THE N-TERMINUS OF UNC-53 INTERACTS WITH REGULATORS OF THE ACTIN CYTOSKELETON

by

Adetayo Olumide Adeleye

B.Sc., Trinity Western University, 2000

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

In the Department
of
Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

August 2003

APPROVAL

Name: **Adetayo Olumide Adeleye**

Degree: Master of Science

Title of Thesis: **THE N-TERMINUS OF UNC-53 INTERACTS WITH REGULATORS OF THE ACTIN CYTOSKELETON**

Examining Committee:

Chair: **Dr. Glen F. Tibbits**

Dr. David L. Baillie
Senior Supervisor
Professor

Dr. Eve G. Stringham
Co-Supervisor, Associate Professor of Biology,
Trinity Western University

Dr. Nicholas Harden
Assistant Professor
Department of Molecular Biology and Biochemistry

Dr. Esther M. Verheyen
Assistant Professor
Department of Molecular Biology and Biochemistry

Dr. Nancy Hawkins
Public Examiner

Date Approved: August 20, 2003
~~September 15~~ 2003

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay:

The N-Terminus of UNC-53 Interacts With Regulators of the Actin Cytoskeleton

Author:

(signature)

Adetayo Adeleye

(name)

15/9/03

(date)

ABSTRACT

The *unc-53* gene is required for the guidance and extension of a subset of cells along the anterior-posterior axis of *Caenorhabditis elegans*, including several neurons, the sex muscles and the excretory canals. Immunofluorescence of adult hermaphrodites with antisera raised against the calponin homology (CH) domain in the amino terminus of UNC-53 stains the excretory canals, vulva and coelomocytes, suggesting that the full length UNC-53 protein is required in these cells after migration and outgrowth. Staining in *n152* and *n166* mutants indicates that these animals express truncated UNC-53 products.

In order to identify proteins that interact with UNC-53, a yeast two-hybrid screen was conducted. Using the N-terminal CH domain as 'bait' and a *C. elegans* cDNA library as 'prey', 18 strong candidate interactors were identified. Seven of these correspond to the *C. elegans* genomic locus B0336.6, which encodes a SH3 containing protein suggesting a role in signal transduction. The SH3 domain of the *Drosophila* homologue, Abi, binds to the polyproline motif of Abelson tyrosine kinase and the yeast homologue, Ysc84p, is thought to couple the actin cytoskeleton to the endocytic machinery. Four candidate cDNAs encode *C. elegans* REF(2) P like protein that contains a zinc finger motif of the ZZ type found in the cytoskeletal protein dystrophin, and a ubiquitin associated domain. The yeast homologue, verprolin, is an actin binding protein that presumably modulates polarization of the actin cytoskeleton. The yeast protein interaction database identified yeast verprolin/ *C. elegans* T12G3.1 and yeast Ysc84p/*Drosophila* Abi /*C. elegans* B0336.6 as interactors with yeast Sla1p, a multifunctional protein that couples the yeast endocytic machinery to proteins regulating actin dynamics. These data suggest that UNC-53, together with verprolin/T12G3.1, Ysc84p/Abi/B0336.6 and Sla1p may form a complex at the cytoskeleton that stabilizes or promotes actin polymerization.

The remaining 7 candidates define molecules with diverse cellular activities. Interestingly, none of the candidate UNC-53 interactors have been discovered in previous genetic screens aimed at identifying genes involved in cell migration and outgrowth.

DEDICATION

To my parents, Edward Adeniran and Victoria Eyinade Adeleye, for your support, love and prayers. I am grateful to God and thankful to you for your untiring support. You have given me an excellent foundation and afforded me the rudiments to build this edifice called 'life.'

ACKNOWLEDGEMENTS

I would like to thank Dr Eve Stringham, for giving me the opportunity to work with her. My experience has been marked by mentorship, guidance, support and encouragement. I am grateful for the patience and direction you gave me all through this program. I count myself very fortunate to have had this opportunity to be your graduate student. I also thank Dr David Baillie for direction and encouragement. I have always been encouraged by your sincere interest in the progress of your students and I am thankful for the opportunity I had to work with you. I would also like to thank Drs. Nick Harden and Esther Verheyen for their active roles in my research progress as members of my supervisory committee. I am very thankful for the support and guidance you gave to me. I also express my gratitude to all members of the Baillie lab for the warm acceptance and support I received at SFU. You all created a relaxed atmosphere which made working in the lab enjoyable. You have made my experience at SFU truly remarkable. In addition, I would like thank John Tyson for technical assistance with my experiments, encouragement and helpful discussions, Zhao Zhongying for your friendship and helpful discussions and Allan Mah for performing injections, providing strains and encouragement. I also thank Dr Candido and members of the Candido Lab at UBC for reagents, yeast strains, cDNA library, and bacteria strains. I thank Don Jones for technical assistance, insightful discussions and Marieke Groombridge for advice, encouragement and assistance with my thesis. I would also like to thank Samantha Grainger for much help I received with making media and solutions. I have also appreciated your friendship and encouragement. I thank all members of the MBB department for their support and friendship. I thank the laboratory technicians at TWU, Darcy Kehler, Lisa Chattel, Kim Siemens and Jemma Tjosvold for giving me accessibility to their offices and computers, as well as free reign in the lab. I also thank Cathy White for assistance with computer and formatting issues. To my family and friends who have supported me in diverse ways, thank you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

3-AT	3-aminotriazole
Abi	Abelson interacting protein
Abl	Abelson protein kinase
Abp	actin binding protein
Arp	actin related protein
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BLAST	basic local alignment search tool
bp	base pairs
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleoside triphosphates
EDTA	ethylene diamine tetra-acetic acid
EMS	ethylmethanesulfonate
GST	glutathione s-transferase
HIS	histidine
IPTG	isopropyl-β-D-thiogalactosidase
LiAc	lithium acetate
NBT	nitroblue tetrazolium
OD	optical density
OTU	Ovarian Tumour
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PEG	polyethylene glycol
RNA	ribonucleic acid
RNAi	RNA interference
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SL	spliced leader

TCOF	Treacher-Collins syndrome
Tris	tris(hydroxymethyl)aminomethane
unc	uncoordinated
VRP1	verprolin
WASP	Wiskott-Aldrich Syndrome protein
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase

1. INTRODUCTION

1.1 *Caenorhabditis elegans* as a model system

The small nematode worm, *Caenorhabditis elegans*, has proved not only to be an intriguing organism of study, but also an excellent model system for the study of higher eukaryotes since its inception as a model system in the mid sixties by Sydney Brenner. Many researchers have adopted it as a model system of choice for reasons known to Brenner then and that have been unearthed during the past 4 decades of research on the tiny worm. Its relatively short (3 days) life cycle, ease of cultivation under laboratory conditions and small size (only 1mm in length), has made *C. elegans* highly amenable to genetic analysis. A comparatively simple anatomy, comprised of 959 somatic body cells that make up an invariant cell lineage that has been delineated from fertilized egg to adult have made it ideal for the study of developmental processes (Sulston et al., 1983). Add to this the transparency of its outer cuticle throughout all developmental stages, which allows visualization with a simple light microscope. *C. elegans* popularity has only heightened with the recent completion of the genomic sequence. The 97-megabase genomic sequence of the worm codes for over 19,000 genes (*C.elegans* Sequencing Consortium, 1998), which is about three times the number of genes in bakers yeast, *Sacharomyces cerevisiae*, 1.3 times that of *Drosophila melanogaster* and ten times that of *Haemophilus influenzae* (Rubin et al., 2000). Initial estimates by the *C. elegans* Sequencing Consortium suggested that the number of genes in *C. elegans* were between one-fifth to one-third the number predicted in humans and 74% of human gene sequences had nematode matches (*C. elegans* Sequencing Consortium, 1998). The human genome project revealed that there were 2031 "strict" human-worm orthologues (i.e. proteins with unambiguous one-to-one relationships), 1523 of which were also present in *D. melanogaster* and thus classified as being evolutionary conserved (Venter et al., 2001). In an *in silico* search to identify *C. elegans* genes that were potential homologues of human genes associated with inborn errors of metabolism, Kuwabara and O'Neil, discovered that of the 90 human genes with OMIM (Online Mendelian Inheritance in Man) identifiers, 78% of these genes defined a potential *C. elegans* homologue (Kuwabara and O'Neil,

2001). These kinds of findings have proven the worm to be a valuable functional tool for the discovery and characterization of eukaryotic genes.

