produced the expected pattern

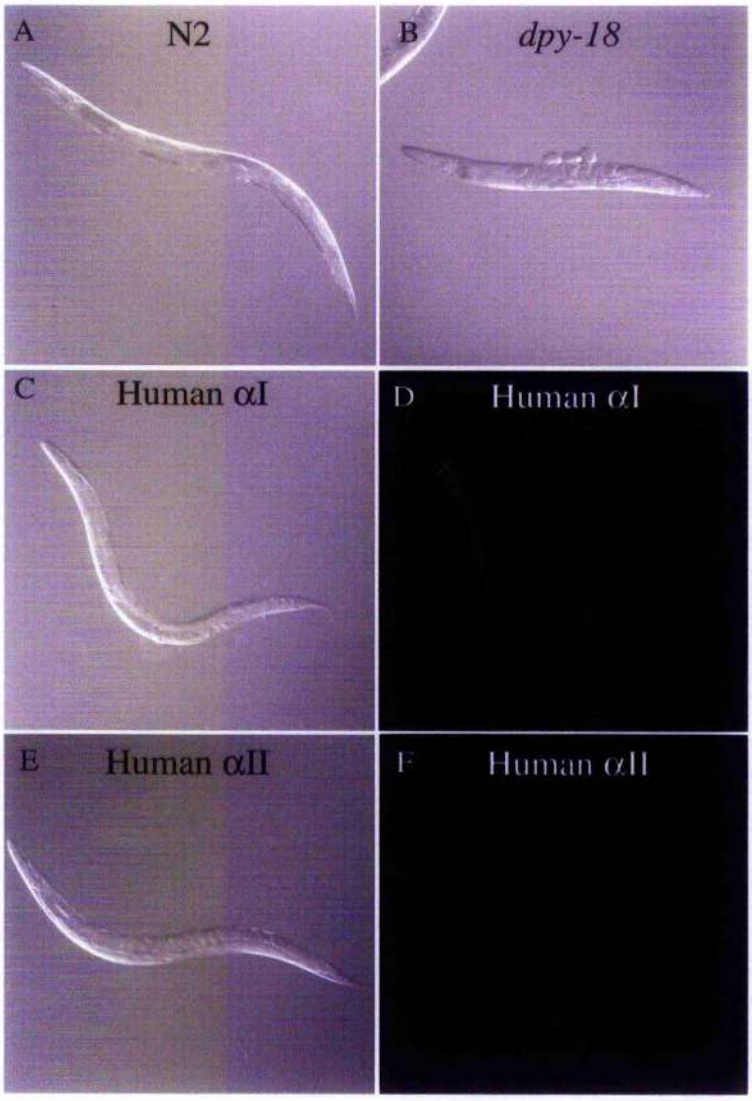
of fragments indicating that no recombination had occurred by growth in the methylase deficient strain (data not shown). The αII rescue construct was digested with restriction endonuclease *Stu* I to provide a blunt ended site for insertion of the *C. elegans* synthetic intron sequence and construct sequenced. The intronic sequence was inserted 837 bp into the αII cDNA coding sequence.

Both αI and αII synthetic intron-containing rescue constructs, αI SI and αII SI, Figure 3.23, were injected separately into *dpy-18*(*e364*) at a concentration of 100 ng/μl with dpy-7-GFP (5 ng/μl) and pBluescript SKM (50 ng/μl). Injection of the αI SI plasmid was found to partially repair the *dpy-18* phenotype. Five transgenic lines were examined and rescue to wild type body length and shape was found to be incomplete, although, as can be seen in Figure 3.24, transformed worms were noticeably longer than *dpy-18* mutants. Not all transformed worms were rescued, although no untransformed nematodes displayed more wild type body shape. The lines were not however particularly healthy, and varied slightly in the degree of repair, some also displayed occasional sterility in individuals. Only two lines could be generated with αII SI rescue plasmid as it was difficult to generate sufficient quantities of F1s. These lines showed all the same effects as the αI rescue, although repair was not quite as pronounced. To be capable of partial repair of the *dpy-18* phenotype, the human α subunits must be expressed from the transgene and associate with the nematode β subunit *Ce*-PDI-2. αI had already been shown to form an active subunit with this protein in an insect cell system. These results indicate strongly that human αII can also associate with *Ce*-PDI-2 to form an active enzyme. Additionally these hybrid complexes must then recognise *C. elegans* collagens as a substrate and be able to perform hydroxylating and chaperone functions. Despite their diverse forms human collagens are predominantly modified by human αI and αII complexes. The *C. elegans* collagens are however distinct from most vertebrate collagens although they are similar in some respects to vertebrate FACIT collagens. Hybrid complexes must be able to modify these nematode collagens pointing to a high degree of functional conservation between these highly divergent organisms. As much of our understanding of these enzymes and their substrates has come from examination of vertebrate systems, this conservation of function justifies the parallels drawn between systems. The partial rescue found suggests that if a fully active complex was formed the difference in substrates resulted in collagens modified in a non-wild type manner. Alternatively a complex may be formed that did not provide high activity

Figure 3.24. *dpy-18* mutant rescue with human α subunits containing intronic sequences

Panel A shows an N2, panel B a *dpy-18* mutant. Panels C (DIC) and D (GFP expression under UV) show partial repair of the *dpy-18* phenotype with the human αI P4H subunit containing a synthetic intron expressed from the construct pAW1. Panels E (DIC) and F (GFP expression under UV) show partial repair of the *dpy-18* phenotype with the human αII P4H subunit containing a synthetic intron expressed from the construct pAW1.

Facing page 141


A
N2
B
dpy-18
C
Human αI
D
Human αI
E
Human αII
F
Human αII

levels possibly due to the α subunits not associating efficiently with *Ce*-PDI-2 as suggested by insect cell expression data (Veijola *et al.*, 1996a).

3.2.11.3. *dpy-18* rescue with *Ce-phy-2* coding sequence

Ce-phy-2 sequences were also expressed in this system to determine if this would provide any degree of repair of the *Ce-phy-1* null *dpy-18* phenotype. The form of complexes containing the *Ce*-PHY-2 subunit was not known at this point although the genetic and RNAi data indicated that *Ce*-PDI-2 was the common subunit for both *Ce*-PHY-1 and *Ce*-PHY-2 (see Chapter 4). It was therefore thought that the *Ce*-PHY-2 could be found in a separate complex to *Ce*-PHY-1 with a common β subunit, *Ce*-PDI-2, in an analogous fashion to the human αI and αII complexes which are each formed separately in combination with a common β subunit. The *Ce-phy-2* cDNA coding sequences were therefore cloned between the *Ce-phy-1* control sequences in vector pAW1, Figure 3.23, to determine if either the slight differences observed in timing of subunit expression could influence phenotype or if simple overexpression of this protein could rescue the deficiency of *Ce*-PHY-1. Therefore if different developmental timing was required expression of *Ce*-PHY-2 under control of the *Ce*-PHY-1 promoter from a transgene could provide sufficient protein at the appropriate developmental time to modify collagens. Plasmid was injected into *C. elegans dpy-18* mutants at 100 μg/ml with transformation marker dpy-7-GFP at 5 μg/ml, and the final concentration made up to 150 μg/ml with pBluescript SKM. Transformation at this concentration of rescue plasmid was toxic with transformed animals dying as embryos. Transformants were generated with injection at 10 μg/ml which was slightly toxic but does not rescue the *dpy-18* phenotype. Insertion of synthetic introns was not possible for *Ce-phy-2* due to the lack of appropriate restriction sites.

3.2.11.4. Detection of transgene expression

Expression of the human α subunit transformed *dpy-18*(*e364*) lines was assessed using RT-PCR from single worms. Transformed lines containing the cDNA without the synthetic intron construct were not assessed for expression. The constructs with synthetic introns were examined, although differentiation of spliced and unspliced transcripts, or product arising from contaminating plasmid DNA, was not practical due to the small size of the intron. Instead primers were designed to span the synthetic intron-exon boundaries in such a way that only 5-6 bp of the 25-26 bp primers lay to the

5' end of the intron. On un-spliced mRNA derived product or contaminating DNA, the primer should anneal beyond the 3' end of the intron sequence and so not yield a product as the 3' bases of the primer would be unable to anneal (see Figure 3.25). If the gene is transcribed and the intron correctly spliced this primer should anneal and used in conjunction with a standard forward primer should produce a product. PCR product arising from correctly spliced RNA could also be confirmed by digestion of the product with *Bsa* BI (αI) or *Stu* I (αII) as these site would be reformed upon correct processing of the RNA.

Primers were tested on αI and αII plasmids with and without synthetic introns. Primers gave the correct sized band from the cDNA clones and gave no product from the clones containing the synthetic intron as expected (data not shown). Therefore in single worm RT-PCR a product should only result from reverse transcribed mRNA and not from unspliced mRNA or any plasmid remaining after DNase treatment. RT-PCRs were performed on two transformed lines and untransformed controls, no expression of transgenes could be detected in the transformed lines using this method (data not shown. Western analysis was also attempted to determine if transgenic lines were expressing the human proteins. The antibody used for detection of αI was a rabbit polyclonal (Veijola *et al.*, 1996b), with αII being a mouse monoclonal (Annunen *et al.*, 1997), both were a gift from J. Myllyharju (Collagen Research Unit, University of Oulu, Finland). Progeny were collected from transformed *dpy-18* lines containing the synthetic intron rescue constructs. No expression of αI or αII could be confirmed at the protein level (data not shown).

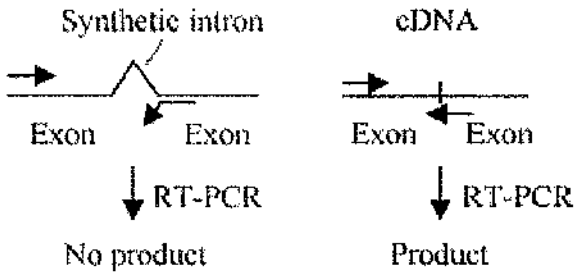


Figure 3.25. Transgene RT-PCR primer design

Depiction of predicted effect of PCR with primers designed to detect spliced product on intron and non-intron containing templates.

3.3. Discussion

3.3.1. Expression of *Ce-phy* genes

The first aim of this chapter was to define the temporal and tissue-specific expression of *Ce-phy-1* and *Ce-phy-2* in order to assess the likelihood of a role for the encoded enzymes in the biogenesis of the *C. elegans* cuticular ECM. Expression of these genes was demonstrated in the cuticle synthesising hypodermis, in all the identifiable cells of this tissue. This spatial expression pattern was confirmed using two different marker plasmids. Both hypodermally (*rol-6*) and non-hypodermally expressed (*unc-76*) transformation markers were employed. No differences were observed in the expression patterns generated with the different markers. This indicated that transcriptional regulatory units within marker plasmids were not driving reporter gene expression and producing the expression patterns observed. Thus the expression of *Ce-phy-1* and *Ce-phy-2* in the hypodermis was likely to reflect their endogenous expression pattern. The expression of both genes throughout post-embryonic development was examined using a semi-quantitative RT-PCR approach. This process has been used to very finely define the temporal expression of a number of collagen genes (Johnstone and Barry, 1996; McMahon *et al.*, 2003). The expression of the *Ce-phy* genes was found to be very similar to the collagens examined with expression peaking once within each larval stage and falling to near starting levels at times around the larval moults. Low level expression was also found in the adult stage for both genes. Some growth of the adult does occur and some collagens are expressed at this stage (Johnstone and Barry, 1996), which would require accompanying expression of P4H. The first larval stage cuticle is synthesised during embryogenesis and expression of both *Ce-phy* genes was demonstrated in the embryo, consistent with the encoded enzymes being required throughout development for the formation of cuticle collagens of all stages. The almost identical expression pattern of these two genes suggested overlapping or combined roles in modification of these collagens.

3.3.2 Single and combined mutant phenotypes indicated that the function of *Ce-phy* genes are essential for development and body morphology

The *C. elegans dpy-18* strain was identified as being mutant in *Ce-phy-1*, and was the first identified mutant in the P4H class of enzyme for any organism. Three alleles of *dpy-18* were identified in this study, two were point mutations that introduced an early stop codon, predicted to cause premature termination of translation and production of

non-functional peptides. One allele was a deletion mutant missing the promoter and N-terminal coding sequences which would cause loss of the ER signal peptide. No missense point mutations were identified that would cause amino acid replacement mutations of essential residues. A number of residues required for P4H function have been identified in vertebrate genes including the conserved active site histidine, aspartic acid and lysine residues (Lamberg *et al.*, 1995; Myllyharju and Kivirikko, 1997) and intra-molecular disulphide bonding cysteine residues (John and Bulleid, 1994; Lamberg *et al.*, 1995). These residues are present in both *Ce-phy-1* and *Ce-phy-2,* however in this analysis, and two other studies (Friedman *et al.*, 2000; Hill *et al.*, 2000), confirmation of their requirement in *Ce*-PHY-1 was not demonstrated by specific mutation in any of the mutants examined. However, the eight remaining uncharacterised mutant alleles available for *dpy-18* may contain some mutants for which these residues may be substituted. Additionally, analysis of this kind may reveal other residues essential for functional enzyme. Therefore identification of *dpy-18* as a *Ce-phy-1* mutant enables the analysis of the molecular changes in these mutants which could possibly reveal residues involved in subunit associations or substrate interactions, the later of which could be important for both the enzymatic and chaperone functions of these enzymes.

The mutant body morphology of *dpy-18* was similar to other cuticle-related defects from mutant collagens (Johnstone, 2000) and collagen processing enzymes. Mutation of certain cuticle collagen genes such as *dpy-7* (Johnstone *et al.*, 1992) and *dpy-10* (Levy *et al.*, 1993) result in a similar medium Dpy phenotype to that observed for *dpy-18*. The null phenotype of these genes is also Dpy indicating that the mutant phenotype resulting is due to lack of collagen in these cases. Thus a reduction in the function of a hypodermally expressed P4H enzyme has a similar effect on overall body morphology as the removal of certain key collagens. The degree to which the removal of P4H activity resulting from the loss of the *Ce*-PHY-1 protein affected the level of 4-hydroxproline in the cuticle collagens was assessed by amino acid analysis of mutant and wild type cuticle (Friedman *et al.*, 2000; Winter and Page, 2000). These studies showed that the cuticles of *dpy-18* animals had 4HP levels of around 30% the level of wild type. This confirmed that removal of *Ce*-PHY-1 had the predicted effect on the cuticle collagens and also indicated that there was a residual level of P4H activity in these mutants. This remaining activity could be predicted to result from the activity of the most similar enzyme in the *C. elegans* genome, *Ce*-PHY-2.

The *Ce-phy-2* gene was previously uncharacterised and for which no biochemical, functional or genetic data existed. Single removal of this gene's function by RNAi revealed no gross phenotype. This was later confirmed by analysis of a genetic mutant which was wild type in appearance (Friedman *et al.*, 2000). Therefore the combined removal of both *Ce-phy* genes was examined to determine if disruption of *Ce-phy-2* would have an additional effect in the background of the medium Dpy *Ce-phy-1* null phenotype. This was done in two ways, by simultaneous injection of dsRNA corresponding to both sequences, and by disruption of *Ce-phy-2* sequences in a *dpy-18*(*e364*) genetic background. Both approaches produced similar though distinct results. The effect of combined removal/reduction was much more severe than the medium dumpy appearance of *dpy-18* with severe malformation of body shape evident. The effect of combined RNAi injection resulted in a severe Dpy appearance with worms remaining in a coiled position. Removal of gene function in a *dpy-18* background resulted in 90% embryonic lethality in the progeny of injected animals. The development of these embryos was observed over time to determine what stage in embryonic development they reached. *Ce-phy-2* RNAi/*dpy-18* embryos were photographed every 30 minutes from the comma stage and were shown to develop normally until after the 3-fold stage. After this time the worm has elongated and the cuticle becomes responsible for maintenance of this shape. At this stage the embryos were wild type in appearance and movement, however after this point the worm gradually retracted back from their elongated form, ceased movement, failed to hatch and died. The progeny from combined injections were slightly less severe than these worms and did manage to hatch but otherwise the phenotype of the hatched coiled larvae and unhatched embryos was comparable. This similarity excluded any likelihood of the embryonic lethal effect being a non-specific lethal effect as a result of increased sensitivity to RNAi, rather than the effect on the cuticle collagens of removal of both conserved *Ce-phy* genes. It also showed the incomplete penetrance of combined RNAi as the effect was less severe using this method. Interference of more than one gene by RNAi may therefore result in a reduced effectiveness of the interference process. The hydroxyproline levels of severely Dpy larvae and embryos was not determined but would be predicted to be negligible due to the extreme effect on the cuticle observed.

The embryonic lethal phenotype of these P4H disrupted nematodes is similar to that described for a mutant allele of the *C. elegans* cuticle collagen *sqt-3* (Priess and Hirsh, 1986). *sqt-3*(*e2117*) is a temperature sensitive mutation with embryos elongating and retaining their elongated shape at the lower permissive temperature but which at higher growth temperatures retract. Comparison of embryos raised at the different temperatures demonstrated that the major structural striated layer is missing at the restrictive temperature, confirming a cuticle specific defect in these animals. Both the retraction after normal elongation and the terminal phenotype of these embryos are very similar to those resulting from the combined removal of *Ce-phy-1* and *Ce-phy-2*. The detailed embryonic lethal phenotype of the four cuticle collagen genes; D2023.7, C39E9.3, F38B6.5 and C09G5.6 (*bli-1*), identified by the genome-wide RNAi screens have not been analysed in detail. However one could hypothesise that these would display a similar elongation and retraction embryonic lethal phenotype.

RNAi feeding phenotypes of each *Ce-phy* gene in strains bearing a null mutation of the other gave a less severe phenotype than RNAi injection of *Ce-phy-2* in a *dpy-18* background. However these were also informative in a manner analogous to that of less severe alleles of embryonic lethal genes. The reason proposed for the embryonic lethality of double *Ce-phy* removal/disruption was the generation of a malformed cuticle that was unable to maintain the normal elongated form. This proposal is strongly supported by the viable feeding RNAi phenotypes, which were extremely malformed. The phenotype of these mutants was also examined by scanning electron microscopy which reveals the lack of properly formed surface cuticle structures such alae and annuli.

3.3.3. Disruption of gene function is associated with disruption of cuticle collagens

The specific effect of removal/reduction of *Ce-phy-1* and *Ce-phy-2* was examined using the localisation and expression of two collagens as a marker. The adult specific cuticle collagen COL-19 was examined via a GFP fusion transgene which was integrated into the wild type N2 strain (generating strain TP12) and *dpy-18*(*e364*) (generating strain TP13) and analysed from a free-array in a *Ce-phy-2* deletion strain. This analysis showed that loss of *Ce-phy-2* alone gave no change in COL-19 expression from that found in TP12. However, loss of *Ce-phy-1* caused mis-localisation of the collagen within the cuticle, as it is no longer uniformly localised to the annuli. Additionally

COL-19 takes on a fragmented thread-like appearance in some individuals. These findings were supported by analysis of DPY-7 using a monoclonal antibody. Again the *Ce-phy-2* deletion strain phenotype was identical to wild type while the *dpy-18* mutant showed mutant localisation of this collagen. Analysis of RNAi disrupted mutants showed the severity of the DPY-7 phenotype to reflect the more severe body morphology defects found in these animals. An advantage of using the RNAi feeding approach to analyse these phenotypes was that viable animals were produced; thus degradation of collagen that could occur in analysis of dead embryos was not a consideration. These studies supported the role of both the *Ce*-PHYs in modification of the nematode cuticle collagens and suggested, as was indicated from both the genetic/RNAi evidence presented here and the biochemical evidence (Winter and Page, 2000), that *Ce*-PHY-1 was the more important enzyme for this process.

3.3.4. *In vivo* evidence of the essential nature of P4H identifies this enzyme as a target for anti-nematode drug design

All previous work defining this class of enzyme as central to collagen biosynthesis was based on biochemical observations. Description of the *dpy-18* mutant and the effects of removal of *Ce-phy-2* in this genetic background as embryonic lethal demonstrated *in vivo* the essential function of this enzymatic step in the synthesis of collagens of the nematode cuticle and suggested that this is likely to be true in all species. Also as disruption of *Ce-phy-1* and *Ce-phy-2* was sufficient to induce this effect indicated that these encoded subunits of the two major enzymes involved in the biosynthesis of cuticle collagens in *C. elegans* and suggested homologues of these enzymes would have similar roles in other nematode species. Studies in the parasitic nematode *O. volvulus* identified a P4H α subunit and PDI which when expressed formed an active enzyme (Merriweather *et al.*, 2001; Wilson *et al.*, 1994), however no data was provided on the formation of this complex. Both these subunits of *O. volvulus* are highly conserved with *C. elegans* suggesting that these enzymes may play a similar essential role in filarial parasitic species as they do in *C. elegans*. Also like *C. elegans* the existence of more than one α subunit seems likely in this organism as other homologous ESTs were identified (Merriweather *et al.*, 2001). The essential role of these enzymes, identified in *C. elegans* development due to their role in modification of the nematode cuticle, identifies them as potential target for development of anti-nematode drugs. It has been shown *in vivo* for *B. malayi* (Merriweather *et al.*, 2001) and *C. elegans* (Friedman *et al.*,

2000; Myllyharju *et al.*, 2002) that inhibitors of P4H result in cuticle specific defects, highlighting the potential for development of P4H inhibitory compounds for the control of parasitic nematode species. Treatment of adult *B. malayi* parasites resulted in apparent separation of the epicuticle from the underlying structure and reduced ratios of 4 hydroxyproline were found especially from isolated embryos (Merriweather *et al.*, 2001). Additionally, a novel series of compounds have also been developed that inhibit P4H compounds both *in vitro* and *in vivo* (Franklin *et al.*, 2001) and these could be investigated for potential anti-nematode activity.

3.3.5. Collagen chaperone function of P4Hs

It is important in the analysis of the *C. elegans* P4H enzymes to take into account the chaperone function as well as the enzymatic properties of these enzymes. When hydroxylation was inhibited for human P4Hs, this enzyme bound to unhydroxylated, non-triple helical collagen chains thereby preventing their secretion (Walmsley *et al.*, 1999). This chaperone function is likely to also be found for the *C. elegans* cuticle collagen P4Hs examined here. The effects of removal of the chaperone function of these enzymes could therefore not be separated from the removal of enzymatic activity. Thus whether the effect of P4H disruption would be as severe without the combined loss of the chaperone function is difficult to establish using the methods employed here. A prediction of the result of chaperone function loss would be that partially folded or incompletely modified collagens would be secreted and not retained. However studies on P4H double mutants presented here suggest less collagen is secreted suggesting retention of incompletely modified protein. This raises the question that if P4H is removed, and hence collagen molecules incompletely modified, what molecules are involved in retaining the collagens. The *C. elegans* genome does not encode for Hsp47, the other major collagen chaperone identified. Thus other collagen specific chaperones must be involved in the process of collagen biosynthesis in the nematode. One important identified chaperone is the enzyme PDI (Wilson *et al.*, 1998) which can function independently of its role in P4H complexes as a collagen chaperone.

3.3.6. Human P4H rescue

The identification of the P4H mutant *dpy-18* also provided a means by which to assess the function of other P4H encoding genes by expressing them in this genetic background. The human P4H α subunits have been well characterised biochemically

(Kivirikko and Myllyharju, 1998) however, the ability of both to partially repair the phenotype of this mutant illustrates a remarkable degree of evolutionary conservation of function. This is especially notable considering the very different forms of collagens between nematodes and vertebrates. Repair of mutant phenotype using the human α subunits suggested that similar experiments from more closely related nematodes, such as the filarial nematode *B. malayi*, would be possible upon identification of such genes. Thus the function of these putative chemotherapeutic targets could be assessed *in vivo*. If capable of functional rescue, the effect of inhibitors on the activity of the enzyme could be rapidly assessed by analysing the effect of drug treatment on transgenic *C. elegans* body shape. Studies such as these would have the potential to complement *in vitro* studies on inhibition of recombinant enzymes with the added advantage that the *in vivo* problems of drug uptake and crossing cuticle and cell boundaries would also be assessed. Additionally, the effect on the human enzymes could be assessed in this system using the same body morphology assay as a guide to inhibitory function, with the aim of finding possible compounds that would inhibit nematode but not vertebrate enzymes. This assay would provide a means by which to extremely rapidly assess the specific inhibitory properties of compounds and identify those for further *in vitro* characterisation. Rescue of *C. elegans dpy-18* mutants with a *phy* gene homologue from *B. malayi* is addressed in Chapter 6.

3.3.7. *Ce*-PHY-1 and *Ce*-PHY-2 complexes appear not to modify type IV collagens

Although faint expression in body muscle cells was observed for *Ce-phy-2* and no expression was found in this tissue for *Ce-phy-1* it was possible that the complete expression pattern of these genes was not uncovered when transgenically determined. It was possible that all regulatory regions required for the endogenous expression pattern were not contained within the constructs tested. Thus although faint, the muscle-specific expression for *Ce-phy-2* may have reflected real and possibly much stronger expression of the endogenous gene in muscle cells. Therefore the possibility of either or both of these genes being involved in modification of the type IV collagens produced by body wall muscle cells had to be considered. The body musculature is arranged as longitudinal bands of muscle cells, with one running in each quadrant of the body. The four bands of muscle lie in grooves in the hypodermis, and at the regions where the hypodermal cells are overlaid by muscle cells they are particularly thin. Fibrous elements that extend through the cytoplasm of the hypodermal cells at these points

anchor the muscle to the cuticle and transmit the forces of muscle contraction to the cuticle to enable movement. Body muscle cells also generate type IV collagen both for regions of their own basement membranes and for the basement membranes of other tissues (Graham *et al.*, 1997).

Basement membranes are thin sheets of ECM which surround the major organs of *C. elegans*. Type IV collagen is found in nearly all regions of basement membranes, exceptions being the pseudocoelomic face of the body wall muscle and on the regions of hypodermis between body wall muscle quadrants (Graham *et al.*, 1997). Although the pharynx and intestine are covered with type IV collagen-containing basement membranes, these tissues do not express the type IV collagen genes (Graham *et al.*, 1997). Type IV collagen made in body wall muscle cells can assemble into pharyngeal, intestinal and gonadal basement membranes. These nematode type IV collagens could be predicted to contain 4-hydroxproline and hydroxylysine in the Y positions of Gly-X-Y repeats as has been shown for vertebrate type IV collagens (Kivirikko *et al.*, 1992). Severe embryonic lethal phenotypes are found in *C. elegans* from mutation of type IV collagens (Guo *et al.*, 1991; Gupta *et al.*, 1997; Sibley *et al.*, 1994; Sibley *et al.*, 1993) and their processing enzymes (Norman and Moerman, 2000). Thus a similar severe phenotype could be predicted from lack of 4-hydroxyproline in these collagens. Mutations in another component of the basement membrane causes an uncoordinated (Unc) phenotype. Products of the *C. elegans unc-52* gene are homologous to mammalian perlecan, a component of basement membranes (Rogalski *et al.*, 1993). One class of mutants in this gene are viable but develop progressive paralysis. The basement membrane has been shown to be critical for the assembly of myofilaments within the body wall muscle as in the absence of perlecan in *unc-52* mutants myofilament assembly does not occur (Rogalski *et al.*, 1993).

The lack of these phenotypes combined with the specific embryonic lethal phenotype displayed upon the combined *Ce-phy* gene removal/disruption indicated that neither gene was involved in modification of type IV collagens. The movement of developing embryos and attainment of a fully elongated form distinguishes them from the mutant phenotypes of genes involved in the formation of the basement membranes, typically two-fold paralysis. Additionally the *dpy-18* mutant strains had no detectable basement membrane associated defects. Thus while not exhaustively examined by for example

antibody analysis, the *Ce*-PHY-1 and -2 subunits were considered not to affect type IV collagen modification and thus formation of the basement membrane ECM of these nematodes.

3.3.8. Possible formations of P4H in *C. elegans*

The data at this stage suggested that two different P4H enzymes were present in *C. elegans* that modified the collagens to be incorporated into the cuticle. The genetic evidence suggested that a *Ce*-PHY-1 P4H, possibly the already identified dimer, was more active than the proposed *Ce*-PHY-2 P4H enzyme. This was verified by the biochemical data from 4HP levels of cuticle extracts of *dpy-18* (Friedman *et al.*, 2000; Winter and Page, 2000) and the *Ce-phy-2* deletion strain (Friedman *et al.*, 2000). *dpy-18* cuticle collagens had a greater reduction in 4-hydroxproline levels of than those from the *Ce-phy-2* mutant strain. The greater importance of *Ce*-PHY-1 was also indicated by the body morphology defect of nematodes mutant for this enzyme, compared to the wild type body form of the *Ce-phy-2* mutant. The effect on localisation of collagens in the cuticle also indicated a greater role for *Ce*-PHY-1. However what form either of these enzymes took *in vivo* had not been established although *in vitro* data indicated that a *Ce*-PHY-1/*Ce*-PDI-2 dimer was likely to be present (Veijola *et al.*, 1996a). *Ce*-PHY-2 could however be hypothesised to be formed from a dimer or tetramer containing this subunit in combination with a *Ce*-PDI, or could possibly be active as a monomer. Active P4Hs monomers have been characterised from *A. thaliana* (Hieta and Myllyharju, 2002), *Paramecium bursaria Chlorella* virus-1 (Eriksson *et al.*, 1999), and from unicellular and multicellular green algae (Kaska *et al.*, 1987; Kaska *et al.*, 1988). Alternatively, if a β subunit were required for *Ce*-PHY-2 to form an active enzyme, the *Ce*-PDIs would be obvious candidates. *Ce*-PDI-1 was shown not to form a complex with *Ce*-PHY-1 (Veijola *et al.*, 1996a) but was still a possible candidate for forming a complex with *Ce*-PHY-2. The RNAi phenotype of the *Ce-pdi-1* gene displayed no phenotype (Gönczy *et al.*, 2000) (A. Page, personal communication) and studies had suggested a non-P4H role for this enzyme (Page, 1997; Veijola *et al.*, 1996a). Thus the defined P4H subunit *Ce*-PDI-2 was considered a more likely candidate for further analysis. In humans the αI and αII subunits both form active $\alpha_2\beta_2$ tetramers with PDI, with evidence suggesting that complexes containing both α subunits do not form (Annunen *et al.*, 1997; Vuori *et al.*, 1992a). The mouse αI and αII subunit also individually combined with human PDI to form active P4H tetramers (Helaakoski *et al.*,

1995). Thus the most obvious hypothesis from the *C. elegans* data was a common PDI subunit, possibly *Ce*-PDI-2, for two separate P4H complexes, with the complexes being in the form of $\alpha\beta$ dimers.

Therefore this chapter determined the essential cuticle collagen modifying role for *Ce*-PHY-1 and *Ce*-PHY-2 but what remained unresolved was what the forms of P4H enzyme were which contained these subunits, and if these were the only cuticle collagen modifying P4Hs found in this nematode. These points were analysed in the following chapters, firstly by examination of the *Ce-pdi-2* gene to determine if the enzyme encoded by this gene was the likely *in vivo* subunit for either or both of *Ce*-PHY-1 and *Ce*-PHY-2 (Chapter 4). Secondly *Ce*-PHY-2 was expressed in an insect cell co-expression system to determine associations formed with identified and potential P4H subunits, and the presence of these complexes confirmed from *in vivo* extracts (Chapter 5). Finally three divergent *Ce*-PHY encoding genes were examined for a role in cuticle biogenesis (Chapter 5).

Chapter 4

C. elegans PDI-2 functions as the single P4H β subunit and is essential for nematode development

4.1. Introduction

Ce-pdi-2 was first identified from an EST sequence (McCombie *et al.*, 1992) that was then used to screen a *C. elegans* cDNA library, identifying all but the two most N-terminal amino acids (Veijola *et al.*, 1996a). The gene was subsequently sequenced by the *C. elegans* Genome Sequencing Consortium (Consortium, 1998) and assigned to cosmid C07A12.4 and its localisation on the X chromosome was confirmed. The predicted protein is highly homologous to the human P4H β subunit, or human PDI (Pihlajaniemi *et al.*, 1987). Insect cell expression of *Ce*-PDI-2 confirmed it as an active disulphide isomerase with activity similar to the human PDI (Veijola *et al.*, 1996a) based on an assay for the renaturation of reduced, denatured RNase (Lyles and Gilbert, 1991). Co-expression studies using this system demonstrated that this peptide was capable of forming active P4H complexes with α subunits from *C. elegans* and human. *Ce*-PDI-2 had been shown to form an $\alpha_2\beta_2$ tetramer with human αI, and a dimer with *Ce*-PHY-1 (Veijola *et al.*, 1996a). Another isoform of PDI exists in *C. elegans*, *Ce*-PDI-1, which has also been confirmed as an active disulphide isomerase (Veijola *et al.*, 1996a). *Ce*-PDI-1 displays lower amino acid homologies with *Ce*-PDI-2 (48% identity and 66% similarity) than *Ce*-PDI-2 shares with human PDI, suggesting a different role for *Ce*-PDI-1. *Ce*-PDI-1 has been examined in detail and was found not to form P4H complexes with any α subunits tested (Veijola *et al.*, 1996a). *Ce*-PDI-1 has been linked to the process of cuticle collagen folding by its expression pattern and its genomic organisation within an operon (Page, 1997). The downstream gene in this operon is a peptidyl prolyl *cis-trans* isomerase, a class of enzyme which is thought could be involved in the folding of the collagen triple helix (Page, 1997). This arrangement of genes is also conserved in the close relative of *C. elegans*, *C. briggsae*, further suggesting a functional relationship between these two genes (Page, 1999). RNAi phenotype of *Ce-pdi-1* gene displayed no phenotype (Gönczy *et al.*, 2000) (A. Page, personal communication) which when compared to the effects of single and combined removal of *Ce-phy-1* and *Ce-phy-2* further indicated that this gene was not a P4H subunit. Another enzymatically active PDI from *C. elegans*, *Ce*-PDI-3, has been

recently identified which has been shown to have transglutaminase activity (Eschenlauer and Page, 2003). *Ce*-PDI-3 has, like its homologue in the parasite *D. immitis* (Chandrashekar *et al.*, 1998), been defined as an ERp60-like protein. ERp60 proteins are related to PDIs but have been shown to be incapable of functionally replacing PDIs in a P4H complex (Koivunen *et al.*, 1996). These findings suggested a non-P4H role for this enzyme, although, like *Ce*-PDI-1, it is likely to have a collagen-related function, possibly being involved in cross-linking these molecules.

Ce-PDI-2 was therefore analysed to determine a role for this enzyme in the formation of P4H complexes *in vivo* with *Ce*-PHY-1 and *Ce*-PHY-2. In this chapter, expression pattern studies and analysis of RNAi knockouts were performed and compared to the results for the *Ce-phy* genes. Further characterisation of the *Ce-pdi-2* partial disruption RNAi phenotype was performed at the level of specific collagens by immunolocalisation and SEM. The locus *let-44* was examined to determine if this represented a *Ce-pdi-2* mutant.

4.2. Results

4.2.1. Examination of *trans*-splicing

The four exons gene structure of *Ce-pdi-2* is shown in Figure 4.1. The size of the 3' UTR is estimated on the basis of the ESTs with the longest 3' sequence and the likely position of a polyadenylation signal sequence in this region. Primers F(pMal) and R(pMal), indicated on Figure 4.1, were used in combination with SL1 and SL2 to examine *trans*-splicing, the results of which are shown in Figure 4.2. Lane 1 shows the product generated by gene primers used to amplify *Ce-pdi-2* from mixed stage cDNA. Figure 4.2 lanes 2 and 3 show lack of appropriately sized product being amplified by use of the SL1 and SL2 primers respectively with the gene antisense primer. This shows that unlike *Ce-phy-1* and *Ce-phy-2* this gene is not SL1 *trans*-spliced, and in addition does not appear to be *trans*-spliced by SL2. *Ce-pdi-2* is therefore included in the small group of approximately 20% of the *C. elegans* genes that are not *trans*-spliced by either SL1 or SL2 (Zorio *et al.*, 1994).

4.2.2. Protein sequence

The completed sequence predicts a protein of 493 amino acids with a proposed signal peptide of 16 amino acids, giving a mature peptide length of 477 residues. The signal

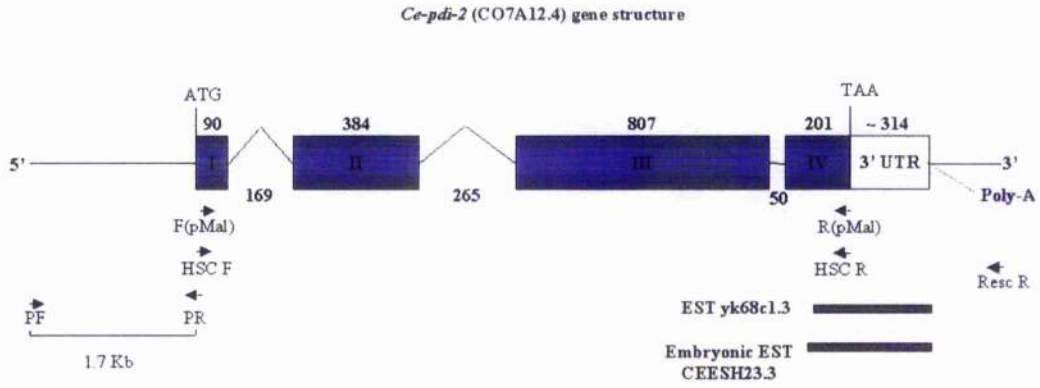


Figure 4.1. Gene structure of *Ce-pdi-2*

Gene structure of *Ce-pdi-2* is depicted with exons shown as coloured boxes, exon numbers are indicated with roman numerals, intron regions are shown as lines and 3' UTR is an open box. Sizes of intron, exon and 3' UTR regions are given in base pairs. Positions of the most 3' EST sequences are given that indicated the likely position of the polyadenylation site. Promoter region and primers are given (region not drawn to scale). The positions of important primers are illustrated.

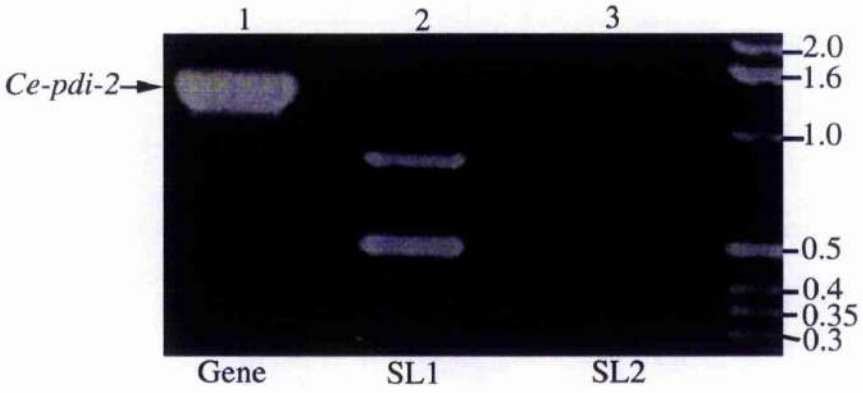


Figure 4.2. SL RT-PCR of *Ce-pdi-2*

Agarose gel of products resulting from SL RT-PCR of the *C. elegans pdi-2* gene. Lane 1 shows the product amplified using gene specific primers for *Ce-pdi-2*, lane 2 shows no product amplified by PCR using SL1 specific sense primer and the antisense gene primer, lane 3 shows no product amplified by PCR SL2 specific and antisense gene primer. The positions of molecular size standards are indicated with sizes given in kilobases.

peptide cleavage site is in agreement with the signal peptide predicted from the initial incomplete N-terminal protein sequence (Veijola *et al.*, 1996a). The C-terminus encodes an HTEL that is a variant of the KDEL ER retention signal encoded by the human PDI. The *Ce*-PHY-1 and *Ce*-PHY-2 polypeptides, like most P4II α subunits, do not contain an ER retention signal and so could be retained in this subcellular compartment by a retention signal contained within the C-terminus of *Ce*-PDI-2. The critical role of vertebrate PDI in this process is demonstrated by deletion of this region which causes the secretion of the tetramer and free PDI polypeptide from the ER (Vuori *et al.*, 1992b). A comparison of amino acid sequences of *Ce*-PDI-2 and identified P4H β subunits from *O. volvulus* (Wilson *et al.*, 1994) and human (Pihlajaniemi *et al.*, 1987) are shown in Figure 4.3 (signal peptides have been removed and numbering refers to the mature processed peptides). *Ce*-PDI-2 shows identity of 71% and similarity of 84% to *Ov*-PDI, and 59% identity and 78% similarity with human PDI. The conserved active sites residues are indicated with an asterisk. This protein therefore demonstrates a high degree of homology with other defined P4H β subunit PDIs supporting the proposed involvement of this gene as a *C. elegans* β subunit.

4.2.3. Spatial expression pattern

The putative promoter region from *Ce-pdi-2* were generated by PCR and cloned into the reporter gene vector pPD95.03. This plasmid contains *lacZ* with multi-introns and a nuclear localisation signal and has been described in Figure 3.2. Approximately 2.6 kb of upstream region plus the first five bases of coding sequence were fused in-frame with *lacZ* and the construct transformed into *C. elegans* by microinjection. Primers used to produce this fragment are illustrated in Figure 4.1. Two different strains were used for injection with two different corresponding markers. A semi-dominant allele of the collagen gene *rol-6* was used as a marker in the wild type strain, allowing identification of transformants on the basis of the roller phenotype (Mello *et al.*, 1991). Wild type *unc-76* was used as a marker for an *unc-76* mutant stain enabling selection of transformants on the basis of repair of uncoordinated movement to wild type mobility (Bloom and Horvitz, 1997). The construct was injected into both strains at a concentration of 20 μg/ml with markers at 100 μg/ml. Transformants were selected and semi-stable lines established. For each marker three or more lines were maintained and examined for βGal activity using the standard concentration of the artificial substrate

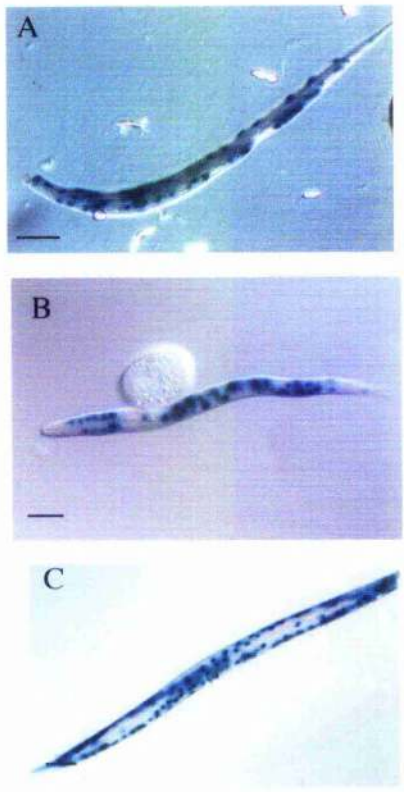
Figure 4.3. Amino acid alignment of *Ce*-PDI-2 with *O. volvulus* and human P4H subunit PDIs

Amino acid alignment of *Ce*-PDI-2 with *O. volvulus* (*Ov*) and human PDI subunits of P4H enzymes. Gaps (-) were introduced for maximal alignment and signal peptides were removed, therefore numbering refers to the mature processed proteins. Active site residues are indicated with asterisks. Genbank Accession numbers- *Ce*-PDI-2 (U41542), *Ov*-PDI (U12440) and human PDI (X05130).

X-gal. Many different nematodes were examined and all the life cycle stages were observed to determine which cells were expressing the reporter gene. No differences were noted between the different marker and strains used. *Ce-pdi-2* driven *lacZ* expression was seen in all stages from embryo to adult. Figure 4.4 panels A and B shows L1 larvae expressing *lacZ* in their hypodermal cell nuclei. Expression was detected in the L1 hypodermal cells; including the anterior H0L, H1L, hyp3, 4, 5, 6 and 7; posterior TL, hyp7, 8, 9, 10, 11; and the mid body hyp7 and lateral P, V, H2R and H2L cells. Mosaicism in expression was evident, especially in the posterior hyp and anterior hyp3 and hyp4 cells. This hypodermal pattern of expression is almost identical to the *Ce-phy* studies. In the late larval and adult stages the expression pattern becomes increasingly more complex Figure 4.4 panel C. The function of PDIs in modifying collagens would predict expression of *Ce*-PDI-2 in the hypodermis where the cuticular collagens are synthesised and secreted. No additional staining in the vulva or in muscles, as seen for *Ce-phy-1* and *Ce-phy-2,* was observed. Other than these differences, the *Ce-pdi-2* expression pattern was identical to the *Ce-phys* supporting the proposed association of *Ce*-PDI-2 in P4H complexes with both these proteins.

4.2.4. Temporal expression of *Ce-pdi-2*

To examine the expression of *Ce-pdi-2* through development, RT-PCR was performed on staged cDNA sampled from regular intervals throughout post-embryonic development. RT-PCR was performed on two hourly samples using a set of primers specific to *Ce-pdi*-2 and a control set of primers to the large subunit of RNA polymerase II encoding gene *ama-1* (Bird and Riddle, 1989), Figure 4.5A. Both sets of primers span an intron region of the gene thus enabling the cDNA signal to be distinguished from products arising from any genomic DNA contamination. Quantifying the strength of signal produced from *Ce-pdi-2* compared to *ama-1* allowed the relative abundance of the *Ce-pdi-2* transcript to be determined for each time point. The *Ce-pdi-2* temporal expression profile, Figure 4.5B, revealed that the transcript was expressed throughout larval development. Within larval stages expression levels after the moult are low, these rise during each stage peaking at approximately the mid-point. Levels then fall to near their starting point, before rising in the next stage. The single peak of expression once within each larval stage corresponds to the synthesis of a new cuticle once within each intermoult period. Cuticle collagens which have been examined in this way show a similar oscillating pattern of expression (Johnstone *et al.*, 1996; McMahon *et al.*, 2003).