1.2 Cell guidance; migration and outgrowth

C. elegans, with its invariant somatic cell lineage and the presence of a self-fertilizing XX hermaphrodite that contributes to its genetic tractability, can lead to a gross oversimplification of the overall complexity of the worm. Despite its ostensible simplicity, the worm undergoes several developmental changes ranging from molting to the development of complex organs including a nervous system, gonad, coelomocytes, and an excretory/secretory system (Riddle et al., 1997). A key biological process that accompanies these developmental changes is cell migration. An understanding of the processes that occur during cell migration and outgrowth are indispensable and paramount to understanding the developmental processes that occur in both vertebrates and invertebrates. During development, many cells migrate through the complex milieu of the extracellular matrix and undergo extensive morphological changes essential for the establishment of functional connections. Others remain in the locality where they are born, making functional connections with their neighbors. How does this all happen? Hynes and Lander (1992) comment that, irrespective of cell type, migration must be promoted and must also be guided. Mechanisms effecting these processes include chemotaxis, in which cells migrate in response to a diffusible gradient, haptotaxis, where migration is directional in response to substrata deposited by molecules of the extracellular matrix and by contact guidance. Importantly these migratory cues may be positive, initiating and promoting migration in a specified direction, but may also be repulsive or inhibitory preventing entry into some environments (Hynes and Lander, 1992). These migratory or inhibitory cues are encoded by biological molecules that include ligands such as laminins, semaphorins, netrins and transmembrane receptors such as integrins and some receptor tyrosine kinases such as the fibroblast growth factor receptor (FGFR) (Tessier-Lavigne and Goodman, 1996). It is presumed that information received at the leading edge of the migrating cell (or growth cone) is integrated to produce an appropriate internal response that ultimately results in a change in the

orientation of the cell. Emanating at the membrane front of the migrating or outgrowing cell are filopodia (thin cylindrical needle-like projections) and lamellipodia (broad flat sheet-like structures) that branch out in several directions and extend in response to migratory stimuli. It has also been observed that the extension of lamellipodia and filopodia is usually coupled with local actin polymerization (Lauffenburger and Horwitz, 1996). In line with this view, Bentley and O'Connor (1994) have observed that in response to an extracellular signal, F-actin accumulates at the leading edge of the cell and a rearrangement of the actin cytoskeleton occurs which favors cell growth in the direction of greatest actin accumulation (Bentley and O'Connor, 1994). In addition, the role of the *rho* subfamily (*cdc42*, *rac* and *rho*) of the *ras* family of GTP-binding proteins in modulating actin cytoskeletal dynamics has been well documented (Lauffenburger and Horwitz, 1996, Steven et al., 1998, Hall, 1998). Rho has been shown to stimulate the assembly of contractile actin-myosin fibers in fibroblast cells in response to the presence of extracellular ligands such as lysophosphatidic acid. Rac activation by platelet-derived growth factor or insulin leads to the accumulation of a mesh work of actin filaments at the cell cortex to produce lamellipodia and membrane ruffles, while cdc42 induces the formation of actin-rich filopodia (Hall, et al., 1998). Also there is evidence that significant cross-talk occurs between these GTPases such that, for example, Cdc42 can activate Rac, which implies formation of filopodia is tied to the establishment of the lamellipodia and vice versa (Hall et al., 1998). Taken together, this infers that the mechanisms that initiate and sustain cell migration are directly linked to signaling processes that modulate the actin cytoskeleton.

1.3 Dorsal - ventral vs. Anterior – Posterior Guidance

In *C. elegans,* cell migrations occur in both the anterior-posterior (i.e. in the longitudinal axis) and the dorsal-ventral directions and certain genes specify migrations in a particular axis. Three such genes shown to function in a common pathway to control axonal guidance along the dorso-ventral axis in *C. elegans* include *unc-5*, *unc-40* and *unc-6*. *unc-5* and *unc-40* are required for dorsal and ventral neuronal migrations respectively, while *unc-6* is required for both dorsal and ventral migrations (Hedgecock et

al., 1990). In agreement with these roles, mutants in *unc-5*, *unc-6*, and *unc-40* show a combination of developmental defects that affect the establishment of a proper circumferential network of neurons. Genetic and molecular evidences point to a signaling system in which *unc-5* and *unc-40* encode cell surface receptors that interact directly with *unc-6*. *unc-5* encodes a transmembrane receptor that is composed of thrombospondin and immunoglobulin repeats in its extracellular region (Serafini et al., 1994). Likewise, *unc-40* encodes a transmembrane receptor that is comprised of immunoglobulin and fibronectin type III repeats and is highly related to the human DCC (deleted in colorectal cancer) tumour suppressor gene (Chan et al., 1996). *unc-6*, like its mammalian homologues, the netrins, codes for an extracellular matrix protein with significant similarities to the laminin family of proteins (Serafini et al., 1994). The laminins through their short range signaling have been shown to promote outgrowth of the nervous system (Hynes and Lander, 1992). More recently crosstalk between pathways that define dorsal ventral migrations have been reported (Hao et al., 2001). The *C. elegans sax-3* gene, which is homologous to the fly *robo*, directs ventral axon guidance and guidance at the midline, whereas *sax-3/robo* mutants display indiscriminate crossing over of axons across the midline (Hao et al., 2001). SLT1 is the ligand for SAX-3 and mutants for the *slt-1* gene also show defects in axon midline crossing. Hao et al. (2001) observed an interplay in the pathways that guide ventral migration of the AVM sensory neuron in cells that expressed both SAX-3 and UNC-40. It was shown that the attractive guidance factor UNC-6, expressed by the ventral cord axons attracts the AVM ventrally and away from the SLT-1 cue emanating from the dorsal muscle cells, thus suggesting that Slt-1 promotes the UNC-40, UNC-6 pathway by acting as a repellant that forces the AVM migration in the ventral direction (Hao et al., 2001).

Another mechanism, albeit not as clearly defined, exists for migrations in the anterior-posterior direction of *C. elegans*. Evidence for such a position comes from analysis of genes such as *vab-8*, *mig-10* and *unc-53*, all genes, which seem to predominantly affect migrations along the longitudinal axis of the worm. Mutations at the *vab-8* locus have been shown to disrupt migratory events in both neuronal and non-neuronal cell types, particularly posteriorly directed cell migrations (Wightman et al., 1996). Recent studies have indicated that *vab-8* encodes a kinesin like motor domain. It

has been postulated that *vab-8*, via its kinesin like domain, might target proteins to the neuronal growth cone by propelling cargo along the microtubule core of the axon toward the end (Wolf et al., 1998).

Another gene modulating outgrowth and extensions along the anteroposterior axis is the *mig-10* gene. The *mig-10* gene is required for the migration of embryonic neurons such as the CAN, ALM, and HSN and proper development of the excretory canals. Molecular characterization of *mig-10* revealed the presence of Pleckstrin Homology domains akin to those found in the mammalian SH2 domain proteins, Grb7 and Grb10. In addition *mig-10* also contains polyproline repeats, which are commonly found in the SH3 binding domains of several signal transduction molecules (Manser et al., 1997).