D

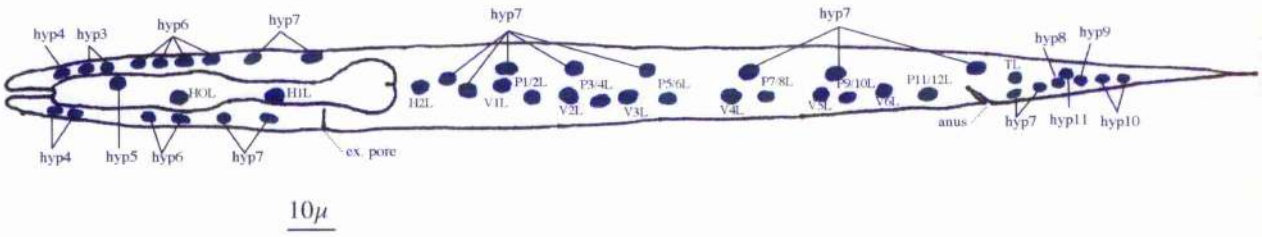


Figure 4.4. *lacZ* expression from the *Ce-pdi-2* gene promoter

Images show the *Ce-pdi-2* promoter-driven *lacZ* expression in the hypodermal cells of *C. elegans*. Panel A, *lacZ* expression in an L1 co-transformed with the *rol-6* marker plasmid; panel B, *lacZ* expression in an L1 co-transformed with the *unc-76* marker plasmid.; panel C, *lacZ* expression in an adult co-transformed with the *rol-6* marker. Scale bar on panels A and B and equals 10 μm, bar for C equals 100 μm. Panel D shows a representation of hypodermal cell nuclei in an L1 larvae.

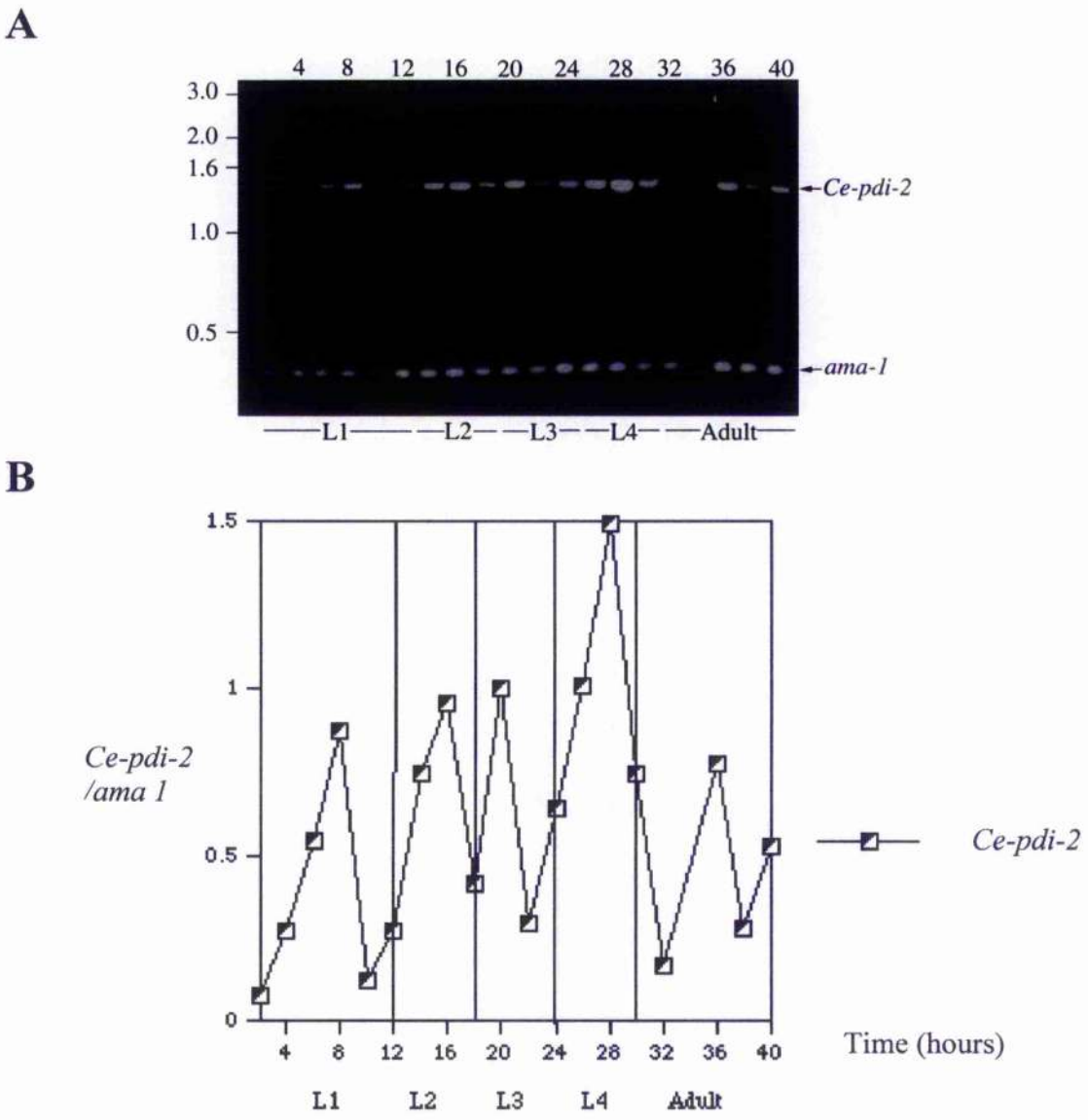


Figure 4.5. Temporal expression of *Ce-pdi-2*

(A) Agarose gel showing the timecourse RT-PCR of *Ce-pdi-2* with the control gene *ama-1* amplified simultaneously. The positions of molecular size standards are indicated with sizes given in kilobases. Along the top of the gel picture the time in hours is given which represents hours post-hatch. The developmental stages from L1 to adult that the samples represent are indicated below the gel. (B) Semi-quantititive RT-PCR graph. Expression levels of *Ce-pdi-2* transcripts in relation to *ama-1* throughout post-embryonic development are expressed in arbritary units along the y-axis. The x-axis is the time in hours post-hatch. The vertical bars represent the timing of the moults with the corresponding developmental stage indicated. A single peak of expression is found within each developmental stage.

Unlike *C. elegans phy-1* and *phy-2*, *Ce-pdi-2* has a much more pronounced peak of expression within the section of adult development examined. Some collagen genes are expressed in the adult stage and the *Ce-pdi-2* peak observed in the adult stage may correspond to the growth of the cuticle in adults. This growth would require collagen gene expression and the corresponding requirement for disulphide isomerase activity alone and/or for its function as a subunit of a P4H. It is extremely unlikely that the peak of expression seen in the adult stage could result from expression of this gene from embryos contained within the adults as the method used to prepare RNA from these nematodes does not enable extraction from embryos (Johnstone and Barry, 1996). Additionally the peak of expression in the adult stage could be due to other functions of this multi-functional enzyme. Comparison of the *Ce-pdi-2* expression with *Ce-phy-1* and *-2* shows an extremely similar profile within each cuticle synthesising larval stage. This similarity is in keeping with the proposed association of these proteins in P4H complexes and the function of these enzymes in modifying the nematode cuticle. Expression levels are seen generally, for all three subunits, to increase through the larval stages, followed by substantial reduction in adults. Less P4H activity would be required in adults as growth is accomplished by modification of the existing cuticle compared to synthesis of a complete new cuticle for each larval stage. Expression of *Ce-pdi-2* was also detected by PCR in embryo cDNA and is shown in Figure 4.6A compared to pooled samples for each of the other developmental stages. Expression in the embryo of *Ce-pdi-2* is confirmed by reporter gene analysis, Figure 4.6B.

4.2.5. RNAi injection

The function of *Ce-pdi-2* was examined in a similar manner to that of *Ce-phy-1* and *Ce-phy-2* using RNA mediated interference. The possibility of *Ce-pdi-2* sequences cross-interfering with similar genes was considered. At the nucleotide level the most similar gene to *Ce-pdi-2* is *Ce-pdi-1* but these only share 60% identity, less than that proposed to cause cross-interference (Tabara *et al.*, 1998). Additionally *Ce*-PDI-1 has been covered by one of the large scale RNAi screens and produces no phenotype (Gönczy *et al.*, 2000), indicating that any phenotype produced by interference of *Ce-pdi-2* would not be a result of disruption of the other isoform. dsRNA was produced *in vitro* for *Ce-pdi-2* by cloning almost full length coding sequence (1.4 kb) into a pBluescript like vector .

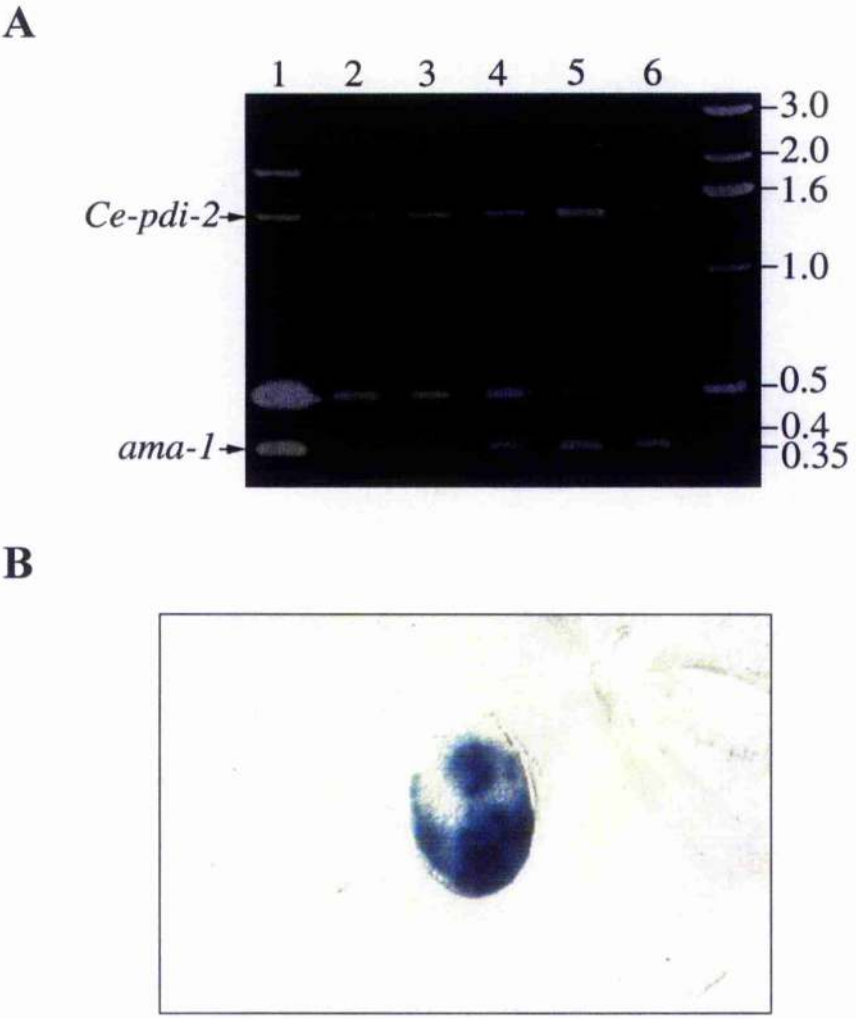


Figure 4.6. Embryonic expression of *Ce-pdi-2*

A) Agarose gel showing a comparison of embryonic expression of *Ce-pdi-2* with samples representing developmental stages L1 to adult. Lane 1 is an embryo sample, lane 2 is an L1 sample, lane 3 is an L2 sample, lane 4 is an L3 sample, lane 5 is an L4 sample, and lane 6 is an adult sample. Expression can clearly be seen in the embryo stage. (B) *Ce-pdi-2* promoter driven reporter gene expression in the developing embryo, showing expression in the hypodermal cells.

Ce-pdi-2 dsRNA was injected into the gonad of young adult hermaphrodite N2s. After a period of recovery, from 4 hours to overnight, nematodes were transferred singly to fresh plates. This period also allowed for partially developed embryos to be cleared, and progeny scored were exposed to dsRNA for their entire development. Injected worms were then transferred daily to fresh plates and progeny from previous plates scored. Injection of dsRNA for *Ce-pdi-2* resulted in embryonic lethality. Plates were scored as having dead embryos if embryos were present when the adult was removed but had not hatched after an overnight period at 20°C. Progeny were counted and dead embryo numbers compared to hatched progeny. These embryos were checked the next day to ensure no more progeny had hatched. Only plates where an effect was observed were scored as some injections were unsuccessful. Scoring of progeny showed that RNAi via injection of dsRNA for *Ce-pdi-2* gave 99% (1431 dead out of 1432 embryos laid) lethality in a wild type genetic background. Therefore single removal of *Ce-pdi-2* generated similar levels of embryonic lethality as the *Ce-phy-2* RNAi/*dpy-18* mutant combination. This suggested that this gene could encode a P4H β subunit involved in both proposed *C. elegans* P4H complexes.

4.2.6. Developmental timecourse analysis of the *Ce-pdi-2* RNAi embryonic lethal phenotype

The lethal embryonic phenotype of *Ce-pdi-2* RNAi injected nematodes was examined in more detail by monitoring their appearance throughout embryonic development. This was accomplished by injecting animals with *Ce-pdi-2* dsRNA, and then after a period of 10 hours, to ensure RNAi was effective, pre-comma stage embryos were collected and photographed at 30 minute intervals throughout development. Specific time points during this process are shown in Figure 4.7. Comma stage embryos developed normally to the three-fold stage, where elongation was completed, transforming a ball of cells into a worm shape. Embryos at this stage were active and moved, indicting correct formation of muscular structures and attachments. After the three-fold stage the mutant phenotype becomes apparent, as embryos become less active and gradually lose their elongated shape. Examination of the timing of loss of worm body morphology reveals that this corresponds exactly with the time with which the cuticle becomes the support to maintain the nematodes elongated body form, see Figure 3.16. The *Ce-pdi-2* RNAi cuticles were proposed to be weakened due to improper modification and are therefore

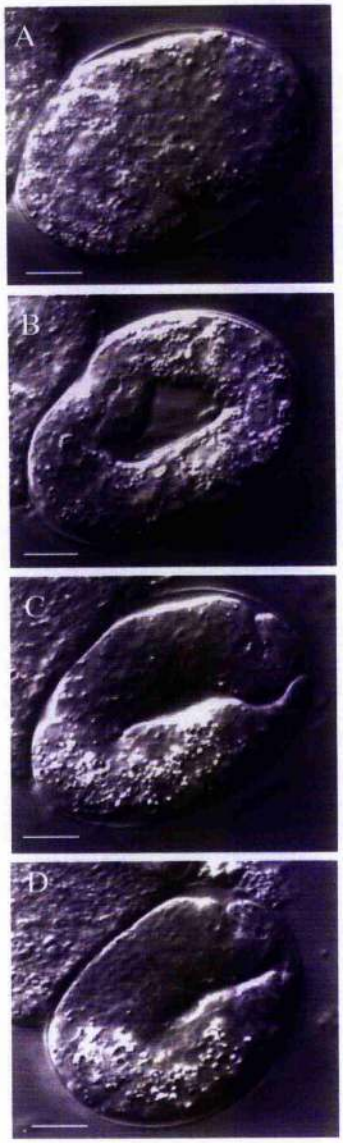


Figure 4.7. RNA interference of *Ce-pdi-2*

Development timecourse of wild type N2 injected with dsRNA for *Ce-pdi-2*. Bar represents 10 μm. (A) 1.5 fold embryo (430 minutes). (B) 3-fold elongated embryo (560 minutes) with the head and tail out of focal plane. (C) Retracting embryo (700 minutes). (D) Terminal phenotype (approximately 960 minutes); a fully retracted dying embryo.

unable to fulfil their role of supporting the nematodes vermiform shape. RNAi treated embryos continue to collapse back over time to an even more disorganised state. The terminal phenotype is variable but was typically reminiscent of the two-fold stage of embryonic development. These animals did not hatch but gradually ceased movement, vacuolated and died. The exact similarity in the embryonic lethal phenotype resulting from *Ce-pdi-2* RNAi and that of *Ce-phy-2* RNAi in *dpy-18,* compare Figure 3.15, mutants strongly supported the contention that *Ce*-PDI-2 is the only β subunit in P4H complexes containing *Ce*-PHY-1 and *Ce*-PHY-2. The contribution of other possible functions of this PDI in cuticle collagen modification could not be easily assessed as it was not possible to determine if the cuticle was further weakened by loss of disulphide bonding activity in addition to the proposed loss of P4H activity resulting from *Ce-pdi-2* interference.

4.2.7. RNAi feeding

RNAi can also be accomplished by feeding nematodes bacteria expressing dsRNA. The same fragment used for *in vitro* production was cloned into the double T7 vector pPD129.36. This was then transformed into a strain of bacteria that contains an IPTG inducible T7 polymerase. Bacteria containing the annealed single strands were then fed to wild-type *C. elegans*. The results for feeding were much less pronounced than those found for injection. Phenotypes were extremely variable with a single experiment and between experiments, ranging from slightly to extremely Dpy progeny. The most consistent phenotype observed was the extreme dumpy phenotype. These nematodes were so fragile that mounting them for images was extremely difficult as they burst when placed on the slide and it was not possible to capture images of the individuals. *Ce-phy-2* RNAi feeding in *dpy-18* mutants (and *Ce-phy-1* RNAi feeding in the *phy-2* deletion mutant) also produced severely Dpy progeny, however the "coiled" position mutants (Figure 3.17) typical of these experiments was not found for *Ce-pdi-2* RNAi. The increased fragility of these mutants also provided a very crude assessment of the more weakened state of the cuticles of *Ce-pdi-2* RNAi-treated nematodes possibly due to the loss of other functions of this gene. Additionally RNAi feeding was found to be temperature dependent. More pronounced RNAi effects were seen at the higher temperature of 25°C compared to 20°C. Whether this was a function of the RNA effect being stronger at this temperature or whether nematode collagens are more fragile at

higher temperatures and thus more susceptible to disruption was not determined. The phenotypes seen for feeding were thought to be the result of incomplete disruption of the *Ce-pdi-2* gene. The body shape defect are consistent with a role for this gene in the modification of the cuticular collagen ECM and are similar to the incomplete joint disruption of *Ce-phy-1/Ce-phy-2* by double injection or RNA feeding of one *Ce-phy* gene in the mutant background of the other.

4.2.8. Analysis of the RNAi feeding phenotype

The viable RNAi phenotypes resulting from RNAi feeding provided a useful tool with which to examine the proposed collagen and cuticle malformations proposed for *Ce-pdi-2* disruption. The Dpy phenotypes were consistent with a role in collagen modification, as this phenotype often results from collagen mutation. Additionally the embryonic lethal phenotype of *Ce-pdi-2* RNAi injected worms indicated strong cuticle abnormalities. To specifically examine the cuticle and its collagens two approaches were taken, firstly to disrupt *Ce-pdi-2* function and use a cuticular collagen specific marker to assess the effect on collagen. Secondly to look in detail at the surface of the disrupted cuticle using scanning electron microscopy (SEM) to observe specific surface cuticular structures.

4.2.8.1. DPY-7 staining

The antibody DPY-7 localises the DPY-7 protein to the circumferential annular furrows of the cuticle and is found throughout *C. elegans* development. Wild type nematodes fed *Ce-pdi-2* dsRNA expressing bacteria were stained with the DPY-7 antibody. Only a few animals were found to contain detectable levels of DPY-7 protein in their cuticles, shown in Figure 4.8. Staining in RNAi nematodes, panels C and D, is only found in small areas of the cuticle in a disordered manner when compared to the regular and discrete localisation in the wild type, panels A and B.

4.2.8.2. Scanning electron microscopy

The surface of *Ce-pdi-2* RNAi fed animals was examined by scanning electron microscopy (SEM). It was not possible to examine worms treated in this way by conventional mirosopy due to their extremely fragile cuticles. Figure 4. 9 show the effect on the cuticle of *Ce-pdi-2* disruption when examined by SEM. Panel A shows the mutant body shape while panel B illustrates the effect on the cuticle with its highly

Figure 4.8. DPY-7 immunolocalisation in *Ce-pdi-2* RNAi fed nematodes
DPY-7 immunolocalisation in wild type N2s and RNAi affected animals showing the collagen specifically in the annular furrows of the wild type adult cuticle (panels A and B) compared to the disruptive effect of RNAi of *Ce-pdi-2* feeding (panel C and D).

Facing page 170

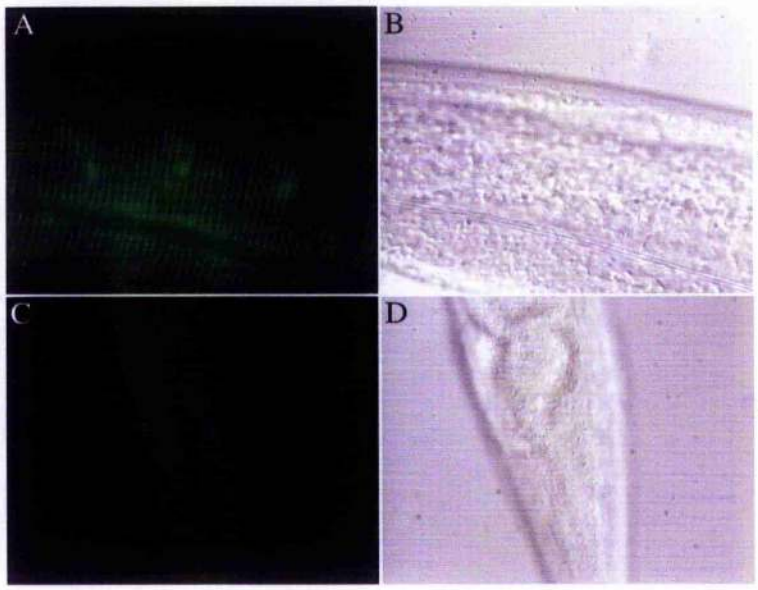
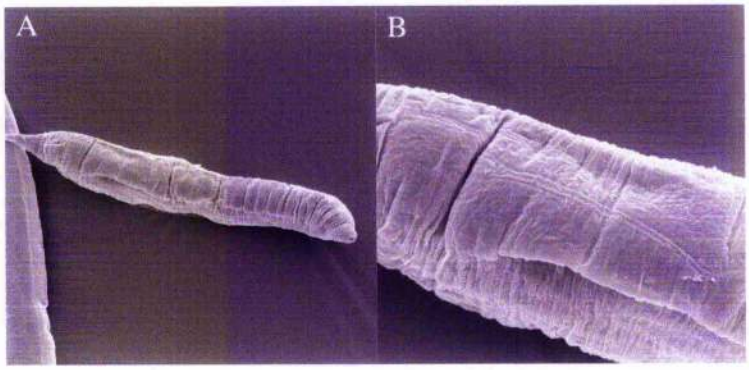

A
B
C
D
N2
Ce-pdi-2
RNAi

Figure 4.9. Scanning electron micrographs

SEMs of wild type C. elegans fed with bacteria expressing *Ce-pdi-2* dsRNA. Panel magnification x 100, panel B magnification x400.

Facing page 171


A
B

convoluted surface, loss of annulae and distorted alae. These effects are similar to the results found for the *Ce-phy* genes examined.

4.2.9. Examination of *let-44* as a candidate *Ce-pdi-2* mutant

4.2.9.1. *let-44* developmental timecourse analysis

The locus *let-44* maps to position –8.53 on chromosome X in comparison with the *Ce-pdi-2* containing cosmid C07A12.4 which has an interpolated map position of –7.63 (Figure 4.10). No other embryonic lethal phenotype is present within a 3-4 cM region. The strain GR1029 was investigated to determine if the locus *let-44* was mutant in *Ce-pdi-2*. Strain GR1029 has the genotype *let-44*(*mg41*) *lon-2*(*e678*)X; mnDp31(X;f). Animals that retain the free duplication are wild type and segregate wild type and dead embryos. Approximately a third of animals lose the free duplication and arrest as dead embryos. The phenotype of arrested embryos when examined through time intervals during development, shown in Figure 4.11, is similar to that described for *Ce-pdi-2* RNAi lethal embryos. Mutant embryos developed to the fully elongated three-fold stage of development. After this stage, at a time consistent with the cuticle becoming the mechanical support maintaining the elongated form of the worm, embryos collapsed back from the three-fold stage to a more disorganised state that failed to hatch and subsequently died. This similarity in phenotype and difference from the most common embryonic lethal phenotype seen in *C. elegans*, the paralysed arrest at two-fold or *pat* mutants, suggested *let-44* was a good candidate for a *Ce-pdi-2* mutant.

4.2.9.2. Rescue of *let-44* phenotype

Repair of the *let-44* phenotype by microinjection of the wild type copy of the mutated gene would result in the production of viable worms that have lost the free duplication but contain the introduced free array. Progeny from injections in which this has occurred would be identifiable due to the presence of the *lon-2* genetic locus that would result in worms with a phenotypically long appearance. *lon-2* worms are typically 50% longer than wild-type worms at all stages. To investigate the ability of *Ce-pdi-2* sequences to rescue the *let-44* phenotype the strain GR1029 was injected with wild type *Ce-pdi-2* and progeny scored for Lon. Cosmid C07A12 contains the entire coding region and promoter and 3' UTR regions for *Ce-pdi-2* at the position C07A12.4. Cosmid DNA was prepared and checked for the presence of *Ce-pdi-2* sequence by PCR. The cosmid was co-injected with the GFP expressing plasmid dpy-7-GFP. Most injections

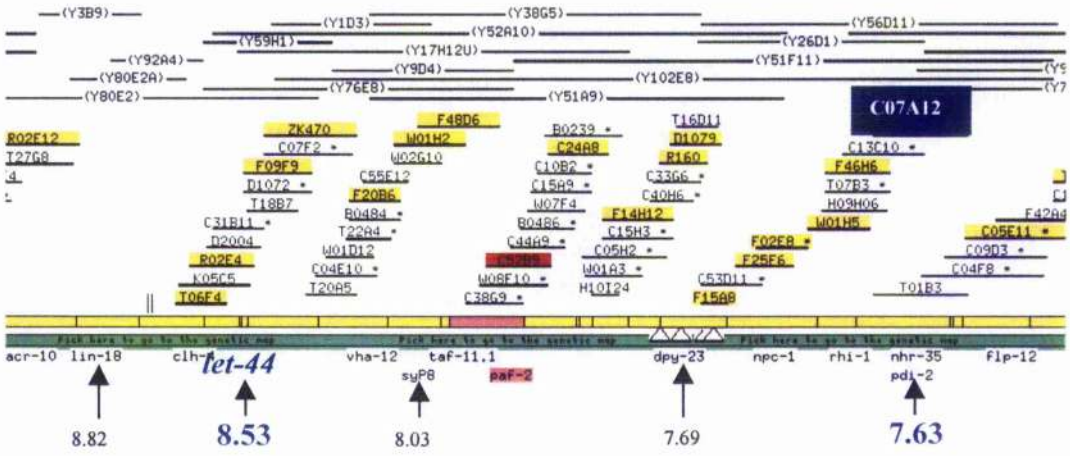


Figure 4.10. Comparison of physical and genetic maps in the region of *Ce-pdi-2*
Positions of *let-44* (blue writing) locus at 8.53 on chromosome X on the genetic map and the *Ce-pdi-2* encoding cosmid C07A12 (shown with a blue background and white writing). The interpolated genetic map co-ordinate of *Ce-pdi-2* at position 7.63 is indicated. Figure adapted from WormBase data.

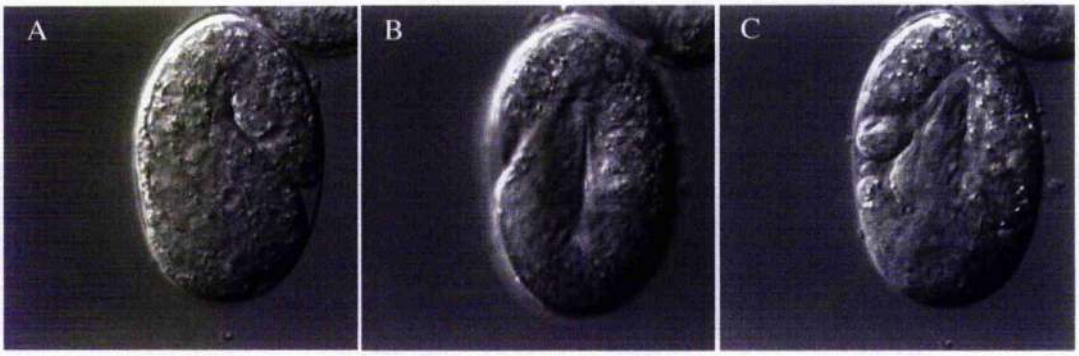


Figure 4.11. Development timecourse analysis of a *let-44* embryo

(A) 1.5 fold embryo (400 minutes). (B) 3-fold elongated embryo (620 minutes). (C) Retracted embryo (735 minutes).

of the cosmid at a higher concentration (5 μg/ml) resulted in very few live transformed progeny with dead GFP expressing embryos present. After repeated attempts at this concentration one line was obtained which gave no *lon* progeny and dead GFP positive embryos. The cosmid was therefore toxic at this concentration and so was re-injected at a concentration of 1 μg/ml. Four lines were established but none of these showed any evidence of phenotypic rescue of *let-44* (data not shown). The toxicity of the plasmid at anything but low concentrations could have been due to the overexpression of other genes contained in this cosmid. The low concentration of cosmid required to give viable transgenic progeny could have resulted in a level of *Ce-pdi-2* expression insufficient to achieve rescue. Therefore transgenic rescue was also attempted with a PCR generated fragment containing only *Ce-pdi-2* sequences cloned into a plasmid. The fragment was produced with the same primer, pdi-2 PF indicated in Figure 4.1, used as the sense primer and a antisense primer down stream of the predicted polyadenylation sequence, pdi-2 Recs R also shown on Figure 4.1. The promoter region of this fragment was therefore identical to the promoter used in reporter gene experiments which produced strong hypodermal staining in all developmental stages, and should have been sufficient to provide high level expression of the transgenic *Ce-pdi-2* gene. This plasmid was injected into the strain carrying mutant *let-44* at a concentration of 5 μg/ml with dpy-7-GFP. The combination of the higher concentration used and the increase in copy number of the *Ce-pdi-2* gene per microgram should have resulted in higher expression of the desired gene without the possible negative effects of overexpression of other genes. Viable progeny were produced with this method although no evidence of rescue of the *let-44* was found.

4.2.9.3. *Ce-pdi-2* sequence from *let-44* mutants

The lack of repair of *let-44* with *Ce-pdi-2* indicated, but did not prove, that *let-44* was not mutant in this gene. In order to investigate this further *Ce-pdi-2* sequences were cloned from *let-44* embryos. Sequences from both mRNA and genomic sources were amplified, as shown in Figure 4.12, cloned and sequences and *let-44* and wild type *Ce-pdi-2* sequences compared. Primers pdi-2 HSC F and pdi-2 HSC R were used (Figure 4.1), which cover from the translational start to stop of this gene. No consistent change in *Ce-pdi-2* sequence was found between any of the six cDNA and three genomic DNA

PCR products examined. The inconsistent changes that were present were likely to be a result of the high number of amplification cycles taken to amplify the cDNA sequences.

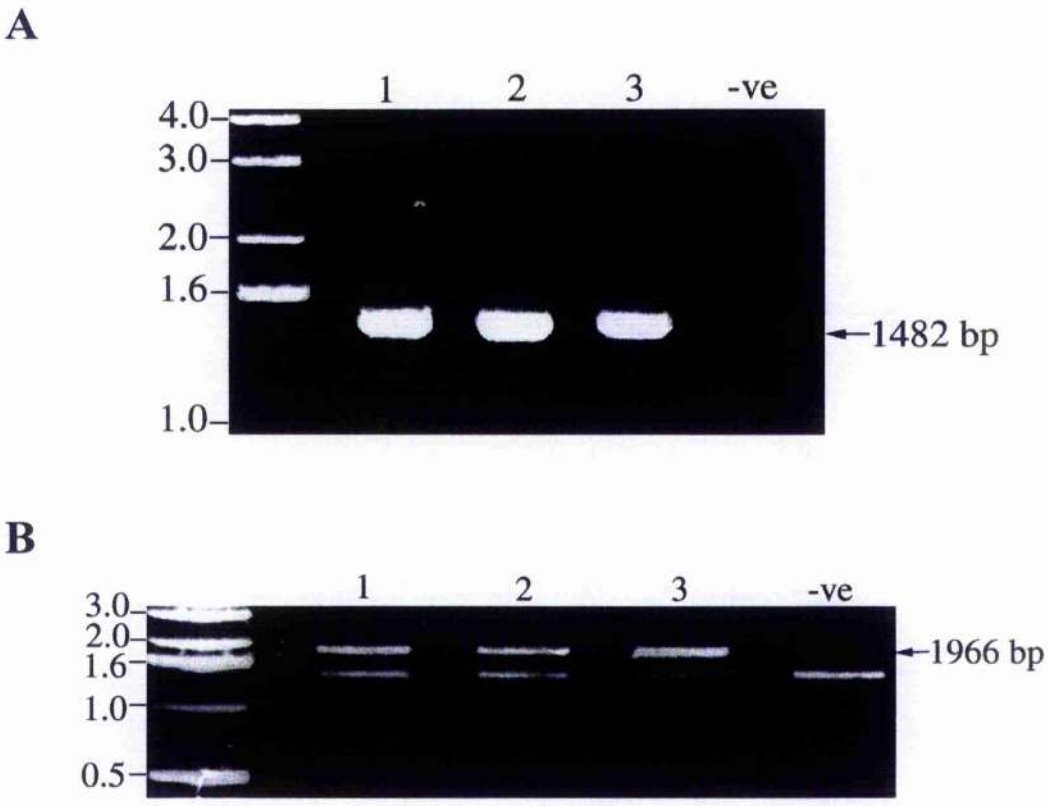


Figure 4.12. *Ce-pdi-2* genomic and cDNA cloning from *let-44* embryos

(A) Agarose gel showing products resulting from RT-PCR of *Ce-pdi-2* from *let-44* embryos. Lanes 1-3 show amplification of wild type sized 1482 bp product. –ve indicates the negative control PCR lane. Products from this and a repeat set of PCRs were cloned and sequenced (B) Agarose gel showing products resulting from PCR of genomic *Ce-pdi-2* sequence from *let-44* embryos showing amplification of wild type sized 1966 bp product. –ve indicates negative control PCR lane. Products were cloned and sequenced. The positions of molecular size standards are indicated with sizes given in kilobases.

4.3. Discussion

4.3.1. Essential function of *Ce-pdi-2* proposed to be the formation of P4H complexes

This chapter set out to establish the likelihood of *Ce*-PDI-2 representing an *in vivo* partner for the *Ce*-PHYs discussed in chapter 3. A dimer had already been shown to form between *Ce*-PHY-1 and this enzyme *in vitro* (Veijola *et al.*, 1996a). The results in this chapter support this finding and further suggest that this may be the only cuticle collagen-specific PDI in this nematode. Other functions are proposed for two other related PDIs in *C. elegans* (Eschenlauer and Page, 2003; Page, 1997), and in humans a single PDI forms separate complexes with two different α subunits (Annunen *et al.*, 1997; Vuori *et al.*, 1992a). Analysis of *Ce-pdi-2*, in a similar manner to analysis of the *Ce-phy* genes, demonstrated that this gene had all the characteristics expected to encode the only P4H β subunit associated with *Ce*-PHY-1 and *Ce*-PHY-2. Their expression patterns were virtually identical both in terms of the cells expressed and the timing of expression. As was proposed for the *Ce-phy* genes, *Ce-pdi-2* did not seem to be involved in modification of the type IV collagens and was instead cuticle collagen specific. The identical phenotype displayed by RNAi embryos also suggested that the main role of *Ce*-PDI-2 is as a subunit for the *Ce*-PHYs. DPY-7 staining of *Ce-pdi-2* RNAi fed nematodes also showed collagen disruption. The model that was therefore proposed at this stage was two separate cuticle collagen P4Hs, both containing *Ce*-PDI-2 in combination with either *Ce*-PHY-1 or *Ce*-PHY-2. As a dimer had already been described, this was a possible form these proposed complexes could take, although the ability of *Ce*-PDI-2 to form tetramers with human subunits had been established, indicating that this could also be the case in the nematode.

4.3.2. Other roles of PDI in collagen synthesis

PDI also has additional roles in collagen biosynthesis as it acts independently of P4H to catalyse disulphide bond formation and also acts as a chaperone. Reducible bonds are formed between collagens that form the nematode cuticle which possibly occur at the level of trimer formation or during the assembly of higher order structures (Cox *et al.*, 1981a). These bonds are likely to form between the conserved C-terminal cysteine residues present in *C. elegans* collagens. Replacement of one of these conserved C-terminal cysteines in the *C. elegans* collagen SQT-1 gives a roller mutant phenotype

(Kramer and Johnson, 1993). Therefore disulphide bonding is required for normal collagen function and so loss of this function may have an additional effect to that resulting from P4H removal. It was not possible in the work presented in this chapter to separate these two effects although a method to define this is discussed in Section 4.3. PDI also has a role as a chaperone in collagen biosynthesis to assist the correct folding and assembly of these proteins. P4H also performs chaperone functions during collagen biosynthesis (Walmsley *et al.*, 1999; Wilson *et al.*, 1998), thus disruption of PDI by RNAi would effect both the individual PDI chaperone functions and those of the P4H complex, which could no longer form. Removal of these chaperone functions would also be predicted to affect collagen biosynthesis.

4.3.3. *let-44* is not a *Ce-pdi-2* mutant

To confirm the RNAi mutant phenotype possible genetic mutants which mapped nearby were examined to determine if any could represent a *Ce-pdi-2* mutant. A strain bearing the *let-44* mutation was investigated as potentially being a *Ce-pdi-2* mutant as to date no other loci in this area displayed either body morphology or embryonic lethal phenotypes. This mutant was proposed based on the comparable map positions of the *Ce-pdi-2* gene and the mutant locus, and the similarities between mutant and gene RNAi induced phenotypes; both of which showed elongation followed by retraction after cuticle synthesis. Transgenic rescue experiments and molecular analysis of *Ce-pdi-2* sequences in this strain suggested that this was not a *Ce-pdi-2* mutant as no repair of phenotype was observed and no molecular mutation was isolated. The possibility existed that a mutation could lie outwith the examined coding sequence of the gene and could still affect gene function. Although possible this type of mutation, in for example the promoter region, is not found as often due to the greater probability of generating an effect by changes within the coding region compared to within control sequences. Taken together the lack of rescue with both cosmid and plasmid bearing wild type *Ce-pdi-2* sequences and the lack of detection of a molecular lesion in this gene indicated that this locus did not represent a *Ce-pdi-2* mutant.

The phenotype of *let-44* is indicative of a role in cuticle collagen modification due to its similarity to the genes discussed here and the cuticle collagen mutant *sqt-3*(*e2117*) (Priess and Hirsh, 1986). Nearby genes were therefore examined within a 3 cM region either side of *let-44* (position –8.53 on X) to establish if any other candidates genes

existed within this area. Nearby genes were therefore examined that could be involved as either a component of the cuticle or affect the generation of the cuticle. Two nuclear hormone receptor (*nhr*) genes (C07A12.3 and F16H11.5) lie to the right of *let-44*. Disruption of *nhr* genes have been described which affect body morphology and moulting (Asahina *et al.*, 2000; Gissendanner and Sluder, 2000; Kostrouchova *et al.*, 1998; Miyabayashi *et al.*, 1999). However both of the genes near *let-44* have been covered in a RNAi screen and do not display embryonic lethality or body shape mutations, with one appearing wild type after RNAi the other sterile (Maeda *et al.*, 2001). Interestingly, to the left-hand side of *let-44* at position –8.15 a collagen is encoded by the gene F14H12.1 (see Figure 4.10) the RNAi phenotype of which has not been determined. The genome wide RNAi screens (Fraser *et al.*, 2000; Gönczy *et al.*, 2000; Maeda *et al.*, 2001) have uncovered embryonic or immediately post hatch lethal phenotypes for collagen genes (see Section 1.6.3). When these RNAi phenotypes are positioned along with the existing collagen genetic mutants (see Section 1.6.1. and Table 1.1) on collagens grouped by similarity (Johnstone, 2000), a cluster of genes (6 out of 7 nearest neighbours) occurs in group 1 for which a mutant phenotype has been described. The *let-44* candidate F14H12.1 is a member of the group 2 collagens for which *rol-8* and a RNAi post-hatch lethal collagen, F38B6.5, are grouped next to each other. F14H12.1 positioning within group 2 places it directly next to these other two collagens, supporting its role as a candidate *let-44* mutant and perhaps indicating clustering of mutations on the basis of similarity also occurring in group 2 collagens. Thus on the basis of the comparative map position and its phenotype this collagen may represent the *let-44* mutation, adding to *sqt-3*(*e2117*), and possibly other embryonic lethal RNAi mutants which have not been studied in detail, exhibiting the elongation and retraction embryonic lethal phenotype displayed by *Ce-pdi-2* RNAi mutants and *Ce-phy* double mutants.

4.3.4. Uses of a *Ce-pdi-2* genetic mutant

A project to produce null alleles of all the genes identified/predicted in the *C. elegans* genome is currently being undertaken by a consortium of labs (http://elegans.bcgsc.bc.ca/knockout.shtml). These labs screen randomly mutagenised nematode libraries by PCR to identify deletions in specific genes of known sequence. The consortium aim is to provide a putative null allele of any gene requested by a lab with *C. elegans* resources. However a request in 1999 for a C07A12.4 knockout to be

made by the *C. elegans* knockout consortium has yielded no results to date. Other approaches to generate a strain are detailed in Section 4.4. Production of a genetic mutant would confirm the RNAi phenotype of this gene and provide a suitable strain for which to assess interspecies conservation of PDI function and *in vivo* assessment of *in vitro* mutagenised PDIs. *pdi* genes from parasitic nematodes such as *B. malayi* could be identified as true orthologues if they were able to replace *C. elegans pdi-2* function, and an essential function for development and cuticle biogenesis of the parasite gene inferred. Additionally, a genetic *Ce-pdi-2* mutant would provide a strain in which *in vitro* mutagenised PDIs could be assessed for their ability to function both as a PDI individually and as a P4H subunit. This would provide a method to elucidate the cause of the *Ce-pdi-2* embryonic lethality by assessing whether this is due to loss of PDIs independent roles, its subunit functions or a combination of these. PDIs can still function as P4H subunits *in vitro* when both PDI active sites have been mutagenised (Vuori *et al.*, 1992b). Therefore mutagenised enzymatically inactive *Ce*-PDI-2 could be used to rescue the mutant strain and determine if return of P4H activity alone is sufficient to repair the phenotype completely. Similar studies could be performed on PDIs from other species if rescue of the *C. elegans* mutant with the wild type genes was first established. However resolving the effects of disruption of *Ce*-PDI-2 due to complex/enzymatic functions and PDIs chaperone functions, which are described above, could not be addressed in this way. Humans PDI mutants containing double active site mutations and so lacking isomerase activity still retained chaperone activity (Hayano *et al.*, 1995). Thus transgenic rescue of a *Ce-pdi-2* mutant with an enzymatically inactive PDI would be likely to produce protein that would act as a chaperone as well as a P4H subunit.

Alternatively if it is not possible to produce a *Ce*-PDI-2 genetic mutant using the process described in Section 4.4, the experiments described above could be performed using a RNAi approach. Functional replacement of gene function with PDIs from other organism could be performed by transformation of the foreign gene in to wild type *C. elegans* followed by RNAi of the endogenous *Ce-pdi-2*. This approach as been recently successfully performed for *C. elegans* and *H. contortus* cysteine protease genes and the results verified in a stable mutant (Britton and Murray, 2002). The *Ce-pdi-2* coding sequence could be used if the foreign gene sequence is sufficiently different to the *C. elegans* gene. Alternatively if the nucleotide coding sequences were too similar the 3'

UTR region of *Ce-pdi-2* could be used to selectively disrupt this genes function. Similar approaches could be used with mutagenised *pdi* genes from *C. elegans* and other species.

4.4. Future work to identify a *Ce-pdi-2* genetic mutant

4.4.1. Screening of a chemically mutagenised nematode library

A potential approach to generate a *Ce-pdi-2* genetic mutant would be to adopt a similar approach to the *C. elegans* gene knockout consortium but perform the library construction and screening independently. This library could be produced either by random Tc1 transposon insertion (Plasterk, 1995) or chemical mutagenesis (Liu *et al.*, 1999). Tc1 insertion does not in itself necessarily result in gene inactivation as insertion often occurs within intronic regions. Therefore, after identification of an insertion in the target gene a subsequent round of screening would be required to identify animals in which the transposon and flanking DNA are deleted through transposon excision. Chemical mutagenesis, which produces small deletions, has the advantage over Tc1 mutagenesis as additional rounds of screening are not required. Chemical mutagenesis can be performed using microtitre plate-based culture for ease of handling. Mutageneised animals (P_0) can be distributed in wells and grown in liquid culture to yield F_2. Mutagenised worms from each well can then be divided into frozen stocks and DNA lysates for PCR analysis. Lysates, or pools of lysates, can then be screened with gene specific primers using conditions where production of product from a gene containing a deletion would be preferentially amplified (Liu *et al.*, 1999).

4.4.2. Precomplementation screen

An alternative, and possibly less time consuming, method for producing a mutation in a specific gene of known sequence with a lethal RNAi phenotype, or predicted lethal phenotype, is to precomplement with the target gene then mutagenise the transgenic strain. For *Ce-pdi-2* the RNAi induced embryonic lethality is predicted to represent the null phenotype. A transgenic strain could be produced by microinjection with wild-type *Ce-pdi-2* coding and promoter sequences. Co-injection with a selectable marker would also be required, this could either be semi-dominant *rol-6* marker, *unc-76* rescue or a GFP expressing transgene. Before progressing an important control would be to confirm expression of the target gene from the transgenic array. A transgenic strain from which transcription had been verified and that showed transmission at a low to medium

frequency (30-50%) would then be mutagenised. Progeny from this population would be screened for requirement of transgene presence for viability. These animals would potentially be the result of target chromosomal gene inactivation by mutation. Due to the essential function of the target gene only individuals carrying the wild-type copy on the transgene would be viable and those where the transgene had been lost would be inviable. Therefore lines that display 100% transmission of the marker gene phenotype and dead embryos would be candidate *Ce-pdi-2* mutants. It would be important to distinguish between other possible outcomes that could produce 100% marker phenotype such as transgene integration events. The presence of dead embryos on the 100% marker plates would also aid distinction of target gene removal from integration events.