1.4 *unc-53* and its role in AP migrations

Another molecule that has been implicated in regulating longitudinal migrations in *C. elegans* is the *unc-53* gene (unc for uncoordinated). The term "uncoordinated" dates from as far back as the early 70's, when Sydney Brenner after subjecting young adult nematodes to ethylmethanesulfonate (EMS) defined all mutant progeny animals that showed a detectable defect from the normal pattern of behavior as "uncoordinated" (Brenner, 1974). Wild type worms display a reversible smooth sinuous movement on agar plates seeded with *E. coli* in the laboratory. Since then several other researchers have mapped mutations involved in developmental processes to *unc-53*. In the early eighties, Carol Trent mapped an egg-laying defective mutation to the *unc-53* locus (Trent et al., 1983). Edward Hedgecock in the mid-late eighties characterized the *unc-53* mutants using Normaski and showed that they possessed truncated excretory canals (Hedgecock et al., 1987). More recently, Stringham and colleagues (2002) observed that the longitudinal migration of the sex muscles, which allows for proper egg-laying at the vulva, does not occur in *unc-53* mutants. In wild type (N2) animals, the sex myoblasts migrate anteriorly from the posterior to the position of the presumptive vulva where they undergo a series of divisions to give rise to the vulva muscle cells. The vulval myoblasts subsequently migrate longitudinally away from the vulva before extending attachments ventrally to the

hypodermis. Simultaneous contraction of these muscles allows the vulva to open and eggs to be laid. Since this latter longitudinal migration does not occur in *unc-53* mutants, the opening of the vulva is constricted and the mutants are unable to lay eggs properly (Stringham et al., 2002). Also, the sex muscle attachments that extend longitudinally to anchor the vulva muscles to the hypodermis are not fully extended in *unc-53(n152)* mutants, whereas, the ventral aspects of these attachments are wild type in the *unc-53* mutants.

In contrast, muscle cells over expressing *unc-53* showed increased growth cone extension. The longitudinal migrations of various neurons are also affected in *unc-53* mutants. For example, the ALN and PLN neurons, which extend processes from the tail ganglion to the head region, terminate prematurely in *unc-53* mutants (Stringham et al., 2002). Together these findings consolidate *unc-53*'s role for directional guidance and extension in a subset of cells and growth cones along the longitudinal axis of the nematode; including several neurons, the sex muscles, and the excretory canals.

The *unc-53* gene (Figure 1a) is large, consisting of 23 exons spanning over 31kb of genomic sequence and the largest transcript encodes a protein of 1583 amino acids. The gene locus contains multiple promoters and *unc-53* transcripts are subject to alternative splicing, suggesting the presence of multiple protein isoforms in vivo (Stringham et al., 2002). Recently three human homologues, *NAV1* (neuron navigator), *NAV2/RAINB1* (retinoic acid inducible in neuroblastoma), *NAV3*, were identified with significant conservation to the nematode *unc-53* gene (Maes et al., 2002). Of the three NAV homologues, UNC-53 is most closely related to NAV2/RAINB1 (Stringham et al., 2002). RAINB1/NAV2 was discovered in a study to identify genes that were up regulated in response to all trans-retinoic acid (atRA), which is required for patterning of the nervous system in mammals (Merill et al., 2002). NAV2 is also expressed in adult kidney and placenta (Maes et al., 2002) and in the heart (Merill et al., 2002). Together, these findings suggest a role for UNC-53/NAV in the navigation of various cells, particularly neurons during development.

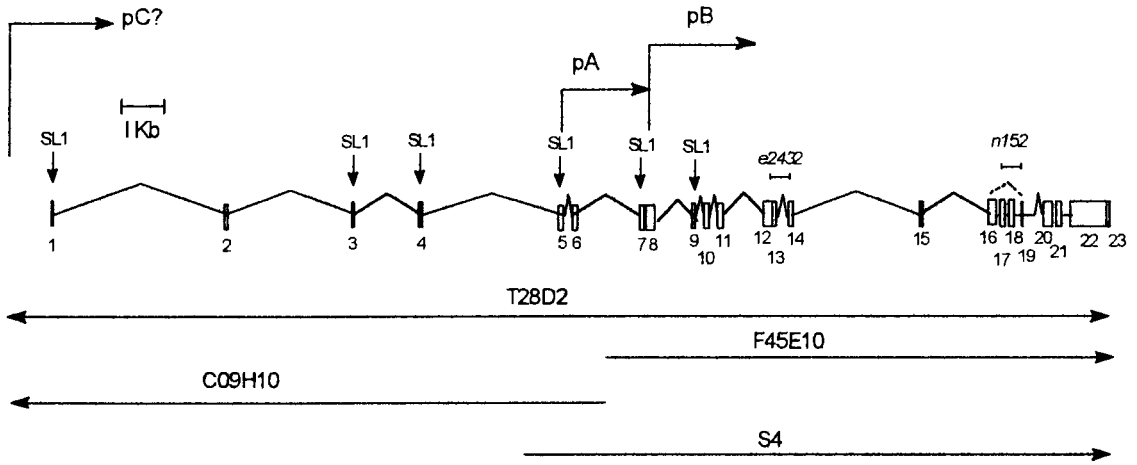
The UNC-53 protein sequence contains domains that are present in a variety of signaling molecules as shown in Figure 1b. UNC-53 contains a single calponin homology

(CH) domain at its N- terminus, situated at amino acids 11 to 109 (Stringham et al., 2002). This CH motif is present in various actin-binding proteins, including α-actinin and dystrophin, which have been implicated in regulating actin and cell shape dynamics (Van Troys et al., 1999). Adjacent to the CH domain is an additional putative actin-binding site of the LKK consensus from amino acids 114 to 133 and a second LKK domain from amino acids 1097 to 1116. Importantly, UNC-53 was shown to co-sediment with F-actin *in vitro* (Stringham, unpublished results).

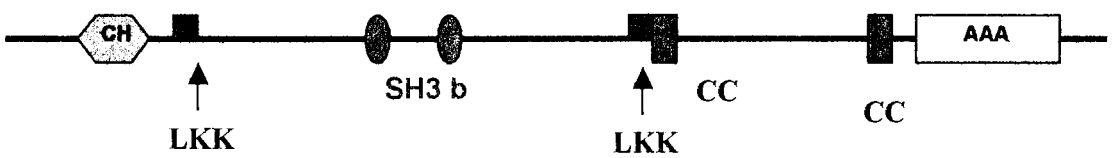
Figure 1: Genomic organization of the *unc-53* locus.

Stringham et al., 2002. (Adapted by permission of Eve Stringham). **(a)** The different SL1 transpliced sites, promoters A, B and C as well as the regions deleted in the *n152* and *e243* alleles are indicated. Phage and rescuing cosmids are shown below. **(b)** Bottom panel shows organization of conserved domains in the UNC-53 protein. A single calponin homology domain (fluorescent green), two leucine rich LKK domains (blue), two polyproline rich domains (green), two coiled-coiled domains (red) and an AAA (ATPases-associated with diverse cellular activities) domain that contains a putative nucleotide (NTP)-binding site (yellow).

(a)


pC?
pB
pA
1Kb
SL1
SL1
SL1
SL1
SL1
SL1
e2432
n152
1
2
3
4
5 6
7 8
9 11
10
12 14
13
15
16 18 20 22 23
17 19 21
T28D2
F45E10
C09H10
S4

(b)


CH
AAA
SH3 b
CC
CC
LKK
LKK

It also contains two polyproline rich sequences (putative SH3 binding domains) at residues 487-495 and 537-545, which are deemed important for binding molecules containing a SH3 domain (Stringham et. al., 2002). These polyproline repeats are found in signaling molecules such as SOS, dynamin and PI3 kinase (Yu et al., 1994). Two regions towards the C-terminus of UNC-53 (890-923 and 1078-1113) define a putative coiled-coil domain that may be able to mediate homomeric or heteromeric protein-protein interactions (Stringham et al., 2002; Maes et al., 2002), in cooperation with a nucleotide (NTP)-binding site contained within an AAA (ATPases-associated with diverse cellular activities) domain (Confalonieri and Duguet, 1995; Stringham et al., 2002). While there is evidence to support a role for UNC-53 in AP migrations, the mechanisms or signaling pathway by which UNC-53 may function are not clearly understood. Genetic analysis has shown that partial loss of function alleles of *unc-53* enhance the sex myoblast migration defect of *sem-5* null mutants.

SEM-5 is the *C. elegans* signaling adaptor homologue of GRB2/DRK and consists exclusively of SH2 and SH3 domains (Clark, et al., 1992; Stern et al., 1993). Biochemical evidence suggests that the genetic interaction between *unc-53* and *sem-5* is direct; UNC-53 protein physically associates with SEM-5 and the mammalian homologue, GRB2 *in vitro* (Stringham et al., 2002). In the *sem-5* model, Branda and Stern (2000) showed that a gonad-dependent attraction is required to guide the sex myoblasts to the position of the future vulva; a process mediated through an attractive cue encoded by the extracellular growth factor, EGL-17/FGF. This signal is internalized by the growth factor receptor, EGL-15/FGFR, and a signaling pathway ensues via the GRB2 adapter homologue SEM-5 to a downstream component. Other molecules that participate in this pathway include a receptor tyrosine phosphatase encoded by the *clr-1* (clear) gene (Kokel et al., 1998), and a leucine rich protein encoded by the *soc-2* (Suppressor of clear) gene (Selfors et al.,1998). In addition to the gonad-dependent mechanism for migration of the sex myoblasts, a gonad independent pathway has also been proposed, in which the activity of *unc-53*, *unc-71* and *unc-73* are required (Chen et al., 1997). Interestingly, all these *unc* genes have been implicated in cell guidance processes (Hedgecock et al., 1987). The *unc-71* gene encodes an ADAM (a disintegrin and metalloprotease) implicated in both axon guidance and sex myoblast migration (Huang et al., 2003), while *unc-73* encodes a guanine

nucleotide exchange factor, which activates Rac GTPase *in vitro* and stimulates actin polymerization in mammalian cells, suggesting that it participates in cell guidance by regulating actin dynamics (Steven et al., 1998).

The genetic data show that *unc-53* acts synergistically with *sem-5*, which would suggest that *unc-53* acts in a parallel pathway to *sem-5*, possibly with *unc-71* and *unc-73* as shown in the gonad independent pathway. In contrast, biochemical results suggest that the interaction between UNC-53 and SEM-5 is direct (Stringham et al., 2002). It is possible that while UNC-53 can bind to SEM-5/GRB2 *in vitro*, the actual *in vivo* target may be another SH2SH3 adapter molecule.

Despite the conflicting views, however, the genetics, the sequence data, and the biochemistry point to a role for UNC-53 in both actin binding and signal transduction. These observations suggest that *unc-53*, like *unc-73*, may participate in cell migration and outgrowth by the interpretation or relay of a signal to the actin cortex to promote cytoskeletal rearrangement, which stabilizes a growth cone and favors extension in a specified direction.

1.5 Functional genomic approaches

Genetics, and in particular suppressor analyses, have proved to be instrumental in identifying components of several signaling pathways including the ones already discussed. Recently however, the near completion of several genomes has precipitated the development of new methods in functional genomics, which have added to the genetic and molecular options available to uncover new constituents and partners in cellular processes. For example, full genome DNA microarrays can be used to determine the expression profile of all genes required at a certain stage in development or in a given environment. This approach is amenable to identify genes that are developmentally regulated, or genes that function together in a protein complex or genetic pathway or further still genes that are differentially enriched in males vs. hermaphrodites. Jiang and colleagues used this method to identify components required for Wnt signaling on the premise that proteins that interact with each other or in the same complex/pathway must

also be expressed at the same time. In essence these schemes have yielded a profile of temporal expression that correlates gene expression to time and place during all stages of development (Jiang et al., 2001).

Another method that has gained much recognition, as a functional approach for analyzing gene activity is RNA mediated interference (RNAi). RNAi, which targets the RNA for degradation by the introduction of double stranded RNA (dsRNA) that is specific to the sequence of the targeted gene, is a highly useful and rapid means for ascertaining gene function (Kamath et al., 2003). This method has found wide application from the analysis of single genes (Tsuboi, et al., 2002) to whole chromosome (Fraser, et al., 2000) to that of complete genomes (Kamath, et al., 2003). RNAi analysis reveals the phenotype of the nematode when the function of a specific gene(s) has been disrupted and indicates the developmental stage at which gene function is required.

The popularity of the yeast two-hybrid screen as a functional genomic tool for the identification of protein-protein interaction has continued to increase since its development by Fields and Songs in the late eighties (Fields and Songs, 1989). The idea that led to the adoption of the yeast two-hybrid system as a practical tool emerged from analysis of transcription factors such as the GAL4p in yeast. Transcription factors increase the rate of transcription of their target genes by binding to upstream activating DNA sequences (UAS) and thus 'activating' RNA polymerase II at the corresponding promoters. Keegan et al. (1986) demonstrated that the DNA binding and the activation functions reside in physically separable domains of GAL4p, which are referred to as the DNA-binding domain (DBD) and the activation domain (AD), respectively. In the two-hybrid system, a protein of interest, 'X', is fused to the DNA binding domain (DBD) of a transcription factor and the potential candidates 'Y' are fused to a transcription activation domain (AD). If there is an interaction between X and Y a functional transcription factor is reconstituted which results in the expression of specific reporter genes under the control of the transcriptional factor (Fields and Songs, 1989). Usually the protein fusion consisting of protein X and the DBD domain is referred to as the 'bait' and the fusion made up of protein Y and the AD domain is called the 'prey.' The DBD and AD domains commonly used are those of the GAL4 transcription factor while the reporter gene most

commonly used is *E. coli lacZ*, which encodes beta-galactosidase. Positive interactions are identified as blue colonies on plates or filters containing X-Gal. A second reporter such as *LEU2* or *HIS3* is usually included to allow for selection of colonies that contain interacting proteins (Fields and Strenglanz, 1994). There are several applications for the two hybrid system including direct two hybrid assays, where two known proteins are cloned into the screening vectors, as seen in the interaction between the serine/threonine kinase UNC-51 and the protein product of the *unc-14* gene which is required for axon outgrowth and guidance (Ogura et al., 1997), or a functional approach, where a single "bait" protein of interest is used to identify "preys" from a cDNA library as shown in the screen to identify molecules that interact with GEX-3, a molecule involved in tissue morphogenesis (Tsuboi et al., 2002), and genome wide functional approaches to perform pair-wise analysis in which all the ORF's are used as both 'bait' and 'prey.' Examples of these include an interaction map of proteins required for the development of the hermaphrodite vulva (Walhout et al., 2000) and a comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae* (Uetz et al., 2000).

Examination of the amino acid sequence of UNC-53 indicates that it is probably involved in additional protein-protein interactions as the functional domains identified so far account for only 20% of the amino acid sequence. A better understanding of the role of UNC-53 in cell migration and outgrowth would require the use of both genetics and molecular approaches to uncover novel molecules that participate in this signaling cascade. The aim of the present study was to identify molecular partners of UNC-53 using a two-hybrid screen in which a portion of the *unc-53* cDNA was used as 'bait' and a *C. elegans* cDNA library was used as 'prey.' In this study the bait used contained the N-terminal portion of the UNC-53 gene product inclusive of the CH domain and first LKK motif and an identical region was used to generate polyclonal antibodies to further aid the characterization of UNC-53 gene products.

2. MATERIALS AND METHODS:

2.1 Maintenance of *C. elegans*

Wild type *C. elegans* Bristol strains (N2) (see below for strains used) were maintained on a nutrient growth media (NGM) (0.3% NaCl, 0.25% peptone, 1.7% agar, 5 mg/ml cholesterol, 1 mM $CaCl_2$, 1 mM $MgSO_4$, 25 mM KH_2PO_4 pH 6) seeded with OP50, a uracil-requiring mutant of *E. coli* (Brenner 1974). Nematodes were maintained at 15 °C or 20 °C.

Other strains used in this study include the following:

unc-53(n152)
unc-53(n166)
BC10729 *dpy-5/dpy-5*;Sex10729 [rCes T12G3.1-GFP+pCeh361]
BC03923 *unc-22(s7)let-52*(s2381)lev-1(x22)/Nt1(IV); +/Nt1(V)
BC10730 *dpy-5/dpy-5*;Sex10730 [rCes ZC404.8-GFP+pCeh361]
BC10129 *dpy-5/dpy-5*;Sex10129 [rCes B0336.6-GFP+pCeh361]

2.2 DNA Techniques:

2.2.1 Agarose gel electrophoresis

Agarose gels were prepared by adding agarose powder to TBE buffer (2.5 mM NaOH, 89.2 mM of Tris base, 89 mM of boric acid, 2 mM of ethylene diamine tetraacetic acid [EDTA, disodium salt]). The solution was heated to boiling to dissolve the agarose completely. On cooling the solution to 50 – 60 °C, ethidium bromide was added to a final concentration of 1.15 mM. Concentrated loading dye (bromophenol blue) was added to the DNA sample to a final concentration of 0.025%.