Chapter 5

C. elegans PHY-1, PHY-2 and PDI-2 combine in unique ways to form enzyme complexes and are the only cuticle collagen modifying P4H subunits

5.1. Introduction

The enzymes encoded by *C. elegans phy-1*, *phy-2* and *pdi-2* genes have been shown in the preceding chapters to have a central role in modification of the cuticular ECM of this nematode. These genes displayed shared tissue specific and developmental expression profiles which were characteristic of their proposed function. The combined effect of removal/disruption of both *Ce-phy* genes gave an identical embryonic lethal expression pattern to that resulting from the single removal of *Ce*-PDI-2. The specific timing where abnormal development became apparent in these embryos was also identical, and was consistent with modification of the cuticular ECM. Biochemical analysis of *dpy-18* (*Ce-phy-1* mutant) strains indicated the presence of residual hydroxylation of their cuticle collagens (Friedman *et al.*, 2000; Winter and Page, 2000). Analysis of a *Ce-phy-2* mutant strain showed cuticle collagen hydroxylation levels lower than wild type but twice as high as *dpy-18* mutants (Friedman *et al.*, 2000). These results suggested the presence of two P4H enzymes in this nematode. In vertebrates, characterised complexes are of the form $\alpha_2\beta_2$ tetramers (Helaakoski *et al.*, 1995; Helaakoski *et al.*, 1989; Vuori *et al.*, 1992a) and an $\alpha_2\beta_2$ tetramer had been described from *Drosophila* (Annunen *et al.*, 1999). The form of a P4H formed from subunits from the parasitic filarial nematode *O. volvulus* (Merriweather *et al.*, 2001) was not determined but was likely to be either an $\alpha_2\beta_2$tetramer or an $\alpha\beta$ dimer. An interpretation of the *C. elegans* data at this stage suggested that, similar to other species examined, the two *Ce*-PHYs were in two separate P4H complexes; with the more active enzyme containing *Ce*-PHY-1. Analysis of *Ce*-PDI-2 indicated both predicted complexes would contain this protein as a β subunit. Studies had suggested a non-P4H subunit role for another *C. elegans* PDI, *Ce*-PDI-1 (Page, 1997; Veijola *et al.*, 1996a) which supported the contention that *Ce*-PDI-2 was the common β subunit. The exact form(s) of these putative complexes was however not known. A *C. elegans* P4H complex had been defined previously *in vitro*, where co-expression of *Ce*-PHY-1 and *Ce*-PDI-2 (but not *Ce*-PDI-1) in an insect cell expression system had shown that these two polypeptides combined to form an active dimer (Veijola *et al.*, 1996a). In an effort

to determine the forms of the P4H enzyme found in *C. elegans*, *Ce*-PHY-2 was expressed in an insect cell co-expression system with possible partner proteins and the resulting associations determined. Analysis of these associations was confirmed *in vivo* and suggested that, contrary to expectation, a third enzyme form was present in *C. elegans*; which was hypothesised to contain *Ce*-PHY-2 and *Ce*-PDI-2 and, it was initially thought, an additional subunit. This possibility was investigated further by defining the expression and function of three divergent *Ce-phy*-like genes with a view to their possible inclusion in this hypothetical complex, and to assess any other involvement they may have in nematode ECM formation. In addition, preliminary experiments on *C. briggsae* homologues of *C. elegans phy-1*, *phy-2* and *pdi-2* were initiated to determine if these genes performed a similar essential function in this species.

5.2. Results

5.2.1. Baculovirus expression of *Ce*-PHY-2

The baculovirus expression system enables multiple proteins to be co-expressed in insect cells and subsequent protein associations determined (Vuori *et al.*, 1992a). Co-expression of multiple proteins is accomplished by infection of cells with multiple recombinant viruses. The Pharmingen system (Crossen and Gruenwald, 1998) utilises the *Autographa californica* nuclear polyhedrous virus (AcNPV). In tissue culture, baculovirus genes such as polyhedrin are non-essential and can be replaced by heterologous genes via homologous recombination from a transfer vector. The heterologous genes are expressed from a strong promoter resulting in high levels of expression. Expression of *Ce*-PHY-2 was carried out in collaboration with J. Myllyharju at the Collagen Research Unit, University of Oulu, Finland to determine the form of any complex containing this subunit. The only previously described *C. elegans* P4H complex was a *Ce*-PHY-1 dimer with *Ce*-PDI-2 (Veijola *et al.*, 1996a). On the basis of the genetic/RNAi data, *Ce*-PDI-2 was considered to be the common subunit in complexes containing both *Ce*-PHY subunits, therefore *Ce*-PHY-2 was assessed for its ability to form P4H complexes with this and other subunits by co-expression.

Ce-PHY-2 was cloned into the baculovirus expression vector and fully sequenced to detect any errors in the PCR generated sequence. The clone was determined to be identical to the sequence predicted from cosmid data confirming this data and the

intron/exon boundary predictions made. P4H α subunits expressed alone in this system are usually highly insoluble (Vuori *et al.*, 1992a), with one of the primary functions of the PDI subunit being to prevent aggregation (John *et al.*, 1993). Expression of *Ce*-PHY-2 in this system alone was insoluble as expected, with efficient extraction of protein from cells requiring 1% SDS buffer (J. Myllyharju, personal communication). This indicated that, as predicted from RNAi/mutant analysis, *Ce*-PHY-2 was not active alone as a monomer and required the presence of other subunit(s). Surprisingly *Ce*-PHY-2 was found not to associate with *Ce*-PDI-2, *Ce*-PDI-1 or human PDI individually when co-expression cell lines were analysed by native western analysis using subunit specific antibodies (J. Myllyharju, personal communication). This was an unexpected finding as it was initially predicted that *Ce*-PHY-*2* may form a dimer with *Ce*-PDI-2 in a similar manner to *Ce*-PHY-1 (Veijola *et al.*, 1996b), and the RNAi data had indicated that both predicted *Ce*-PHYs complexes would have a common *Ce*-PDI-2 β subunit. However, when *Ce*-PHY-2 was expressed in combination with both *Ce*-PHY-1 and *Ce*-PDI-2 subunits, an active tetramer was formed with subunit specific antibodies indicating the presence of all three polypeptides in the complex (J. Myllyharju, personal communication). Expression of *Ce*-PHY-2 with *Ce*-PHY-1 and Ce-PDI-1 did not produce any complex. The tetramer was therefore predicted to be of the form *Ce*-PHY-1/*Ce*-PHY-2(*Ce*-PDI-2)$_2$ and was shown to be over 10-fold more enzymatically active than the previously described *C. elegans* dimer, suggesting that this may be the principle form of the enzyme in this organism. No other mixed α subunit P4H tetramer had been described from any other organism and data from humans indicated that they were not formed in vertebrates (Annunen *et al.*, 1997), suggesting this form of P4H complex, if verified *in vivo*, may be nematode-specific.

5.2.2. Native extract analysis of P4H complexes

To analyse the forms of the P4H enzymes from *C. elegans in vivo*, extracts were made from wild type N2, CB364 [*dpy-18(e364)*; *Ce-phy-1* predicted null], and JK2757 [*phy-2(ok177)*; *Ce-phy-2* deletion] strains and compared to the triple expressing insect cells. These insect cell lines contain the *Ce*-PHY-1/CePDI-2 dimer along with the mixed α subunit tetramer, *Ce*-PHY-1/*Ce*-PHY-2(*Ce*-PDI-2)$_2$. This analysis was performed to determine if the *C. elegans* tetramer and dimer predicted from the insect cell data were present in the nematode *in vivo,* and what the corresponding complexes were in *C.*

elegans P4H mutant strains. Nematode extracts were examined by nondenaturing PAGE, Western blotted and probed with subunit specific antibodies to *Ce*-PHY-1 and *Ce*-PHY-2. Although the *Ce*-PDI-2 antibody recognised the dimer and tetramer complexes in insect cells (data not shown) this antibody did not recognise these complexes efficiently in nematode extracts, and thus the data on complex formation using this antibody is not included here. *In vivo* PDI subunits of P4H complexes are made in large excess to the α subunits (Kivirikko *et al.*, 1992), reflecting the additional roles of these multi-functional PDI enzymes. Thus the free *Ce*-PDI-2 was found in these samples to react well with the antibody (data not shown) but the dimer and tetramer P4H complexes were extremely poorly detected. An excess of free *Ce*-PDI-2 inhibiting complex detection would not have occurred in the insect cell expression system since all subunits would have been expressed at similar levels.

The described tetramer and dimer forms of the P4H complex were present in wild type N2 extracts when detected with an anti-*Ce*-PHY-1 antibody, with the tetramer form being more abundant than the dimer (Figure 5.1, lane 1). The insect cell extract is shown in lane 4 for comparison. Examination of the wild type extracts with an anti-*Ce*-PHY-2 antibody showed only the presence of the tetramer, lane 5, confirming the co-expression data. Therefore the *in vivo* data confirms the predictions from co-expression studies; wild type *C. elegans* contains two P4H complexes, a *Ce*-PHY-1/*Ce*-PDI-2 dimer and a more abundant *Ce*-PHY-1/*Ce*-PHY-2/(*Ce*-PDI-2)$_2$ tetramer. *dpy-18*(*e364*) extracts showed no immunoreactive bands using anti-*Ce*-PHY-1 antibody, lane 2, (identical results were also found for the strain bearing allele *e1096*, not shown), confirming the lack of full-length expression in both these strains. *dpy-18*(*e364*) was predicted to produce a severely truncated protein, however the antibody is active against the extreme C-terminal amino acids of *Ce*-PHY-1. The lack of detection of protein does not confirm this strain as a null but does demonstrate lack of full-length protein production. However, as all the active site residues are missing in the mutant protein produced from this strain, and read-through translation was not detected, it was considered to be a functional null. The strain with allele *dpy-18*(*e1096*) did encode the region detected by the antibody but lacked the promoter region and protein signal peptide encoding sequences. Analysis by RT-PCR suggested transcription of *Ce-phy-1* was taking place and, if translated, the protein from these mutant transcripts would lack

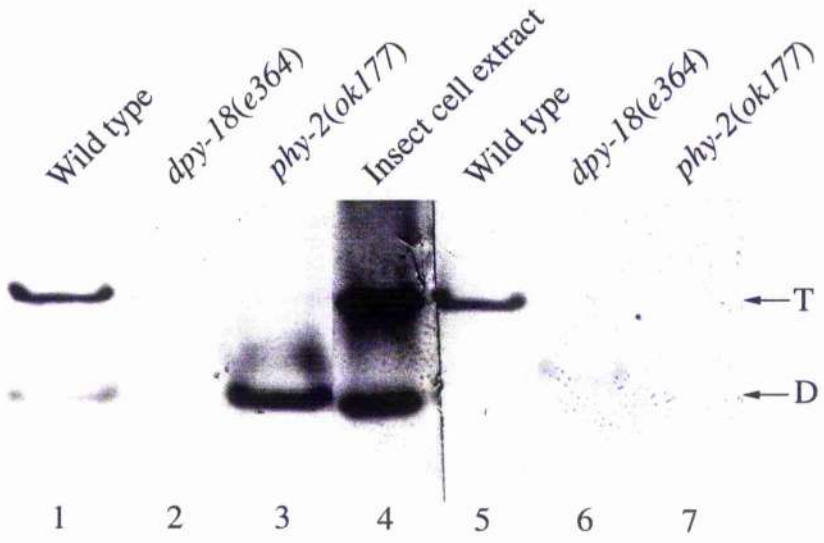


Figure 5.1. Analysis of P4H complexes from *C. elegans* extracts

Triton X-100 soluble nematodes extracts from wild type and mutant *C. elegans* strains analysed by nondenaturing PAGE and Western blotting. Lane 1 and 5 show extracts from the wild type N2 strain, lanes 2 and 5 are from a *dpy-18*(*e364*) mutant strain, lanes 3 and 7 are form the *phy-2* deletion mutant strain and lane 4 is the Tx-100 extract from insect cells co-expressing *Ce*-PHY-1, *Ce*-PHY-2 and *Ce*-PDI-2. Lanes 1-4 were probed with anti-*Ce*-PHY-1 antibody, lanes 5-7 were probed with anti-*Ce*-PHY-2 antibody. The positions of the identified P4H dimer (D) and tetramer (T) complexes are indicated with arrows.

a signal peptide and would therefore be non-functional. Additionally the loss of 5' non-coding sequence would affect the SL1 *trans*-splicing of *Ce-phy-1* gene transcripts from this strain. The SL sequences are proposed to be involved in translation initiation (Blumenthal and Steward, 1997) thus loss of the 5' UTR sequence would stop addition of the SL1 sequence and translation. The lack of protein detection suggested that the abnormal transcript was either not translated or that the abnormal protein was degraded; either possibility indicates *dpy-18*(*e1096*) was also a null allele of *Ce-phy-1*. The phenotypes of *dpy-18*(*e364*) and *dpy-18*(*e1096*) were identical and so both were considered true genetic nulls. Analysis of the *dpy-18* mutant extracts with anti-*Ce*-PHY-2 showed the presence of a faint band, lane 6, although found consistently this band was not always as strong or the background staining as low as in the data shown. This band did not appear to represent a dimer or tetramer as it migrated above the size of the dimer indicated. Additionally in the wild type extract probed with anti-*Ce*-PHY-2 antibody, lane 5, there was no corresponding band found. Therefore this *in vivo* analysis initially appeared to support the original insect cell data, that *Ce*-PHY-2 only formed a P4H complex in the described mixed PHY subunit tetramer.

Analysis of the strain bearing allele *phy-2*(*ok177*) (*Ce-phy-2* deletion) with anti-*Ce*-PHY-1 antibody did not detect a tetramer complex, lane 3, therefore the formation of a tetramer required the presence of *Ce*-PHY-2. However, despite the absence of this highly active complex, nematodes from this stain were phenotypically wild type. This could be explained by the increased levels of the *Ce*-PHY-1/*Ce*-PDI-2 dimer formed, compare lanes 1 and 3, in this strain. As approximately equal protein amounts were loaded, the abundance of the dimer in wild type and the *phy-2*(*ok177*) strain can be directly compared. An increase in the formation of the less active dimer complex that can be seen in *phy-2*(*ok177*) must be sufficient to produce wild type body shape by compensating for lack of a tetramer complex. When *phy-2*(*ok177*) animals were assayed for levels of 4-hydroxproline, they were found to be 60% of the level of wild type (Friedman *et al.*, 2000), indicating that although levels were lower, this level is sufficient to provide wild type body shape under laboratory conditions. No reactivity was found in *Ce-phy-2* deletion extracts using an anti-*Ce*-PHY-2 antibody, lane 7, confirming the predicted absence of full-length protein in this strain.

5.2.3. Possible additional *C. elegans* P4H complex(es)

Examination of nematode extracts provided *in vivo* confirmation of the presence of a *Ce*-PHY-1/*Ce*-PHY-2(*Ce*-PDI-2)$_2$ tetramer and a *Ce*-PHY-1/*Ce*-PDI-2 dimer. It also initially appeared to be in support of these being the only forms of P4H complex in *C. elegans* formed from only these three subunits. This raised the question of how *dpy-18* mutant strains were viable. As mentioned previously when cuticle extracts from *dpy-18* mutant strains were assayed for 4-hydroxyproline content the levels found were 30% that of wild type (Friedman *et al.*, 2000; Winter and Page, 2000). These nematodes are deficient in *Ce*-PHY-1, and lacking this subunit would mean neither the tetramer nor the dimer would form. Lack of these two complexes was confirmed from *in vivo* extracts from this strain. *Ce*-PHY-2 did not appear to form any complex independently of *Ce*-PHY-1 and *Ce*-PDI-2, thus an undefined complex was thought to exist which provided the P4H activity in *dpy-18* nematodes to modify their collagens to the levels noted above. Figure 5.2A shows a cartoon of the P4H complexes present in wild type and mutant strains, suggesting the presence of a third complex.

Several possibilities for another type of complex existed. However, if the RNAi data was considered accurate, the undefined complex must contain both *Ce*-PHY-2 and *Ce*-PDI-2 as additional disruption of *Ce*-PHY-2 causes death of *dpy-18* nematodes and single disruption of *Ce*-PDI-2 gives an identical lethal phenotype suggesting that a common β subunit was present in all complexes. Initial results from insect cell co-expression data indicated that a dimer or tetramer did not form from these two subunits alone, and neither type of complex was initially considered to be present on native Western analysis of nematode extracts. It was therefore hypothesised that another complex, containing *Ce*-PHY-2 and *Ce*-PDI-2, was present that required an additional unidentified subunit to form. Although, as already indicated, no additional tetramer sized *Ce*-PHY-2 complex was present in native Western analysis of worm extracts. It was considered a possibility that this was a result of the *Ce*-PHY-2 epitope, while being accessible in the characterised tetramer, was inaccessible in this putative complex. Therefore the complex proposed, on the basis of the RNAi and insect cell co-expression data, was another mixed α subunit tetramer, containing *Ce*-PHY-2, *Ce*-PDI-2 and an additional undefined *Ce*-PHY (Figure 5.2B). Other possible complexes were also conceivable, and if mixed β subunits complexes were hypothesised as being able to

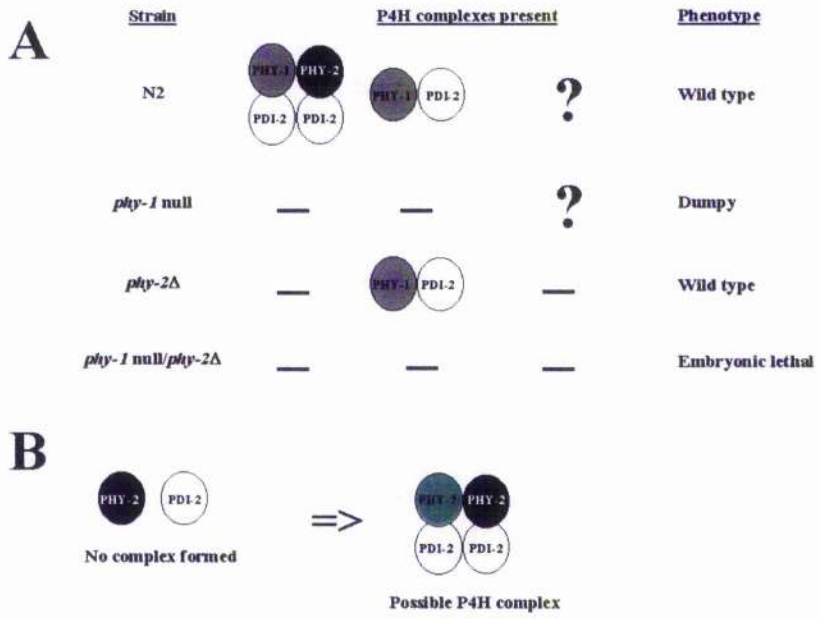


Figure 5.2. Diagram of *C. elegans* P4H complex forms

A) A depiction of the forms of P4H complexes found in wild type and mutant *C. elegans*. The wild type N2 strain is shown with the identified tetramer and dimer forms of the P4H complex. Mutant strains are shown with the complex forms they possess depicted and those that are absent indicated with a dash. The phenotype resulting from the missing forms of the complexes are given. This analysis suggested the presence of a third uncharacterised P4H complex, indicated, ?.

B) Diagram of a proposed third P4H complex based on the initial finding that *Ce*-PHY-2 and *Ce*-PDI-2 did not form a complex independently when co-expressed in an insect cell expression system. A third enzyme form was therefore predicted which contained both these proteins in a tetrameric complex in combination with a third unidentified *Ce*-PHY subunit, represented as PHY-? in the diagram.

form, another two complexes containing *Ce*-PHY-2 and *Ce*-PDI-2 could be predicted; a homo-α-subunit/mixed β subunit tetramer (*Ce*-PHY-2/*Ce*-PHY-2/*Ce*-PDI-2/*Ce*-PDI-?) and a mixed α subunit/mixed β subunit tetramer (*Ce*-PHY-2/*Ce*-PHY-?/*Ce*-PDI-2/*Ce*-PDI-?). Yet more complexes could be envisioned if the RNAi/genetic data was not accurate as these need not contain *Ce*-PHY-2 and *Ce*-PDI-2 and could be formed from every possible combination of putative α and β subunits in tetramer, dimer and possibly monomer forms. The mixed α subunit tetramer containing *Ce*-PHY-2/*Ce*-PHY-?(*Ce*-PDI-2)$_2$ was however the favoured model for a third possible P4H enzyme as the RNAi data was considered reliable, the embryonic lethality of combined *Ce-phy-1* and *Ce-phy-2* removal had been confirmed genetically (Friedman *et al.*, 2000), and a mixed α subunit P4H had been described already in this organism in the course of this work but no mixed β subunit P4H had been described for any organism.

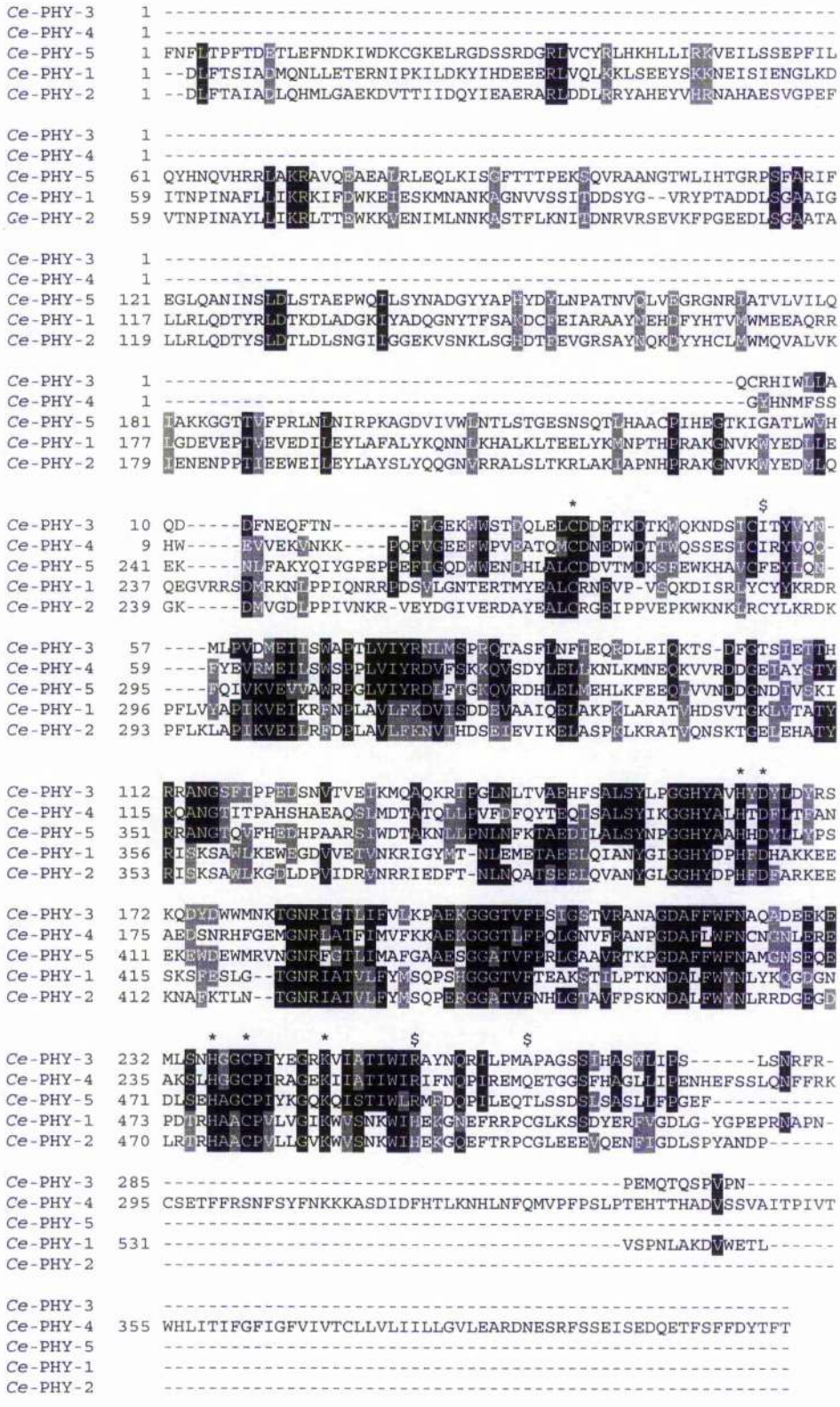
Access to the complete *C. elegans* genome sequence enabled identification of three predicted divergent *Ce-phy*-like genes (identified by J. Myllyharju, personal communication). These were further characterised to determine any role they may have in collagen modification. If one of these was involved in a mixed α subunit tetramer with *Ce*-PHY-2 and *Ce*-PDI-2, reduction of these genes functions in a *dpy-18* genetic background should prevent formation of the proposed tetramer, and would be predicted to result in embryonic lethality. In addition characterisation of these genes was also performed with a view to assessing their capacity, singly or in combination, for involvement in biogenesis of *C. elegans* ECMs of either the cuticle, possibly in a role different to the one proposed, or the basement membrane.

5.2.4. Analysis of divergent *Ce-phy* genes

A comparison of the amino acid sequence of the proteins from the three predicted *Ce*-PHYs compared to *Ce*-PHY-1 and -2 is shown in Figure 5.3 (the amino acid sequence of *Ce*-PHY-3 shown is the experimentally identified form described later). This shows that while highly diverged, and much smaller in the case of *Ce*-PHY-3 and *Ce*-PHY-4, the active site histidine, aspartic acid and lysine residues (Lamberg *et al.*, 1995; Myllyharju and Kivirikko, 1997) are all present, indicated with asterisks. The first and third conserved cysteine residues (John and Bulleid, 1994; Lamberg *et al.*, 1995) are also present in all sequences, indicated with an asterisk, as is probably the second. The

Figure 5.3. Amino acid alignment of *C. elegans* PHY subunits

An alignment of *Ce*-PHY subunits, *Ce*-PHY-1 and *Ce*-PHY-2, and divergent uncharacterised subunits, *Ce*-PHY-3, *Ce*-PHY-4 and *Ce*-PHY-5, is shown. Gaps (-) were introduced for maximal alignment and signal peptides were removed, therefore numbering refers to the mature processed proteins. Conserved cysteines (C), aspartate (D), histidine (H) and lysine (K) residues are indicated with asterisks. Positions of residues highly conserved between *Ce*-PHY-1, *Ce*-PHY-2 and other characterised P4H α subunits are indicated with a dollar sign ($), the first cysteine indicated in this fashion is present in the divergent *Ce*-PHY subunits immediately adjacent to the marked cysteine. The histidine marked in this fashion has been replaced with arginine in all three divergent polpeptides. Accession numbers *Ce*-PHY-1 (Z81134), *Ce*-PHY-2 (Z69637), *Ce*-PHY-3 (Z81593), *Ce*-PHY-4 (U46671) and *Ce*-PHY-5 (AL032623). The *Ce*-PHY-3 depicted is the experimental identified form, the accession number given is for the Genbank prediction, see Riihimaa *et al.*, 2002 for the modified gene sequence. The predictions used for *Ce*-PHY-4 and *Ce*-PHY-5 are those from Genebank.

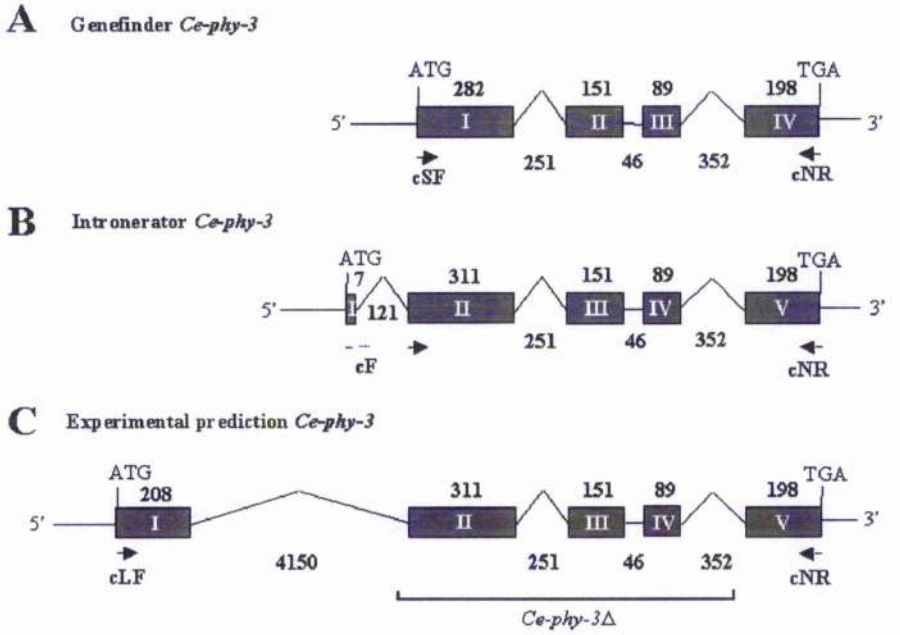


position of this second conserved cysteine in *Ce*-PHY-1 and -2 is indicated with a dollar sign, and is immediately adjacent to cysteine residues found in each of the divergent polypeptides. The fourth conserved cysteine and the position of a catalytically important histidine are also indicated for *Ce*-PHY-1 and -2 with a dollar sign. This conserved cysteine is not present in the divergent *Ce*-PHYs. In human P4H subunit αI, the histidine corresponding to this position can be partly replaced by arginine (Myllyharju and Kivirikko, 1997) as is the case in all three divergent *C. elegans* PHY polypeptides.

5.2.4.1. *Ce-phy-3* gene mapping

A predicted protein with low homology, less than 12% identity, to *Ce*-PHY-1 and -2 was identified, and termed *Ce*-PHY-3. This putative P4H subunit was encoded by the gene T20B3.7 which maps to chromosome V. Analysis of the coding sequences revealed two different gene structures, one predicted by Genefinder and the other by the Intronerator computer programs. No EST data was available to confirm or exclude either gene model. Genefinder predicted a four exon gene, Figure 5.4A, compared to Intronerator which predicted an extra very short 5' sequence as the first exon, with the second exon being a larger version of Genefinder exon one Figure 5.4B. Both the predicted proteins from these proposed genes lacked a signal peptide. Therefore an additional prediction was produced from analysis of possible coding sequences upstream of the 5' end of T20B3.7. This identified a potential 5' exon that encoded a signal peptide between S_{23} and Q_{24} (J. Myllyharju, personal communication). The predicted gene was therefore five exons in length with exons 2-5 corresponding to exons 2-5 of the Intronerator gene prediction. The first exon was however different and was predicted to be 208 bp in length, Figure 5.4C. This resulted in a large first intron of 4150 kb, intronic sequences of this size are uncommon in *C. elegans*, where the majority of introns are less than 60 bp.

The three possible predictions were tested by PCR to determine the true 5' end coding sequence of *Ce-phy-3*. An antisense primer designed just upstream of the translational stop codon, which was common to all predictions, was used with three different upstream primers corresponding to the three different gene models (Figure 5.4A). All three primer combinations were tested on genomic DNA to confirm that the primer combinations were able to produce products. However the sense primer T20B3.7 cF,



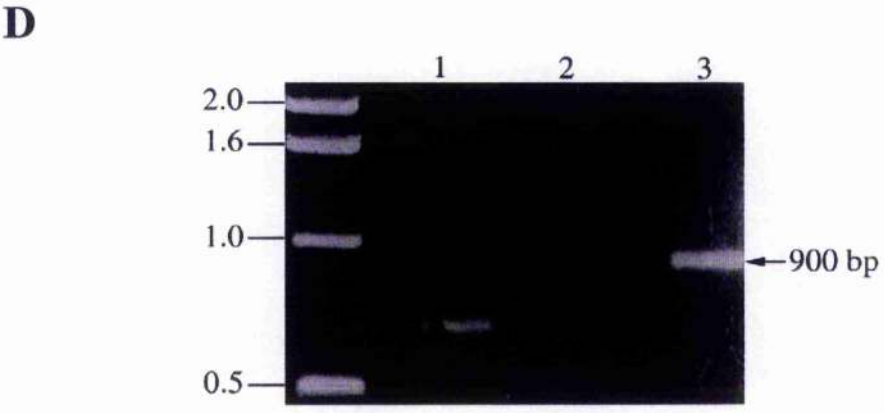


Figure 5.4. Gene structure predictions and mapping of *Ce-phy-3*

The gene structure for *Ce-phy-3* predicted by Genefinder (A), Intronerator (B) and an experimental prediction (C) are shown. Exon sequences are represented by filled boxes with sizes of exons given in base pairs above. Exon numbers are indicated by roman numerals. Intronic sequences are shown as lines with the sizes in base pairs indicated. The positions of primers used to map this gene are shown with arrows. The sequences removed in a *Ce-phy-3* deletion stain are shown. (D) RT-PCR of *Ce-phy-3* sequences to determine which gene model was correct with primers corresponding to the Genefinder prediction (lane 1), the intronerator prediction (lane2) and the experimental prediction (lane 3). The size of the amplified fragment in lane 3 is given in base pairs. The positions of molecular size standards are indicated with sizes given in kilobases.

corresponding to the Intronerator prediction, could not be tested in this way as it spanned an intron and gave no product. The other two primer combinations produced PCR products of the expected size. PCR on N2 mixed stage cDNA gave a product from both the Genefinder and the signal peptide search prediction, but no product from the Intronerator based primers, Figure 5.4B. This indicated that the signal peptide containing prediction was the real gene and the presence of product from the Genefinder prediction was due to this sequence also being present in the larger prediction. The signal peptide containing prediction was confirmed by sequencing of the PCR product and will be referred to as *Ce-phy-3*.

5.2.4.2. *Ce-phy-3* RNAi in N2

The effect of removal of *Ce-phy-3* was assessed by disruption of gene function using both RNAi feeding and microinjection in a wild type genetic background. Both approaches were taken as injection normally provide the most potent interference technique but feeding can be effective in identifying additional post-embryonic phenotypes (Fraser *et al.*, 2000) and for disrupting late acting genes. The 900 bp product from PCR with primers T20B3.7 LF and T20B3.7 NR was cloned into pPD129.36 and was used to produce RNA *in vitro* from the two T7 promoters. Double-stranded RNA was injected into N2 hermaphrodites at a concentration of 0.5 mg/ml and progeny assessed for phenotype. *Ce-phy-3* RNAi-treated nematodes were phenotypically wild type when viewed in the laboratory and analysed for body shape, movement and fecundity. Results from RNAi feeding were identical. The *Ce-phy-3* disruption phenotype was confirmed by analysis of a *Ce-phy-3* deletion strain, TP7 produced by the *C. elegans* Gene knockout consortium (Riihimaa *et al.*, 2002). This strain bears a 1241 bp deletion removing exon 2, 3, and 4, and a region of exon 5 (Figure 5.4), and was also wild type in appearance. Deletion of this region does not remove any other predicted coding sequences.

5.2.4.3. *Ce-phy-3* reporter expression pattern

A *Ce-phy-3* promoter/reporter fusion construct was made in the *lacZ*/GFP reporter gene plasmid pPD 96.04 by P. Riihimaa (Collagen Research Unit, Oulu, Finland). The pPD series of vectors are described in Chapter 3 with the vectors used being derivatives of those described by A. Fire (Fire *et al.*, 1990). The vector pPD96.04 contains a 5' decoy sequence, an upstream synthetic intron, nuclear localisation signal (NLS), and the 3'

region from *unc-54*. The reporter gene is a fusion of GFP (Chalfie *et al.*, 1994; Heim *et al.*, 1995), containing three synthetic introns, and *lacZ* containing a further eleven introns. NLS-containing GFP reporter constructs had been previously found to result in incompletely nuclear localised protein. However the degree of retention was improved using *gfp/lacZ* fusion genes, suggesting that the poor localisation of GFP alone may be the result of the relatively small size of this protein. The promoter region used incorporated *Ce-phy-3* sequences from –1480 to +8 relative to the translational start ATG. The construct was injected in to N2 hermaphrodites with the marker plasmid pRF-4 (*rol-6*) and transgenic lines selected. Lines were then stained for β-galactosidase activity (βGal). *Ce-phy-3* promoter driven expression of βGal was found in the spermatheca of adult hermaphrodites, Figure 5.5. The spermatheca is a specialised region of the gonad and is the site of oocyte fertilisation. Expression in an organ of this function indicates strongly that *Ce-phy-3* is not involved in biogenesis of either of the major ECM forming *C. elegans* collagens. However there are limitations to this technique and the transgenic expression pattern of a gene may not fully reflect the expression of the endogenous gene as the putative promoter region used may not containing all the *cis*-acting sequences necessary for native gene expression.

5.2.4.4. Effect of *Ce-phy-3* interference/removal on cuticle collagen localisation

To further determine if *Ce*-PHY-3 had any cuticular collagen-modifying role the expression pattern of the COL-19::GFP fusion was determined after RNAi disruption of the *Ce-phy-3* in the strain TP12 (an integrated line expressing COL-19 fused to GFP), and DPY-7 immunolocalisations were performed in the *Ce-phy-3* deletion strain. Strain TP12 was fed dsRNA corresponding to the *Ce-phy-3* gene and the effect on COL-19-GFP distribution determined. Mutation or interference of the cuticle collagen associated *Ce*-PHY-1 was found to affect COL-19-GFP localisation. However, no effect was found on COL-19-GFP localisation after *Ce-phy-3* RNAi disruption, as the pattern was completely consistent with that found for wild type nematodes (data not shown). This is also in accord with a non-cuticular collagen role for this enzyme although it should be noted that the single removal of *Ce*-PHY-2, as determined using the deletion strain, also had a wild-type COL-19-GFP transgene expression pattern. Likewise the localisation of the cuticle collagen DPY-7 in a *Ce-phy-3* deletion mutant strain was found to be wild type (data not shown).

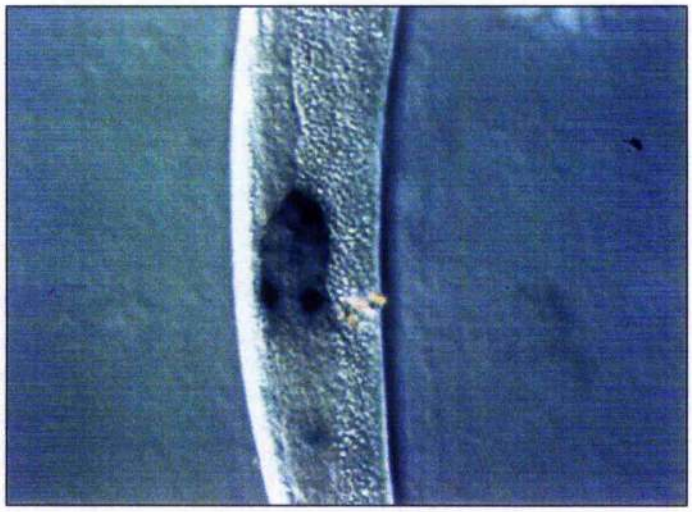


Figure 5.5. Tissue-specific localisation of *Ce-phy-3*

Sensitive X-gal staining for β-galactosidase expression in a transgenic *C. elegans* expressing *lacZ* from the *Ce-phy-3* promoter. This is shown to direct expression of the reporter gene to the spermatheca.

5.2.5.5. *Ce-phy-4* PCR mapping

The Genefinder and Intronerator programs also made two different predictions of exon structure for the gene encoded by cosmid C14E2.4, provisionally named *Ce-phy-4*, which are shown in Figure 5.6A and B. These gene structures differed only in the size of the last of the six exons. No EST data was available to corroborate either gene model. Primers were designed against sequences common to both gene predictions, and wild type N2 genomic DNA was tested to verify the primer combinations; these were then used on N2 cDNA. Although no product was generated using a standard number of amplification rounds, one was obtained using 40 cycles with the primers F1 and R3. This 484 bp product, Figure 5.6C, lane 3, corresponds to the sequence from the translational start ATG to the 3' end of exon 3. This was the largest cDNA amplified from the predicted sequences, no product was produced with primer F1 in combination with R1 or R2 (Figure 5.6C, lanes 1 and 2). Attempts were made to map the 3' end using a poly-T primer in a first round amplification with F1 followed by a semi-nested PCR using the poly-T and F2 primers. The product amplified from this reaction was found to be due to contamination of reactions with DNA containing T-rich sequences. Further attempts to map the gene by SL PCR were unsuccessful, possibly due to the high number of amplification cycles required. Further mapping of this gene would require multiple primers to be designed particularly over exon 4 to see if perhaps this region encodes two genes. 5' and 3' RACE could then be used to verify transcriptional start and stop sites. As the extent of the gene could not be defined further the prediction made by Genefinder was referred to as *Ce-phy-4*. The predicted protein from this gene contained an ER signal peptide (cleavage predicted between Y_{19} and G_{20}) and many of the conserved residues present in P4H subunits but over all shows less than 11% identity and 20% similarity with *Ce*-PHY-1 and *Ce*-PHY-2.

5.2.4.6. *Ce-phy-4* RNAi in N2

The 484 bp fragment isolated by PCR was cloned in to plasmid vectors and used to produce dsRNA in *vitro* and in bacteria in a similar manner to that described earlier. This was introduced to N2 hermaphrodites by microinjection and feeding using *E. coli* producing dsRNA. The progeny produced from both interference methods were wild type as determined by observation using standard light microscopy. During the process of this work, a *Ce-phy-4* deletion strain was produced bearing a deletion covering exons 4 and 5 (see Section 2.3). Deletion of this region does not remove any other gene coding

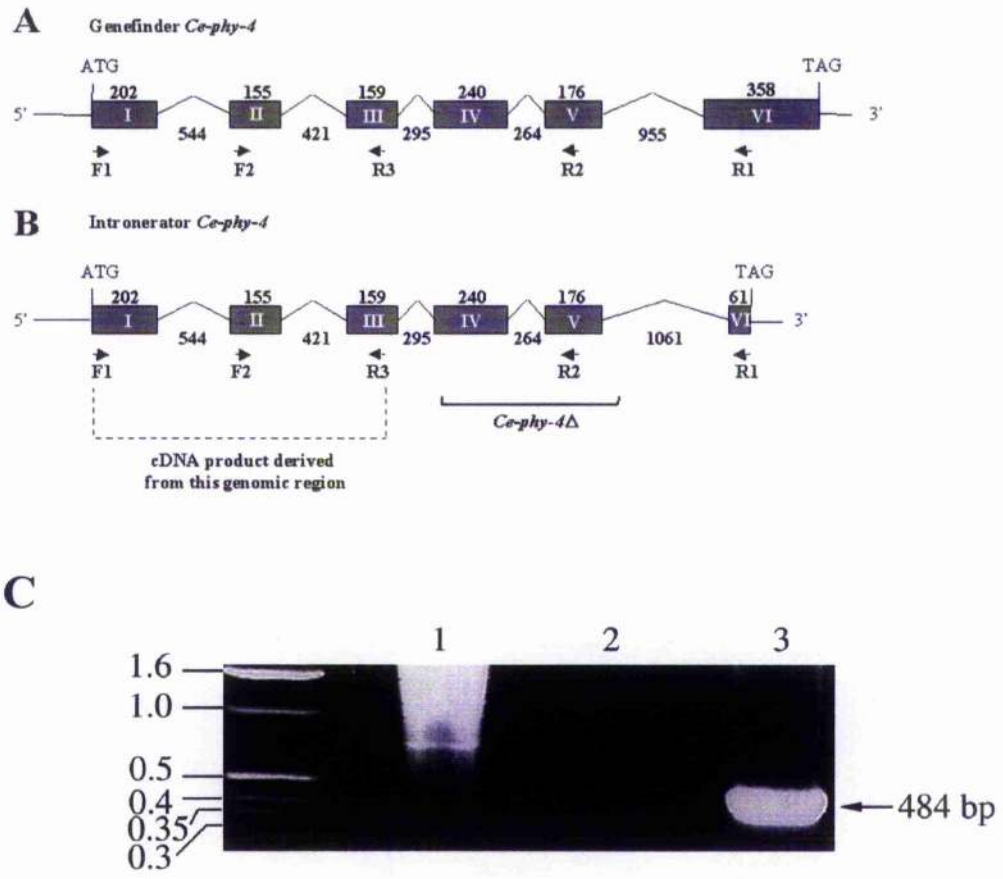


Figure 5.6. Gene structure predictions and mapping for *Ce-phy-4*

The gene structure for *Ce-phy-4* predicted by Genefinder (A) and Intronerator (B). Exon sequences are represented by filled boxes with sizes of exons given in base pairs above. Exon numbers are indicated by roman numerals. Intronic sequences are shown as lines with the size in base pairs indicated. The positions of primers used to attempt mapping of this gene are shown with arrows. The genomic sequences removed in a *Ce-phy-4* deletion stain are shown, as is the corresponding genomic sequence of the area which was amplifiable from cDNA. (C) RT-PCR of *Ce-phy-4* sequences to map the transcribed region of the gene, a common sense primer (F1) was used in combination with antisense primers R1 (lane 1), R2 (lane2) and R3 (lane 3). The size of the band amplified in lane 3 is indicated in base pairs. The positions of molecular size standards are indicated with sizes given in kilobases.

sequence. This region lies outwith the area amplifiable by RT-PCR. These nematodes were not backcrossed but visual inspection revealed no mutant phenotype.

5.4.2.7. *Ce-phy-4* expression pattern analysis

A promoter construct produced by L. Kukkola (Collagen Research Unit, Oulu, Finland) was injected into N2 hermaphrodites, 2.1 kb of upstream region, -2138-+6, was used in a fusion with *lacZ* and GFP in the reporter gene vector pPD96.04. Transmitting lines were analysed for βgalactosidase expression using sensitive staining techniques and staining reactions left for at least 16 hours. No activity was identified using this promoter region to drive reporter gene expression.

5.2.4.8. Effect of *Ce-phy-4* interference/removal on cuticle collagen localisation

To further examine the function of *Ce-phy-4*, RNAi feeding experiments were carried out in an TP12 (N2, COL-19::GFP) genetic background, to assess the effect of gene disruption on the expression of this collagen fusion protein. No difference was observed using this analysis between wild type and RNAi treated nematodes (data not shown). Similarly DPY-7 localisation was normal in the *Ce-phy-4* deletion mutant strain (data not shown).

5.2.4.9. *Ce-phy-5* PCR mapping

C. elegans gene Y43F8B.4 was named *Ce-phy-5*, as the predicted protein from the Genefinder structure displayed approximately 17% identity and 24-29% similarity with the two conserved *C. elegans* PHYs; the protein was also of a similar predicted size. An ER signal peptide is predicted with cleavage occurring between I_{14} and F_{15}. The Genefinder structure predicted for this gene comprised seven exons shown in Figure 5.7A. This gene lies completely within the intronic region of the gene Bovine pancreatic trypsin inhibitory domain (Y43F8B.3). PCR primers designed to *Ce-phy-5* full-length sequence failed to produce the correct product using mixed stage N2 cDNA (control primers on genomic DNA were successful). A range of primers were designed and applied to the same cDNA template. The largest amplifiable region was produced using the intron spanning primer combination F1 and R3, which gave a product of 731 bp, Figure 5.7D lane 3. A previous Genefinder prediction, Y43F8B.k, corresponded closely to the region amplified, Figure 5.7B. This gene had a longer predicted exon 4 that

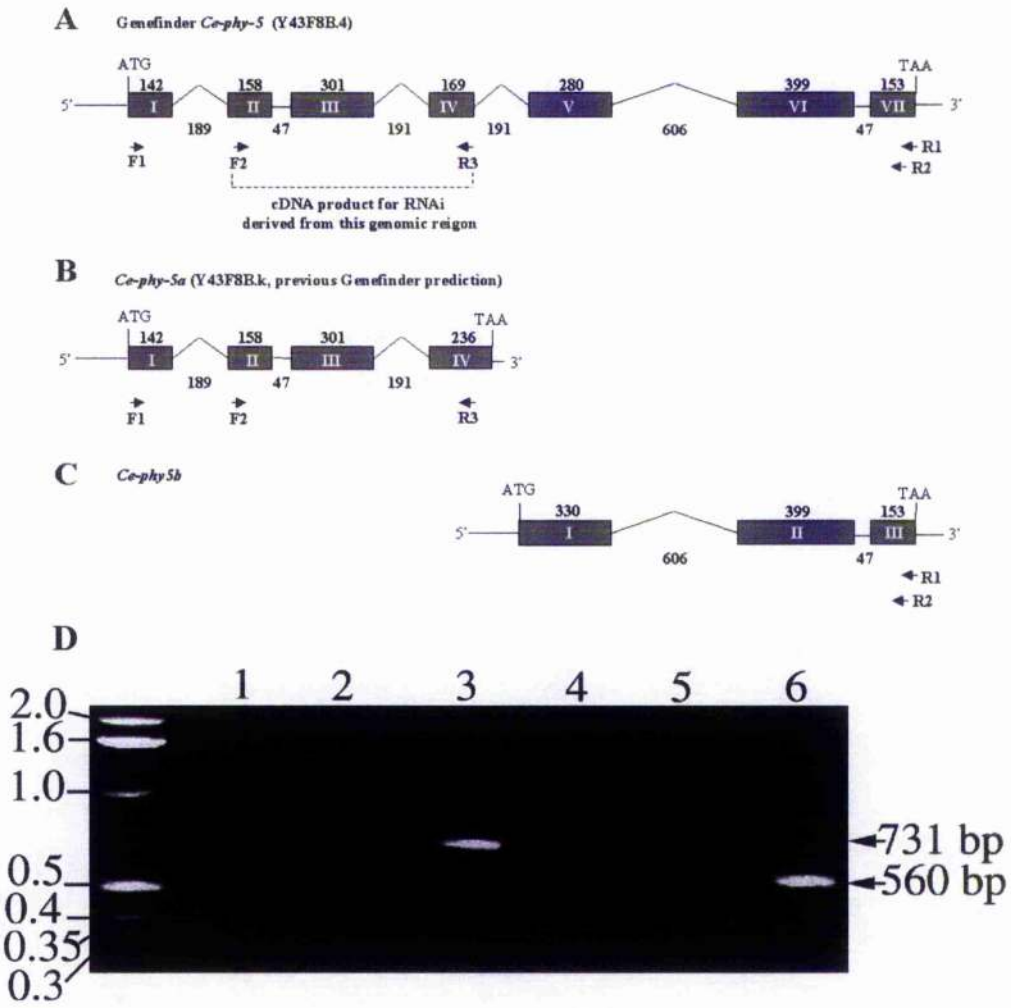


Figure 5.7. Gene structure predictions and mapping for *Ce-phy-5*

The gene structure for *Ce-phy-5* from Genefinder (A), an alternative Genefinder prediction (B) and that of a possible second gene, *Ce-phy-5b* (C). The cosmid gene name for *Ce-phy-5*, Y43F8B.4, is also indicated. Exon sequences are represented by filled boxes with sizes of exons given in base pairs above. Exon numbers are indicated by roman numerals. Intronic sequences are shown as lines with the size in base pairs indicated. The positions of primers used to attempt mapping of this gene are shown with arrows. The genomic region corresponding to that amplifiable from cDNA is shown. (D) RT-PCR of *Ce-phy-5* sequences to map the transcribed region of the gene, primer F1 in combination with R1 (lane 1), R2 (lane 2), and R3 (lane 3); and primer F2 in combination with R1 (lane 4) R2 (lane 5) and R3 (lane 6). The positions of molecular size standards are indicated with sizes given in kilobases. The sizes of amplified bands are indicated in base pairs.

encoded a translational stop codon. Re-examination of the *Ce-phy-5* genomic region suggested that this area could encode two genes, these were provisionally named *Ce-phy-5a* and *Ce-phy-5b*, Figure 5.7B and C. *Ce-phy-5a* corresponds to the Genefinder prediction Y43F8B.k and *Ce-phy-5b* is encoded by the remaining three exons in this area with a larger predicted first exon (corresponding to previous exon 5). The differences in predicted sequences between exon 4-5 of Genefinder compared to possible *Ce-phy-5a* and *Ce-phy-5b* genes are shown in detail in Figure 5.8. The predicted *Ce-phy-5b* product also contains a predicted signal peptide cleavage site between S_{17} and L_{18}. No attempt was made to confirm the presence of this *Ce-phy-5b* gene and no EST data was available to verify any of the gene predictions made.

5.2.4.10. *Ce-phy-5* RNAi in N2

A 560 bp fragment amplified using primers F2 and R3, corresponding to the *Ce-phy-5a* gene, was cloned, sequenced and used to produce dsRNA *in vitro* and for bacterial RNAi feeding. Interference using this sequence in N2 hermaphrodites produced no identifiable mutant phenotype.

5.2.4.11. *Ce-phy-5* reporter gene expression

A PCR fragment corresponding to upstream sequences of the *Ce-phy-5a* from -1402 to +125 was cloned in frame with *lacZ* reporter gene in the plasmid pPD95.03. Transgenic lines produced carrying this plasmid, injected at 20 mg/ml along with the pRF-4 (*rol-6*) plasmid marker at 100 mg/ml, did not produce any staining pattern. Injection of a very similar construct, made by L. Kukkola, Oulu, Finland, at the same concentration did however give a reproducible expression pattern using sensitive staining techniques, shown in Figure 5.9. Three cells of the pharynx stain in larval and adult stages. These cells would appear to be the pharyngeal gland cells, right and left subventral and dorsal gland cells (Bird and Bird, 1991).

5.2.4.12. Effect of *Ce-phy-5* interference on cuticle collagen localisation

The function of *Ce-phy-5* was further examined by RNAi feeding experiments carried out with strain TP12 (N2, COL-19::GFP) to assess the effect of gene disruption on the expression of this collagen fusion protein. No differences were observed between wild type and RNAi treated nematodes using this technique (data not shown).

***Ce-phy-5* (Y43F8.4)**

```
          H  E  G   T  K  I  G   A  T  L   W  V  H   E  K  N
1151      CACGAAGGCA CCAAGATCGG TGCGACTTTA TGGGTCCACG AGAAGAAgca
1201      ggaattttcg cttccctgtg cactgaccaa caaactctac aacacaaaac
1251      acttttctc ataaaatcgt atttttcggg gcaataaaat cgataaaata
1301      ggcgttttct tttccattcc catttttcta tttgcaaaat gttctatata
                                                   L  F  A  K
1351      cacatttttc ttctattttg ctttattatc gctttcagTC TTTTCGCAAA
```

End of exon 4

Start of exon 5

Ce-phy-5a* (Y43F8.k) and *Ce-phy-5b

```
          H  E  G   T  K  I  G   A  T  L   W  V  H   E  K  K  Q
1151      CACGAAGGCA CCAAGATCGG TGCGACTTTA TGGGTCCACG AGAAGAAGCA
            E  F  S   L  P  C   A  L  T  N   K  L  Y   N  T  K
1201      GGAATTTTCG CTTCCCTGTG CACTGACCAA CAAACTCTAC AACACAAAAC
          H  F  F  S  *
1251      ACTTTTTCTC ATAAaatcgt atttttcggg gcaataaaat cgataaaata
                                                   M  F  Y  I
1301      ggcgttttct tttccattcc catttttcta tttgcaaaAT GTTCTATATA
          H  I  F  L   L  F  C   F  I  I   A  F  S  /  L  F  A  K
1351      CACATTTTTC TTCTATTTTG CTTTATTATC GCTTTCAGTC TTTTCGCAAA
```

End of *Ce-phy-5a*

Start of *Ce-phy-5b*

Figure 5.8. Predicted sequence of a second gene from the *Ce-phy-5* locus

The sequence from *Ce-phy-5* at the junction of exons 4 to 5 is shown in comparison to an alternative model where this region encodes two genes. Extension of the end of exon 4 to include a stop codon provides the end of the new *Ce-phy-5a* (which correspond to the previous prediction of this gene by Genefinder) and extension of the beginning of exon 5 provided the start of the new *Ce-phy5b*. The predicted signal peptide cleavage site of the protein encoded by this gene is indicated by a forward slash.

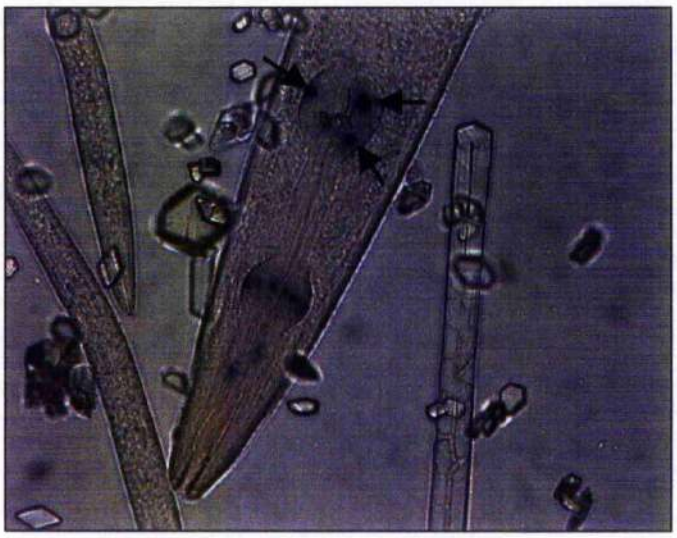


Figure 5.9. Tissue-specific localisation of *Ce-phy-5*

Sensitive X-gal staining for β-galactosidase expression in transgenic *C. elegans*. The *Ce-phy-5* promoter is shown to direct expression of the reporter gene to the gland cells at the back of the pharynx (arrowed).

5.2.4.13 Combined genetic/RNAi disruption of *Ce-phy* genes

No phenotypes were evident from single disruption of any of the divergent *Ce-phy* gene functions in a wild type genetic background. Removal/reduction of all pairwise combinations of *Ce-phy* genes was then attempted by RNAi interference of each gene in all the available P4H genetic mutant backgrounds. The effects of RNAi of *Ce-phy-2* in a *dpy-18* mutant background, and of *Ce-phy-1* in a *Ce-phy-2* deletion mutant background had already been determined (Chapter 3). Although less severe than the effects of injection, the RNAi feeding procedure for both *Ce-phy-1 and Ce-phy-2* had been demonstrated to be effective. Therefore, RNAi in the *Ce-phy-3* and *Ce-phy-4* deletion mutant strains was performed using only this delivery method. Additionally removal/reduction of each combination were also determined from RNAi injections and feeding of *Ce-phy-3* and *Ce-phy-4* in the *dpy-18* and *Ce-phy-2Δ* strains. Analysis of the disruption of *Ce-phy-3,-4* and *-5* in the *Ce-phy-1* null mutant strain *dpy-18* (and the complementary disruptions of *Ce-phy-1* in the *Ce-phy-3* and *Ce-phy-4* deletion strains) would reveal if any of these genes were involved in formation of a proposed third P4H isoform. This complex, containing *Ce*-PHY-2/Ce-PHY-?/(*Ce*-PDI-2)$_2$, was postulated to account for the survival and residual 4-hydroxproline content of *dpy-18* mutants. However, none of these combinations resulted in a phenotype different from the single removal of *Ce-phy-1*, which indicated that neither *Ce*-PHY-3, *Ce*-PHY -4 or *Ce*-PHY-5 was involved in this type of complex. Additionally no other combination of gene disruption affected this or any other observable aspect of nematode development. The results from all RNAi experiments are shown in Table 5.1. This is particularly interesting with regard to the type IV or basement membrane collagen of *C. elegans*. Based on mutant phenotype and expression pattern data, no major role could be identified for *Ce*-PHY-1 and -2 in modification of these collagens. If not involved in hydroxylation of the cuticular collagens, either alone or in a complex of the type proposed, a function for the divergent *Ce*-PHYs could have been in type IV collagen biosynthesis. From the data produced from this analysis this does not seem to be the case as phenotypes indicative of disruption of the basement membrane components such as embryonic lethality (Guo *et al.*, 1991; Norman and Moerman, 2000) or uncoordinated movement (Rogalski *et al.*, 1993) were not seen.

Gene/RNAi method		Strains and phenotypes				
		N2	*dpy-18*	*phy-2Δ*	*phy-3Δ*	*phy-4Δ*
Ce-phy-1	inj.	Dpy	–	Emb	–	–
	fed	Dpy		sDpy	w.t.	w.t.
Ce-phy-2	inj.	w.t.	Emb	–	–	–
	fed	w.t.	sDpy		w.t.	w.t.
Ce-phy-3	inj.	w.t.	Dpy	w.t.	–	w.t
	fed	w.t.	Dpy	w.t.		w.t
Ce-phy-4	inj.	w.t.	Dpy	w.t.	w.t	–
	fed	w.t.	Dpy	w.t.	w.t.	
Ce-phy-5	inj.	w.t.	Dpy	w.t.	w.t.	–
	fed	w.t.	Dpy	w.t.	w.t.	w.t.

Table 5.1. Combined *Ce-phy* gene removal/disruption

The combined removal/disruption of the *Ce-phy* genes was performed by RNAi delivered via injection (inj.) and bacterial feeding (fed) of each gene in the single mutant genetic backgrounds; *dpy-18* (*phy-1* mutant) and deletion mutants (Δ) of *Ce-phy-2*, *-3* and *-4*. The results for single RNAi of each gene in a wild type (N2) background are also shown. The resulting phenotypes are given; w.t. (phenotypically wild type) Dpy (dumpy), Emb (embryonic lethal) sDpy (severely dumpy). A dash indicates combinations that were not performed.

5.2.5. Identification of a third P4H complex in *C. elegans*

The finding that no additional phenotype was produced on interference of divergent *Ce-phy* genes in *dpy-18* mutants suggested that their gene products had no involvement with *Ce*-PHY-2 and *Ce*-PDI-2 in a third *C. elegans* P4H complex. These results therefore forced a re-assessment of the existing data that demonstrated that, although not readily detectable by Western analysis, a complex must be formed between *Ce*-PHY-2 and *Ce*-PDI-2. When insect cell extracts co-expressing these two proteins were re-examined for P4H activity using an assay based on the hydroxylation coupled decarboxylation of 2-oxo-[1-^{14}C]glutarate (Kivirikko and Myllylä, 1982) enzymatic activity was detectable at a low level (J. Myllyharju, personal communication). When the wild type and mutant strain complex data (Figure 5.1), was re-assessed in light of these new findings, the previously disregarded band in *dpy-18* mutants detected with anti-*Ce*-PHY-2 antibody, suggests a complex can be formed from *Ce*-PHY-2 and *Ce*-PDI-2. The complex formed is most likely to be a dimer with the differences in migration between this dimer and the *Ce*-PHY-1/*Ce*-PDI-2 dimer being possibly accounted for by differences in the isoelectric points between the two PHY subunits. Due to the limitations discussed earlier for the anti-*Ce*-PDI-2 it was not possible to accurately determine whether this band is observed in *dpy-18* samples using this antibody. Thus it has not been conclusively confirmed that this is the other subunit found in this complex. However, the genetic/RNAi data and biochemical activity of the co-expression extracts would indicate that this is the case. The proposed *Ce*-PHY-2/*Ce*-PDI-2 dimer however would not appear to form under conditions where either two of the main configuration of P4H in *C. elegans* can form. It is not found in the wild type extracts detected using anti-*Ce*-PHY-2, and it would therefore appear to be a forced association occurring in the absence of appropriate partners. A summary of the forms of complexes thought to exist in wild type and mutant *C. elegans* strains is shown in Figure 5.10. Wild type N2s have two major forms of complex, the described dimer and mixed tetramer. The additional dimer forming from *Ce*-PHY-2 and *Ce*-PDI-2 is illustrated in this figure in brackets to indicate that while this complex has the potential to make an active P4H enzyme it does not form under normal conditions. In *dpy-18* mutant strains neither the tetramer or *Ce*-PHY-1/*Ce*-PDI-2 dimer complexes are produced due to the absence of *Ce*-PHY-1. In this case the second dimer forms which hydroxylates collagens to a level sufficient to allow survival in a morphologically mutant form. In the *Ce-phy-2* deletion mutant strain the tetramer cannot form and the

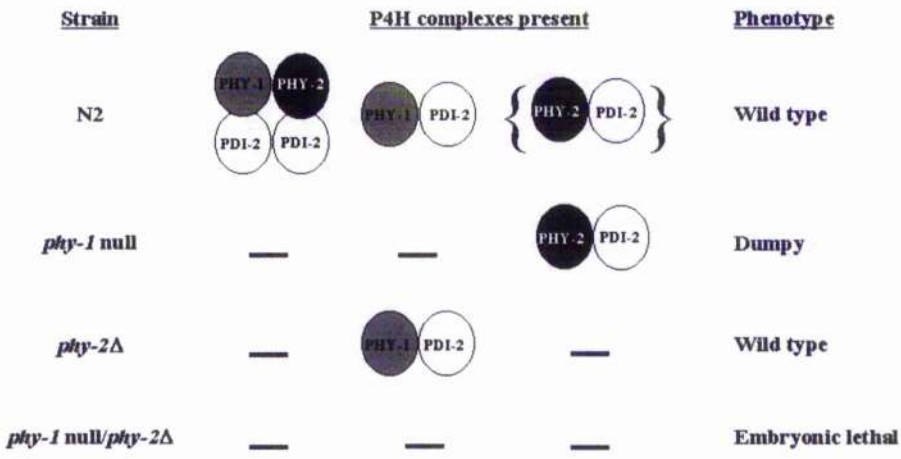


Figure 5.10. Diagram of characterised *C. elegans* P4H complex forms

A depiction of the forms of P4H complexes found in wild type and mutant *C. elegans*. The wild type N2 strain is shown with the identified tetramer and dimer forms of the P4H complex. Mutant strains are shown with the complex forms they possess depicted and those that are absent indicated with a dash. The phenotype resulting from the missing forms of the complexes are given. A third complex is indicated in brackets in the wild type as it is thought not to form under wild type conditions but provides the P4H activity in a *dpy-18* mutant that lacks the two other forms of the enzyme due to mutation in *Ce-phy-1*. This third form of the complex was formed from *Ce*-PHY-2 and *Ce*-PDI-2 combining as a dimer which, in the initially analysis, was thought not to form.

activity of the *Ce*-PHY-1/*Ce*-PDI-2 dimer enables worms to develop with wild type appearance. The effects of mutation are only detectable in this strain on biochemical analysis of their cuticle collagens. When both *Ce-phy* genes are mutated/disrupted none of the three complexes can form thus causing embryonic lethality due to loss of modification of collagens forming the first cuticle. The effect of single removal of *Ce-pdi-2* has an identical phenotype supporting the co-expression data indicating that this is the common β subunit for all the enzyme complexes. All three of the enzyme forms described here for *C. elegans* are unique as no other mixed PHY or α subunit tetramer has been identified, and no P4H dimers have been characterised for any other species. These combinations of subunits may prove to be nematode specific and inhibition of their formation could provide a possible basis for design of anti-nematode drugs.

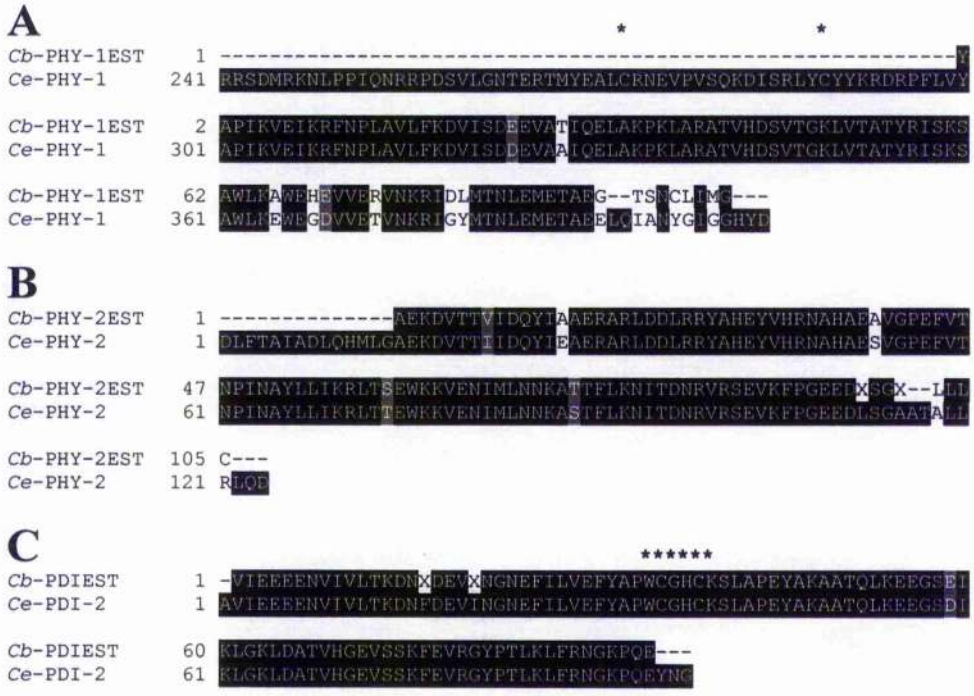
5.2.6. *C. briggsae* P4H subunit homologues

At this point in time *C. elegans* was only nematode from which the form and function of P4Hs had only been analysed in detail. Characterisation of an active P4H from *O. volvulus* did not include an analysis of what form the enzyme was found *in vitro*, or if this was reflected *in vivo* (Merriweather *et al.*, 2001). The related nematode C. *briggsae* was therefore examined to determine if homologues of the three P4H forming genes were present in this organism and to attempt to establish if they were performing a similar essential function. *C. briggsae* is also a free-living nematode with the two species having diverged around 40 million years ago. Additionally *C. briggsae* has its own genome-sequencing project thus enabling identification of putative homologues of interesting genes in this organism. Although both species appear essentially morphologically identical their genomes have diverged. Conservation of sequence, proposed function and genome organisation have been found for many genes however including collagens (Gilleard *et al.*, 1997) and collagen-related enzymes (Page, 1999; Thacker *et al.*, 1999).

Three EST sequences corresponding to possible P4H subunits were identified from *C. elegans* closest relative, the free-living nematode *C. briggsae*. These sequences showed a high degree of homology to *Ce-phy-1*, *Ce-phy-2* and *Ce-pdi-2*. EST pk34g06.s1 (accession number R04908), identified by M. A. Marra from a mixed stage G16 Lambda cDNA library, is 82% identical at the nucleotide level to *Ce-phy-1*. Translation

of this sequence produces a hypothetical protein sequence with 91% identity and 93% similarity to *Ce*-PHY-1. A second EST thought to correspond to *Ce-phy-2* was also identified from the same library. EST pk31f10.r1 (accession number R04761) was 84% identical to a region of *Ce-phy-2*, with the proposed translation showing 81% identity and 85% similarity to N-terminal sequences from *Ce*-PHY-2. A potential homologue of *Ce-pdi-2* was identified again by M. A. Marra. EST pk13h05.r1 (accession number R04009) which was 93% identical to the *C. elegans* nucleotide sequence. This sequence would not however translate in a single reading frame, and as the sequence in the database is raw data it was likely to have contained errors. In order to translate the sequence and compare protein sequences, bases for which a gap was introduced in the nucleotide alignment between the two genes were removed. Removal of five bases in this manner enabled translation of the sequence in a single reading frame. This showed 96% identity and 97% similarity to N-terminal sequences from *Ce*-PDI-2. These EST clones were not available for further analysis and SL PCR using antisense primers designed against the EST sequence was not successful in generating further data from *C. briggsae* mixed stage cDNA. Amino acid sequence comparisons of the putative proteins with their likely *C. elegans* counterpart are shown in Figure 5.11A-C.

Therefore, at that time, in the absence of full-length products, RNAi experiments were performed to determine if these enzymes functioned similarly to their *C. elegans* counterparts. Primers were designed based on the available sequence data and expression of these sequences was verified by RT-PCR from mixed stage G16 cDNA, which produced products of the expected sizes (Figure 5.11D). These were cloned in the RNAi feeding vector and bacterial cell line and fed to *C. briggsae*. However no effects were visible in these nematodes. After completion of these experiments, the complete genome sequence for this organism became available. The availability of the genomic sequences corresponding to the ESTs described enabled the cloning and sequencing of the full length cDNAs for each of these genes (performed by G. McCormack, WCMP, Glasgow). Amino acid comparisons for *Ce*- and *Cb*-PHY proteins are shown in Figure 5.12, and *Ce*-PDI with P4H PDIs from other organisms shown in Figure 5.13.



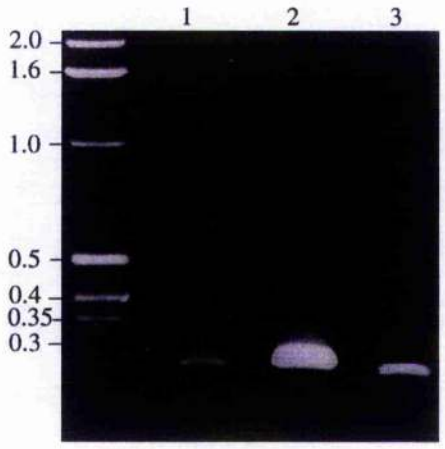


Figure 5.11. *C. briggsae* P4H subunit encoding ESTs

A) Amino acid comparison of a translation of EST pk34g06.s1 (accession number R04908), predicted to encode *C. briggsae* (*Cb*) PHY-1, with the corresponding region of *C e*-PHY-1 (signal peptide removed) (accession number Z81134). Asterisks indicate highly conserved cysteine residues.

B) Amino acid comparison of a translation of EST pk31f10.r1 (accession number R04761), predicted to encode *Cb*-PHY-2, with *Ce*-PHY-2 (accession number Z69637).

C) Amino acid comparison of a translation of EST pk13h05.r1 (accession number R04009), predicted to encode *Cb*-PDI, with *Ce*-PDI-2 (accession number U41542). Asterisks indicate active site residues.

D) Agarose gel showing products from RT-PCR of *Cb-phy-1* (lane 1), *Cb-phy-2* (lane 2) and *Cb-pdi* (lane 3) sequences from wild type *C. briggsa*e confirming expression of these genes. The products indicated were used to generate dsRNA for subsequent RNAi experiments. The positions of molecular size standards are indicated with sizes given in kilobases.

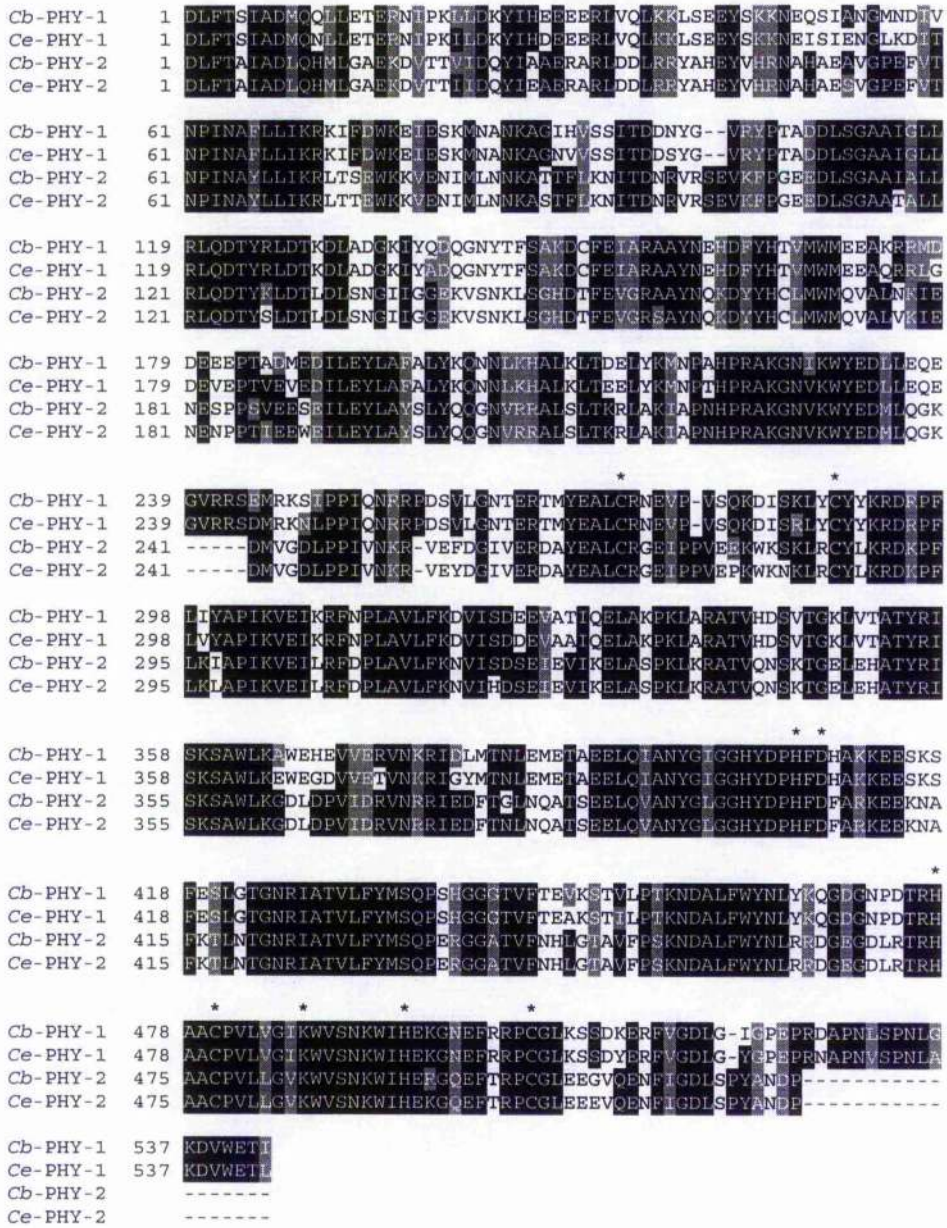


Figure 5.12. Alignment of *C. elegans* and *C. briggsae* PHYs

Amino acid comparison of predicted *C. briggsae* (*Cb*) PHY-1 and *Cb*-PHY-2 with *Ce*-PHY-1 (accession number Z81134) and *Ce*-PHY-2 (accession number Z69637). The signal peptide regions for all polypeptides were removed and conserved residues indicated with an asterisk. Gaps (-) were introduced for maximal alignment.

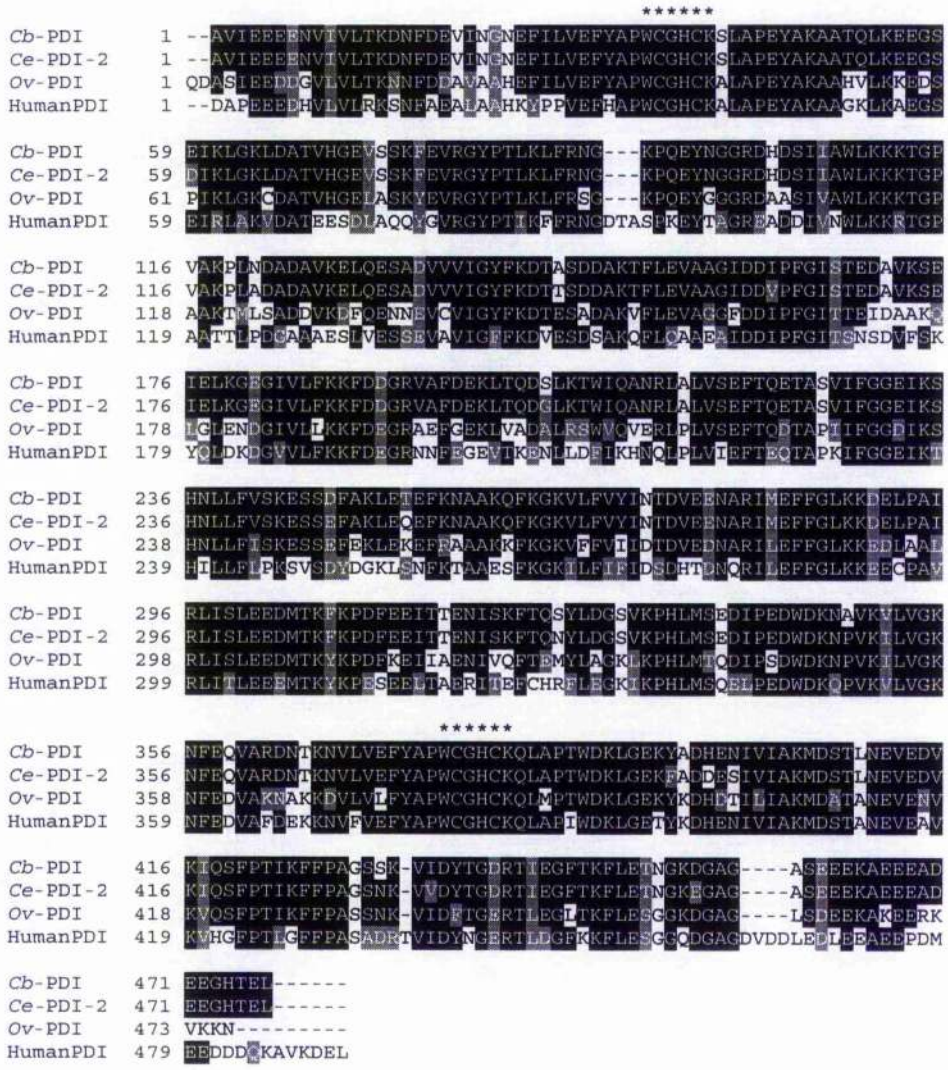


Figure 5.13. Alignment of nematode and human PDIs

Amino acid alignment of *Cb*-PDI, *Ce*-PDI-2, *O. volvulus* (*Ov*) and human PDI subunits of P4H enzymes. Gaps (-) were introduced for maximal alignment and signal peptides were removed, therefore numbering refers to the mature processed proteins. Active site residues are indicated with asterisks. Genbank Accession numbers- *Ce*-PDI-2 (U41542), *Ov*-PDI (U12440) and human PDI (X05130).

5.3. Discussion

5.3.1. *Ce-phy-1*, *Ce-phy-2* and *Ce-pdi-2* form unique cuticle collagen modifying P4H complexes

The first hypothesis on the forms of the P4H complexes in this organism was that two independent enzymes existed, one a complex of *Ce*-PHY-1 with *Ce*-PDI-2, and a second less active enzyme formed from *Ce*-PHY-2 and *Ce*-PDI-2. A combination of *in vitro* and *in vivo* analysis using subunit specific antibodies revealed unexpectedly that *Ce*-PHY-1 was in fact in involved in two complexes, one a dimer with *Ce*-PDI-2 and another a tetramer containing all three subunits. In the *Ce-phy-1* mutant strain *dpy-18* neither of these complexes could form. However the presence of 4-hydroxyproline had been identified in the cuticle collagens of *dpy-18* mutant strains (Friedman *et al.*, 2000; Winter and Page, 2000) and P4H enzyme activity was demonstrated in extracts from this mutant (Myllyharju *et al.*, 2002). An additional enzyme must therefore have been active in the absence of *Ce*-PHY-1. Data strongly favoured the view that this must contain *Ce*-PHY-2 and *Ce*-PDI-2, however these were thought not to form a complex when only these two proteins were co-expressed in an insect cell system. The possibility was that another subunit was required to form an additional mixed PHY subunit tetramer. Three divergent predicted *Ce*-PHY encoding genes were investigated to find if they were involved in a complex of this form.

All the data from these genes indicated that they were not involved in a cuticle collagen modifying P4H complex of the proposed form. The data also suggested these putative enzymes did not modify the nematode cuticle in any other way, or, it would appear the other major ECM of nematodes, the basement membrane. Re-analysis of the insect cell data revealed that *Ce*-PHY-2 and *Ce*-PDI-2 alone could, under certain conditions, form an active enzyme. This analysis has therefore proved that, as was originally thought, *Ce*-PHY-1, Ce-PHY-2 and *Ce*-PDI-2 are the only cuticle collagen modifying P4H subunit forming enzymes in this nematode. Additionally the manner in which these subunits combine has not been described for any other organism and clearly differs from that of the human enzymes. The central importance of *Ce*-PHY-1 particularly was also highlighted, as removal of this single gene removes the only two naturally occurring forms of the enzyme. The third form of the enzyme would appear to be an association that occurs *in vivo* only under the forced conditions of the *dpy-18* mutant background. If similar forms of the complex were present in parasitic nematodes the homologue of

PHY-2 may not be able to form any complex at all in the absence of a PHY-1 homologue. The ability of a *Ce*-PHY homologue to form an independent complex may have been completely lost in other species, as no selective pressure would force its maintenance. Thus identification of a PHY-1 homologue in these species and discovery of compounds capable of inhibiting this enzyme alone may be sufficient for loss of all the essential cuticle modifying activity in these nematodes.

5.3.2. Further work on *C. briggsae* P4H subunits

The lack of effect that was found from the disruption of proposed P4H subunit-encoding homologues in *C. briggsae* was surprising given the close evolutionary relationship between these nematodes. This could be a result of the observation that certain fragments do not seem capable of inducing an interference effect where others from the same gene are functional (Tabara *et al.*, 1998). The RNAi effects found in *C. elegans* for the homologous genes by feeding compared to injection delivery methods was less pronounced, perhaps the region of interfering dsRNA combined with, for these genes, a less effective delivery method, could result in lack of mutant phenotype. *In vitro* transcription and injection of the *C. briggsae* sequences would determine if the feeding method used for RNAi delivery was a factor in the lack of observable mutant phenotypes. The recent identification of the full-length sequences enables these RNAi experiments to be repeated using either full-length gene sequence, as was successful for the *C. elegans* genes, or different gene regions. This will reveal if these genes perform the same essential functions as their homologues do in *C. elegans*. This could be confirmed by inter-species rescue using *C. briggsae* subunits to attempt repair of *C. elegans dpy-18* mutants. The complementary experiment may also be possible; rescuing a *C. briggsae* mutant with *C. elegans* subunits.

A number of Dpy genetic mutants have been identified in *C. briggsae* (by R. Johnsen and D. Baillie, Simon Fraser University, British Columbia, Canada) with similar phenotypes to *C. elegans dpy-18*. Several approaches are possible to establish which, if any, of these *C.briggsae dpy* (*cby*) strains correspond to *C. elegans*. Firstly, if the RNAi of *Cb-phy-1* was Dpy but *Cb-phy-2* was wild type, *Cb-phy-2* RNAi could be used to rapidly screen the strains for additional lethality when this gene is disrupted. If embryonic lethality was observed this would indicated that the strain was mutant in *Cb-phy-1*, which could be verified molecularly, and would also indicate that the subunits

could be found in similar complexes as in *C. elegans*. Alternatively, transgenic rescue could be attempted with the *Cb-phy-1* gene and then the *C. elegans* homologue. These studies would reveal whether the essential functions of these P4H genes are conserved. Additional insect cell co-expression and *in vivo* analysis of wild type and mutant strains (if identified) would determine if the subunits form in similar ways to that described for *C. elegans*.

5.3.3. Tetratrico peptide motifs

The only functional domain so far identified for the α subunits of P4H enzymes is the human αI peptide-binding domain which has been defined as beginning at G138 and ending at approximately S244 of this protein (Myllyharju and Kivirikko, 1999). Additionally, in *C. elegans* residues have been identified that are required for formation of the mixed PHY subunit tetramer, where the first 122 N-terminal amino acid residues of the mature processed *Ce*-PHY-2 protein, and residues Q121 to A271 of *Ce*-PHY-1 are critical for assembly of the tetramer (Myllyharju *et al.*, 2002). This was established by producing hybrid polypeptides *in vitro* with the critical residues swapped between proteins. Formation of a tetramer occurred only if the regions were present on separate proteins, suggesting these regions contain elements which are required for interaction of the subunits. Interestingly, within the *Ce*-PHY-1 critical region is a single tetratrico peptide repeat (TPR) motif (Lamb *et al.*, 1995). These motifs are usually found in repeated units in proteins and are involved in protein-protein interactions. Analysis of *Ce*-PHY-1 shows the presence of a single TPR motif at position E188 to H221, however no TPR is found by a similar analysis of *Ce*-PHY-2. This suggests that this TPR motif may be the essential feature of the region identified above for *Ce*-PHY-1 to form a tetramer. Lacking this TPR motif may be the reason that *Ce*-PHY-2 cannot form a complex efficiently without Ce-PHY-1. In support of this hypothesis, analysis of the human α subunits, both of which are capable of forming an active tetramer, showed that both contain a single TPR motif at comparable positions to *Ce*-PHY-1. The determination of the potential importance of these motifs for subunit interactions and characterisation of the *C. briggsae* genes will also be informative, as the TPR motif is also found in subunits from this organism. However, in this case it is *Cb*-PHY-2 that possesses the motif and it is not present in *Cb*-PHY-1. An understanding of the function and associations between P4H enzyme subunits from this organism should reveal

whether these motifs are likely to be functional. Also it could be that as well as or in addition to subunit association the TPR motifs are involved in interactions with the collagen substrate as the positioning of these motifs for these proteins falls within the region defined for human α subunit I as representing the peptide binding domain.

5.3.4. Functional role for *Ce-phy-3*

As described above no phenotype was detected by visual inspection of *Ce-phy-3* RNAi treated nematodes, or in the mutant strain, suggesting no role in modification of the cuticle collagens of *C. elegans*. Additionally, RNAi of this gene in P4H mutant backgrounds produced no further detectable phenotypes, indicating it did not work in combination with another gene in this process. However, the spatial expression pattern in the spermatheca of late larval and adult stages suggested a possible role for the product of this gene in either the modification of type IV collagens produced by the spermatheca (Graham *et al.*, 1997) or of collagens of the egg shell. Analysis of the hydroxyproline content of the *Ce-phy-3* deletion strain embryos demonstrated a reduction in the hydroxyproline content of these embryos which were otherwise phenotypically normal (Riihimaa *et al.*, 2002). Although not ruling out a role for *Ce*-PHY-3 in modification of type IV collagens expressed in the spermatheca these findings suggests that the role of this enzyme is in biosynthesis of collagens in the early embryo most likely those contained in the egg shell (Riihimaa *et al.*, 2002). Interestingly baculovirus expression in insect cells of this protein showed that by itself it was insoluble and inactive, formed no association with *Ce*-PDI-2 or human PDI, but did form an active enzyme when co-expressed with *Ce*-PDI-1 (Riihimaa *et al.*, 2002). This PDI isoform does not form complexes with any other α subunit tested (Myllyharju *et al.*, 2002; Veijola *et al.*, 1996a) and is found in an operon with *C. elegans cyp-9* gene (Page, 1997). However what form any complex with *Ce*-PHY-3 this protein made could not be determined (Riihimaa *et al.*, 2002).

5.3.2. *Ce-phy-4* deletion

The *Ce-phy-4* deletion strain was not backcrossed to remove any additional mutations from the genome before being used in these RNAi studies. Therefore should any phenotype have resulted from any of the combinations using this strain the results would have been considered preliminary findings. In that case the experiment would

have to be repeated once this strain had been backcrossed to N2. Then, if reproducible, the results could be compared to the complementary experiment using *Ce-phy-4* RNAi in wild type and mutant strains. The region deleted in the *Ce-phy-4* deletion strain was not amplifiable by RT-PCR and so may not represent the *Ce-phy-4* gene. This could therefore indicate that two genes were present or if only one this gene was not covered by the mutation. If the second of these possibilities was true then experiments performed which involved the disruption of the other *Ce-phy* genes in this deletions background may not represent the removal/disruption of the two genes. However these combinations should have been assessed by *Ce-phy-4* RNAi in strains *dpy-18*, *phy-2*Δ and *phy-3*Δ. If this were the case, then the combined disruption of *Ce-phy-4* and *Ce-phy-5* would however still be undetermined from this study. This could be resolved by double injection of both these sequences although, as found for the combined RNAi knock-down of *Ce-phy-1* and *-2,* this can result in a reduced effectiveness compared to the actual result of removal of the two genes. Additionally, if the *Ce-phy-4* locus was defined as producing more than one gene then all the possible combinations of putative P4H subunits would not have been tested.

5.3.3. *Ce-phy-5* gene structure

The possibility of additional genes also exists for the *Ce-phy-5* region. Although this was not tested, two genes each with an expected signal peptide can be predicted in this region. PCR has confirmed that regions of the first gene are expressed in *C. elegans* although the presence of the extended region of the fourth exon that would complete this gene was not determined. Likewise, primer combinations specific for amplifying a potential second gene in this region were not used. If more than one gene was transcribed from this locus, then all the combined RNAi disruptions of divergent *Ce-phy* genes would not have been performed, as the region used for RNAi corresponded only to the *Ce-phy-5a* region.

5.3.4. Function of divergent *Ce-phy* genes

All evidence presented here would lead to the conclusion that none of the divergent *Ce-phy* genes are involved in the generation of the *C. elegans* cuticle collagens. However no role in any another biological process could be conclusively determined for *Ce-phy-4* and *-5* by the RNAi data. Tissue-specific expression patterns were identified for *Ce-*

phy-3 and *-5* suggesting possible functions. A possible approach to gain more insight into the function of the divergent *Ce-phy* would be to use a hypersensitive RNAi strain containing *rrf-3* mutant alleles. Use of strains containing mutants alleles of this putative RNA-directed RNA polymerase produced approximately twice the number of mutant phenotypes when compared to interference in the wild type N2 strain (Simmer *et al.*, 2002). Many of these genes had known genetic mutant phenotypes for the target gene that were not previously detected by RNAi. Although these were neuronally expressed genes, which are known to be refractory to RNAi by injection and feeding, the *Ce-phy-4* expression pattern has not been determined and other specific cell types not yet defined, may also be resistant to RNAi in a similar manner to cells of the nervous system. Thus RNAi of *Ce-phy-3*, *-4*, and *-5* in this hypersensitive RNAi strain may reveal a further function for these genes. A possible function for these genes at the time of starting this work could have been in regulation of the hypoxia inducible factor (HIF). In mammalian cells the transcription factor (HIF) is post-translationally regulated by prolyl hydroxylation (Ivan *et al.*, 2001; Jaakkola *et al.*, 2001; Yu *et al.*, 2001). HIF is a central regulator of oxygen homeostasis with the majority of the transcriptional responses to hypoxia regulated by it. HIF operates as a heterodimer of HIF-α and HIF-β subunits, where HIF-α is the regulated component. Under normal oxygen conditions the von Hippel-Lindau tumour suppresser protein (VHL) binds directly to HIF-α and targets it for degradation. In hypoxic conditions degradation of HIF is inhibited and HIF-α can bind to HIF-β and activate transcription. Targeting of HIF-α for VHL-mediated degradation is regulated by prolyl hydroxylation (Ivan *et al.*, 2001; Jaakkola *et al.*, 2001). In hypoxic conditions HIF-α is not hydroxylated and is thus not bound by VHL and is able to activate target genes. Mutants in the *C. elegans* HIF-α encoding gene exhibit no defects at normal oxygen levels but are unable to survive under hypoxic conditions (Jiang *et al.*, 2001) normally tolerated by wild types (van Voorhies and Ward, 2000). A role for the divergent *Ce-phy* genes described here in this process now seems unlikely as the *C. elegans* gene *egl-9* (*egg-l*aying defective) was identified as the HIF-α hydroxylating enzyme. EGL-9 acts as the oxygen sensor and represents a novel functional group of 2-oxoglutarate-dependent oxygenases (Epstein *et al.*, 2001).