2.2.2 Restriction analysis and DNA extraction from agarose gels

DNA fragments were generated by restriction digest with a variety of restriction enzymes obtained from New England Biolabs, Amersham Pharmacia and GIBCO/BRL.

Digested DNA samples were analyzed on agarose gels and fragments of interest were eluted using the Qiaex II Kit (Qiagen, Inc.).

2.2.3 Ligation of endonuclease digested DNA fragments

For cloning purposes, endonuclease digested DNA was ligated using T4 DNA ligase (Amersham Pharmacia). Ligations were performed based on standard protocols (Ausubel et al., 1993).

2.2.4 Transformation of DNA into Escherichia coli

Purified plasmids and ligation reactions were transformed into competent *E. coli* Sure cells (Invitrogen) according to the manufacturer's instructions or into $CaCl_2$ treated competent cells according to standard protocols (Ausubel et al., 1993). Plasmid DNA extracted from yeast cells was transformed into competent *E. coli* DH5α cells by electroporation.

2.2.5 Transformation of plasmid DNA into Saccharomyces cerevisiae

S. cerevisiae cells were grown overnight at 30 °C in 25 ml of YEPD (1% Bacto yeast extract, 2% bacto peptone, 2% glucose). 50 ml of fresh YEPD was inoculated to an OD_{600} of 0.05 – 0.1 using the overnight turbid culture and grown for about 6 hrs. The cells were pelleted at 3000 rpm for 5 minutes and re-suspended in 50 ml of sterile water. After the second wash the supernatant was discarded and the pelleted cells re-suspended in 1ml of 100 mM lithium acetate (LiAc). These cells were pelleted at 5000 rpm using a microfuge and re-suspended in 400-500 µl of 100 mM LiAc. The cells were divided into 50 µl aliquots in fresh microfuge tubes and spun down at 5000 rpm. The supernatant was discarded and the undisturbed pellet of cells was layered on top in the following order: 240 µl of 50% polyethylene glycol (PEG), 36 µl of 1M LiAc, 5 µl of sheared salmon sperm DNA (10 mg/ml), 4 µg of plasmid DNA and 74 µl of water. Each reaction tube was vortexed for 30 seconds and incubated for 1hr at 30 °C. After this incubation period, the cells were vortexed again for 10 seconds and incubated for 30 minutes at 42 °C. The

cells were pelleted and re-suspended in 100 µl of water and spread on the appropriate dropout media.

2.2.6 Plasmid preparations from E. coli

DNA plasmid preparations were performed using the alkaline lysis method (Birnboim and Doly 1979). Plasmids required for DNA sequencing and other processes requiring high purity DNA was performed using the Qiagen kits and following manufacturer's instructions (Qiagen Inc.)

2.2.7 Plasmid extraction from Yeast and transformation of E. coli

Plasmid extraction from yeast cells was performed as previously described (Hoffman and Winston 1987), with the following modification; an extra precipitation step with 95% and 70% ethanol was included before transforming competent *E. coli* cells with yeast DNA. Electrocompetent *E. coli* cells were prepared by growing *E. coli* DH5α cells to log-phase stage, washing twice with and re-suspending in a 10% glycerol mixture. Electroporation was performed following manufacturer's directives, using 1 mm cuvettes (VWR) and instrument settings of 2.5 kV, capacitance of 2 µF and resistance of 4 kΩ.

2.2.8 Cloning of the UNC-53 bait plasmids

pVA200: A *Nde*I-*Nco*I fragment of the *unc-53* cDNA (nucleotides 64 – 480 which corresponds to amino acids 1 to 139) was excised from pTB57 (Stringham et al., 2002) and subcloned into the *Nde*I-*Nco*I restriction endonuclease sites of the pAS2 DNA binding domain vector (Matchmaker) to generate construct pVA200. **pVA201:** A *Nde*I-*Sma*I fragment of the *unc-53* cDNA (nucleotides 64 – 2350) which corresponds to amino acids 1 to 769) was subcloned into the *Nde*I-*Sma*I restriction sites of pAS2 to generate construct pVA201. **pVA202:** A *Nde*I-*Nco*I fragment of the *unc-53* cDNA (nucleotides 64 – 5060) which corresponds to amino acids 1 to 1583 was subcloned into the *Nde*I-*Nco*I restriction endonuclease sites of pAS2 to make construct pVA 202.

2.2.9 Polymerase chain reaction (PCR)

PCR was performed using the RoboCycler Gradient 40 (Stratagene). Plasmid inserts were amplified in a total reaction volume of 50 µl. The mixture included the following: PCR buffer (10 mM of Tris pH 8.3, 50 mM of KCL), 10 mM each of the 5' and 3' primers (see appendix for primer sequences), 5 mM of $MgCl_2$, 0.4 mM of deoxynucleotide triphosphates (dNTP), 4 units of Taq polymerase and 15 ng of template DNA. Reactions were executed at 96 °C for 1 minute for 1 cycle, followed by 30 cycles of 96 °C for 90 seconds, 56 °C for 30 seconds, 68 °C for 2 minutes and 1 cycle at 68 °C for 10 minutes. Annealing temperatures were determined based on the melting points of the individual primers, which were between 52 and 68 °C.

2.2.10 DNA Sequencing

DNA samples of positive candidates were prepared using the Qiagen Miniprep Kit and re-suspended in sterile deionized water. Sequencing reactions were performed by technicians at the Nucleic Acid Protein Services (NAPS) Unit at the University of British Columbia, Vancouver, B.C. Canada. DNA samples were supplied at a concentration of 100 ng/ul and PCR products at a concentration of 25 ng/ul. The primers (see appendix for sequence of primers and cDNAs) were supplied at a concentration of 3.2 pmol/ul.

2.3 Protein and Immunological Techniques

2.3.1 Generation of an anti-UNC-53 polyclonal antibody, PAB-UNC-53N

A *Sac*I-*Nco*I fragment of the *unc-53* cDNA, nucleotides 64 to 480 that corresponds to amino acids 1 to 139 was subcloned into the prokaryotic expression vector pRSET (Amersham Pharmacia) to generate pTB63 and expressed in *E. coli* BL21 cells for protein expression. The corresponding HIS Tag recombinant fusion protein was purified over a nickel based 'ProBond' column according to manufacturer's protocols (Invitrogen). Purified protein was emulsified in Titre Max and injected into a female New Zealand white rabbit at the Simon Fraser Animal Care Facility. The antiserum was shown to be active at titers of 1:30,000 on western blots of extracts from bacteria expressing the

purified recombinant fusion protein. To remove unwanted antibodies against contaminating bacterial proteins, an acetone powder was made from a bacterial culture of *E. coli* BL21 cells expressing the vector alone and this was used to adsorp the polyclonal antiserum (Epstein, H.F. et al., 1995).

2.3.2 Preparation of protein extract from S. cerevisiae

Yeast cells were grown to an OD_{600} of 0.5 harvested by centrifugation at 3000 rpm for 5 minutes and re-suspended in 1ml of sterile deionized water. The samples were transferred to sterile microfuge tubes and centrifuged at 10,000 rpm for 5minutes. The pellet was boiled for 2 minutes, re-suspended in 100 ml of Laemmli buffer, and sterilized acid washed glass beads were added just below the level of the meniscus in the tube. The sample was vortexed vigorously for 30 seconds and then boiled for 3 minutes. This was repeated a total of three times. The sample was centrifuged at 13,000 rpm for 1 minute and the supernatant transferred to a fresh microfuge tube and used immediately or stored at –20 °C.

2.3.3 Determination of protein concentration

Protein concentration was determined using a Bradford assay according to manufacturer's protocol (Invitrogen).