Chapter 6

Cloning of a hypodermally expressed independently active *Brugia malayi phy* homologue

6.1. Introduction

Brugia malayi along with *B. timori* and *Wuchereria bancrofti* are the causative agents of lymphatic filariasis in humans, with over 120 million people infected and over 1 billion people at risk of infection in over 80 countries (Michael and Bundy, 1997; WHO, 2000). In humans the pathologies associated with infection are rarely related to the larvae but instead to the adult nematodes. Although lymphatic filariasis is rarely fatal it is a debilitating and often disfiguring disease. Approximately a third of those infected are incapacitated and/or disfigured by the infection, while about another half have internal damage to their renal and lymphatic systems (Ottesen *et al.*, 1997). Recently lymphatic filariasis has been recognised as the second leading cause of long-term disability worldwide. As such the Filarial Genome Project (FGP) was initiated in response to the World Health Organisations (WHO) call for a genomics approach to the causative agents of lymphatic filariasis, with a view to better control or possible eradication of the parasite. The complex life cycle of these filarial nematodes involves human infection by mosquito transmitted L3 larvae. These larvae then invade the body and develop to adults that, for *B. malayi*, reside in the lymphatic vessels of the human host. Here the long-lived adults (up to five years) produce millions of live L1 larvae, termed microfilaria, that may then be transmitted to the mosquito via blood. Microfilariae then undergo two moults within the mosquito host to become human infective L3 larvae.

Parasitic nematode species are particularly intractable to genetic and molecular analysis due to their complex life cycles, difficulty of maintenance in the lab and lack of transgenic techniques. *B. malayi* can however undergo its entire life cycle in the lab using a rodent host. In addition, a genome project exists through the FGP, consisting of EST analysis and genomic mapping projects. FGP labs have also constructed and made available reagents to aid molecular analysis such as filarial cDNA and genomic libraries. The detailed knowledge and techniques available for the study of the free-living nematode *C. elegans* are also extremely useful for the study of filarial nematodes in two ways. Firstly, the homologue(s) of a parasitic nematode target gene may be

examined in this organism using the full range of methodologies available for *C. elegans*. The advantages of genetic and transgenic manipulation along with ease of culture, the complete genome sequence and the defined development, makes for rapid and more complete study in this organism. Knowledge of the gene function in *C. elegans* can then be applied to the parasite homologues. Secondly, the free-living species can be used more directly to examine parasite genes. Using *C. elegans* as a heterologous expression system can reveal functional roles and control elements of parasite genes (Hashmi *et al.*, 2001). Additionally, for foreign protein expression, use of *C. elegans* is advantageous as it is a higher eukaryote and expression of filarial parasite proteins in *C. elegans* is more likely to result in proteins in their native form than expression in other systems. Expression of parasite protein in *C. elegans* has proved possible (Krause *et al.*, 2001; Redmond *et al.*, 2001) and is important where a protein to be produced as a vaccine has conformational epitopes, or where glycosylation and/or enzymatic activity are important for the proteins effectiveness as a vaccine.

Results in *C. elegans* have shown the essential nature of P4H complexes in development and body morphology as a result of the cuticular collagen modifying activity of this enzyme. Additionally, the three subunits comprising the complex forms in this nematode, *Ce*-PHY-1, *Ce*-PHY-2 and *Ce*-PDI-2, combine in unique ways that may be nematode specific. Examination of three other divergent *phy*-like genes in *C. elegans* showed they were not to be involved in cuticular collagen modification, thereby underlining the importance of *Ce-phy-1*, *Ce-phy-2* and *Ce-pdi-2* in this essential process. The characterisation of a viable *Ce-phy-1* knockout also provides the potential for functional studies of other P4Hs by transformation into *C. elegans*. Commercially available inhibitors of P4H function have been shown to function in the destruction of *B. malayi* adults, with what appears to be cuticular defects (Merriweather *et al.*, 2001), thus highlighting this enzyme as a drug target for control of infection. A proyl 4-hydroxylase *phy* gene from *B. malayi* was therefore cloned and characterised and *C. elegans* used as an expression system to examine more directly some aspects of gene function and control.

6.2. Results

6.2.1. Cloning of *Bm-phy-1* cDNA

6.2.1.1. Identification and sequencing of expressed sequence tag clones

Expressed sequence tags (ESTs) are single sequence reads from randomly selected cDNA clones with data from EST projects available through publicly accessible databases. Sequence similarity search tools such as BLAST (Altschul *et al.*, 1990) can then be used to survey this data for sequences of interest. In addition to giving a representation of the genes expressed in an organism, EST data also provides information on levels and stage specificity of expression. An impression of the expression level of a particular gene can be gained from the number of representative ESTs found. The expression of a gene in a particular developmental stage and/or tissue may also be represented in the EST dataset as many are derived from tissue and stage-specific libraries. *B. malayi* has an ongoing EST based genome project with over 22,000 ESTs currently in the database (Williams *et al.*, 2000). These represent about 8000 genes or approximately 40% of the predicted genes in the *B. malayi* genome (Williams *et al.*, 2000).

To determine if any ESTs were present in the *B. malayi* database that represented potential prolyl 4-hydroxylase homologues the database was searched with complete *Ce*-PHY-1 and *Ce*-PHY-2 protein sequences. When these were compared to a translated *B. malayi* EST database using tBLASTn three ESTs were identified that were homologous to the *C. elegans* proteins. SW3D9CA480SK (accession number AA585698), (submitted by Steven A.Williams, FGP, Smith College) from a *B. malayi* L3 moulting day 9 larval cDNA library (SW97WLMBmL3d9), gave 67% identity and 77% similarity to *C. elegans* *Ce*-PHY-1 over a 53 amino acid region. MBAFCX8G05T3 (accession number AA509222), (submitted by Mark Blaxter, FGP, Edinburgh University) from an adult female cDNA library (SAW96MLWBmAF), gave 68% identity and 85% similarity over a 69 amino acid region. MBAFCZ7H09T3 (accession number AA406985), submitted by Mark Blaxter from the same library, gave 46% identity and 71% similarity over a 122 amino acid region.

EST clones were received from the FGP and subcloned via PCR for further sequence analysis, see Figure 6.1. SW3D9CA480SK was a 500 bp clone that gave very poor sequence data. The clone MBAFCX8G05T3 gave a product of 1.5 kb by PCR.

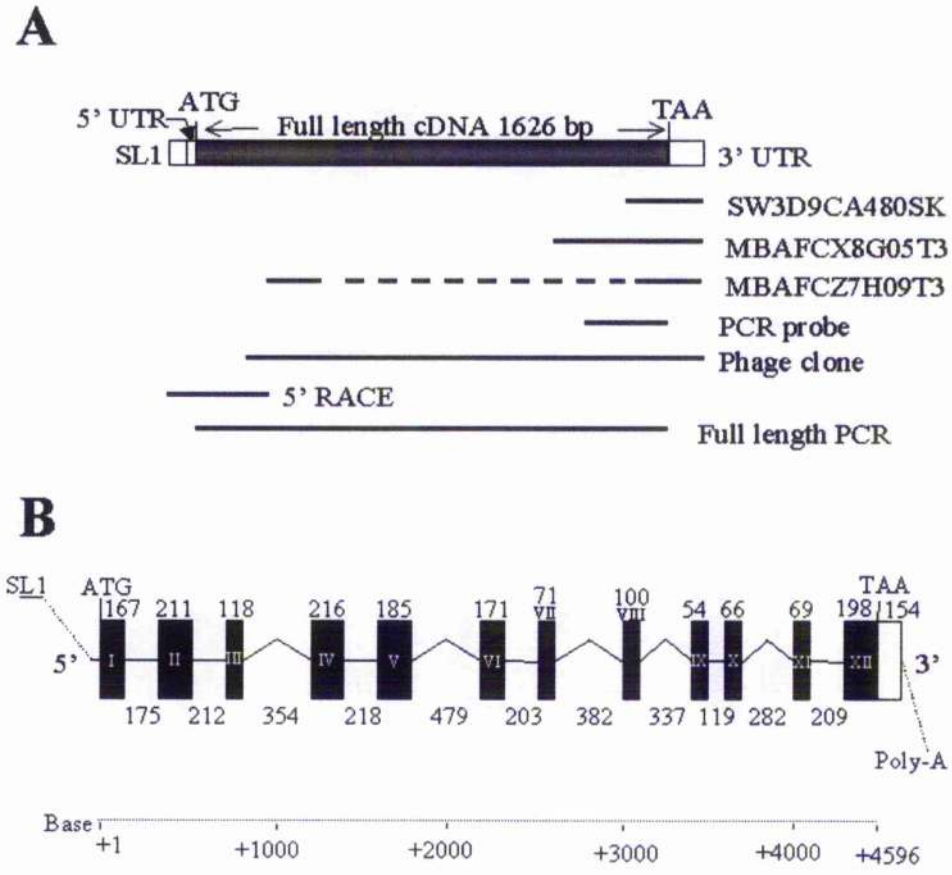


Figure 6.1. Cloning and gene structure of *Bm-phy-1*

A. Cloning of the full-length *Bm-phy-1* cDNA sequence. Open boxes indicate the SL1 trans-spliced leader and 5' and 3' UTRs. The 1626 bp coding sequence is represented by a filled box, ATG and TAA indicate, respectively, the translational start and stop codons. The positions of the ESTs identified and areas sequenced are shown along with the probe used for library screening. The position of the largest phage clones isolated, the 5' RACE PCR product used to complete the cDNA sequence and the full length PCR which was completely sequenced to confirm the cDNA are also represented. B. Gene structure of *Bm-phy-1*. Exon sequences are represented by filled boxes with sizes of exons given in base pairs above. Exon numbers are indicated by roman numerals. Lines represent intronic sequences with the sizes in base pairs indicated. The 3' UTR is depicted as an open box and the positions of SL1 *trans*-splicing and polyadenylation signals are indicated.

Sequence from only one end of this clone was homologous to prolyl 4-hydroxylase however this homology ended approximately 600 bp through the clone after which no significant homology was found. This clone was thought to be chimeric, a problem that has been documented previously in the libraries used for production of EST data (Williams *et al.*, 2000). The 1.4 kb clone MBAFCZ7H09T3 was sequenced at its 5' and 3' ends only. Both ends gave homology to prolyl 4-hydroxylase but did not appear to represent the full-length gene. Due to the sequencing and chimeric clone problems, and the incomplete nature of this clone, further sequence analysis was not continued. Comparison of the sequence generated from all three ESTs (for SW3D9CA480SK original sequence was used) showed that they all contained the predicted 3' coding sequence and 3' UTR of a *phy*-like gene, termed *Bm-phy-1*.

6.2.1.2. PCR based attempts to clone *Bm-phy-1*

Attempts were made to clone *Bm-phy-1* gene using 3' end sequence from the ESTs and the splice leader sequences (SL1 and SL2) as 5' primers. An estimated 70% of *C. elegans* genes are SL *trans*-spliced (Zorio *et al.*, 1994) and *trans*-splicing has been identified in *B. malayi* using an SL1 sequence identical to that found in *C. elegans* (Takacs *et al.*, 1988). Primer BM N Phy Reverse was used in combination with primers SL1 and SL2 on a variety of cDNA templates. None of these combinations gave products. cDNA libraries were also screened by PCR with vector specific primers along with BM N Phy Reverse to attempt amplification of full-length cDNA. Library PCRs did not amplify any *Bm-phy-1* sequences and attempts at 5' RACE were also unsuccessful at this stage.

6.2.1.3. Library screens to identify *Bm-phy-1*

A *B. malayi* adult male cDNA library SAW94NLBmAm (from Steven A.Williams, FGP, Smith College) was screened using a probe representing 312 bp of 3' coding sequence from *Bm-phy-1*, identified from the EST sequences. The highest homologies between prolyl 4-hydroxylase genes are found in this region and it was hoped that this approach could potentially identify more than one *Bm-phy* gene. The library was tested by PCR and found to be positive for the region used for the probe, indicating that the gene would be represented in this library. From screens of this library, eleven positive clones were identified and were taken to the tertiary or quaternary screening level, subcloned into plasmids and sequenced.

Full-length sequencing was performed on two 1.6 kb clones both of which were found to represent *Bm-phy-1* (see Figure 6.1). The nine other clones identified contained inserts with sizes ranging from 0.6 kb to 1.6 kb and were all also from *Bm-phy-1*. Sequence was identical from all clones at their 3' ends, the 5' ends of four clones were shorter but identical to the consensus sequence. Three other clones had large non-P4H sequence at their 5' end. These sequences showed 98% homology to 28S Ribosomal RNA from *Brugia pahangi*. These sequences were judged, therefore, to be from chimeric clones containing sequences from more than one gene. The 5' end of two clones had an unusual form, the structure of which could only be determined once the entire cDNA sequence was assembled (after 5' RACE, Section 6.2.1.4). They were eventually shown to represent the 5' end of the gene in one reading frame, a 170 bp deletion, followed by the remainder of the consensus sequence in a different reading frame. These clones were thought to be artefacts from library construction or to be derived from aberrant transcripts. Excluding the 5' ends of these two clones and the chimeric clones, the consensus sequence generated from library screening therefore provided 1669 bp of sequence, with a 1515 bp open reading frame, 154 bases of untranslated 3' sequence and a poly(A) tract.

No consensus AATAAA polyadenylation signal sequence was identified within the 3' UTR before the poly(A) sequence. A non-consensus signal sequence of GATAAA was however found as shown in the complete genomic sequence (Appendix 1). In *C. elegans* the requirement for consensus AATAAA consensus sequence is not as stringent as in vertebrates (Blumenthal and Steward, 1997). Approximately half of *C. elegans* genes have the consensus sequence with the remainder having a limited number of variants that occur in a very defined region, normally 11-17 bp upstream, from the cleavage site. The GATAAA variant signal sequence is found in approximately 5% of the *C. elegans* genes examined (Blumenthal and Steward, 1997). Of the tolerated variations of the consensus sequence found in *C. elegans* only GATAAA was found in the 3' UTR of *Bm-phy-1*. As 11 base pairs separate the 3' end of this motif and the poly(A) cleavage site the positioning of this motif in *Bm-phy-1* is also identical to the restricted positioning found in *C. elegans* and is thus the proposed polyadenylation signal sequence of *Bm-phy-1*.

6.2.1.4 Cloning of the *Bm-phy-1* 5' end sequence

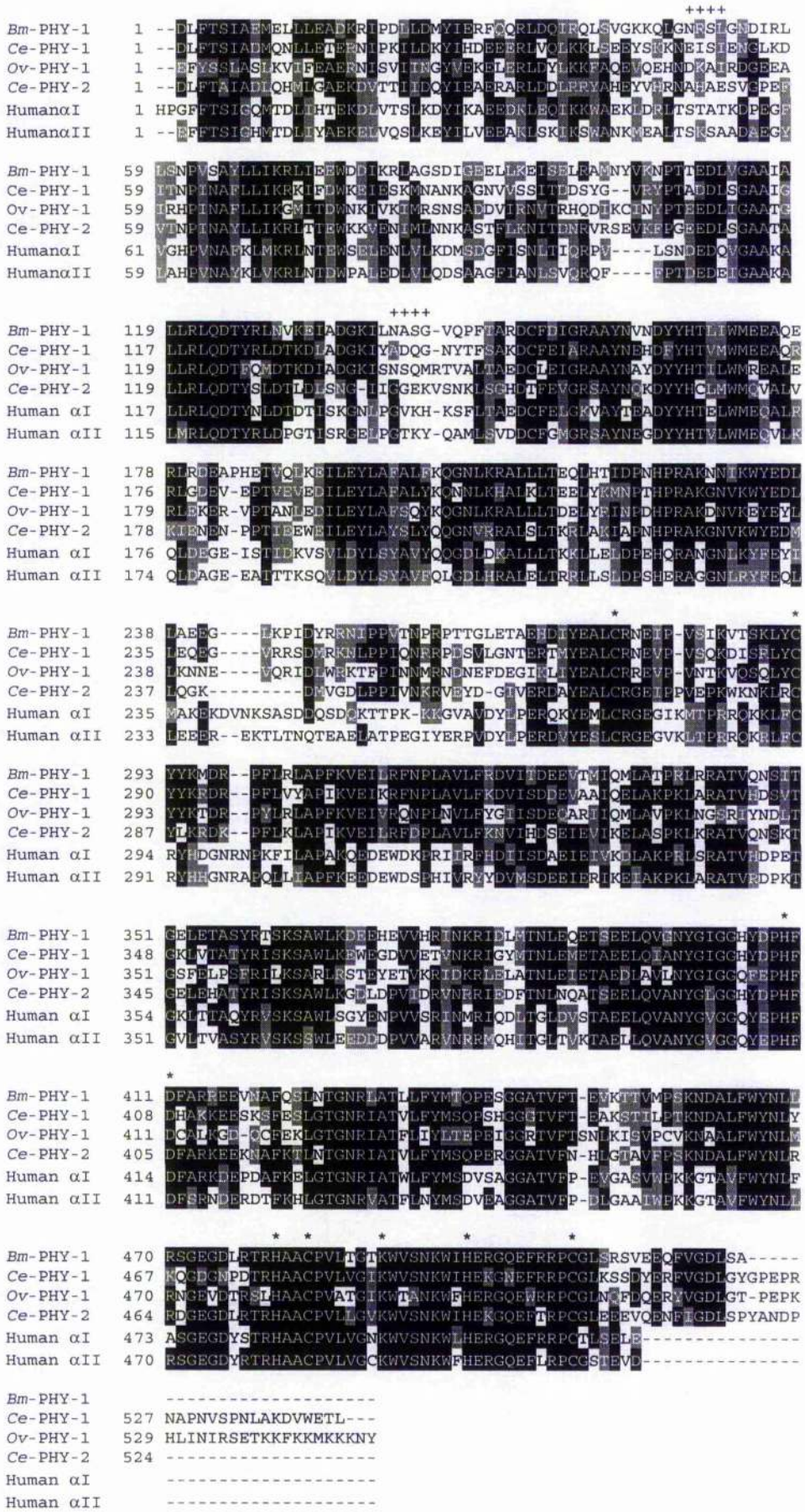
Comparison of putative *Bm*-PHY-1 amino acid sequence with known prolyl 4-hydroxylase sequence suggested the 5' end sequence was incomplete. The N-terminal amino acids also did not contain an expected ER signal sequence when analysed using Signal P (Nielsen *et al.*, 1997). This suggested that none of the clones sequenced contained the sequence from the extreme 5' end. The 5' RACE system for rapid amplification of cDNA ends was used to determine the *Bm-phy-1* 5' coding sequence (see Figure 6.1). This technique involves the transfer of a homopolymeric tail to the end of the unknown DNA sequence. PCR primers based on known sequence and the artificially linked sequence were then used to amplify the 5' end of the cDNA. A total of three clones from two separate PCRs were sequenced and provided information on *trans*-splicing, 5' UTR sequence and 111 bp of 5' protein coding sequence. The 5' RACE data was assembled with the existing *Bm-phy-1* sequence to give the full-length cDNA. The linkage between the two sets of sequences and the overall sequence of the gene was confirmed by complete sequencing of a full-length PCR product (see Figure 6.1).

6.2.1.5 Complete sequence of *Bm-phy-1* cDNA and predicted protein

The completed cDNA *Bm-phy-1* sequence was submitted to the EMBL database under accession number AJ297845. *Bm-phy-1* is *trans*-spliced by a consensus 22 bp SL1 splice leader sequence. The presence of this *trans*-spliced leader sequence does not allow for the identification of the transcriptional start site by this analysis. An 8 bp 5' UTR is then followed by a single open-reading frame of 1626 bp and a 3' UTR of 154 bp, giving a predicted protein of 541 amino acids. The N-terminal amino acid sequence was analysed for an ER signal sequence as predicted using the Signal P server (Nielsen *et al.*, 1997). A signal peptide was predicted in *Bm*-PHY-1 with cleavage most likely to occur between amino acids A17 and D18 of the protein sequence giving a mature protein size of 524 amino acids. Possession of a signal peptide by *Bm*-PHY-1 is expected from comparison with other prolyl 4-hydroxylase subunits and the expected ER localisation of this enzyme. An alignment of the derived mature amino acid sequence of *Bm*-PHY-1 (signal peptide removed) in comparison with mature prolyl 4-hydroxylase α subunits from other organisms is shown in Figure 6.2. *Bm*-PHY-1 was most similar to *Ce*-PHY-1 with 59% identity and 76% similarity. Next most similar was

Figure 6.2. Amino acid alignment of *Bm*-PHY-1 with *C. elegans*, *O. volvulus* and human α subunits

Amino acid alignment of *B. malayi* (*Bm*), *C. elegans* (*Ce*) and *O. volvulus* (*Ov*) PHY proteins, and human P4H α subunits. Gaps (-) were introduced for maximal alignment and signal peptides were removed, therefore numbering refers to the mature processed proteins. The two predicted N-glycosylation sites of *Bm*-PHY-1 are marked (++++) and refer only to this polypeptide. Conserved cysteines (C), aspartate (D), histidine (H) and lysine (K) residues are indicated with an asterisk. Genbank Accession numbers- *Bm*-PHY-1 (AJ297845), *Ce*-PHY-1 (Z81134), *Ce*-PHY-2 (Z69637), *Ov*-PHY-1 (AF369787) Human αI (M24486) and Human αII (U90441).



Ce-PHY-2 with which *Bm*-PHY-1 shows 53% identity and 71% similarity. The filarial parasitic nematode *Onchocerca volvulus* protein *Ov*-PHY-1 (Merriweather *et al.*, 2001) shows 49% identity and 70% similarity. The human subunits αI (Helaakoski *et al.*, 1989) and αII (Annunen *et al.*, 1997) are 45% and 44% identical, and 62% and 63% similar respectively. *Ce*-PHY-3 (Riihimaa *et al.*, 2002), *Ce*-PHY-4 and *Ce*-PHY-5 are not included in the alignment analysis, these shared from between 10-18% identity and 19-31% similarity with *Bm*-PHY-1. The conserved cysteine residues required for intrachain disulphide bonding (John and Bulleid, 1994; Lamberg *et al.*, 1995) and the active site histidine, aspartic acid and lysine residues (Lamberg *et al.*, 1995; Myllyharju and Kivirikko, 1997) are all found in *Bm*-PHY-1 (indicated with an asterisk in Figure 6.2). The predicted peptide-binding domain of *Bm*-PHY-1, which was defined for human αI (Myllyharju and Kivirikko, 1999), extends from N140 to P260 of the processed protein. The extended C-terminal region found in *Ce*-PHY-1 and *Ov*-PHY-1, which, for *Ce*-PHY-1, is implicated in PDI binding (Veijola *et al.*, 1996a), is not present in *Bm*-PHY-1. *Bm*-PHY-1 has two predicted N-linked glycosylation sites at N49-L52 and N140-G143 (indicated with +, Figure 6.2), a predicted mature molecular mass of 60.4 kDa and isoelectric point of 5.48. The protein is thus comparable to *Ce*-PHY-1 whose 543 amino acids (minus 16 amino acid signal peptide) have a isoelectric point of 5.83, a mass of 62 kDa, and one N-linked glycosylation site (Veijola *et al.*, 1994) in a similar position (N142-F145) to the second *Bm*-PHY-1 site. A tetratrico peptide repeat motif (Lamb *et al.*, 1995) is also found in *Bm*-PHY-1 at a similar position to the motif identified in *Ce*-PHY-1.

6.2.2. *Bm-phy-1* genomic coding sequence

The consensus sequence from three separate PCR products was used to determine the 4596 base pair full length genomic coding sequence for *Bm-phy-1*. The sequence spanning 12 exons and 11 introns was submitted to EMBL with accession number AJ421993 (see Appendix 1). Exonic sequences were identical to those derived from mRNA. Introns ranged in size from 119 bp to 479 bp with an average intronic size of 270 bp. Larger intronic size, as compared to *C. elegans*, is a general feature of the *B. malayi* genomic sequences available (Blaxter, 1998). A diagram of gene structure with exon and intron sizes is shown in Figure 6.1. All intron sequences began with a GT

dinucleotide and ended with an AG in keeping with the consensus sequence for intron-exon boundaries (Blumenthal and Steward, 1997).

6.2.3. Baculovirus expressed *Bm*-PHY-1

Baculovirus expression enables multiple proteins to be expressed at the same time in insect cells and protein associations to be determined (Vuori *et al.*, 1992a). Co-expression of multiple proteins is accomplished by infection of cells with multiple recombinant viruses. Expression of *Bm*-PHY-1 was carried out in collaboration with J. Myllyharju at the Collagen Research Unit, University of Oulu, Finland. α subunits expressed alone in this system are usually highly insoluble (Vuori *et al.*, 1992a), one of the primary functions of the PDI subunit being to prevent aggregation (John *et al.*, 1993). The other defined function of PDI subunits in P4H complexes is to retain the α subunits within the ER (Vuori *et al.*, 1992b). Protein was extracted from insect cells expressing *Bm*-PHY-1 alone in a Triton-X 100 buffer, surprisingly *Bm*-PHY-1 was found within this fraction and did not require further solubilisation in 1% SDS, a treatment normally needed for singly expressed α subunits. Therefore *Bm*-PHY-1 is soluble when expressed alone from baculovirus in an insect cell expression system and does not aggregate in the absence of PDI. Triton-X 100 extracts run on native gels show three *Bm*-PHY-1 immunoreactive bands, Figure 6.3. P4H activity of insect cell expressed *Bm*-PHY-1 was analysed using a method based on the hydroxylation-coupled decarboxylation of 2-oxo-[1-^{14}C]glutarate (Kivirikko and Myllylä, 1982) (performed by J. Myllyharju). This demonstrated that *Bm*-PHY-1 is an active P4H enzyme with activity of about half that shown for the *C. elegans* tetramer (J. Myllyharju personal communication). Other examples of P4H which are active in the absence of a PDI subunit exist (Eriksson *et al.*, 1999; Hieta and Myllyharju, 2002; Kaska *et al.*, 1987; Kaska *et al.*, 1988), however none, with the exception of the HIF-α modifying hydroxylases (Bruick and McKnight, 2001; Epstein *et al.*, 2001), are from the animal kingdom. This unique feature of the *Bm*-PHY-1 of forming an active enzyme in the absence of a PDI subunit, provides another example of the diverse forms of these enzymes in free-living and parasitic nematode species. The potential of *Bm*-PHY-1 to bind to other PDI proteins was also examined by co-expressing the protein in the insect cell system with PDI subunits from a variety of species (co-infections performed by Johanna Myllyharju) including *C. elegans* and human. Surprisingly the BM protein does

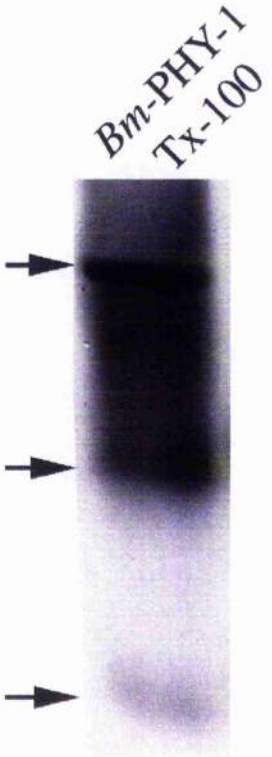


Figure 6.3. Native PAGE of a Triton X-100 extract of insect cell expressed *Bm*-PHY-1. Triton X-100 extracts from insect cells expressing Bm-PHY-1. Native Western analysis of this extract shows the presence of three *Bm*-PHY-1 immunoreactive bands (arrowed).

not associate with any of the β-subunits tested. This is in contrast to the *C. elegans* and human α subunits which can both associate with the other PDI subunit (Veijola *et al.*, 1996a; Veijola *et al.*, 1994).

6.2.4. Comparison of nematode and insect cell extracted *Bm*-PHY-1

A comparison of *Bm*-PHY-1 from nematode extracts and insect cell expressed proteins was performed. Nematode protein was extracted from approximately 100 adult female *Brugia malayi* parasites. The procedure was designed to keep the complexes in their native state by preserving protein-protein interactions thus enabling analysis of potential complexes. When compared directly by native PAGE and Western blotting both extracts show two major *Bm*-PHY-1 immunoreactive bands (Figure 6.4). The bands from different sources do not migrate at comparable positions, with bands from parasite extracted material migrating higher than the insect cell expressed protein. In Figure 6.4 lane 3 a *C. elegans* native extract probed with anti-*Ce*-PHY-1 antibody is shown for comparison, with sizes of the dimer and tetramer bands indicated. Bands from both worm extracts can be seen to have a more similar profile than either does to the *Bm*-PHY-1 insect cell sample. Although extractions from nematodes and insect cells were performed using different buffers, differences in buffer systems are unlikely to account for the size differences as *C. elegans* complexes extracted from worm and cell sources using these buffers migrated in identical positions to each other (Figure 5.1)

6.2.5. Glycosylation levels in native and baculovirus expressed *Bm*-PHY-1

The *Bm*-PHY-1 protein contains two predicted N-glycosylation sites (see Figure 6.2). Extracts from nematode and insect cells were treated with N-glycosidase F and analysed by SDS PAGE and Western blotting to determine whether these sites are modified, and whether the native size difference found between these samples was due to differences in glycosylation levels. Figure 6.5 shows glycosidase and non-glycosidase treated samples from insect cell extract (lanes 1 and 2) and worm extract (3 and 4). With glycosidase treatment a single band of approximately 60 kDa, representing unglycosylated protein, is observed for both samples (lanes 1 and 3). The untreated insect cell extract (lane 2) displays the 60 kDa unglycosylated band with an additional band of approximately the same intensity migrating slightly higher. Unglycosylated *Bm*-PHY-1 is not found in the untreated *B. malayi* worm extract, the nematode extract lacks

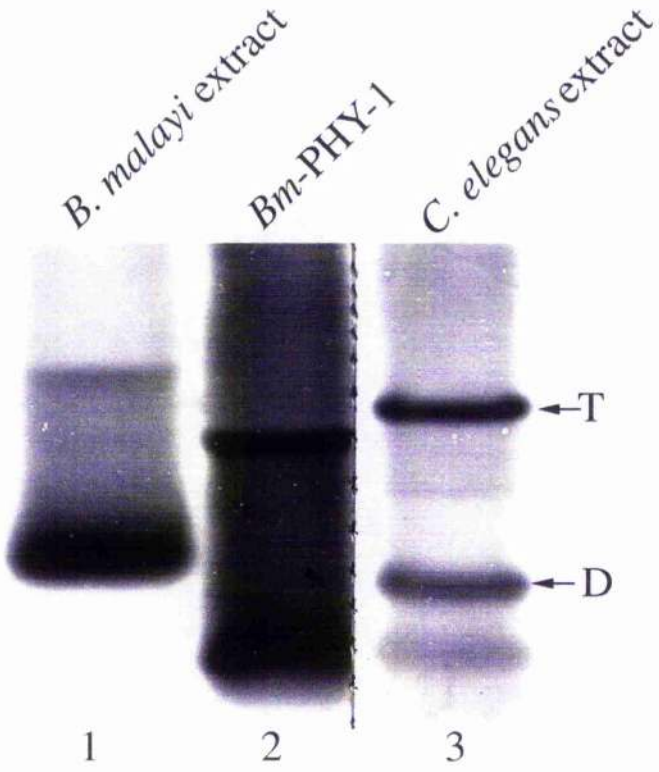


Figure 6.4. Native PAGE comparison of nematode extracted and recombinant *Bm*-PHY-1

Bm-PHY-1 extracted from *B. malayi* parasites, lane 1, and from insect cells, lane 2, was compared by Native PAGE probed with anti-*Bm*-PHY-1, along with *C. elegans* extract (run on the same gel then separately probed with anti-*Ce*-PHY-1) in lane 3. Positions of tetramer (T) and dimer (D) bands are arrowed for the *C. elegans* extract.

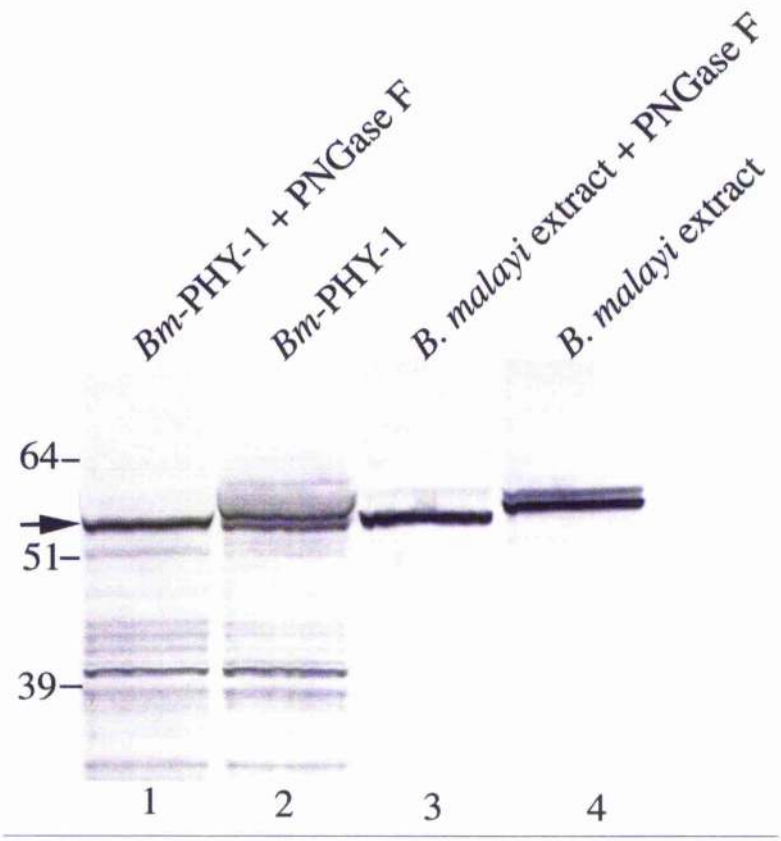


Figure 6.5. Analysis of glycosylation levels in nematode and insect cell extracts

Glycosylation levels of Bm-PHY-1 in extracts from *B. malayi* parasites and insect cells were determined by comparing N-glycosidase F treated and untreated samples from both sources. Lane 1 - insect cell extract glycosidase treated, lane 2 - insect cell extract untreated, lane 3 - worm extract glycosidase treated, lane D - worm extract untreated. The arrow indicates the 60 kDa unglycosylated form of *Bm*-PHY-1 found in lanes 1 and 3. In lane 2 the 60 kDa band is found along with one higher band, representing the single glycosylated form. In lane 4 the single glycosylated form is found most abundantly along with a higher, twice glycosylated form.

the 60 kDa band but has two higher bands. The smaller of these is the more prominent and corresponds to the larger band found in lane 2. This band probably represents the mono-glycosylated form of *Bm*-PHY-1 with the higher less abundant band in lane 4 representing *Bm*-PHY-1 glycosylated at both sites. Therefore compared to the worm extracts, insect cell extracts are under glycosylated containing approximately equal quantities of single glycosylated and non-glycosylated protein. In contrast, nematode derived *Bm*-PHY-1 is always glycosylated, predominantly singly, with a lesser amount modified at both sites. Differences in glycosylation may account for some or all of the size difference found between worm and insect samples analysed under native conditions.

6.2.6. Functional analysis of *Bm*-PHY-1

6.2.6.1. Vector for expression of *Bm*-PHY-1 in *C. elegans*

Phenotypic rescue of a *C. elegans* P4H mutant with *Bm-phy-1* was attempted in order to assess interspecies conservation of gene function and gain insight into the native function of the *B. malayi* gene. Attempts were made to rescue the *C. elegans phy-1* null mutant [*dpy-18(e364)*] using the *Bm-phy-1* coding sequence expressed under control of a *C. elegans* promoter. A vector, pAW1, was constructed which contained the *C. elegans phy-1* promoter and 3' UTR. The 3' UTR region incorporated the polyadenylation signal sequence and poly(A) transfer site. For a further description of the vector pAW1 see Section 3.2.11.1. A similar approach had been successfully applied to rescue the *dpy-18* mutant phenotype with the endogenous *C. elegans* gene and to partially rescue with human P4H α subunits (Figure 6.6A and Figure 3.24). Full-length *Bm-phy-1* coding sequence, in the form of cDNA or genomic DNA, was inserted between the two *C. elegans* sequences as depicted in Figure 6.6B. The junction between the *Ce-phy-1* promoter sequence and the *Bm-phy-1* coding sequence encodes four extra amino acids at the N-terminus of the protein. These result from the first five bases of the *Ce-phy-1* coding sequence being included with the promoter sequence, the *Bam* HI restriction site used to fuse both fragments, and an additional base engineered in the *Bm-phy-1* sense primer to adjust the reading frame. The addition of these amino acids was predicted not to interfere with the function of the N-terminal ER signal sequence of *Bm*-PHY-1 by using the Signal P signal peptide prediction program. An ER signal peptide is still identified with cleavage occurring at the same site. The additional N-

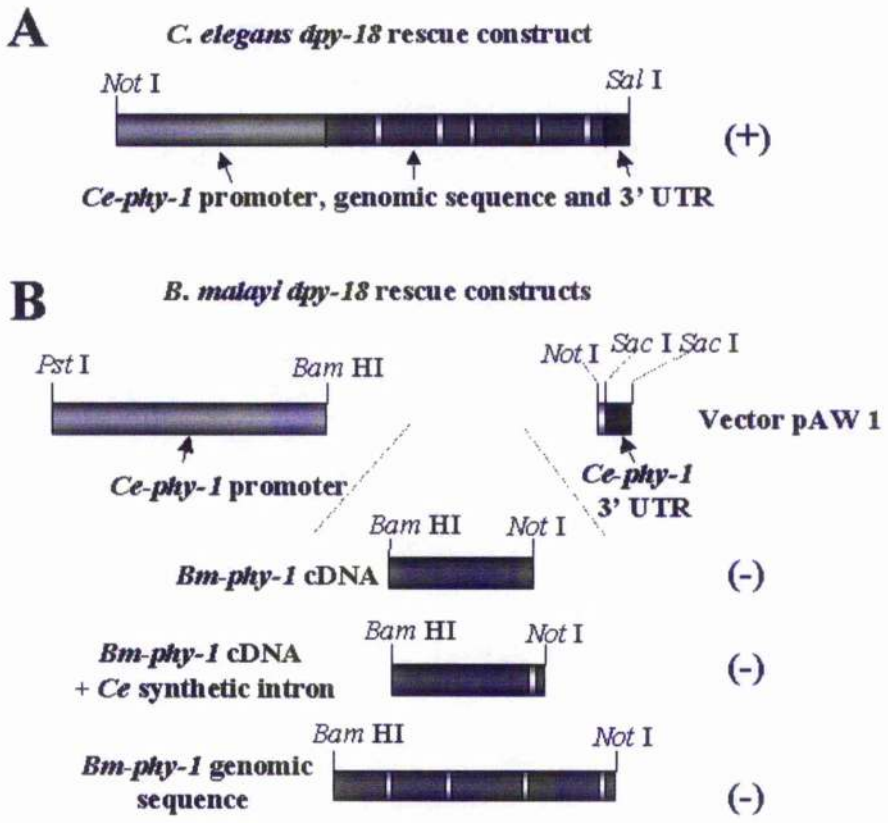


Figure 6.6. Diagram of *Bm-phy-1 dpy-18* mutant rescue constructs

(A) Depiction of the *C. elegans phy-1* gene construct (containing promoter, genomic and 3' UTR sequences) used to successfully rescue *dpy-18* mutants, indicated with, (+). (B) Representation of vector pAW1 with inserted *Bm-phy-1* sequences. Vector pAW1 consists of the *Ce-phy-1* promoter and 3' UTR. Between these sequences *B. malayi phy-1* cDNA, cDNA with a single *C. elegans* synthetic intron, and genomic sequences were inserted. Sizes of fragments are not drawn to scale and the full number of introns for the genomic fragment not depicted. Constructs were injected into the *C. elegans* P4H mutant *dpy-18(e364)*. Lack of rescue with transformed *Bm-phy-1* sequences is indicated, (-).

terminal amino acids were therefore predicted not to interfere with native protein function. Similar extra amino acids were added to the α subunits of humans which still retained activity (see Section 3.2.11.2).

6.2.6.2. Rescue attempts with *Bm-phy-1* cDNA derived constructs

Constructs were microinjected into a strain carrying a *dpy-18(e364)* mutation and transgene function assessed by the ability to repair the medium Dpy phenotype of this mutant. Transformed animals from multiple semi-stable lines made by transformation with high (100 μg/ml) and low (10 μg/ml) concentrations of the rescue construct were examined but no phenotypic rescue was observed. The plasmid dpy-7-GFP was co-injected as a marker of transformation. Figure 6.7 shows *C. elegans dpy-18*(*e364*) mutants transformed with the rescue construct (panels C, DIC image, and D, showing expression of marker plasmid under UV light) in comparison to wild type N2 (panel A) and a *dpy-18*(*e364*) mutant (panel B) with no repair to wild type body morphology evident in transformed *dpy-18*.

Lack of introns in this construct could result in low levels of transgene expression as a comparison of intron containing and intron-less transgenes in transgenic animals and tissue culture demonstrated increased expression levels of around 100-fold from the spliced transcripts (Brinster *et al.*, 1988; Buchman and Berg, 1988). An artificial intron based on typical *C. elegans* introns (Blumenthal and Steward, 1997) was synthesised and inserted in to the *Bm-phy-1* cDNA coding sequence using a blunt ended restriction site. The 51 bp *C. elegans* intron was inserted into a *Stu* I restriction recognition site located 57 bp from the end of *Bm-phy-1* coding sequence. Microinjection of this synthetic intron-containing *Bm-phy-1* construct into *dpy-18* also failed to rescue the Dpy phenotype, with results identical to those shown in Figure 6.7 found using different concentrations and with multiple lines.

6.2.6.3. Rescue attempts with *Bm-phy-1* genomic constructs

Genomic *Bm-phy-1* coding sequence constructs were used to determine if the multiple intron containing genomic sequence was capable of rescuing *dpy-18*. High concentrations (100 μg/ml) of this construct were toxic and no transformed lines could be generated as transformed nematodes died during embryonic development.

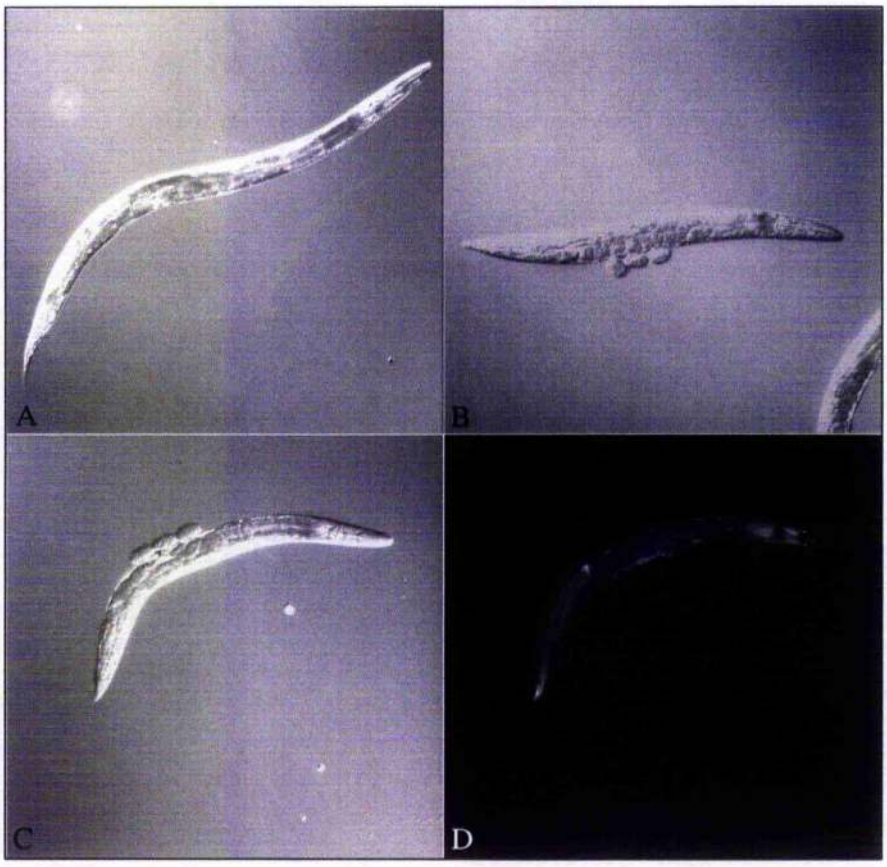


Figure 6.7. Comparison of body shapes of wild type and *dpy-18 C. elegans* strains with transformed *dpy-18* transgenic lines

Typical results from *dpy-18* rescue attempts with *Bm-phy-1*. Panel A, wild-type N2 *C. elegans* animal; panel B, *Ce-phy-1* null strain *dpy-18(e364)* with shorter fatter body shape than the wild type; panel C, DIC image of a transformed *dpy-18* line with panel D showing corresponding fluorescence from the marker plasmid dpy-7-GFP. Panels C and D show the failure of Bm-*phy-1* transformation to repair the *dpy-18* phenotype. These results were consistent for all three types of construct injected; cDNA, cDNA with synthetic intron and genomic, and for all lines examined and at all concentrations attempted.

Examination of multiple lines generated from lower, non-lethal, injection concentrations showed no repair of the Dpy phenotype in transformed worms compared to non-transformed controls. A possible reason for the failure of the *Bm-phy-1* genomic construct to rescue could be the presence of the eleven *B. malayi* introns within the coding sequence of *Bm-phy-1*. Although all of these introns seem to conform to the consensus splicing sequences of *C. elegans* (Blumenthal and Steward, 1997) this organism may not be able to splice introns properly and/or efficiently for other reasons. Improperly spliced *Bm-phy-1* mRNA could result in either a truncated protein from the creation of a premature stop codon, a protein missing domain(s) due to exonic regions being spliced out, or lack of splicing which would either not be exported or would result in a non-functional protein. Splicing in *C. elegans* of heterologous genes from *H. contortus* (Kwa *et al.*, 1995; Redmond *et al.*, 2001) occurs correctly, however no data is yet available on the expression and splicing in *C. elegans* of *B. malayi* genes.

6.2.6.4. Detection of transgene expression

Expression of *Bm-phy-1* transformed *dpy-18(e364)* lines was assessed using RT-PCR from single worms. Transformed lines containing just the cDNA construct were not assessed for transgene expression. For the cDNA with synthetic intron constructs discrimination on the basis of size between spliced and unspliced transcripts (and product arising from contaminating plasmid DNA) was not practical due to the small size of the intron. Instead primer BM phy-1 SWRT R was designed to span the synthetic intron-exon boundary in such a way that only 6 bp of the 26 bp primer lay to the 5' end of the intron. On un-spliced mRNA derived product or contaminating DNA the primer should anneal beyond the 3' end of the intron sequence and so be unable to yield a product due to bases at the 3' end of the primer being unable to anneal (see Figure 3.25). If the gene is transcribed and the intron correctly spliced this primer should anneal and, used in conjunction with a standard forward primer, should generate a product. PCR product arising from correctly spliced RNA could also be confirmed by digestion of the product with *Stu* I, the site which would be reformed upon correct processing of the RNA.

Primers BM phy Res F and BM phy-1 SWRT R were tested on *Bm-phy-1* cDNA and cDNA with synthetic intron plasmid DNA. Primers gave the correct sized band of 1.6 kb from the cDNA clone and gave no product from the clone containing the synthetic

intron as expected (data not shown). Therefore in single worm RT-PCR a product should only result from reverse transcribed mRNA and not from unspliced mRNA or any plasmid remaining after DNase treatment. RT-PCRs were performed on two transformed lines and untransformed controls, no expression of transgenes could be detected in the transformed lines using this method (data not shown). A *Bm-phy-1* genomic DNA construct line was also examined for transcription of the transgene by single worm RT-PCR using full-length gene cloning primers in comparison to control untransformed worms. Correct splicing and discrimination of contaminating DNA was based on the large size difference between spliced product (1.6 kb) to unspliced product (4.5 kb). No product was obtained for the genomic *Bm-phy-1* containing line (data not shown). Western analysis was also attempted to determine if transgenic lines were expressing the *Brugia* protein. As the antibody to the *B. malayi* protein is reactive to the last 20 C-terminus amino acids, detection of expression by this means would indicate that expression and accurate splicing had occurred. Worms were collected from transformed *dpy-18* lines containing *Bm-phy-1* cDNA with synthetic intron and genomic constructs. Two different synthetic intron lines were tested along with one genomic line with experiments repeated twice. No *Bm*-PHY-1 protein could be detected however using this method (data not shown).

6.2.7. Temporal expression

The EST data and library screens provided information on the developmental stages where expression of *Bm-phy-1* is found. ESTs were identified from adult females and L3 larvae, with library screening demonstrating expression in adult males. For a more detailed analysis of developmental timing of expression, RNA extracted from different points in the nematode life cycle was examined for the presence of the *Bm-phy-1* transcript. RT-PCR was performed using cDNA (a gift from Bill Gregory of ICAPB in the University of Edinburgh) made from daily extracts of infected jirds (daily extract taken up to day 14 post infection, after which 2-4 day extracts were taken) (Gregory *et al.*, 2000). These samples cover the L3/L4 and L4/adult moults. PCR was performed using the primers BM phy 1.1 IS 1F and BM phy 1.1 IS 2R. This primer pair span an intron and produce a product of 655 bp which can be distinguished from the larger product produced from genomic DNA. A second set of primers to the *B. malayi* tubulin gene, which also span an intron, were simultaneously applied as an internal control. PCR was performed on the staged cDNA samples with gene specific and control

primers (Figure 6.8). Gels were also Southern blotted, radioactively probed and quantified by scintillation counting. Bands from *Bm-phy-1* and *B. malayi* tubulin from each time point were excised and the relative abundance of *Bm-phy-1* mRNA levels was determined by subtracting the value for tubulin from the same time point (data not shown). Although this process was repeated on three independent sets of PCRs no repeatable pattern of values from quantification could be generated. Expression was also found by RT-PCR from an L1 (microfilaria) extract (data not shown). Taken together with the EST and library screening data these results show that *Bm-phy-1* is expressed in microfilaria, at all time points examined through out L3 and L4 development and in both adult sexes. Due to the inability to generate reliable data from quantification the fluctuation of expression levels within these larval stages and between the L3/L4 moult and L4/adult moult could not be assayed although visual examination does suggest a high level of expression during the moulting period.

6.2.8. Analysis of the promoter region from *Bm-phy-1*

6.2.8.1. Cloning of the putative *Bm-phy-1* promoter

Two methods were used to identify sequences lying upstream from the 5' end of *Bm-phy-1* coding sequence. A combination of a linker based PCR genomic walking technique and isolation of a BAC from a genomic *Brugia* library generated data for 2.2 kb of putative promoter sequence. This was submitted to the EMBL database with accession number AJ421994. Upstream sequences were confirmed as being linked to the coding region by PCR on genomic DNA. A 3' *trans*-splice acceptor site sequence, TTTCCAGA, is located at position –15 to –8 with respect to the ATG with splicing of the SL1 sequence occurring after the G and before the A nucleotide (see Appendix 1). The sequence differs from the consensus *trans*-splice acceptor site of TTTTCAGA found most commonly in *C. elegans* (Blumenthal and Steward, 1997). Located 130 bp from the splice acceptor site is a TATAA box (indicated in Appendix 1). This is the binding site for the RNA polymerase II TATA box binding protein. Three heat shock elements, two of which are also found in the promoter of *Ce-phy-1*, are present in the upstream sequence of the *Brugia* gene. The significance of these elements has not been determined for either gene. No other motifs could be identified and no major regions of similarity could be identified between the *Bm-phy-1* promoter and those from the *C. elegans phy* genes when Dot plot analysis was performed (data not shown). The 2.2 kb

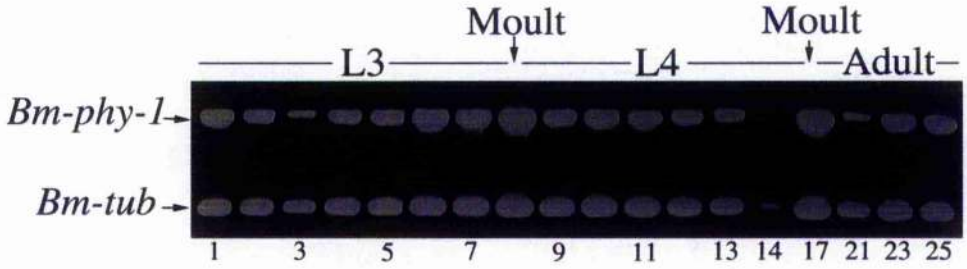


Figure 6.8. Analysis of *Bm-phy-1* expression through L3 to adult development

Expression of the *Bm-phy-1* transcript was compared to the constitutively expressed tubulin transcript through L3 to L4 and L4 to adult moults. Extracts made from parasites were from daily samples from infected jirds through L3 and L4 then at longer intervals in the adult. Numbers refer to days post infection. PCR was performed on these samples simultaneously with primer pairs for both genes.

region identified does not appear to contain any other coding sequence as determined by BLASTx analysis.

6.2.8.2. *Bm-phy-1* promoter-driven reporter gene expression in *C. elegans*

PCR generated upstream sequences were translationally fused to multi-intron containing, nuclear localised reporter genes *lacZ*/GFP in the vector pPD96.04 (described in Figure 3.2). This construct, fusing approximately 2.2 kb of *Bm-phy-1* upstream sequence and the first three N-terminal amino acids from *Bm*-PHY-1, was transformed into the *C. elegans* strain DR96 at 100 µg/ml. Multiple lines were examined with expression being found consistently in the hypodermal cells; shown in Figure 6.9, panels A-D. Reporter GFP expression was particularly prominent in the hypodermal cells hyp5, 6 and 7 of the head (panel A), and in the pair of hyp7 cells in the tail of adult worms (panel B). Using sensitive staining techniques for the detection of β-galactosidase activity, expression can be seen in larval stages again strongly in the head and tail (panels C and D). Less pronounced expression can also be seen in the hypodermal cells hyp4, 5, 6 and 7 of the body (panel C). With prolonged staining *lacZ* expression was also observed in the vulval cells of the adult (data not shown), which are of ectodermal origin, and in adult body-wall muscle cells (data not shown). Expression was found in embryos from the 3-fold stage of development (data not shown). This pattern of expression predominantly in cells of hypodermal origin is consistent with this gene functioning to modify cuticular collagens. The secondary staining of muscle cells and vulval cells is identical to that found for the *C. elegans* homologues *Ce-phy-1* and *Ce-phy-2,* which both express in vulval cells with additional muscle cell staining found for *Ce-phy-2*. Both *C. elegans* genes showed much stronger and more complete hypodermal staining under less sensitive conditions than is observed here.

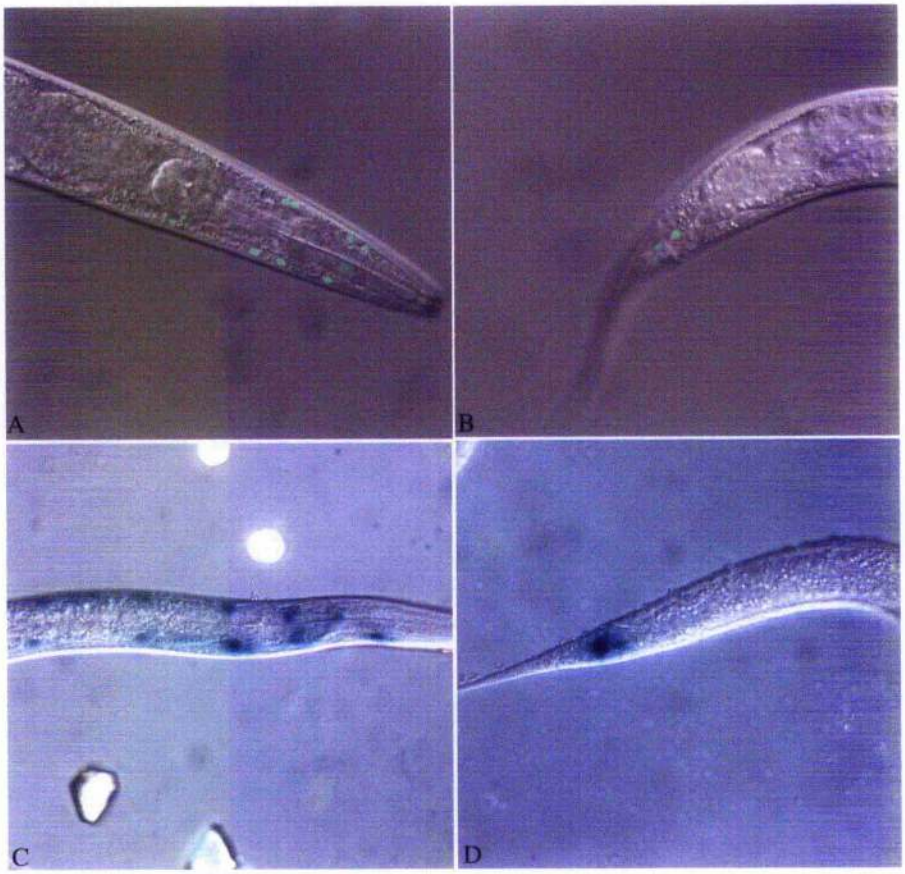
6.2.9. Identification of other potential *B. malayi* P4H subunits

In order to more fully understand the nature and composition of P4H complexes in *Brugia malayi*, additional *phy* isoforms and a *pdi* homologue, if they are present in the genome, should to be identified and characterised. A candidate *B. malayi* PDI was identified using the program tBLASTn with full length *Ce*-PDI-2 protein sequence compared against the *Brugia malayi* nucleic acid database. An EST sequence was identified the derived amino acid sequence for which displayed high identity to PDI

Figure 6.9. *Bm-phy-1* promoter driven expression in the hypodermal cells of transgenic *C. elegans*

Transgenic expression patterns in *C. elegans* of reporter genes expressed from the *Bm-phy-1* promoter. Panel A is a merge of DIC and UV images showing expression of GFP in the hypodermal cells hyp5, 6 and 7 in the head. Panel B is a merge showing expression of GFP in the hyp7 cell of the tail. Panels C-D show *lacZ* staining, with panel C showing transgene expression in the hypodermal cells hyp4, 5, 6 and 7 of the head and body; panel D indicated reporter expression in hyp7 cells of the tail (compare panel B).

Facing page 243


A
B
C
D

protein sequences. EST SWAMCAC30B11SK (accession number AI784701) was submitted by Steven A. Williams, FGP, Smith College. This EST would appear to represent the incomplete N-terminal region of a potential *Brugia* PDI protein, see Figure 6.10A. The predicted translation of 92 amino acids from this EST, putatively termed *Bm*-PDI, showed 94% identity and 96% similarity with *Ov*-PDI (Wilson *et al.*, 1994). With *Ce*-PDI-2 *Bm*-PDI shows 76% identity and 86% similarity. This sequence then shows higher identity to the human PDI (63% identity, 75% similarity) than to the other defined non-P4H associated PDI isoform from *C. elegans*, PDI-1 (Veijola *et al.*, 1996a) (61% identity, 78% similarity). This EST clone was received from the FGP however cloning and sequencing of it did not provide any additional data as the insert of the clone was under 400 bp in size and did not generate good sequence data.

To date, the available ESTs in the *B. malayi* database showing homology with prolyl 4-hydroxylases are the three ESTs described which represented *Bm-phy-1* (an additional EST, SWD25CAU08C01SK, is 97% identical with *Bm-phy-1* and most probably also represents this gene). Screening of a *Brugia* bacteriophage cDNA library with *Bm-phy-1* 3' coding sequence did not identify any additional *phy* genes. To identify other potential *Bm-phy* genes tBLASTn was performed using full-length *Ce*-PHY-2 protein sequence compared against a translated database of genomic sequences. A *B. malayi* BAC end sequence, submitted by the Mark Blaxter, FGP, Edinburgh as a Genome Survey Sequence (GSS), was identified that was not identical to *Bm-phy-1*. BMBAC305D10T7 (accession number BH615947) when translated in frame -2 shows homology with PHY subunits. The derived 79 amino acid sequence, provisionally named *Bm*-PHY-2, shows 42% identity (60% similarity) with *Ce*-PHY-2. A low identity of approximately 30% is shared with *Bm*-PHY-1, however an identity of 76% is found between *Bm*-PHY-2 and *Ov*-PHY-1. An alignment of protein sequences from both *Brugia* PHY proteins and *Ov*-PHY-1 is shown in Figure 6. 10B.

The BAC from which the *Bm-phy-2* sequence was identified is not useful for identification of the complete sequence of this gene. *Bm-phy-2* resides at the end of this BAC and appears to represent the incomplete 5' end of the coding sequence. The orientation of *Bm-phy-2* is however 3' to 5' with respect to this end of the clone, thus continued sequencing would only complete the 5' end of the gene. Therefore the nature

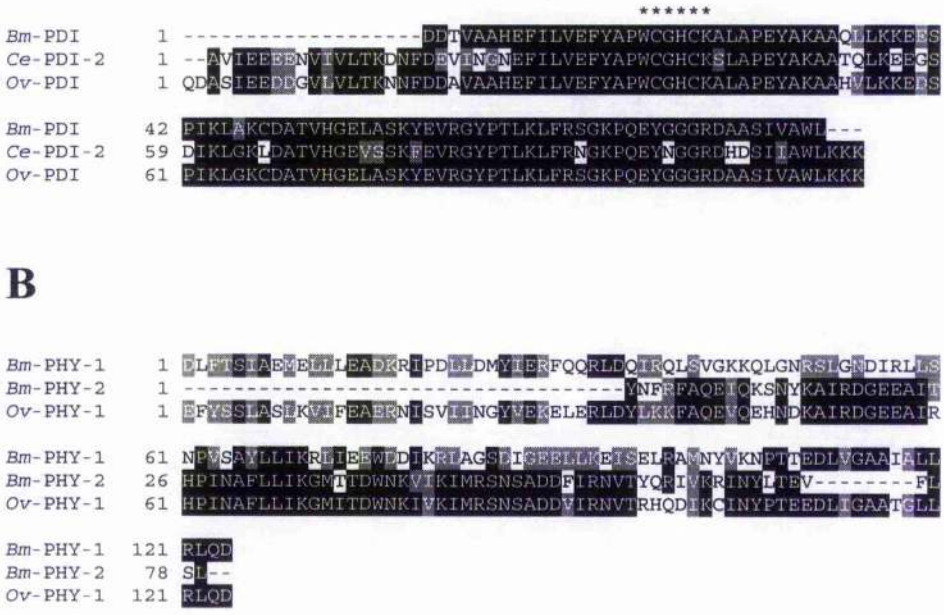


Figure 6.10. Amino acid comparisons of *Bm*-PDI with *C. elegans* and *O. volvulus* PDIs.

A) Comparison of nematode PDIs. Amino acid alignments of the predicted 92 amino acid region from the EST SWAMCAC30B11SK termed *Bm*-PDI with nematode PDIs from *C. elegans* (*Ce*) and *O. volvulus* (*Ov*). Active site residues are indicated with asterisks. Accession numbers, *Bm*-PDI EST (AI784701), *Ce*-PDI-2 (U41542) and *Ov*-PDI (U12440).

B) Comparison of nematode PHYs. Amino acid alignments of the derived 79 amino acid sequence from GSS sequence BMBAC305D10T7 named here *Bm*-PHY-2 with PHYs from filarial nematodes *B. malayi* (*Bm*) and *O. volvulus* (*Ov*). Accession numbers *Bm*-PHY-2 EST (BH615947), *Bm*-PHY-1 (AJ297845) and *Ov*-PHY (AF369787).

and orientation of the *Bm-phy-2* sequence within this clone does not facilitate the generation of 3' sequence through further analysis of the BAC clone.

6.3. Discussion

6.3.1. Possible forms of the *Bm*-PHY-1 enzyme

In what form(s) the *Bm*-PHY-1 enzyme occurs has not yet been conclusively established. *Bm*-PHY-1 may not require association with any other protein to form an active enzyme and in this case the presence of multiple bands found in native analysis could represent multimers of the *Bm*-PHY-1 protein. In this case the native size differences between insect and nematode derived proteins would arise only from differing glycosylation modification levels. Whether monomer, multimeric or both forms of the enzyme were active would also have to be established in this scenario. Gel filtration studies using recombinant *Bm*-PHY-1 indicated that P4H is found only in fractions corresponding to a tetramer sized complex, indicating that *Bm*-PHY-1 may possibly self-associate into active tetramers which partially dissociate into enzymatically inactive dimer and monomer forms (J. Myllyharju, personal communication). Alternatively, if partner(s) are required for *Bm*-PHY-1, the size differences between native samples may reflect the differences in the size/charge/glycosylation of the natural partners in the nematode compared to insect cell PDI-like proteins with which *Bm*-PHY-1 may be forming complexes with when expressed in this system. Regardless of whether *Bm*-PHY-1 functions alone or forms an active complex with endogenous insect cell protein(s), it differs from its homologous enzymes in nematodes, vertebrates and *Drosophila* characterised to date, as all these enzymes require a PDI subunit for activity. P4Hs which do not require additional subunits, but function as monomers, have however been identified in *A. thaliana* (Hieta and Myllyharju, 2002), algae (Kaska *et al.*, 1987; Kaska *et al.*, 1988) and a virus (Eriksson *et al.*, 1999). The nematode and vertebrate HIF-α modifying hydroxylases also function in this manner (Bruick and McKnight, 2001; Epstein *et al.*, 2001). *Bm*-PHY-1 does not posses a recognisable ER retention signal and if acting without additional proteins would in theory not be retained within its expected sub-cellular localisation. However it would not be unique in this respect either, as neither the *Ov*-PHY-1, nor its partner, *Ov*-PDI, possess an identifiable ER retention motif (Merriweather *et al.*, 2001; Wilson *et al.*, 1994). P4H tetramers are usually formed from

$[(\alpha I)_2]\beta_2$ or $[(\alpha II)_2]\beta_2$, uniquely however in *C. elegans* a tetramer can only form if all three subunits (*Ce*-PHY-1, *Ce*-PHY-2 and *Ce*-PDI-2) are present. To determine if *Bm*-PHY-1 is capable of functioning in a mixed PHY subunit tetramer it could be expressed in combination with *Ce*-PDI-2 and either *Ce*-PHY-1 or *Ce*-PHY-2. This analysis would reveal whether *Bm*-PHY-1 requires more than one additional protein to form a complex with other subunits and may indicate whether this type of complex is perhaps nematode specific. This property of *Bm*-PHY-1, if identified, would then make it similar to *Ce*-PHY-2, as it requires both *Ce*-PHY-1 and *Ce*-PDI-2 to form a P4H complex. However, a clear difference however between the two proteins in this case would be that unlike *Ce*-PHY-2, *Bm*-PHY-1 is also a soluble, active P4H enzyme when expressed alone in a recombinant system.

6.3.2. Function of *Bm-phy-1*

The *Bm-phy-1* gene was found to be unable to rescue the phenotype of the *C. elegans phy-1* mutant, *dpy-18*. Thus the function of this gene is either different to the essential collagen modifying function defined for *Ce-phy-1* or the conditions of the experiment were not conducive to correct function of the *Brugia* gene. Identical approaches were employed for detection of expression from *Bm-phy-1* transgenes as were used in detection of human α subunit transgene expression (see chapter 3). The human α subunits were determined to be active due to their repair of *dpy-18* body morphology but despite this, no transgene expression was detected at either the protein or mRNA level. Therefore although no expression was detected for *Bm-phy-1* in *C. elegans* this did not rule out that the possibility the gene was being expressed. If the gene was expressed, codon usage differences between different species could be a factor determining protein production, and thus rescue efficiency. Examination of the *Bm-phy-1* coding sequence however does not reveal any unusual codons that are not utilised in *C. elegans*, nor does *Bm-phy-1* use any codons that are found rarely in *C. elegans*. If expression of the protein in its native form was occurring at sufficient levels lack of rescue could be attributable to other reasons. Firstly the protein when expressed by baculovirus had a lower activity level than the *C. elegans* tetramer. If this reflects the natural level of activity of the *Bm*-PHY-1 protein this may not be sufficient to compensate for the P4H activity needed to produce wild-type body shape *C. elegans*. Alternatively, if sufficient levels of active protein were present in transformed

nematodes, the *C. elegans* cuticular collagens may provide an unsuitable substrate for this enzyme and would not be correctly modified. If *Bm*-PHY-1 modifies cuticular collagens the available data would suggest that the *Brugia* substrate is similar. Only one cuticular collagen, *Bm*-COL-2 (Scott *et al.*, 1995), has been cloned from *B. malayi*. Analysis suggests that, like the *C. elegans* collagens, this collagen is a member of a multi-gene family (Selkirk *et al.*, 1989) although the exact size of this family is not known. The 33 kDa size of *Bm*-COL-2 protein is typical of the relatively small 26-35 kDa size of *C. elegans* collagens. *Bm*-COL-2 shows similarities with the *C. elegans* cuticular collagens, particularly to those of group 2. Comparison at the amino acid level gives 46% identity and 50% similarity with *C. elegans* collagen T07H6.3 and similar values for other collagens from group 2. Precise positioning of cysteine residues, required for registering and stabilising the collagen triple helix, is also conserved between *Bm*-COL-2 and group 2 collagens from *C. elegans*. Additionally, regions of conserved sequence in the amino terminal non-Gly-X-Y domains of *C. elegans* collagens (homology blocks A-D) (Kramer, 1997) are also found in *Bm*-COL-2, which contains homology domains B, C and D although it appears to lack the subtilisin-like protease recognition site (Scott *et al.*, 1995). Lastly, rescue may not be observed in *C. elegans* using *Bm-phy-1* as the protein may occur as part of a complex in its native from. If additional subunits are needed for a fully functional *Bm*-PHY-1 enzyme, these would not be provided by transformed *C. elegans* as the most likely partner in this nematode, *Ce*-PDI-2, does not form a complex *in vitro* with *Bm*-PHY-1.

6.3.4. *Bm*-PHY-2 and *Bm*-PDI, potential P4H subunits

In order to understand more fully the nature and composition of P4H complexes in *B. malayi*, additional *Bm-phy* and *Bm-pdi* homologues would have to be identified and characterised. Through this process the full range of possible P4H complexes in this nematode could be examined. Two sequences were identified from *B. malayi* that are proposed to represent a second *Bm-phy* isoform and a *Bm-pdi*. However the size of the *Bm-pdi* EST clone and the form of the *Bm-phy-2* BAC did not permit further sequence analysis. Using primers designed to the existing sequence data full-length sequence from both these genes could potentially be identified by PCR screening the BAC library used to identify the promoter of *Bm-phy-1*. BACs positive for the sequence of interest could be obtained from the FGP labs. Should the sequences identified represent *Bm-pdi* and *Bm-phy-2* as proposed, cDNA from these genes could be cloned and sequenced to

confirm/redefine predicted intron/exon boundaries. To ascertain if these proteins are P4H subunits and determine the types of complexes present in *B. malayi* they could then be expressed in an insect cell system and associations between them and *Bm*-PHY-1 determined by co-expression studies. Other potential approaches to study these genes are addressed in Section 6.3.5.

6.3.5. *C. elegans* as an expression system for parasite proteins

The ability of these potential *B. malayi* P4H genes to functionally compensate for loss of P4H activity in *C. elegans* could also be determined in a manner similar to that already described. Although *Bm-phy-1* was not able to rescue the *C. elegans dpy-18* mutant, rescue was achieved by expression of human α subunits in this mutant. To be enzymatically active, the human subunits must first presumably combine with a partner β subunit from *C. elegans,* almost certainly *Ce*-PDI-2. This hybrid complex must then be capable of hydroxylating *C. elegans* collagens that are markedly different from most human collagens. Rescue of a P4H mutant in *C. elegans* with *B. malayi* genes is therefore conceivable due to the much closer evolutionary relationship between these two nematodes. Although no data currently exists for rescue of *C. elegans* mutants with *B. malayi* genes this process has been successful using genes from the parasitic nematode *H. contortus*. The *ben-1* gene from *C. elegans* encodes β-tubulin with loss of this gene leading to resistance to the benzimidazole (BZ) drugs. Parasite β-tubulin was shown to function in *C. elegans* by injection into the BZ resistant *Ce-ben-1* mutant strain with *Hc-ben-1* cloned from a BZ sensitive *H. contortus* strain (Kwa *et al.*, 1995). Expression of the parasite gene gave renewed sensitivity to BZ in *C. elegans*. In contrast, transformation of *Ce-ben-1* mutants with a mutated *H. contortus* β-tubulin gene gave no acquired sensitivity. This process therefore provides a method for identification of residues important in drug action/resistance. Heterologous rescue of *C. elegans* by a *H. contortus* gene has also been described for cathepsin L proteases. The embryonic lethality of a *C. elegans cpl-1* deletion strain was rescued by transformation with a homologue from its parasitic relative (Britton and Murray, 2002). This demonstrates the conservation of essential developmental function between these two genes in different species and identifies them as true orthologues. This suggests that the function of this enzyme in *H. contortus* is essential for development of the parasite and as such is a target for chemical control. The degree of amino acid identity found

between *H. contortus* and *C. elegans* for β-tubulin is over 90% (Kwa *et al.*, 1995) with the identity of mature cathepsin L proteases between these two species of 87% (Britton and Murray, 2002). These identities are significantly higher than the 57% identity between the PHY proteins from *C. elegans* and the more distantly related *B. malayi*. However as described partial rescue of a *C. elegans* P4H mutant by human α subunits (43-44 % identity) has been achieved. Heterologous rescue with a *Brugia* P4H protein or complex in *C. elegans* would enable the effects of site-directed mutagenesis and chemical inhibition of the parasite enzyme(s) to be performed *in vivo*. This would allow a much more detailed analysis of *Brugia* P4H function than is currently achievable. It may be potentially possible to examine the function of an entire *B. malayi* complex in the absence of any *C. elegans* cuticle collagen modifying activity. Mutants for *Ce-phy-1* and *Ce-phy-2* have already been identified. Although no mutant currently exists for *Ce-pdi-2* one could be generated using the process discussed earlier (see Chapter 4). Alternatively if no mutant was identified, "pre-rescuing" with the parasite *pdi* gene followed by RNAi to specifically remove *Ce-pdi-2* function could be employed to assess the ability of *Bm-pdi* to replace native enzyme function. This approach has been successfully used in *C. elegans* for rescue with parasite genes (Britton and Murray, 2002).

6.3.6. Expression of *Bm-phy-1*

C. elegans has been used as a heterologous transformation system to determine if promoter elements from parasitic nematodes can direct expression of a reporter gene in this organism. This approach was used here to determine that the *Bm-phy-1* promoter was able to direct tissue-specific expression in *C. elegans*. No published data exists to date for heterologous expression in *C. elegans* of reporter genes driven by *B. malayi* promoters. The promoter region for the *Bm-phy-1* gene was identified and used to drive reporter gene expression in the hypodermal cells of *C. elegans*. Although the native localisation of this protein in the parasite is not known, its P4H activity and the expression patterns of homologous genes from *C. elegans* would predict a hypodermal localisation for expression of this gene. Expression in this cuticle synthesising tissue is in keeping with a proposed role for this gene in the modification of *B. malayi* cuticular collagens. The restricted tissue specific expression by sequences from *B. malayi* in *C. elegans* in the tissue appropriate for the proposed function of this gene demonstrates

that evolutionary distantly related nematodes (Blaxter and Bird, 1997), appear to show conservation of the elements required for directing spatial expression. This method provides a means to predict the native expression pattern of a gene via inference from expression of reporter genes in a heterologous system, and can be used when no other approach to determine spatial expression is available. Examination of other gene regulatory sequences from *B. malayi*, and comparison with function and native expression, will reveal if conservation of expression is found more generally between these two nematodes.

Expression of reporter genes in *C. elegans* has been described for promoters from a number of genes from related parasitic nematodes. The gluteraldehyde-3-phosphate-dehydrogenase promoter from the potato cyst nematode, *Globodera rostochiensis,* directs expression of GFP in *C. elegans* mainly to the body wall muscle cells (Qin *et al.*, 1998). Immunological studies of this protein in the parasite indicate that this is the native localisation site. Upstream sequences from two *H. contortus* genes, the pepsinogen *pep-1* and the cysteine protease gene *ac-2*, directs expression of reporter genes to the gut cells in *C. elegans* (Britton *et al.*, 1999). This gut-specific expression correlates with the localisation of the native proteins in the parasite and with the proposed functions of these genes. The promoter region of a hypodermally expressed cuticular collagen gene from *Ostertagia circumcincta* also directs expression of a reporter gene to the hypodermis in *C. elegans* (Britton *et al.*, 1999). The gluthione S-transferase gene promoter from the filarial nematode *O. volvulus* was found to direct expression in *C. elegans* in a manner consistent with its native expression pattern in the parasite (Krause *et al.*, 2001). One observation from some of these studies is that although tissue specificity of expression is conserved, temporal expression may not be as well conserved, with timing of expression of the reporter genes in *C. elegans* found not to be identical to that described for the native proteins (Britton *et al.*, 1999). This was also found in examination of *Bm-phy-1* expression where although tissue specificity is conserved, expression in larval stages was comparatively weak. This suggests that although promoter elements appear to be recognised by a conserved basal transcription apparatus, the elements that direct spatial expression may be more conserved than those responsible for developmental timing of gene expression.

The *Bm-phy-1* reporter gene expression seen in late larval and adult stages is consistent with the available RT-PCR data. Strong expression was found at all L3, L4 and adult time points examined by RT-PCR from staged cDNA samples. It was not possible however to quantify this data to determine fluctuation of expression levels within a moulting period. Synchrony of the cultures used for cDNA preparation is a factor to be taken into account when interpreting any results from this procedure. This however is not sufficient to explain the differences found between experiments using the same template. A degree of synchrony in cultures can be inferred from other non-quantified RT-PCRs on this same panel of cDNAs. The genes *Bm-alt-1* and *Bm-alt-2*, when examined in this way have clear periods of expression followed by periods of no detectable expression (Gregory *et al.*, 2000). If cultures were extremely asynchronous this pattern of expression would be unlikely. It may be interesting to repeat these experiments with the 60S ribosomal protein encoding gene as an internal control as has successfully been applied in *B. pahangi* (Lewis *et al.*, 1999).

In prolyl 4-hydroxylases we have an enzyme whose function is well understood and whose range of forms are now becoming clear. The essential nature of their activity for development and body morphology has been demonstrated in *C. elegans* both by genetics/RNAi analysis and by chemical inhibition studies. The extrapolation of the knowledge gained in *C. elegans* and other species along with the emerging use of *C. elegans* as a surrogate system for expression of foreign proteins provides powerful approaches for understanding parasite genes and assessing their potential as targets for chemical control.

Chapter 7
General discussion

7.1. Introduction

In this study the role of the P4H enzyme class in formation of nematode ECMs was examined. As is found for all animals, nematode ECMs are formed predominantly from collagen. Nematodes have two forms of ECM known as the cuticle and the basement membranes (Kramer, 1997). The nematode cuticle is an exoskeleton that almost entirely encases the animal and the function of P4Hs in the formation of this structure was the focus of this work. P4H function was studied in the free-living nematode *C. elegans* and in the filarial parasitic nematode *B. malayi*. *C. elegans* has a number of advantages for the study of many biological processes, including ECM formation, due to the range of genetic, transgenic, molecular and biochemical techniques which can be used in this animal. Completion of the entire genome sequence (Consortium, 1998) also enabled identification of entire gene families such as those predicted to encode subunits of P4H enzymes. Due to these advantages *C. elegans* is also being used increasingly as a system to study the less experimentally amenable parasitic nematode species. This can be approached by studying the homologue of a parasitic gene of interest in *C. elegans* and using the knowledge of gene function to infer function of the parasite gene, as well as directly examining parasite genes by using the free-living species as a heterologous expression system (Hashmi *et al.*, 2001).

The nematode cuticle is composed principally of collagen (Cox *et al.*, 1981a) and as well as its many other functions, such as movement and protection from the environment, is critical for nematode body morphology. The cuticle is formed and shed five times during the development of *C. elegans* and all nematodes. The first cuticle is formed in the late stages of embryogenesis and this forms the cuticle of the first larval stage. Cuticles are then shed and re-formed during each of the subsequent four larval moults (Singh and Soulston, 1978). Much of what is known about the structure and formation of the cuticle comes from the study of *C. elegans*. The entire genome of this organism encodes 154 cuticular collagen genes (Johnstone, 2000). These collagens are relatively small compared to vertebrate collagens, and homologues of a similar nature have been identified in a number of parasitic nematode species (Bisoffi and Betschart, 1996; Johnstone *et al.*, 1996; Kingston *et al.*, 1989; Scott *et al.*, 1995; Shamansky *et al.*,

1989). A number of collagen mutants have been described in *C. elegans*, with the abnormal body morphology exhibited by these animals demonstrating the requirement of these molecules for cuticle formation and development (Johnstone, 1994). All collagens are defined by the presence of regions of Gly-X-Y repeats, where Gly is the amino acid glycine and X and Y can be any residue, but are most commonly proline and 4-hydroxyproline respectively. Mature collagen is formed by trimerisation of monomers, which ultimately form a triple helical structure, and a number of enzymatic modification steps. These modifications occur both intra- and extra-cellularly and pre- and post-trimer formation. Steps involved in the formation of mature collagen molecules are; hydroxylation of proline and lysine residues, disulphide bond formation, formation of a triple helix which requires the action of a peptidyl prolyl *cis-trans* isomerase, cleavage of non-triple helical regions of trimers and cross-linking reactions.

In vitro studies of vertebrate collagens indicated that a central processing step was the hydroxylation of Y position proline residues to 4-hydroxyproline. This residue is modified co- and post-translationally with collagens lacking these residues found to be thermally unstable at physiological temperatures (Berg and Prockop, 1973). The enzymes responsible for this modification are the collagen P4Hs which have been extensively characterised in vertebrates (Kivirikko and Myllyharju, 1998). The identified collagen P4Hs are usually multi-subunit enzyme complexes with a catalytically active α subunit, with the β subunit being formed by PDI (Pihlajaniemi *et al.*, 1987). Two characterised α subunits have been described in mouse and humans (Annunen *et al.*, 1997; Helaakoski *et al.*, 1995; Helaakoski *et al.*, 1989) and these form complexes with a common PDI β subunit. The vertebrate enzymes complexes take the form of $\alpha_2\beta_2$ tetramers, with tetramers containing different α subunits thought not to form (Annunen *et al.*, 1997; Vuori *et al.*, 1992a). However, relatively little was understood about the role of these enzymes in formation of nematode ECMs.

7.2. The *C. elegans* cuticle collagen modifying P4H complexes

In this study, two conserved *C. elegans* α subunit-encoding genes, *Ce-phy-1* and *-2*, are described, along with a single associated PDI subunit, *Ce-pdi-2*. Three divergent α subunit-encoding genes, *Ce-phy-3, -4* and *-5*, were also examined to determine any role in formation of nematode ECMs. Expression of *Ce-phy-1, Ce-phy-2* and *Ce-pdi-2*

was demonstrated in the cuticle synthesising ectodermal cells (known as the hypodermis) at times of maximal collagen synthesis. Single disruption of *Ce-phy-1* resulted in nematodes with a morphologically mutant body shape, typical of that found for many collagen gene mutants. Single disruption of *Ce-phy-2* gave no observable phenotype. The mutant *dpy-18* was identified as being the result of lesions in the *Ce-phy-1* gene and represented the first description of a P4H mutant in any organism. The body shape defect combined with the temporal and tissue specific expression pattern of this gene implicated it strongly in the modification of the cuticle collagens of *C. elegans*. This proposed role was strengthened when the combined function of the *Ce-phy* genes was determined to be essential for normal embryonic development. Disruption of *Ce-phy-2* function by RNAi in the *Ce-phy-1* mutant strain, *dpy-18*, resulted in failure of affected progeny to complete embryonic development, embryos developed normally until the cuticle was required to maintain their elongated form. The cuticle of these double mutant/disrupted worms was then too weak to maintain this morphology, and the elongated form of the embryo was lost. An exact copy of this phenotype was exhibited in *C. elegans* embryos with single disruption of *Ce-pdi-2*. Using bacterially mediated RNAi, viable post-embryonic phenotypes of both the double *Ce-phy* and single *Ce-pdi-2* disruption were generated. The phenotypes of these viable mutants confirmed a cuticle collagen-specific role for these genes as they displayed severe body shape malformations. These RNAi feeding mutants, along with the *dpy-18* mutant and a *Ce-phy-2* deletion strain, generated by the *C. elegans* gene knockout consortium, were examined in detail for the effect of enzyme disruption on two collagens. The effect on the production and localisation of *C. elegans* collagens COL-19 and DPY-7 were examined using a GFP fusion and monoclonal antibody respectively. The expression and localisation of collagen was found to be severely affected in *dpy-18* mutants, and in double *Ce-phy* and single *Ce-pdi-2* disrupted animals, although not in the morphologically wild type *Ce-phy-2* deletion strain. The identical temporal and spatial expression patterns, and RNAi and collagen disruption phenotypes, of *Ce-pdi-2* to that of the *Ce-phy* genes suggested that this was the sole β subunit for both *Ce-phy* gene products. The *let-44* locus was investigated as a potential *Ce-pdi-2* mutant due to its position nearby *Ce-pdi-2* on the genetic map and a similar mutant embryonic development profile to that of *Ce-pdi-2* RNAi embryos. Examination of *Ce-pdi-2* genomic and cDNA sequences from *let-44* mutants did not reveal any molecular lesion

in the *Ce-pdi-2* coding sequence. Likewise, repair of the mutant phenotype of *let-44* with wild type *Ce-pdi-2* sequences was not observed. These results indicated that *let-44* was not a *Ce-pdi-2* mutant.

One P4H complex had been described previously at the biochemical level. *Ce*-PHY-1 was observed to form an active dimer when co-expressed with *Ce*-PDI-2 in an insect cell system (Veijola *et al.*, 1996). The genetic evidence above was consistent with the finding that *Ce*-PDI-2 functioned as a P4H β subunit with *Ce*-PHY-1 *in vivo* and also suggested that it was the common subunit for both *Ce*-PHY-1 and -2. To establish what forms these enzymes existed in *Ce*-PHY-2 was expressed in an insect cell expression system (in collaboration with J. Myllyharju). This suggested that *C. elegans*, as well as containing a *Ce*-PHY-1/ *Ce*-PDI-2 dimer, also possessed a unique *Ce*-PHY-1/*Ce*-PHY-2/(*Ce*-PDI-2)$_2$ tetramer. No P4H complex of this form had been previously described. Investigation of *in vivo* extracts from *C. elegans* strains confirmed this finding and demonstrated that the wild type appearance of the *Ce-phy-2* deletion was a result of formation of a larger amount of *Ce*-PHY-1/*Ce*-PDI-2 dimer in the absence of the mixed *Ce*-PHY subunit tetramer. However the viability of *dpy-18* mutants which lacked both the described forms of the complex suggested the existence of an additional complex which contained *Ce*-PHY-2 and *Ce*-PDI-2. However this third complex was at first thought not to be simply formed by only these subunits as they were reported to form no association when co-expressed in an insect cell expression system. The divergent *Ce-phy, -3, -4* and -5, genes were therefore examined to determine if they could be involved in formation of a third P4H complex or alternatively in another aspect of ECM development in this organism. From this study these genes were determined not to have a role in either the modification of the major ECMs or the formation of a third P4H complex in this nematode. No effect was found on collagen markers from the single removal/disruption of each divergent *Ce-phy*. Additionally, no phenotypes were observable from either the single or combined removal/disruption of these genes. Importantly, no further effect was found upon gene function knockdown in a *dpy-18* mutant genetic background. If any of these gene products were required in the hypothesised third P4H complex, transcript interference of this gene in a *Ce-phy-1* null background would have had a similar effect to removal of both *Ce-phy-1* and *-2* or single disruption of *Ce-pdi-2*. Since no embryonic lethality was observed, and no

increase in severity of the mutant morphology noted, these subunits were deemed to not be required for formation of the additional P4H complex. In light of these findings the co-expression data from *Ce*-PHY-2 with *Ce*-PDI-2 was re-examined, revealing P4H activity in these extracts. This data and analysis of *in vivo* extracts indicated that this third complex was a *Ce*-PHY-2/*Ce*-PDI-2 dimer which is probably only formed in the absence of *Ce*-PHY-1. Although the activity of this complex is low it must be sufficient to hydroxylate the cuticle collagens in *dpy-18* mutant strains to enable these nematodes to exist, albeit in a morphologically aberrant form. These findings therefore confirmed that *Ce*-PHY-1, *Ce*-PHY-2 and *Ce*-PDI-2 are the sole subunits responsible for modification of the *C. elegans* cuticle collagens and associate in unique, possibly nematode-specific, combinations. The identification of the essential nature of these genes highlighted the P4H class of enzymes as potential targets for design of anti-nematode compounds. The identification of a TPR protein-protein interaction motif (Lamb *et al.*, 1995) within a region of *Ce*-PHY-1 which had previously been described as being critical for complex formation (Myllyharju *et al.*, 2002) suggests a possible way in which subunits could associate. *Ce*-PHY-2 does not have a TPR motif and does not form complexes alone under wild type conditions. Examination of the homologues of *Ce*-PHY-1 and -2 in the nematode *C. briggsae* showed that in this case *Cb*-PHY-2 does contain a TPR motif but *Cb*-PHY-1 does not, suggesting that in this organism *Cb*-PHY-1 may require *Cb*-PHY-2 to form its natural complexes.

7.3. A *phy* gene homologue from *B. malayi*

The parasitic nematode *Brugia malayi* is one of the causative agents of lymphatic filariasis in humans. Approximately a third of the 120 million people infected are severely incapacitated by the infection which may include damage to their renal and lymphatic systems. (Michael and Bundy, 1997; Ottesen *et al.*, 1997; WHO, 2000). The possibility of using P4H inhibition to control these nematodes was also supported by the ability if commercially available inhibitors to disrupt *B. malayi* adult cuticles (Merriweather *et al.*, 2001). This finding together with the essential function of *C. elegans* P4Hs in cuticle collagen modification highlights this enzyme class as a potential drug target for control of parasitic nematode infections. Studies in *B. malayi* are facilitated by the Filarial Genome Project (FGP), which was initiated in response to the World Health Organisations call for a genomics approach to a the causative agents of lymphatic filariasis. The FGP have an ongoing Expressed Sequence Tag (EST) and

genomic sequencing projects as well as having constructed and made available filarial cDNA and genomic libraries. Additionally, *B. malayi* can undergo its entire life cycle in the laboratory, thereby making it possible to collect fresh material for analysis. The characterisation of a viable *Ce-phy-1* knockout also provided the potential for functional studies of other P4Hs by transformation into *C. elegans*. A P4H *phy* gene from *B. malayi* was therefore cloned and characterised and *C. elegans* used as an expression system to examine more directly some aspects of gene function and control. The *B. malayi* PHY homologue, *Bm*-PHY-1, is unusually a soluble and active P4H in the absence of a PDI subunit. Most catalytic subunits of P4H enzymes examined require a PDI subunit to maintain them in a soluble and active form (John *et al.*, 1993). *Bm*-PHY-1 was also found not to form any complexes with P4H subunits from other organisms. This is another unusual property of this protein not found for subunits from other organisms. Expression of the *Brugia* gene was examined using *C. elegans* as a heterologous expression system where it demonstrated expression in the cuticle synthesising tissue, suggesting this homologue would also function to modify cuticle collagens. Interestingly, despite this proposed function and the independently active form of the enzyme it was not able to replace function in a *C. elegans phy-1* mutant. A similar analysis with both α subunits from humans demonstrated that these could both partially functionally replace *Ce-phy-1*. The possibility exists that Bm-PHY-1 requires other subunits for full activity/function *in vivo* and sequences from two other possible P4H subunit-encoding genes were identified in this study.