2.3.4 Preparation of protein extract from C. elegans

Wild type N2 Bristol strain worms were harvested using 0.14 M NaCl or M9 solutions and washed several times to remove bacterial and other dead debris. 30 µl of packed worms was made up to 100 µl with sterile distilled water, 2X SDS sample buffer (0.5 M Tris.Cl pH 6.8, 20 % glycerol, 4 % SDS, 250 mM 2-Mercaptoethanol, 1 mg bromophenol blue), and boiled for 5 minutes. The sample was centrifuged at 13, 000 rpm for 5 minutes and the supernatant was transferred to a fresh tube and used immediately or stored at –20 °C.

2.3.5 *SDS-polyacrylamide gel electrophoresis*

Proteins were separated based on size using the denaturing (SDS) discontinuous gel electrophoresis method (Laemmli 1970). Protein samples were resolved on single (12 %) or gradient (5 - 25 %) polyacrylamide gels, which were set using the Mini Protean II Cell system and manufacturer's instructions (BioRad). Samples were re-suspended in 2X SDS sample buffer (0.5 M Tris.Cl pH 6.8, 20 % glycerol, 4 % SDS, 250 mM 2-Mercaptoethanol, 1 mg bromophenol blue) and boiled at 95 °C for 5 minutes before loading into the gel lanes.

2.3.6 *Western blot analysis*

Protein samples were electrophoresed on a SDS gel along with a pre-stained molecular weight marker. The gel was soaked in transfer buffer (15% methanol, 25 mM Tris base pH 8, 192 mM glycine, 0.05% SDS) with gentle agitation for 15 minutes before transfer to a PVDF membrane (BioRad). The PVDF membrane (BioRad) was cut to the size of the gel (7 cm [L] X 8 cm [W]), soaked in methanol, rinsed in deionised water for 5 minutes and then agitated in transfer buffer for 15 minutes. An assembly of filter pads, filter paper that had been soaked in transfer buffer was used to form a sandwich around the gel and PVDF membrane. The gel was placed on the negative anode side and the PVDF membrane in contact with the gel on the positive cathode side of the cassette. The cassette holding the sandwich was placed in the Mini Trans – Blot Cell system (BioRad), where the proteins were transferred at 250 mA for 120 minutes.

Following transfer, the sandwich was disassembled and the PVDF membrane rinsed in transfer buffer for 5 minutes and incubated with gentle agitation in blocking buffer which contains TBST (20 mM Tris.Cl pH 7.5, 140 mM NaCl, 0.1% Tween20) and 2% milk for 30 minutes. The membrane was washed three times in TBST with gentle agitation. This wash buffer was then replaced with blocking buffer containing the primary antibody at an appropriate dilution with gentle agitation for 120 minutes. The wash step was repeated again with gentle agitation and replaced with the blocking buffer containing the Anti Ig G secondary antibody (1:30,000) for 90 minutes. The membrane was washed

again in TBST and then developed using the BCIP-NBT color development mixture according to manufacturer's protocols (Amersham Pharmacia).

2.4 Immunohistochemistry

In order to characterize the expression pattern of the UNC-53 protein in N2 and *unc-53* (*n152*), *unc-53* (*n166*), immunohistochemistry was employed. Nematodes were harvested using 0.14 M NaCl or M9 solutions, rinsed several times with the same buffer to remove bacteria and debris or bleached (20% bleach, 0.5M NaOH) to obtain eggs. The freeze-crack method was used to permeabilize eggs while the Ruvkin fixation method was used to permeabilize whole animals (Epstein, H.F. et al., 1995). UNC-53 was detected using the polyclonal antibody raised against the N-terminal calponin region corresponding to amino acids 1 to 139 in this study. Fixed specimens were incubated overnight in primary antibody solution (1:100 dilution of PAB UNC-53N;), washed in AbB buffer (AbA [1X PBS, 1% bovine serum albumin, 0.5% Triton, 0.05% sodium azide, 1 mM EDTA], 0.1% BSA) with several changes of the buffer over a 15 minutes period and incubated overnight with the secondary antibody solution (1:30 dilution of Cy3-conjugated sheep anti-rabbit; Sigma). The nematodes were washed again in AbB buffer, mounted in a commercially available mounting medium (Vectashield, Vector Laboratories) and viewed with a Zeiss compound microscope with rhodamine filters. GFP was viewed with a Leica compound microscope with GFP filters.

2.5 Maintenance and mating of *S. cerevisiae*

The Yeasts strains, Y187 (MATα *gal4 gal80 his3 trp1*-901 *ade2*-101 *ura3*-52 *leu2*-3,-112 *met*$^-$ URA3::GAL *lac*Z) and Y190 (MATa *gal4 gal80 his3 trp1*-901 *ade2*-101 *ura3*-52 *leu2*-3,-112 + URA3::GAL *lac*Z LYS2::GAL(UAS) HIS3 *cyh*^{r}) were maintained on YEPD plates (10 mg/ml Bacto yeast extract, 20 mg/ml Bacto peptone, 20 mg/ml glucose and 2% agarose) or cultured in liquid media (10 mg/ml Bacto yeast extract, 20 mg/ml Bacto peptone, 20 mg/ml glucose) at 30 °C. In addition yeast was

grown on drop out (DO) media for selection purposes. DO media contained 20 mg/ml glucose, yeast nitrogen base without amino acids (Difco), 2% agarose and the appropriate synthetic complete (SC) mixture (note: the following concentrations stated in brackets are in respect to the synthetic complete mixture and not the amino acids that are excluded from the mixture) of amino acids lacking leucine (-Leu) (0.7 mg/ml), lacking tryptophan (-Trp) (0.7 mg/ml), lacking leucine and tryptophan (-Leu-Trp) (0.6 mg/ml), or lacking leucine, tryptophan and histidine (-Leu-Trp-His) (0.6 mg/ml) (Bio101). Yeast transformed with a pACT::*cDNA* plasmid or pSE1111 control plasmid was selected for on DO media lacking leucine (SC –Leu) and DO media lacking tryptophan (SC –Trp) was used to select for yeast transformed with pVA200 (pAS2::*unc-53*) or pSE1112 control plasmid. Doubly transformed or mated yeast cells were selected for on DO media lacking both tryptophan and leucine (Sc–Trp-Leu). Triple DO media lacking leucine, tryptophan and histidine (Sc-Leu-Trp-His), as well as supplemented with 25 mM 3-aminotriazole (3AT) was used to select for doubly transformed yeast expressing potentially interacting proteins. 3AT is a competitive inhibitor of the yeast *HIS3* protein (His3p); consequently it is used to inhibit the low levels of His3p expressed in some reporter strains (Durfee et al., 1993).

After selecting a positive colony that was potentially expressing interacting proteins, the colony was cured of the bait plasmid by growing the colonies in Sc–Leu media and spreading an appropriate dilution of cells on Sc–Leu plates that had been supplemented with 2.5 μg/ml of cycloheximide. The bait plasmid, pAS2 carries a drug recessive gene, *cyh^{r}2*, which makes yeast cells harboring the bait plasmid unable to survive on cycloheximide containing plates. Only yeast cells that have lost the bait plasmid would survive on the cycloheximide plates, leaving only cells containing the prey plasmid (pACT::*cDNA)* alone. Isolating the prey plasmid in this manner allows for further confirmatory tests to be performed and also permits sequencing to determine the molecular identity of the cDNA insert.

Yeast mating experiments with unrelated baits such as pAS1::*snf1* and pAS1::*cdk2* were also performed to ensure that the candidate interactor was not promiscuous in its associations. These unrelated baits were transformed into yeast of the

opposite mating strain, Y187 (MATα) and mated with Y190 (MATa) yeast cells carrying only the prey plasmid pACT::*cDNA*. Using a sterile flat-ended toothpick, a small patch of Y190 cells carrying only the prey plasmid was spread on YEPD plates, followed by a corresponding sized patch of Y187 cells harboring the unrelated plasmids laid overtop to ensure close contact and mating of the yeast cells. This patch was allowed to grow at 30 °C for 2 days and cells from the mated patch were then restreaked on Sc–Leu-Trp double DO plates. Streaking the colonies on the double DO plates verifies successful mating of the yeast cells because only cells that have both prey and bait plasmid would grow. The cells were grown for another 3 days on Sc–Leu-Trp double DO plates before restreaking on Sc–Leu-His-Trp triple DO plates supplemented with 25 mM 3-aminotriazole to select for cells expressing interacting proteins.