7.4. Type IV collagens

The basement membranes form the second major ECM in nematodes and contain type IV collagens. In *C. elegans*, type IV collagen containing basement membranes underlie the hypodermis and surround the pharynx, intestine, gonad and some body wall muscles (Kramer, 1997). Mutation in both *C. elegans* type IV collagen encoding genes *emb-9* and *let-2* results in embryonic arrest in the developing nematode (Guo *et al.*, 1991; Gupta *et al.*, 1997; Sibley *et al.*, 1994; Sibley *et al.*, 1993) with many alleles displaying temperature sensitivity. None of the *C. elegans* P4H genes examined in this study appeared to be involved in the modification of basement membrane type IV collagens. *C. elegans phy-1* and *phy-2* were specific for the modification of the cuticle collagens and no evidence was found for involvement of the divergent *Ce-phy* genes in this

process. Type IV collagens in vertebrates have been demonstrated to contain 4-hydroxproline and hydroxylysine in Y positions of Gly-X-Y repeats (Kivirikko *et al.*, 1992) and must therefore be modified by the hydroxylating enzymes. This raised the question of how *C. elegans* type IV collagens are modified. Examination of the single lysyl hydroxylase enzyme *C. elegans, let-268*, demonstrates that this is specific for the modification of the basement membrane collagens (Norman and Moerman, 2000). This is in agreement with the finding that cuticle collagens of most developmental stages of *C. elegans* do not contain hydroxylysine (Cox *et al.*, 1981b). The protein is an ER localised homodimer with expression of transgenes being detected in the body wall muscle cells and glial like cells (Norman and Moerman, 2000). This expression pattern is similar to that of the type IV collagens but is not seen in all the tissues which express type IV collagens (Graham *et al.*, 1997). A putative null allele of this gene shows embryonic lethality at the two-fold stage of development with the type IV collagens, EMB-9 and LET-2, being intracellularly retained (Norman and Moerman, 2000). Less severe alleles allow some secretion of collagens to the basement membranes and these animals were found to progress further in development. These observations show that correct modification of lysine is required for collagen secretion. As hydroxylysine is not generally found in the *C. elegans* cuticle collagens and no type IV specific function was found for any *Ce-phy* gene suggests that nematodes unlike vertebrates do not require 4-hydroxproline in their type IV collagen, but hydroxylysine is specifically required to stabilise these collagens.

7.5. Human diseases resulting from improperly modified collagen

Some human diseases have been described which are due to improperly modified collagen (Myllyharju and Kivirikko, 2001). Although no genetic disease has been found for P4H, mutation in the human lysyl hydroxylase 1 gene manifest as Ehlers-Danlos syndrome type VI (Yeowell *et al.*, 2000; Yeowell and Walker, 2000). EDS is a heterogeneous group of inherited connective tissue disorders characterised by altered mechanical properties of the skin, joints, blood vessels and ligaments. Lysyl hydroxylase, like its *C. elegans* homologue *let-268* (Norman and Moerman, 2000), catalyses the hydroxylation of lysine residues in procollagen. In vertebrates hydroxylysine is used in the formation of intermolecular cross-links that gives collagen its strength. Many mutations in this gene, but not the genes encoding two other

isoforms, have been identified, with loss or reduction of hydroxylysine resulting in collagens that contain fewer cross-links and therefore show reduced stability.

Vertebrate fibrillar collagens are processed to remove their N- and C-termini. Both these regions must be cleaved for the proteins to assemble into fibrils under physiological conditions. Failure to remove the N-terminal propeptide, either by mutation of the collagen sequence at the processing site (Byers *et al.*, 1997; Vasan *et al.*, 1991) or by mutation of the proteinase enzyme itself (Colige *et al.*, 1999) causes the type VII form of Ehlers-Danlos syndrome (EDS). These mutations are analogous to the mutations in the N-terminal recognition site of *C. elegans* cuticle collagens (Kramer and Johnson, 1993; Levy *et al.*, 1993; Yang and Kramer, 1994; Yang and Kramer, 1999) and *bli-4* processing enzyme (Peters *et al.*, 1991; Thacker *et al.*, 1995). The autosomal dominant EDS type VIIA and type VIIB result from mutations in the type I collagen encoding genes, COL1A1 and COL1A2 respectively, which abolish the N-terminal processing site (Byers *et al.*, 1997; Vasan *et al.*, 1991). The third variant of the condition, the recessively inherited type C, comes from mutations in the N-terminal proteinase gene, procollagen I N-proteinase (Colige *et al.*, 1999) which processes the aminopropeptide of type I and II procollagens (Colige *et al.*, 1995)

7.6. C-terminal collagen processing in vertebrates and nematodes

Enzymes which modify the C-termini of vertebrate fibrillar collagens have also been identified. This proteinase activity, which cleaves the C-terminal propeptides of procollagens I, II and III, was shown to be the enzyme previously known as bone morphogenic protein (Kessler *et al.*, 1996; Li *et al.*, 1996). The metalloprotease BMP-1 and a longer protein, named mammalian tolloid (mTLD), are encoded by alternatively spliced transcripts of the same gene (Takahara *et al.*, 1994). BMP-1/mTLD-null mice are perinatal lethal and show evidence of abnormal collagen processing, but have properly formed bone structures. This suggested possible redundancy and led to the isolation of distinct but related C-terminal protease, mammalian Tolloid-like-1 (mTLL-1) (Takahara *et al.*, 1996) which also specifically process procollagen at the correct C-terminal site (Scott *et al.*, 1999). Interestingly, for the fibril collagen α1(V) the N-terminus is processed by BMP-1 and the C-terminus is cleaved at a different site (Imamura *et al.*, 1998) perhaps demonstrating the potential complexity of collagen

processing. A gene with homology to BMP-1 has been identified in *C. elegans* by characterisation of the mutant *dpy-31* (J. Novelli and J. Hodgkin, personal communication). Processing of the C-termini of *C. elegans* collagens is has yet been confirmed, however a mutant strain has been recently identified that may be defective in processing these regions. *C. elegans* nematodes with the *dpy-31(e2770)* allele are severely dumpy in appearance with the phenotype displaying temperature sensitivity. At the permissive lower temperature half the progeny are lethal and the remainder dumpy. All worms raised at the restrictive temperature are embryonic lethal or arrest at early larval stages. The altered body morphology, embryonic lethality and temperature sensitivity of this locus suggests a role in formation of the nematode cuticle. A cosmid containing a gene that encodes a predicted zinc metalloprotease related to BMP-1 rescues the mutant phenotype of *dpy-31*. This gene is therefore a candidate for involvement in the processing of the C-propeptides of some or all of the *C. elegans* collagens.

7.7. Non-collagen proteins containing hydroxylated residues

Although the vast majority of hydroxylated amino acids occur in collagens, a number of other proteins have been described which contain 4-hydroxyproline and hydroxylysine. Proteins of this nature include acetylcholinesterase, which contains 4-hydroxproline and hydroxylysine, and elastin which only contains hydroxyproline (Kivirikko *et al.*, 1992). Although the proteins are non-collagens these residues are found in collagen-like domains of the polypeptides. Unlike collagens though these proteins do not contain 3-hydroxyproline, which has been found only in collagens, although no gene for a prolyl 3-hydroxylase has been cloned from any organism.

In mammalian cells the transcription factor hypoxia inducible factor (HIF) is post-translationally regulated by prolyl hydroxylation (Ivan *et al.*, 2001; Jaakkola *et al.*, 2001; Yu *et al.*, 2001). HIF is a central regulator of oxygen homeostasis with the majority of the transcriptional responses to hypoxia regulated by it. HIF operates as a heterodimer of HIF-α and HIF-β subunits, where HIF-α is the regulated component. Under normal oxygen conditions the von Hippel-Lindau tumour suppresser protein (VHL) binds directly to HIF-α and targets it for degradation. In hypoxic conditions degradation of HIF is inhibited and HIF-α can bind to HIF-β and activate transcription.

Targeting of HIF-α for VHL-mediated degradation is regulated by prolyl hydroxylation (Ivan *et al.*, 2001; Jaakkola *et al.*, 2001). In hypoxic conditions HIF-α is not hydroxylated and is thus not bound by VHL and is able to activate target genes. Mutants in the *C. elegans* HIF-α encoding gene exhibit no defects at normal oxygen levels but are unable to survive under hypoxic conditions (Jiang *et al.*, 2001) normally tolerated by wild types (van Voorhies and Ward, 2000). The *C. elegans* gene *egl-9* (*egg-laying* defective) was identified as the HIF-α hydroxylating enzyme which acts as the oxygen sensor and represents a novel functional group of 2-oxoglutarate-dependent oxygenases (Epstein *et al.*, 2001). Three human and one *Drosophila* HIF P4Hs have been identified and shown to hydroxylate the HIF-α subunit depending on oxygen concentration (Bruick and McKnight, 2001; Epstein *et al.*, 2001). These proteins are distinct from animal collagen P4Hs both in terms of their hydroxylation amino acid recognition sequence (Ivan *et al.*, 2001; Jaakkola *et al.*, 2001; Yu *et al.*, 2001), their apparent monomeric form (Bruick and McKnight, 2001; Epstein *et al.*, 2001) and presumed cytoplasmic subcellular localisation. Additionally mammalian collagen P4Hs are unable to hydroxylate the HIF-α polypeptide (Jaakkola *et al.*, 2001). Description of this method of regulation of proteolytic targeting via prolyl hydroxylation gives the possibility of other proteins being similarly regulated.

Another interesting protein that contains 4-hydroxproline is PrP^{Sc} or prion protein scrapie isoform. Neurodegenerative diseases such as scrapie in sheep, bovine spongiform encephalopathies (BSE) and Creutzfeldt-Jakob disease (CJD) in humans, are associated with the conversion of a normal, cell surface protein, PrP^{c} (cellular prion protein) to an aggregated conformational isomer PrP^{Sc}. The N-terminal region of this protein contains an X-Pro-Gly consensus sequence for prolyl 4-hydroxylase modification. Recombinant protein is modified at this site to 4-hydroxproline as is protein extracted from scrapie infected mice (Gill *et al.*, 2000). This suggests a possible role for post-translational modification at proline residues by a prolyl 4-hydroxylase enzyme in the control of PrP^{c} function and in development of the disease.

7.8. Future prospects for direct analysis of parasitic nematode genes

Functional analysis of nematode parasite genes has been slowed by the lack of transformation, genetic methods and cell lines available for these organisms. Hence an

approach involving the investigation of homologous genes of interest from free-living species has often been taken, complemented by the emerging use of *C. elegans* as a heterologous host for analysis of both functional and regulation of parasite genes (Hashmi *et al.*, 2001). Recent advances suggest future prospects for the direct analysis of parasitic nematode genes such as the P4H subunits of *B. malayi*.

7.8.1. RNAi

RNAi is powerful and relatively straightforward technique used in *C. elegans* which can be performed by RNA injection (Fire *et al.*, 1998), soaking (Tabara *et al.*, 1998), bacterial feeding (Timmons *et al.*, 2001) or DNA transformation (Tavernarakis *et al.*, 2000). Recently the technique of RNAi soaking has also been adapted to the parasitic nematode *Nippostrongylus brasiliensis* suggesting a possible broader use of this approach in related species (Hussein *et al.*, 2002). *In vitro* transcribed dsRNA corresponding to acetylcholinesterase (AChE) B isoform was used to soak adult parasites for a period of 6 days in culture. Production of the secreted enzymes was monitored by sampling the culture medium daily and assaying for AChE activity. Suppression of enzyme activity of between 80-90% was induced for a period of 4-6 days. Cross-interference was found as secretion of three AChE isoforms (A, B and C) were suppressed by interference with AChE B sequences even though the sequences used were only 67-68% identical between isoforms B and A. This finding suggest a possible downward revision of the nucleotide identity required for cross-interference to occur in. In this study the ability of the parasite to re-establish infection in the host was affected by the conditions required for RNAi treatment, prohibiting an assessment of the effect in the host after RNAi knockdown. Despite this the success of this technique in *N. brasiliensis* suggests this approach may be useful in directly defining roles of other parasite genes. In particular this approach would be applicable to *B. malayi* as this can be cultured in the laboratory. Thus the effect of RNAi inhibition of *Bm-phy-1* and other potential P4H could be assessed. This analysis could be performed even before the full gene sequences of potential subunits had been identified. Thus the large amounts of data produced by the EST projects for this and other organisms (Degrave *et al.*, 2001; Williams *et al.*, 2000) could be used to produce functional data.

7.8.2. DNA transformation

Transformation of parasitic nematode species has until recently not been possible. However the direct analysis of parasite gene function *in vivo* may now be possible due to techniques developed in *C. elegans*. Transformation of *C. elegans* although usually performed by microinjection can also be carried out by particle bombardment, termed biolistics (Wilm *et al.*, 1999). DNA co-precipitated with gold particles can be shot at worms by use of a helium beam to generate semi-stable transformed lines. Biolistic DNA transformation of *Ascaris* embryos demonstrated various promoters including the SL RNA promoter could drive expression of the luciferase reporter gene (Davis *et al.*, 1999). Biolistic transformation of *Ascaris* with luciferase RNA also resulted in enzyme activity (Davis *et al.*, 1999). This technique has also been successful in the filarial parasite *Litomosoides sigmodontis* (Jackstadt *et al.*, 1999), a rat parasite often used as a model for human filarial parasites, and *B. malayi* (Higazi *et al.*, 2002). In *L. sigmodontis* adult worms were biolistically transformed, allowed to recover and reinfected into hosts. *lacZ* expression driven by an actin promoter was observed in these individuals and in a small number of microfilariae *sigmodontis* (Jackstadt *et al.*, 1999). For *B. malayi* embryos, adult female parasites and infective larvae were transformed by bombardment and luciferase activity detected (Higazi *et al.*, 2002). When similar techniques were used with a GFP reporter to view live worms expression was detected in the hypodermis. However, this was the tissue in which the particles had become embeded and when nematodes were microinjected with the same construct expression was again found in the tissue where DNA had been delivered (Higazi *et al.*, 2002). Although semi-stable lines have not been produced in *B. malayi* and the techniques have not been adapted to reveal tissue specific promoter activity, the ability to transiently transfect this organism should make these experiments possible. This along with emerging parasite RNAi techniques should enable a more direct assessment of parasite gene function to complement studies utilising *C. elegans* and could be applied both to the *Bm-phy-1* and the other putative P4H subunit encoding genes in this organism.

7.9. Future prospects for genome wide analysis of cuticular ECM formation in *C. elegans*

In this study a *C. elegans* strain was used that was especially informative for examining ECM formation. A collagen GFP fusion construct was made that, when transformed in to the nematode, produces protein that localises to specific regions of the *C. elegans*

cuticle (Thein *et al.*, 2003). The cuticle collagen COL-19 when fused to GFP localised specifically to the adult annular furrows and alae structures of the cuticle. Integrated lines generated were used in this study to analyse the effect of *Ce-phy* gene disruption or mutation on the localisation of the GFP tagged collagen. For the *Ce-phy-1* gene *Ce-phy-1/Ce-phy*-2 double, clear disruption was evident both at the level of overall morphology and of the individual collagen. However this strain could also be used to uncover subtle phenotypes for genes where RNAi gave no discernible phenotype. These could then be assessed by use of the COL-19::GFP strain to examine cuticle collagen-specific effects of gene disruption. A project is also underway in *C. elegans* to construct a complete genome-wide RNAi bacterial feeding library. As whole chromosomes of this library are now freely available it may be possible to perform a high-throughput screen using this collagen marker strain. This process, as well as identifying collagen phenotypes for genes where a role in cuticle formation could be hypothesised, such as the *Ce-phy* genes, may also identify novel genes either for which no function has be found proposed or could define new roles in ECM formation for genes already described. Many of the *C. elegans* genes have no known homologues in any other species databases (Blaxter, 1998). As the cuticle plays such an important role in the nematode life cycle some of these genes may represent novel components or enzymes not previously identified. This may either provide information for possible additional drug targets exploiting processes specific to nematode development not present in vertebrates, or if homologous genes are found in vertebrates, will identify novel processes common to formation of all ECMs. This process of discovery of new mutants, with possible novel functions, along with the further characterisation of existing genes and mutants identified as being critical for cuticle formation in *C. elegans* will aid the discovery of anti-nematode drugs and understanding of the biosynthesis of ECMs in vertebrates and other animals.

Appendix 1

Compiled *Bm-phy-1* sequences

Complete genomic sequence of *Bm-phy-1* including promoter and 3' untranslated regions. Promoter, intronic and untranslated regions are in lowercase lettering. Translation of the uppercase exonic sequences is given above the DNA sequence with the signal peptide region shown in italics. The signal peptide cleavage site, between residues A17 and D18, is indicated by a (/). Within the promoter region the RNA polymerase II binding site, or TATAA box, is underlined. The arrow indicates the site of *trans*-splicing of the SL1 *trans*-splice leader (sequence shown) with splicing occurring after the final G and before the final A of the *trans*-splicing signal sequence (underlined). Within the 3' UTR the underlined sequence represents the predicted polyadenylation signal of the *Bm-phy-1* transcript.

```
   1    gaatgagaca attgcacaag tattagagcg acttgcatgt aggaaaaggt
  51    ttggttcttt taactattct agctacctac tgccgttaat ttattggcca
 101    gaaattgctt taatgtagaa aactacaaat aatgaaggag catgtactat
 151    aagtaagtga tatgcaattg atatgattca acatatcgtt tgatttatca
 201    atgaatttgt cccaaaaatg acatagaatt atactaaaga ttgttgtgca
 251    acgaatacaa aatggctttg ttactcattg tccccgtcac ttgagtacaa
 301    atttcaatgg tgtgttaaat aaaccttagt tacactcaat tatattaaaa
 351    ttcagcaaac atttgaagac agttcctgtg acaataacaa gacgattaat
 401    aagaaagaca agtaaaaaga tttatacgat taatgggaag tagcaaacat
 451    atgatgactt tatgaaaatt gaacaaaaac gagattytat ggtaagaggc
 501    gaaaactttt tttacaccga ttcaaagact tattttactt gttattgttt
 551    aatgttgtga tggtggctag aagaaagcga aattctattt tagtagcacc
 601    agttttctca ctttgaattg taarwccaat tgaaacttca attttaataa
 651    attttcatta aataatctaa aatctttttc ccactagaga tcaaatagag
 701    cttaaagttc tacatttttc ctgtactcgg tctttctgct caataagcct
 751    caaattgttc attaatttat atttaagaca caaatggatt taatagtaat
 801    attaacaaat ttatcttgtt cgggatcaaa tctgcaggaa ttcattaact
 851    tacggtgagt aaagtgatgt ttctagtcct tttttaataa attcattcca
 901    tatttaagt tatatagtga gaaagggaaa taaatatatt cccgaagaaa
 951    aattaatgtg atgtggtagt gctaaaaatt tagatcacac tacgtacagg
1001    acgacgaaga cttagccccc tttcctcctt cattgttttt ctatttctgt
1051    atttgatttt ctctgtcact tttttaaatc tctccttctc tcttggcawc
1101    tatctattaa aacgacaaat tataatatct caatttcaaa ttttctagga
1151    ttgtaaatgt ttgttgaaga ttagaagaaa tgaaaaaqga ctgaagcttg
1201    gaatatattt ccctctctct attccaatcg tcaagtgcct ggatagttcc
1251    tctccttttc aacagtaaaa ttgtttcact gttgtaggat ttcaaaagag
1301    gttgaaaccc attacggagg aagaatcaga ttatatgaaa acgcgaaaga
1351    ttaaaatgta cgagagtata gtcgccatgt ggttgacagt cttccggaaa
1401    catggaagta ataagcaacc tttcccgtaa tgagtcgtcc ggcagtgcac
1451    ttagaagaag gctactactg ccattgtggt tagcaaatgt catcatccat
1501    cccattacta cagctgactg cagtaactct tcgcttttta cattgtacat
1551    gagagaaaaa gaagaggaag gaaggataga ttgccttctc gcctgccttt
1601    cctttcagca ttttcggctc atatccatca ctttttctct ctctcccttc
1651    ctttcaaatt ccttctttct ttagatctgt tcgtattgat tttatttgac
1701    tgtccgcaat tttgttgcta tttgtctgtc caataataaa tttcccctcc
1751    ttgtttcatc tataatttat agtgacctgt taatgctcaa attaatgtca
1801    tcacttctct cccctaatct tgcagagttt aggattgaaa ttaggacaac
1851    tacgacacta agggcgactt ttgcgccctt agtgtcgtag ttgtccgatt
1901    ttaagttact tccatatttt aatttttgct atagaaaagc agtaaacata
1951    gatgatatgg tatcatttaa atccttcatt taatgagagt tttttaaaaa
2001    acctaaattt tattaccgta gcacaacaaa ggtagtggat tgaagcatct
2051    gaattataat ggacttgaat aaatcaaatg ataacgtttt acagtaattg
```

```
2101    cgctgtttct ttgttacaac gtattcgttg taaaagaaca tatcttcatt

            SL1  ggtttaattacccaagtttgag
                                                    M   I   A   T
2151    atttgttatt ttgatacttc ttatttttcca gagccagtgA TGATAGCTAC

          V   V   F   I   L   A   L   R   S   T   T   H   A / D   L   F
2201    CGTGGTGTTC ATTTTGGCAC TGCGGAGTAC TACACATGCT GATTTATTTA
        T   S   I   A   E   M   E   L   L   L   E   A   D   K   R   I   P
2251    CATCAATCGC AGAAATGGAG TTACTTCTGG AAGCTGATAA GAGAATTCCT
          D   L   L   D   M   Y   I   E   R   F   Q   Q   R   L   D   Q   I
2301    GATTTGCTCG ATATGTACAT CGAAAGGTTC CAACAACGTT TGGATCAAAT
            R   Q
2351    TAGACAgttt gttttatcca tttatgttga aagtttaaaa tttttggatg
2401    cctttaaaat attgcgccag ttggattata atgttcaagt actaaaaatg
2451    gttatcaatt actggaagaa tatgatttca atttcgattc tttcatctta
                                                L   S   V   G   K   K
2501    agtgtcagat cccactaatt ttatttttca gGCTTTCAGT AGGCAAAAAA
          Q   L   G   N   R   S   L   G   N   D   I   R   L   L   S   N   P
2551    CAGCTTGGCA ATCGCAGTTT AGGAAATGAT ATACGATTGC TAAGCAATCC
            V   S   A   Y   L   L   I   K   R   L   I   E   E   W   D   D
2601    AGTCAGTGCT TATCTGCTTA TCAAAAGACT GATTGAGGAA TGGGATGATA
        I   K   R   L   A   G   S   D   I   G   E   E   L   L   K   E   I
2651    TAAAACGGTT GGCAGGAAGT GATATTGGCG AAGAGTTATT GAAGGAAATT
          S   E   L   R   A   M   N   Y   V   K   N   P   T   T
2701    TCAGAATTAC GAGCAATGAA TTATGTGAAA AACCCTACAA CGgtactttt
2751    tcatgagtcc ctccctaact tatcctattc tatgatctta gcagattact
2801    tctcatttct ataatttgca tgaaaagtaa ggttcaaatt atcataatta
2851    tgaatttttt taagatttac ggacacttat aggatggatt ttcagtaatt
2901    tttcatataa aattgggcaa atgacctctt gcttcaataa aatacatttt
              E   D   L   V   G   A   A   I   A   L   L   R   L   Q   D
2951    tgagGAAGAT TTGGTTGGAG CTGCAATAGC CTTACTTCGG CTGCAAGATA
        T   Y   R   L   N   V   K   E   I   A   D   G   K   I   L   N   A
3001    CATATCGGCT AAATGTTAAG GAGATAGCTG ATGGGAAAAT TCTGAATGCT
          S   G   V   Q   P   F   T
3051    TCTGGTGTTC AACCGTTTAC AGgttcgaaa gcaaatgcat tataaaatga
3101    cgtttccagt cctttatttt tgaacacctt aatgaatgtt ttattttagt
3151    acatacattt ctaactgtct gctaaattag ggtctaggag tttgatcaaa
3201    tgtcaagccc tgcttttaaa gttcagcata aaatggatga acaaaaagtt
3251    tgcaatttatt tgaaactgta tttgctgcta ttttttttctt ctgtttaccg
3301    aagaagtttc aacttttcgga tataaaaaag aaagatacag gcatatatat
3351    attttatcat aattagtatt tttttagaac aatttctaaa aattctggta
                                        A   R   D   C   F   D   I   G
3401    tggaatatga ttttatattt tttcagCCAG GGATTGCTTT GATATTGGTC
        R   A   A   Y   N   V   N   D   Y   Y   H   T   L   I   W   M   E
3451    GAGCAGCATA CAATGTGAAT GACTACTACC ATACCTTAAT ATCGATGGAA
          E   A   Q   E   R   L   R   D   E   A   P   H   E   T   V   Q   L
3501    GAGGCACAGG AGCGGTTGCG GGATGAAGCG CCGCACGAAA CTGTGCAATT
            X   E   I   L   E   Y   L   A   F   A   L   F   K   Q   G   N
3551    AAAGGAAATT TTGGAATATT TAGCTTTTGC GCTCTTCAAA CAAGGCAATC
        L   K   R   A   L   L   L   T   E   Q   L   H   T   I
3601    TGAAACGCGC CCTTCTGCTC ACTGAACAGC TGCATACAAT TGgttgtttt
3651    ttactacttg tagtgttact tgaatttttaa tttgaaactg tgtaagttca
3701    cccaaatcta tctaagattt ttgctacgtt gctaccatta aagattatta
3751    tggtatattt tagaaacgtt caatatataa gaagctaaat atagcgaatt
3801    caaatttaga tcaggttact tttcgtaacc gaggttaaat aatgtccttc
                    D   P   N   H   P   R   A   K   N   N   I   K   W   Y
3851    ttcatcacag ATCCAAATCA TCCACGCCCA AAAAACAATA TTAAGTGGTA
          E   D   L   L   A   E   E   G   L   K   P   I   D   Y   R   R
3901    TGAAGACCTA TTGGCTGAGG AAGGTCTGAA GCCAATTGAT TATCGGCGAA
        N   I   P   P   V   T   N   P   R   P   T   T   G   L   E   T   A
3951    ATATTCCGCC GGTAACTAAT CCTCGACCTA CGACTGGTTT GGAAACAGCG
          E   H   D   I   Y   E   A   L   C   R   N   E   I   P   V
4001    GAGCATGATA TCTACGAAGC CTTATGTAGA AACGAAATAC CCGTGgtgtt
4051    ttttctgtat cttactcaca tttgaacaaa actgctttgg ttaaactgac
4101    aaaatttttaa aataagaaaa tctgttgtca aaagctaact aacttccgtt
4151    aggccaatta agaggtaaat taaaaacata attaaactgt ttgcataatt
4201    tcgtttcggg aagtcgatga ttagccgtta ctcgaatttt gctttacgta
4251    actaccttca taaaaaacaa aaaattattc aggaggcaaa aattaatatg
4301    agagacaaac ggtttatcta aggcgttttt ttatggatat gagaagtttc
4351    agtaagcaat tgtaagcctt ttagtcttgt ttcccctcta tttctgggca
4401    aaaaaataaa atctgaaatt tctaaatttt gttattttga aaacttttttg
4451    aactcatctg cgcttcttta ttcactcgta attttttcact ttcatttttt
                                        S   I   K   V   T   S   K   L   Y
4501    cagttgaaac gtcaaaattt tcagAGCATT AAGGTAACGT CAAAACTTTA
          C   Y   Y   K   M   D   R   P   F   L   R   L   A   P   F   K
4551    CTGCTACTAC AAAATGGATC GTCCATTTTT GCGTTTGGCT CCTTTCAAAG
```

```
        V  E  I  L    R  F  N    P  L  A  V    L  F  R    D  V  I
4601    TTGAAATTTT AAGATTTAAT CCACTCGCCG TATTGTTCCG TGATGTTATT
         T  D  E  E    V  T  M    I  Q  M    L  A  T  P    R
4651    ACAGACGAAG AAGTTACAAT GATCCAAATG TTGGCGACAC CGAGAgtatc
4701    cgttatttta ataagcatct taatattgtc cttcctatgt tttctttatt
4751    atattatccc ttataattaa tttctaaatg atattctttt gcaaattaat
4801    tgtaaggatt ttcgaaagcg aaattcaatg tcggtttttt tttttaattt
                                                            L
4851    ttttcttctc tatcccctta ctttttatta tttttattgt tttcatagCT
          R  R  A    T  V  Q  N    S  I  T    G  E  L    E  T  A
4901    GCGAAGAGCT ACTGTGCAGA ACTCGATAAC TGGTGAACTT GAAACAGCAT
        S  Y  R  T    S  K  S
4951    CATACCGAAC AAGTAAAAGg ttggctgtaa acttgctcac attcatgatt
5001    tgctagctta agaagcattt tagctggatg cgaaagtttc cggaatatat
5051    gatggtcaat tgtagggaga catgagcact ttttttttatt tttaaaaaaa
5101    ttaatttttta gatttcgaga attgtgatat atcctatttc aaatagatgt
5151    aaaaaaaatt ttaattgctg atgttttgta cttgccatac ctgctttcta
5201    ctattttgga tgtttgaaag aaagtataag atttctgaat tttattattt
5251    cacaggattt catgaaatta cttcaatttc tttttttctt attcactggt
5301    aagaaataaa ttgcaaggaa aaaagcgaat gctttgttca tttacttata
          A  W  L    K  D  E    E  H  E  V    V  H  R    I  N  K
5351    gTGCATGGTT GAAAGATGAG GAACACGAAG TTGTACATCG GATCAATAAA
         R  I  D  L    M  T  N    L  E  Q    E  T  S  E    E  L  Q
5401    CGAATTGATC TGATGACTAA TTTGGAGCAA GAGACTTCAG AAGAATTGCA

5451    Ggtttgttgc ttttattcaa agacaaattt taaaattaaa ttgctgatgt
5501    ttcttacttg ctttacctgc tttctattac tagttaattt gtagtttaat
5551    taattccaga tttcacatat ataatttgta tttttactac aaatcactgt
5601    ttggaaattt tgtttcttcag taatagtttg cttttaaaaa agtctctaac
5651    tccctaattt cacagtatta ctttgaagat tcacactgtt tagttgtttt
5701    ggctggtatt actgtcatcg tagtgcgatg ctttttttat tcaatagttt
                                                      V  G  N  Y
5751    ctactatttt gtatttttca tgctaaagca aatattagGT TGGAAATTAT
         G  I  C  G    H  Y  D    P  H  F    D  F  A  R
5801    GGTATTGGTG GTCATTATGA TCCTCATTTT GACTTCGCAA GAgtaagttt
5851    cattattttc gattaattaa aatattttta tttaattttt taactctgat
5901    ttatttttcc attattgtac tgagtatttt attgtacata aatgatgtgg
                     R  E  E    V  N  A  F    Q  S  L    N  T  G
5951    cattcattca gAGAGAAGAA GTTAACGCTT TTCAGTCGCT CAACACCGGG
         N  R  L  A    T  L  L    F  Y
6001    AATCGTTTGG CAACATTGCT ATTTTATgtg agtattcaat tcattctgca
6051    taaaagaaat tagaaaagga gggtcgctaa ttttcgctaa tactcgctaa
6101    tttctgtgaa gcccaaacca ccaaaatctc ttgttgtaac tttgatgcat
6151    taaaattgtt tgaaaacata aaatttctaa acgtcatctt tttcaaattt
6201    tctctgctcc tagtttactt tlctctaacc gttaagttta ttcagtccat
6251    agccattgct gaaaaagttg ctttttaaatg aaaatttttt aaaacgaaaa
                  M  T  Q  P    E  S  G    G  A  T  V    F  T  E
6301    aatttacagA TGACACAGCC TGAATCAGGA GGTGCTACTG TATTTACTGA
         V  K  T    T  V  M  P    S  K
6351    GGTTAAAACA ACAGTTATGC CTTCTAAAgt tagtgatttc tacatgtttt
6401    gatcaatatt actatttaaa ctttgccatt ttgattggaa gtggtggtcg
6451    aattggaaaa ttgaagattg atatctgctt aaagtctttt ggatgaagtt
6501    tttcggatac taaattacga atttgaagtt tttctgactt ttagactaat
                                                      N  D  A  L
6551    aatcaagttt ttgaatataaa agaaattttc tttgcagAAC GATGCATTAT
        F  W  Y  N    L  L  R    S  G  E  G    D  L  R    T  R  H
6601    TTTGGTATAA TTTATTGCGC AGCGGTGAGG GTGACTTGAG AACACGTCAC
         A  A  C  P    V  L  T    G  T  K    W  V  S  N    K  W  I
6651    GCTGCGTGTC CTGTGTTAAC CGGTACGAAG TGGGTTTCAA ATAAATGGAT
          H  E  R    G  Q  E  F    R  R  P    C  G  L    S  R  S
6701    ACATGAACGT GGTCAAGAGT TTCGAAGGCC TTGTGGCTTA AGCCGATCAG
        V  E  E  Q    F  V  G    D  L  S  A    *
6751    TGCAAGAACA GTTTGTGGGC GATCTAAGTG CTTAAatact attgggaagt
6801    acactaaaaa gctaagaatt tttgatacat tatatttcaa agtaatataa
6851    ctgttactgc atgttgggaa agccaaaact aaatttataa tatgtccaga
6901    taattttgga ttaattccat atgataaaaa cttgtaggc
```

Appendix 2

Table of accession numbers

Gene	Chromo-some	Cosmid/YAC	Accession number[1]	Protein size[2]	Signal peptide[3]
Ce-phy-1	III	T28D6.1/ Y43D3B.10	Z81134	559	16
Ce-phy-2	IV	F35G2.4	Z69637	539	16
Ce-pdi-2	X	C07A12.4	U41542	493	16
[4]*Ce-phy-3*	V	T20B3.7	Z81593	318	23
[5] *Ce-phy-4*	X	C14E2.4	U46671	429	19
[5] *Ce-phy-5*	V	Y43F8B.4	AL032623	533	14
Bm-phy-1	N.D.	N.A.	AJ297845	541	17
Bm-phy-1 genomic	N.D.	N.A.	AJ421993	N.A.	N.A.
Bm-phy-1 promoter	N.D.	N.A.	AJ421994	N.A.	N.A.
Bm-phy-2 GSS	N.D.	N.A.	BH615947	79	-
Bm-pdi EST	N.D.	N.A.	AI784701	92	-

N.D.- not determined

N.A.- not applicable

[1] Accession numbers are for cosmids/YACS, except for gene *Bm phy-1*, where it corresponds to the cDNA sequence.

[2] Predicted protein sizes are given in amino acids and include the signal peptide regions.

[3] Signal peptides were predicted using Signal P and are confirmed experimentally for *Ce*-PHY-1 only (Veijola *et al.*, 1994).

[4] *Ce*-PHY-3 protein size given is that of the larger experimental identified form, the accession number given is for the Genbank prediction, see Riihimaa *et al.*, 2002 for the modified gene sequence.

[5] The Genebank predictions were used to give the protein sizes for the putative proteins encoded by *Ce-phy-4* and *Ce-phy-5* although these are unconfirmed.

Appendix 3

List of published papers containing work described in this thesis

Winter, A. D., and Page, A. P. (2000). Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode *Caenorhabditis elegans*. *Molecular and Cellular Biology* **20,** 4084-4093.

Riihimaa, P., Nissi, R., Page, A. P., Winter, A. D., Keskiaho, K., Kivirikko, K. I., and Myllyharju, J. (2002). Egg shell collagen formation in *Caenorhabditis elegans* involves a novel prolyl 4-hydroxylase expressed in spermatheca and embryos and possessing many unique properties. *Journal of Biological Chemistry* **277,** 18238–18243.

Myllyharju, J., Kukkola, L., Winter, A. D., and Page, A. P. (2002). The exoskeleton collagens in *Caenorhabditis elegans* are modified by prolyl 4-hydroxylase with unique combinations of subunits. *Journal of Biological Chemistry* **277,** 29187-29196.

Winter, A. D., Myllyharju, J., and Page, A. P. (2003). A hypodermally expressed prolyl 4-hydroxylase from the filarial nematode *Brugia malayi* is soluble and active in the absence of protein disulphide isomerase. *Journal of Biological Chemistry*, in press.

Thein, M. C., McCormack, G., Winter, A. D., Johnstone, I. L., Shoemaker, C. B., and Page, A. P. (2003). The*Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification, assembly and the analysis of organismal morphology. *Developmental Dynamics*, in press.

References

Abrams, E. W., and Andrew, D. J. (2002). Prolyl 4-hydroxylase α-related proteins in *Drosophila melanogaster*: tissue-specific embryonic expression of the 99F8-9 cluster. *Mechanisms of Development* **112,** 165-171.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic alignment search tool. *Journal of Molecular Biology* **215,** 403-410.

Annunen, P., Autio-Harmainen, E., and Kivirikko, K. I. (1998). The novel type II prolyl 4-hydroxylase is the main enzyme form in chondrocytes and capillary endothelial cells, whereas the type I enzyme predominates in most cells. *Journal of Biological Chemistry* **273,** 5989-5992.

Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1997). Cloning of the human prolyl 4-hydroxylase α subunit isoform α(II) and characterization of the type II enzyme tetramer. The α(I) and α(II) subunits do not form a mixed $\alpha(I)\alpha(II)\beta_2$ tetramer. *Journal of Biological Chemistry* **272,** 17342-17348.

Annunen, P., Koivunen, P., and Kivirikko, K. I. (1999). Cloning of the α subunit of prolyl 4-hydroxylase from *Drosophila* and expression and characterization of the corresponding enzyme tetramer with some unique properties. *Journal of Biological Chemistry* **274,** 6790-6796.

Asahina, M., Ishihara, T., Jindra, M., Kohara, Y., Katsura, I., and Hirose, S. (2000). The conserved nuclear receptor Ftz-F1 is required for embryogenesis, moulting and reproduction in *Caenorhabditis elegans*. *Genes to Cells* **5,** 711-723.

Bächinger, H. P. (1987). The influence of peptidyl-prolyl *cis-trans* isomerase on the *in vitro* folding of type III collagen. *Journal of Biological Chemistry* **262,** 17144-17148.

Baird, S. E., and Emmons, S. W. (1990). Properties of a class of genes required for ray morphogenesis in *Caenorhabditis elegans*. *Genetics* **126,** 335-344.

Berg, B. A., and Prockop, D. J. (1973). The thermal transition of a nonhydroxylated form of collagen. Evidence for the role of hydroxyproline in stabilizing the triple-helix of collagen. *Biochemical and Biophysical Research Communications* **52,** 115-120.

Bergmann, D. C., Crew, J. R., Kramer, J. M., and Wood, W. B. (1998). Cuticle chirality and body handedness in *Caenorhabditis elegans*. *Developmental Genetics* **23,** 164-174.

Bird, A. F., and Bird, J. (1991). *The structure of nematodes, Second edition* Academic Press, San Diego.

Bird, D. M., and Riddle, D. L. (1989). Molecular cloning and sequencing of *ama-1*, the gene encoding the largest subunit of *Caenorhabdits elegans* RNA polymerase II. *Molecular and Cellular Biology* **9,** 4119-4130.

Bisoffi, M., and Betschart, B. (1996). Identification and sequence comparison of a cuticular collagen of *Brugia pahangi*. *Parasitology* **113,** 145-155.

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

Blaxter, M., and Bird, D. (1997). Parasitic nematodes. In, *C. elegans II,* (Riddle, D. L., Blumenthal, T., Meyer, B. J., Priess, J. R., eds) Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY. pp. 851-878.

Bloom, L., and Horvitz, H. R. (1997). The *Caenorhabditis elegans* gene *unc-76* and its human homologs define a new gene family involved in axonal outgrowth and fasciculation. *Proceedings of the National Academy of Sciences of the United States of America* **94,** 3414-3419.

Blumenthal, T., Evans, D., Link, C. D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W. L., Duke, K., Kiraly, M., and Kim, S. K. (2002). A global analysis of *Caenorhabditis elegans* operons. *Nature* **417,** 851-854.

Blumenthal, T., and Steward, K. (1997). RNA processing and gene structure. In, *C. elegans II,* (Riddle, D. L., Blumenthal, T., Meyer, B. J., Priess, J. R., eds) Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, pp. 117-145.

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

Brinster, R. L., Allen, J. M., Behringer, R. R., Gelinas, R. E., and Palmiter, R. D. (1988). Introns increase transcriptional efficiency in transgenic mice. *Proceedings of the National Academy of Sciences of the United States of America* **85,** 836-840.

Britton, C., and Murray, L. (2002). A cathepsin L protease essential for *Caenorhabditis elegans* embryogenesis is functionally conserved in parasitic nematodes. *Molecular and Biochemical Parasitology* **122,** 21-33.

Britton, C., Redmond, D. L., Knox, D. P., McKerrow, J. H., and Barry, J. D. (1999). Identification of promoter elements of parasite nematode genes in transgenic

Caenorhabditis elegans. *Molecular and Biochemical Parasitology* **103,** 171-181.

Bruick, R. K., and McKnight, S. L. (2001). A conserved family of prolyl-4-hydroxylases that modify HIF. *Science* **294,** 1337-1340.

Buchman, A. R., and Berg, P. (1988). Comparison of intron-dependent and intron-independent gene expression. *Molecular and Cellular Biology* **8,** 4395-4405.

Bulleid, N. J. (1996). Novel approach to study the initial events in the folding and assembly of procollagen. *Seminars in Cell & Developmental Biology* **7,** 667-672.

Bürglin, T. R., Lobos, E., and Blaxter, M. L. (1998). *Caenorhabditis elegans* as a model for parasitic nematodes. *International Journal for Parasitology* **28,** 395-411.

Byers, P. H., Duvic, M., Atkinson, M., Robinow, M., Smith, L. T., Krane, S. M., Greally, M. T., Ludman, M., Matalon, R., Pauker, S., Quanbeck, D., and Schwarze, U. (1997). Ehlers-Danlos syndrome type VIIA and VIIB result from splice-junction mutations or genomic deletions that involve exon 6 in the *COL1A1* and *COL1A2* genes of type I collagen. *American Journal of Medical Genetics* **72,** 94-105.

Caulagi, V. R., and Rajan, T. V. (1995). The structural organization of an α2 (type IV) basement membrane collagen gene from the filarial nematode *Brugia malayi*. *Molecular and Biochemical Parasitology* **70,** 227-229.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., and Prasher, D. C. (1994). Green fluorescent protein as a marker for gene expression. *Science* **263,** 802-805.

Chandrashekar, R., Devarajan, E., and Mehta, K. (2002). *Dirofilaria immitis*: further characterization of the transglutaminase enzyme and its role in larval molting. *Parasitology Research* **88,** 185-191.

Chandrashekar, R., Tsuji, N., Morales, T., Ozols, V., and Mehta, K. (1998). An ERp60-like protein from the filarial parasite *Dirofilaria immitis* has both transglutaminase and protein disulfide isomerase activity. *Proceedings of the National Academy of Sciences of the United States of America* **95,** 531-536.

Chappell, L. H., and Wastling, J. M. (1992). Cyclosporin A: antiparasite drug, modulator of the host-parasite relationship and immunosuppressant. *Parasitology* **105,** S25-S40.

Colige, A., Beschin, A., Samyn, B., Goebels, Y., Vanbeeumen, J., Nusgens, B. V., and Lapière, C. M. (1995). Characterization and partial amino acid sequencing of a 107-kDa procollagen I N-proteinase purified by affinity chromatography on immobilized type XIV collagen. *Journal of Biological Chemistry* **270,** 16724-16730.

Colige, A., Sieron, A. L., Li, S. W., Schwarze, U., Petty, E., Wertelecki, W., Wilcox, W., Krakow, D., Cohn, D. H., Reardon, W., Byers, P. H., Lapière, C. M., Prockop, D. J., and Nusgens, B. V. (1999). Human Ehlers-Danlos syndrome type VIIC and bovine dermatosparaxis are caused by mutations in the procollagen I N-proteinase gene. *American Journal of Human Genetics* **65,** 308-317.

Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282,** 2012-2018.

Cox, G. N., Kusch, M., and Edgar, R. S. (1981a). Cuticle of *Caenorhabditis elegans*: its isolation and partial characterisation. *Journal of Cell Biology* **90,** 7-17.

Cox, G. N., Staprans, S., and Edgar, R. S. (1981b). The cuticle of *Caenorhabditis elegans* II. Stage-specific changes in ultrastructure and protein composition during postembryonic development. *Developmental Biology* **86,** 456-470.

Crameri, A., Whitehorn, E. A., Tate, E., and Stemmer, W. P. (1996). Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nature Biotechnology* **14,** 315-319.

Crossen, R., and Gruenwald, S. (1998). Baculovirus expression vector system instruction manual. PharMingen, San Diego, CA.

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

Degrave, W. M., Melville, S., Ivens, A., and Aslett, M. (2001). Parasite genome initiatives. *International Journal for Parasitology* **31,** 532-536.

Dornan, J., Page, A. P., Taylor, P., Wu, S. Y., Winter, A. D., Husi, H., and Walkinshaw, M. D. (1999). Biochemical and structural characterization of a divergent loop cyclophilin from *Caenorhabditis elegans*. *Journal of Biological Chemistry* **274,** 34877-34883.

Edens, W. A., Sharling, L., Cheng, G. J., Shapira, R., Kinkade, J. M., Lee, T., Edens, H. A., Tang, X. X., Sullards, C., Flaherty, D. B., Benian, G. M., and Lambeth, J. D. (2001). Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peroxidase with homology to the phagocyte oxidase subunit gp91 *phox*. *Journal of Cell Biology* **154,** 879-891.

Engel, J., and Prockop, D. J. (1991). The zipper-like folding of collagen triple helices and the effects of mutations that disrupt the zipper. *Annual Review of Biophysics and Biophysical Chemistry* **20,** 137- 152.

Epstein, A. C. R., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R., Mukherji, M., Metzen, E., Wilson, M. I., Dhanda, A., Tian, Y. M., Masson, N., Hamilton, D. L., Jaakkola, P., Barstead, R., Hodgkin, J., Maxwell, P. H., Pugh, C. W., Schofield, C. J., and Ratcliffe, P. J. (2001). *C. elegans* EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* **107,** 43-54.

Eriksson, M., Myllyharju, J., Tu, H., Hellman, M., and Kivirikko, K. I. (1999). Evidence for 4-hydroxyproline in viral proteins. Characterization of a viral prolyl 4-hydroxylase and its peptide substrates. *Journal of Biological Chemistry* **274,** 22131-22134.

Eschenlauer, S. C. P., and Page, A. P. (2003). The *Caenorhabditis elegans* ERp60 homolog PDI-3 has disulphide isomerase and transglutaminase-like cross-linking activity and is involved in the maintainance of body morphology. *Journal of Biological Chemistry*, in press.

Ferrari, D. M., and Söling, H. D. (1999). The protein disulphide-isomerase family: unravelling a string of folds. *Biochemical Journal* **339,** 1-10.

Fetterer, R. H., and Rhoads, M. L. (1990). Tyrosine-derived cross-linking amino acids in the sheath of *Haemonchus contortus* infective larvae. *Journal of Parasitology* **76,** 619-624.

Fetterer, R. H., Rhoads, M. L., and Urban, J. F. (1993). Synthesis of tyrosine-derived cross-links in *Ascaris suum* cuticular proteins. *Journal of Parasitology* **79,** 160-166.

Fire, A. (1992). Histochemical techniques for locating *Escherichia coli* β-galactosidase activity in transgenic organisms. *Genetic Analysis-Biomolecular Engineering* **9,** 151-158.

Fire, A., White Harrison, S., and Dixon, D. (1990). A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93,** 189-198.

Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806-811.

Fischer, G., Wittmannliebold, B., Lang, K., Kiefhaber, T., and Schmid, F. X. (1989). Cyclophilin and peptidyl-prolyl *cis-trans* isomerase are probably identical proteins. *Nature* **337,** 476-478.

Fodor, A., Riddle, D. L., Nelson, F. K., and Golden, J. W. (1983). Comparison of a new wild-type *Caenorhabditis briggsae* with laboratory strains of *C. briggsae* and *C. elegans*. *Nematologica* **29,** 203-217.

Franklin, T. J., Morris, W. P., Edwards, P. N., Large, M. S., and Stephenson, R. (2001). Inhibition of prolyl 4-hydroxylase *in vitro* and *in vivo* by members of a novel series of phenanthrolinones. *Biochemical Journal* **353,** 333-338.

Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000). Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408,** 325-330.

Freedman, R. B., Klappa, P., and Ruddock, L. W. (2002). Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO Reports*.

Friedman, L., Higgin, J. J., Moulder, G., Barstead, R., Raines, R. T., and Kimble, J. (2000). Prolyl 4-hydroxylase is required for viability and morphogenesis in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* **97,** 4736-4741.

Fujimoto, D., Horiuchi, K., and Hirama, M. (1981). Isotrityrosine, a new crosslinking amino acid isolated from *Ascaris* cuticle collagen. *Biochemical and Biophysical Research Communications* **99,** 637-643.

Gill, A. C., Ritchie, M. A., Hunt, L. G., Steane, S. E., Davies, K. G., Bocking, S. P., Rhie, A. G. O., Bennett, A. D., and Hope, J. (2000). Post-translational hydroxylation at the N-terminus of the prion protein reveals presence of PPII structure *in vivo*. *EMBO Journal* **19,** 5324-5331.

Gilleard, J. S., Barry, J. D., and Johnstone, I. L. (1997). *cis* regulatory requirements for hypodermal cell-specific expression of the *Caenorhabditis elegans* cuticle collagen gene *dpy-7*. *Molecular and Cellular Biology* **17,** 2301-2311.

Gissendanner, C. R., and Sluder, A. E. (2000). *nhr-25*, the *Caenorhabditis elegans* ortholog of *ftz-f1*, is required for epidermal and somatic gonad development. *Developmental Biology* **221,** 259-272.

Gönczy, P., Echeverri, C., Oegema, K., Coulson, A. R., Jones, S. J. M., Copley, R. R., Duperon, J., Oegema, J., Brehm, M., Cassin, E., Hannak, E., Kirkham, M., Pichler, S. C., Flohrs, K., Goessen, A., Leidel, S., Alleaume, A. M., Martin, C., Ozlu, N., Bork, P., and Hyman, A. A. (2000). Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408,** 331-336.

Göthel, S. F., and Marahiel, M. A. (1999). Peptidyl-prolyl *cis-trans* isomerases, a superfamily of ubiquitous folding catalysts. *Cellular and Molecular Life Sciences* **55,** 423-436.

Graham, P. L., Johnson, J. J., Wang, S. R., Sibley, M. H., Gupta, M. C., and Kramer, J. M. (1997). Type IV collagen is detectable in most, but not all, basement membranes of *Caenorhabditis elegans* and assembles on tissues that do not express it. *Journal of Cell Biology* **137,** 1171-1183.

Gregory, W. F., Atmadja, A. K., Allen, J. E., and Maizels, R. M. (2000). The abundant larval transcript-1 and-2 genes of *Brugia malayi* encode stage-specific candidate vaccine antigens for filariasis. *Infection and Immunity* **68,** 4174-4179.

Guo, S., and Kemphues, K. J. (1995). *par-1*, a gene required for establishing polarity in *C. elegans* embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell* **81,** 611-620.

Guo, X. D., Johnson, J. J., and Kramer, J. M. (1991). Embryonic lethality caused by mutations in basement membrane collagen of *C. elegans*. *Nature* **349,** 707-709.

Guo, X. D., and Kramer, J. M. (1989). The two *Caenorhabditis elegans* basement membrane (type IV) collagen genes are located on separate chromosomes. *Journal of Biological Chemistry* **264,** 17574-17582.

Gupta, M. C., Graham, P. L., and Kramer, J. M. (1997). Characterization of α1(IV) collagen mutations in *Caenorhabditis elegans* and the effects of α 1 and α 2(IV) mutations on type IV collagen distribution. *Journal of Cell Biology* **137,** 1185-1196.

Gupta, R., Jung, E., and Brunak, S. (2002). Prediction of N-glycosylation sites in human proteins. In preparation.

Harding, M. W., Galat, A., Uehling, D. E., and Schreiber, S. L. (1989). A receptor for the immunosuppressant FK506 is a *cis-trans* peptidyl-prolyl isomerase. *Nature* **341,** 758-760.

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

Hayano, T., Hirose, M., and Kikuchi, M. (1995). Protein disulfide isomerase mutant lacking its isomerase activity accelerates protein folding in the cell. *FEBS Letters* **377,** 505-511.

Heim, R., Cubitt, A. B., and Tsien, R. Y. (1995). Improved green fluorescence. *Nature* **373,** 663-664.

Helaakoski, T., Annunen, P., Vuori, K., Macneil, I. A., Pihlajaniemi, T., and Kivirikko, K. I. (1995). Cloning, baculovirus expression, and characterization of a second mouse prolyl 4-hydroxylase α subunit isoform: formation of an $\alpha_2\beta_2$ tetramer with the protein disulfide isomerase/β subunit. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 4427-4431.

Helaakoski, T., Veijola, J., Vuori, K., Rehn, M., Chow, L. T., Taillonmiller, P., Kivirikko, K. I., and Pihlajaniemi, T. (1994). Structure and expression of the human gene for the α subunit of prolyl 4-hydroxylase. The two alternatively spliced types of mRNA correspond to two homologous exons the sequences of which are expressed in a variety of tissues. *Journal of Biological Chemistry* **269,** 27847-27854.

Helaakoski, T., Vuori, K., Myllylä, R., Kivirikko, K. I., and Pihlajaniemi, T. (1989). Molecular cloning of the α subunit of human prolyl 4-hydroxylase: the complete cDNA-derived amino acid sequence and evidence for alternative splicing of RNA transcripts. *Proceedings of the National Academy of Sciences of the United States of America* **86,** 4392-4396.

Hieta, R., and Myllyharju, J. (2002). Cloning and characterization of a low molecular weight prolyl 4-hydroxylase from *Arabidopsis thaliana*. Effective hydroxylation of proline-rich, collagen-like and hypoxia -inducible transcription factor α-like peptides. *Journal of Biological Chemistry* **277,** 23965–23971.

Higazi, T. B., Merriweather, A., Shu, L., Davis, R., and Unnasch, T. R. (2002). *Brugia malayi*: transient transfection by microinjection and particle bombardment. *Experimental Parasitology* **100,** 95-102.

Hill, K. L., Harfe, B. D., Dobbins, C. A., and L'Hernault, S. W. (2000). *dpy-18* encodes an α-subunit of prolyl-4-hydroxylase in *Caenorhabditis elegans*. *Genetics* **155,** 1139-1148.

Hodgkin, J. A., and Brenner, S. (1977). Mutations causing transformation of sexual phenotype in the nematode *C. elegans*. *Genetics* **86,** 275-287.

Hong, X. Q., Ma, D., and Carlow, C. K. S. (1998a). Cloning, expression and characterization of a new filarial cyclophilin. *Molecular and Biochemical Parasitology* **91,** 353-358.

Hong, X. Q., Ma, D., Page, A. P., Kumar, S., and Carlow, C. K. S. (1998b). Highly conserved large molecular weight cyclophilin of filarial parasites. *Experimental Parasitology* **88,** 246-251.

Hussein, A. S., Kilchenin, K., and Selkirk, M. E. (2002). Suppression of secreted acetylcholinesterase expression in *Nippostrongylus brasiliensis* by RNA interference. *Molecular and Biochemical Parasitology* **122,** 91-94.

Imamura, Y., Steiglitz, B. M., and Greenspan, D. S. (1998). Bone morphogenetic protein-1 processes the NH2-terminal propeptide, and a furin-like proprotein convertase processes the COOH-terminal propeptide of pro-α1(V) collagen. *Journal of Biological Chemistry* **273,** 27511-27517.

Ivan, M., Kondo, K., Yang, H. F., Kim, W., Valiando, J., Ohh, M., Salic, A., Asara, J. M., Lane, W. S., and Kaelin, W. G. (2001). HIFα targeted for VHL-mediated destruction by proline hydroxylation: implications for O_2 sensing. *Science* **292,** 464-468.

Jaakkola, P., Mole, D. R., Tian, Y. M., Wilson, M. I., Gielbert, J., Gaskell, S. J., von Kriegsheim, A., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, C. W., and Ratcliffe, P. J. (2001). Targeting of HIFα to the von Hippel-Lindau ubiquitylation complex by O_2-regulated prolyl hydroxylation. *Science* **292,** 468-472.

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

Jiang, H. Q., Guo, R., and Powell-Coffman, J. A. (2001). The *Caenorhabditis elegans hif-1* gene encodes a bHLH-PAS protein that is required for adaptation to

hypoxia. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 7916-7921.

John, D. C. A., and Bulleid, N. J. (1994). Prolyl 4-hydroxylase: defective assembly of α-subunit mutants indicates that assembled α-subunits are intramolecularly disulfide bonded. *Biochemistry* **33,** 14018-14025.

John, D. C. A., and Bulleid, N. J. (1996). Intracellular dissociation and reassembly of prolyl 4-hydroxylase: the α-subunits associate with the immunoglobulin-heavy-chain binding protein (BiP) allowing reassembly with the β-subunit. *Biochemical Journal* **317,** 659-665.

John, D. C. A., Grant, M. E., and Bulleid, N. J. (1993). Cell-free synthesis and assembly of prolyl 4-hydroxylase: the role of the β-subunit (PDI) in preventing misfolding and aggregation of the α-subunit. *EMBO Journal* **12,** 1587-1595.

Johnstone, I. L. (1994). The cuticle of the nematode *Caenorhabditis elegans*: a complex collagen structure. *Bioessays* **16,** 171-178.

Johnstone, I. L. (2000). Cuticle collagen genes: expression in *Caenorhabditis elegans*. *Trends in Genetics* **16,** 21-27.

Johnstone, I. L., and Barry, J. D. (1996). Temporal reiteration of a precise gene expression pattern during nematode development. *EMBO Journal* **15,** 3633-3639.

Johnstone, I. L., Shafi, Y., and Barry, J. D. (1992). Molecular analysis of mutations in the *Caenorhabditis elegans* collagen gene *dpy-7*. *EMBO Journal* **11,** 3857-3863.

Johnstone, I. L., Shafi, Y., Majeed, A., and Barry, J. D. (1996). Cuticular collagen genes from the parasitic nematode *Ostertagia circumcincta*. *Molecular and Biochemical Parasitology* **80,** 103-112.

Kamath, R. S., Martinez-Campos, M., Zipperlen, P., Fraser, A. G., and Ahringer, J. (2000). Effectiveness of specific RNA-mediated interference through ingested double-stranded RNA in *Caenorhabditis elegans*. *Genome Biology* **2,** 1-10.

Kaska, D. D., Gunzler, V., Kivirikko, K. I., and Myllylä, R. (1987). Characterization of a low relative-molecular-mass prolyl 4-hydroxylase from the green alga *Chlamydomonas reinhardii*. *Biochemical Journal* **241,** 483-490.

Kaska, D. D., Myllylä, R., Gunzler, V., Gibor, A., and Kivirikko, K. I. (1988). Prolyl 4-hydroxylase from *Volvox carteri*. A low-Mr enzyme antigenically related to the α subunit of the vertebrate enzyme. *Biochemical Journal* **256,** 257-263.

Kent, W. J., and Zahler, A. M. (2000). The intronerator: exploring introns and alternative splicing in *Caenorhabditis elegans*. *Nucleic Acids Research* **28,** 91-93.

Kessler, E., Takahara, K., Biniaminov, L., Brusel, M., and Greenspan, D. S. (1996). Bone morphogenetic protein-1: the type I procollagen C-proteinase. *Science* **271,** 360-362.

Kingston, I. B., Wainwright, S. M., and Cooper, D. (1989). Comparison of collagen gene sequences in *Ascaris suum* and *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology* **37,** 137-146.

Kivirikko, K. I., and Myllyharju, J. (1998). Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biology* **16,** 357-368.

Kivirikko, K. I., and Myllylä, R. (1982). Posttranslational enzymes in the biosynthesis of collagen: intracellular enzymes. *Methods in Enzymology* **82,** 245-304.

Kivirikko, K. I., Myllylä, R., and Pihlajaniemi, T. (1992). Hydroxylation of proline and lysine residues in collagens and other animal and plant proteins. In *Post translational modification of proteins* (Harding, J.J. and Crabe, M. J. C., eds) CRC Press, Boca Raton, pp. 1-51.

Kivirikko, K. I., and Pihlajaniemi, T. (1998). Collagen hydroxylases and the protein disulfide isomerase subunit of prolyl 4-hydroxylases. *Advances in Enzymology and Related Areas of Molecular Biology* **72,** 325-400.

Ko, F. C. F., and Chow, K. L. (2002). A novel thioredoxin-like protein encoded by the *C. elegans dpy-11* gene is required for body and sensory organ morphogenesis. *Development* **129,** 1185-1194.

Koivunen, P., Helaakoski, T., Annunen, P., Veijola, P., Veijola, J., Raisanen, S., Pihlajaniemi, T., and Kivirikko, K. I. (1996). ERp60 does not substitute for protein disulphide isomerase as the β-subunit of prolyl 4-hydroxylase. *Biochemical Journal* **316,** 599-605.

Koivunen, P., Pirneskoski, A., Karvonen, P., Ljung, J., Helaakoski, T., Notbohm, H., and Kivirikko, K. I. (1999). The acidic C-terminal domain of protein disulfide isomerase is not critical for the enzyme subunit function or for the chaperone or disulfide isomerase activities of the polypeptide. *EMBO Journal* **18,** 65-74.

Koltai, H., Chejanovsky, N., Raccah, B., and Spiegel, Y. (1997). The first isolated collagen gene of the root-knot nematode *Meloidogyne javanica* is developmentally regulated. *Gene* **196,** 191-199.

Kostrouchova, M., Krause, M., Kostrouch, Z., and Rall, J. E. (1998). CHR3: a *Caenorhabditis elegans* orphan nuclear hormone receptor required for proper epidermal development and molting. *Development* **125,** 1617-1626.

Kostrouchova, M., Krause, M., Kostrouch, Z., and Rall, J. E. (2001). Nuclear hormone receptor CHR3 is a critical regulator of all four larval molts of the nematode *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 7360-7365.

Kramer, J. M. (1994). Structures and functions of collagens in *Caenorhabditis elegans*. *FASEB Journal* **8,** 329-336.

Kramer, J. M. (1997). Extracellular matrix. In, *C. elegans II*, (Riddle, D. L., Blumenthal, T., Meyer, B. J., Priess, J. R., eds) Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, pp. 471-500.

Kramer, J. M., Cox, G. N., and Hirsh, D. (1985). Expression of the *Caenorhabditis elegans* collagen genes *col-1* and *col-2* is developmentally regulated. *Journal of Biological Chemistry* **260,** 1945-1951.

Kramer, J. M., French, R. P., Park, E. C., and Johnson, J. J. (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,** 2081-2089.

Kramer, J. M., and Johnson, J. J. (1993). Analysis of mutations in the *sqt-1* and *rol-6* collagen genes of *Caenorhabditis elegans*. *Genetics* **135,** 1035-1045.

Kramer, J. M., Johnson, J. J., Edgar, R. S., Basch, C., and Roberts, S. (1988). The *sqt-1* gene of *C. elegans* encodes a collagen critical for organismal morphogenesis. *Cell* **55,** 555-565.

Krause, S., Sommer, A., Fischer, P., Brophy, P. M., Walter, R. D., and Liebau, E. (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.

Kusch, M., and Edgar, R. S. (1986). Genetic studies of unusual loci that affect body shape in the nematode *Caenorhabtidis elegans* and may code for cuticle structural proteins. *Genetics* **113,** 621-639.

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

Laboissière, M. C. A., Sturley, S. L., and Raines, R. T. (1995). The essential function of protein-disulfide isomerase is to unscramble nonnative disulfide bonds. *Journal of Biological Chemistry* **270,** 28006-28009.

Lam, G. T., Jiang, C. A., and Thummel, C. S. (1997). Coordination of larval and prepupal gene expression by the DHR3 orphan receptor during *Drosophila* metamorphosis. *Development* **124,** 1757-1769.

Lamb, J. R., Tugendreich, S., and Hieter, P. (1995). Tetratrico peptide repeat interactions: to TPR or not to TPR? *Trends in Biochemical Sciences* **20,** 257-259.

Lamberg, A., Helaakoski, T., Myllyharju, J., Peltonen, S., Notbohm, H., Pihlajaniemi, T., and Kivirikko, K. I. (1996a). Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. *Journal of Biological Chemistry* **271,** 11988- 11995.

Lamberg, A., Jauhiainen, M., Metso, J., Ehnholm, C., Shoulders, C., Scott, J., Pihlajaniemi, T., and Kivirikko, K. I. (1996b). The role of protein disulphide isomerase in the microsomal triacylglycerol transfer protein does not reside in its isomerase activity. *Biochemical Journal* **315,** 533-536.

Lamberg, A., Pihlajaniemi, T., and Kivirikko, K. I. (1995). Site-directed mutagenesis of the α-subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *Journal of Biological Chemistry* **270,** 9926-9931.

Lassandro, F., Sebastiano, M., Zei, F., and Bazzicalupo, P. (1994). The role of dityrosine formation in the cross-linking of *cut-2*, the product of a second cuticlin gene of *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology* **65,** 147-159.

Levitan, D., Doyle, T. G., Brousseau, D., Lee, M. K., Thinakaran, G., Slunt, H. H., Sisodia, S. S., and Greenwald, I. (1996). Assessment of normal and mutant human presenilin function in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 14940-14944.

Levy, A. D., Yang, J., and Kramer, J. M. (1993). Molecular and genetic analyses of the *Caenorhabditis elegans dpy-2* and *dpy-10* collagen genes: a variety of molecular alterations affect organismal morphology. *Molecular Biology of the Cell* **4,** 803-817.

Lewis, E., Hunter, S., Tetley, L., Nunes, C., Bazzicalupo, P., and Devaney, E. (1999). *cut-1*-like genes are present in the filarial nematodes, *Brugia pahangi* and *Brugia malayi*, and, as in other nematodes, code for components of the cuticle. *Molecular and Biochemical Parasitology* **101,** 173-183.

Li, S. W., Sieron, A. L., Fertala, A., Hojima, Y., Arnold, W. V., and Prockop, D. J. (1996). The C-proteinase that processes procollagens to fibrillar collagens is identical to the protein previously identified as bone morphogenic protein-1. *Proceedings of the National Academy of Sciences of the United States of America* **93,** 5127-5130.

Liu, L. X., Spoerke, J. M., Mulligan, E. L., Chen, J., Reardon, B., Westlund, B., Sun, L., Abel, K., Armstrong, B., Hardiman, G., King, J., McCague, L., Basson, M., Clover, R., and Johnson, C. D. (1999). High-throughput isolation of *Caenorhabditis elegans* deletion mutants. *Genome Research* **9,** 859-867.

Liu, Z. C., Kirch, S., and Ambros, V. (1995). The *Caenorhabditis elegans* heterochronic gene pathway controls stage-specific transcription of collagen genes. *Development* **121,** 2471-2478.

Luo, M. J., and Reed, R. (1999). Splicing is required for rapid and efficient mRNA export in metazoans. *Proceedings of the National Academy of Sciences of the United States of America* **96,** 14937-14942.

Lustigman, S., Brotman, B., Huima, T., Castelhano, A. L., Singh, R. N., Mehta, K., and Prince, A. M. (1995). Transglutaminase-catalyzed reaction is important for molting of *Onchocera volvulus* third-stage larvae. *Antimicrobial Agents and Chemotherapy* **39,** 1913-1919.

Lyles, M. M., and Gilbert, H. F. (1991). Catalysis of the oxidative folding of ribonuclease A by protein disulphide isomerase: dependance of the rate on the composition of the redox buffer. *Biochemistry* **30,** 613-619.

Ma, D., Hong, X. Q., Raghavan, N., Scott, A. L., McCarthy, J. S., Nutman, T. B., Williams, S. A., and Carlow, C. K. S. (1996). A cyclosporin A-sensitive small molecular weight cyclophilin of filarial parasites. *Molecular and Biochemical Parasitology* **79,** 235-241.

Maeda, I., Kohara, Y., Yamamoto, M., and Sugimoto, A. (2001). Large-scale analysis of gene function in *Caenorhabditis elegans* by high-throughput RNAi. *Current Biology* **11,** 171-176.

Matsui, M., Oshima, M., Oshima, H., Takaku, K., Maruyama, T., Yodoi, J., and Taketo, M. M. (1996). Early embryonic lethality caused by targeted disruption of the mouse thioredoxin gene. *Developmental Biology* **178,** 179-185.

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

McMahon, L., Muriel, J. M., Roberts, B., Quinn, M., and Johnstone, I. L. (2003). Two sets of interacting collagens form functionally distinct sub-structures within the *C. elegans* extracellular matrix. *Molecular Biology of the Cell*, in press.

Mehta, K., Rao, U. R., Vickery, A. C., and Birckbichler, P. J. (1990). Significance of transglutaminase-catalyzed reactions in growth and development of filarial parasite, *Brugia malayi*. *Biochemical and Biophysical Research Communications* **173,** 1051-1057.

Mehta, K., Rao, U. R., Vickery, A. C., and Fesus, L. (1992). Identification of a novel transglutaminase from the filarial parasite *Brugia malayi* and its role in growth and development. *Molecular and Biochemical Parasitology* **53,** 1-15.

Mello, C., and Fire, A. (1995). DNA transformation. In, *Methods in cell biology volume 48, Caenorhabditis elegans: modern biological analysis of an organism* (Epstein, H. F. and Shakes, D. C., eds), Academic Press, San Diego, pp. 451-482.

Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO Journal* **10,** 3959- 3970.

Merriweather, A., Guenzler, V., Brenner, M., and Unnasch, T. R. (2001). Characterization and expression of enzymatically active recombinant filarial prolyl 4-hydroxylase. *Molecular and Biochemical Parasitology* **116,** 185-197.

Meyer, B. J. (1997). Sex determination and X chromosome dosage compensation. In, *C. elegans II*, (Riddle, D. L., Blumenthal, T., Meyer, B. J., Priess, J. R., eds) Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY, pp. 209-240.

Michael, E., and Bundy, D. A. P. (1997). Global mapping of lymphatic filariasis. *Parasitology Today* **13,** 472-476.

Miyabayashi, T., Palfreyman, M. T., Sluder, A. E., Slack, F., and Sengupta, P. (1999). Expression and function of members of a divergent nuclear receptor family in *Caenorhabditis elegans*. *Developmental Biology* **215,** 314-331.

Myllyharju, J., and Kivirikko, K. I. (1997). Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *EMBO Journal* **16,** 1173-1180.

Myllyharju, J., and Kivirikko, K. I. (1999). Identification of a novel proline-rich peptide-binding domain in prolyl 4-hydroxylase. *EMBO Journal* **18,** 306-312.

Myllyharju, J., and Kivirikko, K. I. (2001). Collagens and collagen-related diseases. *Annals of Medicine* **33,** 7-21.

Myllyharju, J., Kukkola, L., Winter, A. D., and Page, A. P. (2002). The exoskeleton collagens in *Caenorhabditis elegans* are modified by prolyl 4-hydroxylase with unique combinations of subunits. *Journal of Biological Chemistry* **277,** 29187-29196.

Nagai, N., Hosokawa, M., Itohara, S., Adachi, E., Matsushita, T., Hokawa, N., and Nagata, K. (2000). Embryonic lethality of molecular chaperone Hsp47 knockout mice is associated with defects in collagen biosynthesis. *Journal of Cell Biology* **150,** 1499-1505.

Natsuka, S., Takubo, R., Seki, R., and Ikura, K. (2001). Molecular cloning and expression of *Caenorhabditis elegans* ERp57-homologue with transglutaminase activity. *Journal of Biochemistry* **130,** 731-735.

Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10,** 1-6.

Nokelainen, M., Nissi, R., Kukkola, L., Helaakoski, T., and Myllyharju, J. (2001). Characterization of the human and mouse genes for the α subunit of type II prolyl 4-hydroxylase. Identification of a previously unknown alternatively

spliced exon and its expression in various tissues. *European Journal of Biochemistry* **268,** 5300-5309.

Norman, K. R., and Moerman, D. G. (2000). The *let-268* locus of *Caenorhabditis elegans* encodes a procollagen lysyl hydroxylase that is essential for type IV collagen secretion. *Developmental Biology* **227,** 690-705.

Nyström, J., Shen, Z., Aili, M., Flemming, A. J., Leroi, A., and Tuck, S. (2002). Increased or decreased levels of *Caenorhabditis elegans lon-3*, a gene encoding a collagen, cause reciprocal changes in body length. *Genetics* **161,** 83-97.

Ottesen, E. A., Duke, B. O. L., Karam, M., and Behbehani, K. (1997). Strategies and tools for the control/elimination of lymphatic filariasis. *Bulletin of the World Health Organization* **75,** 491-503.

Page, A. P. (1997). Cyclophilin and protein disulphide isomerase genes are co-transcribed in a functionally related manner in *Caenorhabditis elegans*. *DNA and Cell Biology* **16,** 1335-1343.

Page, A. P. (1999). A highly conserved nematode protein folding operon in *Caenorhabditis elegans* and *Caenorhabditis briggsae*. *Gene* **230,** 267-275.

Page, A. P., Kumar, S., and Carlow, C. K. S. (1995a). Parasite cyclophilins and antiparasite activity of cyclosporine A. *Parasitology Today* **11,** 385-388.

Page, A. P., Landry, D., Wilson, G. G., and Carlow, C. K. S. (1995b). Molecular characterization of a cyclosporine A insensitive cyclophilin from the parasitic nematode *Brugia malayi*. *Biochemistry* **34,** 11545-11550.

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

Page, A. P., and Winter, A. D. (1998). A divergent multi-domain cyclophilin is highly conserved between parasitic and free-living nematode species and is important in larval muscle development. *Molecular and Biochemical Parasitology* **95,** 215-227.

Page, A. P., and Winter, A. D. (1999). Expression pattern and functional significance of a divergent nematode cyclophilin in *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology* **99,** 301-306.

Park, Y. S., and Kramer, J. M. (1994). The *C. elegans sqt-1* and *rol-6* collagen genes are coordinately expressed during development, but not at all stages that display mutant phenotypes. *Developmental Biology* **163,** 112-124.

Peters, K., McDowall, J., and Rose, A. M. (1991). Mutations in the *bli-4* (I) locus of *Caenorhabditis elegans* disrupt both adult cuticle and early larval development. *Genetics* **129,** 95-102.

Pettitt, J., and Kingston, I. B. (1991). The complete primary structure of a nematode α2(IV) collagen and the partial structural organization of its gene. *Journal of Biological Chemistry* **266,** 16149-16156.

Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M.-L., and Koivu, J. K., K. I. (1987). Molecular cloning of the β-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO Journal* **6,** 643-649.

Pirneskoski, A., Ruddock, L. W., Klappa, P., Freedman, R. B., Kivirikko, K. I., and Koivunen, P. (2001). Domains *b'* and *a'* of protein disulfide isomerase fulfill the minimum requirement for function as a subunit of prolyl 4- hydroxylase. The N-terminal domains *a* and *b* enhance this function and can be substituted in part by those of ERp57. *Journal of Biological Chemistry* **276,** 11287-11293.

Plasterk, R. H. A. (1995). Reverse genetics: from gene sequence to mutant worm. In, *Methods in cell biology volume 48, Caenorhabditis elegans: modern biological analysis of an organism* (Epstein, H. F. and Shakes, D. C., eds), Academic Press, San Diego, pp. 59-80.

Priess, J. R., and Hirsh, D. I. (1986). *Caenorhabditis elegans* morphogenesis: the role of the cytoskeleton in elongation of the embryo. *Developmental Biology* **117,** 156-173.

Prockop, D. J. (1998). What holds us together? Why do some of us fall apart? What can we do about it? *Matrix Biology* **16,** 519-528.

Prockop, D. J., and Kivirikko, K. I. (1995). Collagens: molecular biology, diseases, and potentials for therapy. *Annual Review of Biochemistry* **64,** 403-434.

Qin, L., Smant, G., Stokkermans, J., Bakker, J., Schots, A., and Helder, J. (1998). Cloning of a trans-spliced glyceraldehyde-3-phosphate-dehydrogenase gene from the potato cyst nematode *Globodera rostochiensis* and expression of its putative promoter region in *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology* **96,** 59-67.

Rao, U. R., Mehta, K., Subrahmanyam, D., and Vickery, A. C. (1991). *Brugia malayi* and *Acanthocheilonema viteae*: antifilarial activity of transglutaminase inhibitors *in vitro*. *Antimicrobial Agents and Chemotherapy* **35,** 2219-2224.

Redmond, D. L., Clucas, C., Johnstone, I. L., and Knox, D. P. (2001). Expression of *Haemonchus contortus* pepsinogen in *Caenorhabditis elegans*. *Molecular and Biochemical Parasitology* **112,** 125-131.

Riihimaa, P., Nissi, R., Page, A. P., Winter, A. D., Keskiaho, K., Kivirikko, K. I., and Myllyharju, J. (2002). Egg shell collagen formation in *Caenorhabditis elegans* involves a novel prolyl 4-hydroxylase expressed in spermatheca and embryos and possessing many unique properties. *Journal of Biological Chemistry* **277,** 18238–18243.

Ristoratore, F., Cermola, M., Nola, M., Bazzicalupo, P., and Favre, R. (1994). Ultrastructural immuno-localization of *cut-1* and *cut-2* antigenic sites in the cuticles of the nematode *Caenorhabditis elegans*. *Journal of Submicroscopic Cytology and Pathology* **26,** 437-443.

Rocheleau, C. E., Downs, W. D., Lin, R. L., Wittmann, C., Bei, Y. X., Cha, Y. H., Ali, M., Priess, J. R., and Mello, C. C. (1997). Wnt signaling and an APC-related gene specify endoderm in early *C. elegans* embryos. *Cell* **90,** 707-716.

Rogalski, T. M., Williams, B. D., Mullen, G. P., and Moerman, D. G. (1993). Products of the *unc-52* gene in *Caenorhabditis elegans* are homologous to the core protein of the mammalian basement membrane heparan sulphate proteoglycan. *Genes & Development* **7,** 1471-1484.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular cloning: a laboratory manual. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

Scherens, B., Dubois, E., and Messenguy, F. (1991). Determination of the sequence of the yeast *YCL313* gene localized on chromosome *III*. Homology with the protein disulfide isomerase (PDI gene-product) of other organisms. *Yeast* **7,** 185-193.

Schreiber, S. L., and Crabtree, G. R. (1992). The mechanism of action of cyclosporine A and FK506. *Immunology Today* **13,** 136-142.

Scott, A. L., Yenbutr, P., Eisinger, S. W., and Raghavan, N. (1995). Molecular cloning of the cuticular collagen gene *Bmcol-2* from *Brugia malayi*. *Molecular and Biochemical Parasitology* **70,** 221-225.

Scott, I. C., Blitz, I. L., Pappano, W. N., Imamura, Y., Clark, T. G., Steiglitz, B. M., Thomas, C. L., Maas, S. A., Takahara, K., Cho, K. W. Y., and Greenspan, D. S.

(1999). Mammalian BMP-1/tolloid-related metalloproteinases, including novel family member mammalian tolloid-like 2, have differential enzymatic activities and distributions of expression relevant to patterning and skeletogenesis. *Developmental Biology* **213,** 283-300.

Sebastiano, M., Lassandro, F., and Bazzicalupo, P. (1991). *cut-1* a *Caenorhabditis elegans* gene coding for a dauer-specific noncollagenous component of the cuticle. *Developmental Biology* **146,** 519-530.

Selkirk, M. E., Nielsen, L., Kelly, C., Partono, F., Sayers, G., and Maizels, R. M. (1989). Identification, synthesis and immunogenicity of cuticular collagens from the filarial nematodes *Brugia malayi* and *Brugia pahangi. Molecular and Biochemical Parasitology* **32,** 229-246.

Shamansky, M. L., Pratt, P., Boisvenue, R. J., and Cox, G. N. (1989). Cuticle collagen genes of *Haemonchus contortus* and *Caenorhabditis elegans* are highly conserved. *Molecular and Biochemical Parasitology* **37,** 73-86.

Shaw, L. M., and Olsen, B. R. (1991). FACIT collagens: diverse molecular bridges in extracellular matrices. *Trends in Biochemical Sciences* **16,** 191-194.

Sibley, M. H., Graham, P. L., Vonmende, N., and Kramer, J. M. (1994). Mutations in the α2(IV) basement membrane collagen gene of *Caenorhabditis elegans* produce phenotypes of differing severities. *EMBO Journal* **13,** 3278-3285.

Sibley, M. H., Johnson, J. J., Mello, C. C., and Kramer, J. M. (1993). Genetic identification, sequence, and alternative splicing of the *Caenorhabditis elegans* α2(IV) collagen gene. *Journal of Cell Biology* **123,** 255-264.

Siekierka, J. J., Hung, S. H. Y., Poe, M., Lin, C. S., and Sigal, N. H. (1989). A cytosolic binding-protein for the immunosuppressant FK506 has peptidyl-prolyl isomerase activity but is distinct from cyclophilin. *Nature* **341,** 755-757.

Simmer, F., Tijsterman, M., Parrish, S., Koushika, S. P., Nonet, M. L., Fire, A., Ahringer, J., and Plasterk, R. H. A. (2002). Loss of the putative RNA-directed RNA polymerase RRF-3 makes *C. elegans* hypersensitive to RNAi. *Current Biology* **12,** 1317-1319.

Singh, N., and Sulston, J. (1978). Some observations on molting in *C. elegans*. *Nematologica* **24,** 63-71.

Singh, R. N., Chandrashekar, R., and Mehta, K. (1995). Purification and partial characterization of a transglutaminase from dog filarial parasite, *Dirofilaria immitis*. *International Journal of Biochemistry & Cell Biology* **27,** 1285-1291.

Singh, R. N., and Mehta, K. (1994). Purification and characterization of a novel transglutaminase from filarial nematode *Brugia malayi*. *European Journal of Biochemistry* **225,** 625-634.

Sluder, A. E., Mathews, S. W., Hough, D., Yin, V. P., and Maina, C. V. (1999). The nuclear receptor superfamily has undergone extensive proliferation and diversification in nematodes. *Genome Research* **9,** 103-120.

Steinmann, B., Bruckner, P., and Supertifurga, A. (1991). Cyclosporin A slows collagen triple-helix formation *in vivo*: indirect evidence for a physiological role of peptidyl-prolyl *cis-trans* isomerase. *Journal of Biological Chemistry* **266,** 1299-1303.

Stern, M. J., Marengere, L. E. M., Daly, R. J., Lowenstein, E. J., Kokel, M., Batzer, A., Olivier, P., Pawson, T., and Schlessinger, J. (1993). The human *GRB2* and *Drosophila* Drk genes can functionally replace the *Caenorhabditis elegans* cell signaling gene *sem-5*. *Molecular Biology of the Cell* **4,** 1175-1188.

Sugimoto, A., Hozak, R. R., Nakashima, T., Nishimoto, T., and Rothman, J. H. (1995). *dad-1*, an endogenous programmed cell death suppressor in *Caenorhabditis elegans* and vertebrates. *EMBO Journal* **14,** 4434-4441.

Sulston, J. E., and Horvitz, H. R. (1977). Post-embryonic cell lineages of the nematode *Caenorhabditis elegans*. *Developmental Biology* **56,** 110-156.

Sulston, J. E., Schierenberg, E., White, J. G., and Thomson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Developmental Biology* **100,** 64-119.

Tabara, H., Grishok, A., and Mello , C. C. (1998). RNAi in *C. elegans*: soaking in the genome sequence. *Science* **282,** 430-431.

Takacs, A. M., Denker, J. A., Perrine, K. G., Maroney, P. A., and Nilsen, T. W. (1988). A 22-nucleotide spliced leader sequence in the human parasitic nematode *Brugia malayi* is identical to the *trans*-spliced leader exon in *Caenorhabditis elegans*. *Proceedings of the National Academy of Sciences of the United States of America* **85,** 7932-7936.

Takahara, K., Brevard, R., Hoffman, G. G., Suzuki, N., and Greenspan, D. S. (1996). Characterization of a novel gene product (mammalian tolloid-like) with high

sequence similarity to mammalian tolloid/bone morphogenetic protein-1. *Genomics* **34,** 157–165.

Takahara, K., Lyons, G. E., and Greenspan, D. S. (1994). Bone morphogenetic protein-1 and a mammalian tolloid homolog (mTld) are encoded by alternatively spliced transcripts which are differentially expressed in some tissues. *Journal of Biological Chemistry* **269,** 32572-32578.

Takahashi, N., Hayano, T., and Suzuki, M. (1989). Peptidyl-prolyl *cis-trans* isomerase is the cyclosporin A-binding protein cyclophilin. *Nature* **337,** 473-475.

Tasab, M., Batten, M. R., and Bulleid, N. J. (2000). Hsp47: a molecular chaperone that interacts with and stabilizes correctly-folded procollagen. *EMBO Journal* **19,** 2204-2211.

Tavernarakis, N., Wang, S. L., Dorovkov, M., Ryazanov, A., and Driscoll, M. (2000). Heritable and inducible genetic interference by double-stranded RNA encoded by transgenes. *Nature Genetics* **24,** 180-183.

Taylor, P., Page, A. P., Kontopidis, G., Husi, H., and Walkinshaw, M. D. (1998). The X-ray structure of a divergent cyclophilin from the nematode parasite *Brugia malayi*. *FEBS Letters* **425,** 361-366.

Thacker, C., Marra, M. A., Jones, A., Baillie, D. L., and Rose, A. M. (1999). Functional genomics in *Caenorhabditis elegans*: an approach involving comparisons of sequences from related nematodes. *Genome Research* **9,** 348-359.

Thacker, C., Peters, K., Srayko, M., and Rose, A. M. (1995). The *bli-4* locus of *Caenorhabditis elegans* encodes structurally distinct kex2/subtilisin-like endoproteases essential for early development and adult morphology. *Genes & Development* **9,** 956-971.

Thacker, C., and Rose, A. M. (2000). A look at the *Caenorhabditis elegans* kex2/subtilisin-like proprotein convertase family. *Bioessays* **22,** 545-553.

Thein, M. C., McCormack, G., Winter, A. D., Johnstone, I. L., Shoemaker, C. B., and Page, A. P. (2003). The*Caenorhabditis elegans* exoskeleton collagen COL-19: an adult-specific marker for collagen modification, assembly and the analysis of organismal morphology. *Developmental Dynamics*, in press.

Timmons, L., Court, D. L., and Fire, A. (2001). Ingestion of bacterially expressed dRNAs can produce specific and potent genetic interference in *Caenorhabditis elegans*. *Gene* **263,** 103-112.

Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. *Nature* **395,** 854-854.

van der Keyl, H., Kim, H., Espey, R., Oke, C. V., and Edwards, M. K. (1994). *Caenorhabditis elegans sqt-3* mutants have mutations in the *col-1* collagen gene. *Developmental Dynamics* **201,** 86-94.

van Voorhies, W. A., and Ward, S. (2000). Broad oxygen tolerance in the nematode *Caenorhabditis elegans*. *Journal of Experimental Biology* **203,** 2467-2478.

Vasan, N. S., Kuivaniemi, H., Vogel, B. E., Minor, R. R., Wootton, J. A. M., Tromp, G., Weksberg, R., and Prockop, D. J. (1991). A mutation in the pro-α2(I) gene (*COL1A2*) for type I procollagen in Ehlers-Danlos syndrome type VII: evidence suggesting that skipping of exon 6 in RNA splicing may be a common cause of the phenotype. *American Journal of Human Genetics* **48,** 305-317.

Veijola, J., Annunen, P., Koivunen, P., Page, A. P., Philajaniemi, T., and Kivirikko, K. I. (1996a). Baculovirus expression of two protein disulphide isomerase isoforms from *Caenorhabditis elegans* and characterisation of prolyl 4-hydroxylases containing one of these polypeptides as their β subunit. *Biochemical Journal* **317,** 721-729.

Veijola, J., Koivunen, P., Annunen, P., Pihlajaniemi, T., and Kivirikko, K. (1994). Cloning, baculovirus expression, and characterization of the α subunit of prolyl 4-hydroxylase from the nematode *Caenorhabditis elegans*. This α subunit forms an active αβ dimer with the human protein disulfide isomerase/β subunit. *Journal of Biological Chemistry* **269,** 26746-26753.

Veijola, J., Pihlajaniemi, T., and Kivirikko, K. I. (1996b). Co-expression of the α subunit of human prolyl 4-hydroxylase with BiP polypeptide in insect cells leads to the formation of soluble and insoluble complexes. Soluble α subunit-BiP complexes have no prolyl 4-hydroxylase activity. *Biochemical Journal* **315,** 613-618.

von Mende, N., Bird, D. M., Albert, P. S., and Riddle, D. L. (1988). *dpy-13*: a nematode collagen gene that affects body shape. *Cell* **55,** 567-576.

Vuorela, A., Myllyharju, J., Nissi, R., Pihlajaniemi, T., and Kivirikko, K. I. (1997). Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: formation of a stable enzyme tetramer requires coexpression

with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *EMBO Journal* **16,** 6702-6712.

Vuori, K., Pihlajaniemi, T., Marttila, M., and Kivirikko, K. I. (1992a). Characterization of the human prolyl 4-hydroxylase tetramer and its multifunctional protein disulfide isomerase subunit synthesized in a baculovirus expression system. *Proceedings of the National Academy of Sciences of the United States of America* **89,** 7467-7470.

Vuori, K., Pihlajaniemi, T., Myllylä, R., and Kivirikko, K. I. (1992b). Site-directed mutagenesis of human protein disulfide isomerase: effect on the assembly, activity and endoplasmic reticulum retention of human prolyl 4-hydroxylase in *Spodoptera frugiperda* insect cells. *EMBO Journal* **11,** 4213-4217.

Walmsley, A. R., Batten, M. R., Lad, U., and Bulleid, N. J. (1999). Intracellular retention of procollagen within the endoplasmic reticulum is mediated by prolyl 4-hydroxylase. *Journal of Biological Chemistry* **274,** 14884-14892.

Walsh, C. T., Zydowsky, L. D., and McKeon, F. D. (1992). Cyclosporine A, the cyclophilin class of peptidylprolyl isomerases, and blockade of T-cell signal transduction. *Journal of Biological Chemistry* **267,** 13115-13118.

Waterston, R. H., and Brenner, S. (1978). A suppressor mutation in the nematode acting on specific alleles of many genes. *Nature* **275,** 715-719.

White, K. P., Hurban, P., Watanabe, T., and Hogness, D. S. (1997). Coordination of *Drosophila* metamorphosis by two ecdysone-induced nuclear receptors. *Science* **276,** 114-117.

WHO, (2000). Lymphatic filarisis. *Fact sheet no. 102.*

Williams, S. A., Lizotte-Waniewski, M. R., Foster, J., Guiliano, D., Daub, J., Scott, A. L., Slatko, B., and Blaxter, M. L. (2000). The filarial genome project: analysis of the nuclear, mitochondrial and endosymbiont genomes of *Brugia malayi*. *International Journal for Parasitology* **30,** 411-419.

Wilm, T., Demel, P., Koop, H. U., Schnabel, H., and Schnabel, R. (1999). Ballistic transformation of *Caenorhabditis elegans*. *Gene* **229,** 31-35.

Wilson, R., Lees, J. F., and Bulleid, N. J. (1998). Protein disulfide isomerase acts as a molecular chaperone during the assembly of procollagen. *Journal of Biological Chemistry* **273,** 9637-9643.

Wilson, W. R., Tuan, R. S., Shepley, K. J., Freedman, D. O., Greene, B. M., Awadzi, K., and Unnasch, T. R. (1994). The *Onchocerca volvulus* homolog of the

multifunctional polypeptide protein disulfide isomerase. *Molecular and Biochemical Parasitology* **68,** 103-117.

Winter, A. D., and Page, A. P. (2000). Prolyl 4-hydroxylase is an essential procollagen-modifying enzyme required for exoskeleton formation and the maintenance of body shape in the nematode *Caenorhabditis elegans*. *Molecular and Cellular Biology* **20,** 4084-4093.

Yang, J., and Kramer, J. M. (1994). *In vitro* mutagenesis of *Caenorhabditis elegans* cuticle collagens identifies a potential subtilisin-like protease cleavage site and demonstrates that carboxyl domain disulfide bonding is required for normal function but not assembly. *Molecular and Cellular Biology* **14,** 2722-2730.

Yang, J., and Kramer, J. M. (1999). Proteolytic processing of *Caenorhabditis elegans* SQT-1 cuticle collagen is inhibited in right roller mutants whereas cross-linking is inhibited in left roller mutants. *Journal of Biological Chemistry* **274,** 32744-32749.

Yeowell, H. N., Allen, J. D., Walker, L. C., Overstreet, M. A., Murad, S., and Thai, S. F. (2000). Deletion of cysteine 369 in lysyl hydroxylase 1 eliminates enzyme activity and causes Ehlers-Danlos syndrome type VI. *Matrix Biology* **19,** 37-46.

Yeowell, H. N., and Walker, L. C. (2000). Mutations in the lysyl hydroxylase 1 gene that result in enzyme deficiency and the clinical phenotype of Ehlers-Danlos syndrome type VI. *Molecular Genetics and Metabolism* **71,** 212-224.

Yochem, J., Tuck, S., Greenwald, I., and Han, M. (1999). A gp330/megalin-related protein is required in the major epidermis of *Caenorhabditis elegans* for completion of molting. *Development* **126,** 597-606.

Yu, F., White, S. B., Zhao, Q., and Lee, F. S. (2001). HIF-1α binding to VHL is regulated by stimulus-sensitive proline hydroxylation. *Proceedings of the National Academy of Sciences of the United States of America* **98,** 9630-9635.

Zorio, D. A. R., Cheng, N. S. N., Blumenthal, T., and Spieth, J. (1994). Operons as a common form of chromosomal organization in *C. elegans*. *Nature* **372,** 270-272.