2.6 X-gal filter lift assay

A Schleicher and Schuell BA85 45 μm circular nitrocellulose filter was used to obtain an imprint of yeast colonies growing on triple DO plates supplemented with 3AT. After transfer the filter was placed on a liquid nitrogen float to permeabilize the yeast cells and subsequently incubated in a petri-dish containing 3MM chromatography paper (VWR) soaked with 0.30ml/square inch of Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCL, 1 mM $MgSO_4$, 40 mM 2-mercaptoethanol, pH 7.0) containing 1 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactosidase in dimethylformamide at 30 °C for 1 hour to overnight for development of color. Putative positives were picked from the original plate and retested a second time.

2.7 Computer analysis

The BLAST (http://www.ncbi.nlm.nih.gov/BLAST/) program was used to determine the molecular identity of the UNC-53 interacting proteins encoded by the *pACT::cDNA* library plasmid. The Simple Modular Architecture Research Tool (SMART) was used to search for conserved or unique protein domains in the sequences (http://smart.embl-heidelberg.de/) (Schultz et al., 1998; Schultz et al., 2000). Wormbase (www.wormbase.org) was used to find homologues of UNC-53 interacting proteins

(http://www.wormbase.org/db/gene/gene?). CLUSTALX (ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/) was used for alignments and GENEDOC (http://www.psc.edu/biomed/genedoc/) was used to display alignments.

3. RESULTS:

3.1 Novelty of the UNC-53 N-terminus as 'bait' in yeast two hybrid screen

The *unc-53* locus is not only large, containing 23 exons spanning 31kb of genomic DNA, but is also complex, punctuated with several SL1 trans-splice sites and the presence of multiple promoters, which together give credence to the presence of multiple protein isoforms *in vivo* (Stringham et al., 2002). Adding to its complexity is the presence of several motifs shared with other proteins that have been implicated in diverse aspects of signaling. Located in the amino terminus of the UNC-53 protein is a single calponin homology (CH) domain (Figure 1b), a motif found in several cytoskeletal and signal transduction molecules and has been implicated in the regulation of cell shape dynamics (Stradal et al., 1998). The CH signature is present in a variety of actin-binding proteins, including α-actinin and dystrophin, proteins known to crosslink actin filaments into bundles and networks (Van Troys et al., 1999). Situated next to the CH domain in the UNC-53 sequence is a putative actin-binding site of the LKK consensus (Stringham et al., 2002). While α- actinin and dystrophin each contain 2 CH domains, it has been shown that the single CH domain in calponin is neither necessary nor sufficient for actin binding. Instead actin binding is presumed to be facilitated through additional sites and modulated by the CH domain (Gimona and Mital, 1998; Stradal et al., 1998).

Previous work had shown that genomic clones of *unc-53* devoid of the CH domain had the ability to partially rescue the loss of function phenotype of strong *unc-53* alleles. Secondly, this domain is absent in polypeptides translated from UNC-53 messages transcribed from two internal promoters (Figure 1; Stringham et al., 2002). Thirdly, a truncated form of UNC-53 lacking the CH domain is able to co-sediment F-actin *in vitro* while the CH domain alone does not (Stringham, unpublished results). Together these findings prompted the question; what interaction, if any, does the CH domain of UNC-53 mediate? To this end, a yeast two-hybrid screen was employed to identify molecules that interact with the N-terminal portion of UNC-53.

3.2 Yeast two hybrid baits

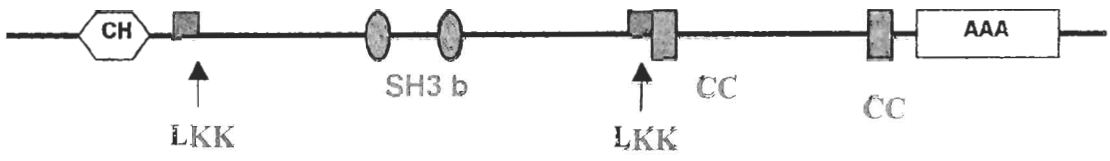
Three UNC-53 bait constructs containing partial cDNAs were generated as shown in Figure 2. They are as follows: pVA200 (pAS2::*unc-53*) was generated by cloning nucleotides 64 – 480 (which corresponds to amino acids 1 – 139) of the *unc-53* cDNA sequence into the GAL4 DNA binding domain vector, pAS2. This construct encodes a fusion protein product consisting of the N-terminal portion of UNC-53, which includes a calponin homology domain (amino acids 11 – 109) and one LKK motif (114 – 133), and the GAL4 DNA binding domain. This construct will be referred to as UNC-53N (UNC-53 N-terminus) hereafter for simplicity (Figure 2).

pVA201 (pAS2::*unc-53*) was generated by cloning nucleotides 64 – 2350 (which corresponds to amino acids 1 – 762) of the *unc-53* cDNA sequence into the GAL4 DNA binding domain vector, pAS2. This construct is a fusion protein product of the first 762 amino acids of UNC-53, which includes a CH domain (amino acids 11-109), one LKK motifs (114 – 133), and two polyproline rich sequences (487 – 495; 537 – 545) and the GAL4 DNA binding domain. This construct will be referred to as UNC-53 (1 – 762) (Figure 2).

pVA202 (pAS2::*unc-53*) was generated by cloning nucleotides 64 – 5060 (which corresponds to amino acids 1 – 1583) of the *unc-53* cDNA sequence into the GAL4 DNA binding domain vector, pAS2. This construct is a fusion protein product of the full length UNC-53 protein, which includes a CH homology domain (amino acids 11-109), two LKK motifs (114 – 133 and 1097 - 1116), two SH3 binding motifs (SH3b; 487 – 495; 537 – 545) two coiled-coil domains (890 – 923 and 1078 – 1113), an AAA (ATPases associated with diverse cellular activities) domain (1292 – 1425) that contains a potential nucleotide binding motif (1300 – 1307) and the GAL4 DNA binding domain. This construct will be referred to as UNC-53 because it encodes the full-length protein (Figure 2).

Figure 2: pAS2::*unc-53* GAL4 DBD fusions.

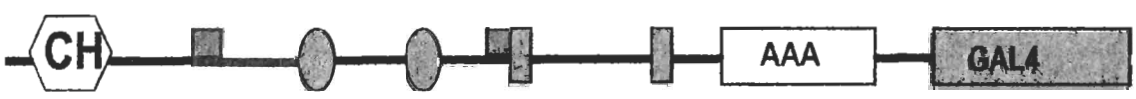
Three bait constructs were generated using different portions of the *unc-53* cDNA sequence. UNC-53N (pAS2::*unc-53*) was generated by cloning nucleotides 64 – 480 (which corresponds to amino acids 1 – 139) of the *unc-53* cDNA sequence into the GAL4 DNA binding domain vector, pAS2. This construct encodes a fusion protein product consisting of the N-terminal portion of UNC-53, which includes a CH homology domain (amino acids 11 – 109) and one LKK motif (114 – 133) and the GAL4 DNA binding domain. UNC-53(1-762) (pAS2::*unc-53*) was generated by cloning nucleotides 64 – 2350 (which corresponds to amino acids 1 – 762) of the *unc-53* cDNA sequence into the GAL4 DNA binding domain vector, pAS2. This construct is a fusion protein product of the first 762 amino acids of UNC-53, which includes a CH homology domain (amino acids 11-109), one LKK motifs (114 – 133), and two SH3 binding motifs (SH3b; 487 – 495; 537 – 545) and the GAL4 DNA binding domain. UNC-53 (pAS2::*unc-53*) was generated by cloning nucleotides 64 – 5060, which corresponds to the complete *unc-53* cDNA sequence (1583 amino acids), into the GAL4 DNA binding domain vector, pAS2. This construct is a fusion protein product of the full length UNC-53 protein, which includes a CH homology domain (amino acids 11-109), two LKK motifs (114 – 133 and 1097 - 1116), two SH3 binding motifs (SH3b; 487 – 495; 537 – 545) two coiled-coil domains (890 – 923 and 1078 – 1113), an AAA (ATPases associated with diverse cellular activities domain) (1292 – 1425) that contains a potential nucleotide binding motif (1300 – 1307) and the GAL4 DNA binding domain.



UNC-53N; pTB63 (420bp)

UNC-53(1-762) (2.3Kb)

UNC-53 (5Kb)



3.3 Yeast two-hybrid scheme for identifying interacting proteins

UNC-53N was used as "bait," to screen a "prey" library of *C. elegans* cDNAs fused to the GAL4 activation domain (AD) (Figure 3). Reporter genes under the transcriptional control of the GAL4 transcription factor include the growth selection marker, *his3*, and the β-galactosidase marker, *lacZ*. This screen was performed in yeast Y190 cells and interactors were identified by their ability to grow on media lacking histidine and their ability to activate the β-galactosidase reporter gene. Figure 4 outlines the scheme for the identification and verification of clones that interact with UNC-53N (Figure 4).

Figure 3: Outline of the two-hybrid system.

Fields et al., 1994. (Adapted by permission of Stanley Fields). (**a**) A fusion protein consisting of a DNA-binding domain (filled circle) and UNC-53N is generated. The fusion binds DNA but is unable to activate transcription in the absence of an activation domain.(**b**) A fusion protein consisting of the Activation domain (open circle) and the *C. elegans* cDNA library is generated.The prey fusion is insufficient to activate transcription since it does not bind to the upstream activating sequence (UAS). (**c**) Interaction between UNC-53N and a DNA library protein activates transcription of the reporter genes.

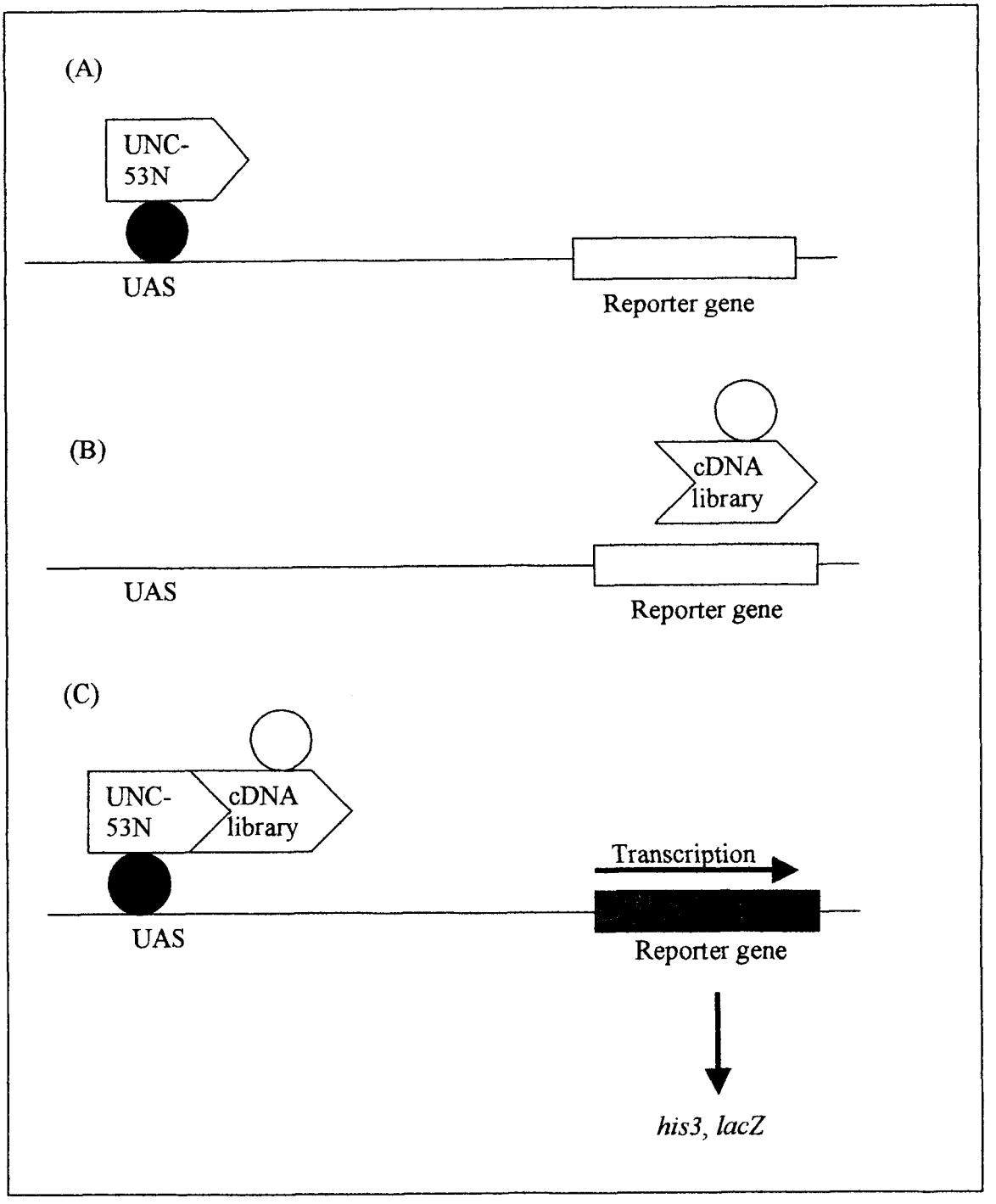

(A)
UNC-
53N
UAS
Reporter gene
(B)
cDNA
library
UAS
Reporter gene
(C)
UNC-
53N
cDNA
library
Transcription
UAS
Reporter gene
his3, lacZ

Figure 4: Yeast two hybrid protein interaction cloning outline for identifying interacting proteins.

Cloning of *unc-53* sequences into the bait plasmid (pAS2) and transformation into yeast Y190

↓

Transformation of library into Y190 cells containing a fusion of the bait plasmid and *unc-53* sequences

↓

Recovery and plating on triple drop out media (-His-Leu-Trp) + 25 mM 3-Aminotriazole

↓

X-GAL filter assay for Beta-galactosidase activity

↓

Blue colonies are selected, streaked and retested for beta-galactosidase activity with X-GAL.

↓

Colonies are seeded on cycloheximide media to select for loss of the bait plasmid pAS-CHY (test for loss of TRP1 by plating on Sc-Trp-Leu or Sc-Trp)

↓

$Trp^{-}Leu^{+}$ colonies mated to pAS-UNC-53/Y187 strains to retest interactions. Colonies were selected by growth on Sc-Leu-Trp. Diploids screened for β-galactosidase activity with X-GAL

↓

Colonies failing to activate non-specific fusions in diploids are chosen for further analysis. Plasmid DNAs are recovered in *E. coli* and retested

↓

Secondary assays to confirm interactions

3.4 Yeast two hybrid controls

The empty vectors without *unc-53* sequence (pAS2) or cDNA inserts, (pACT) and the bait constructs, UNC-53N, UNC-53(1-762), UNC-53, were transformed into Y190 yeast cells. The growth properties of these transformed Y190 cells on different media lacking specific amino acids was verified. For example the GAL4 DBD plasmid vector pAS2 encodes the biosynthetic marker tryptophan, which allows yeast strains carrying this plasmid to survive on media lacking tryptophan. Likewise the GAL4 AD plasmid vector pACT, carries the biosynthetic marker leucine and strains carrying this plasmid should survive on media lacking leucine. Both pAS2 and pACT survive on media lacking tryptophan and leucine respectively as shown in Table1. The bait vectors, UNC-53N, UNC-53(1-762), UNC-53(1-1583) were subcloned into the GAL4 DBD vector and selected for on media lacking tryptophan. The results obtained showed that the bait vectors grew as expected (Table 1).

Secondly, the transformed Y190 cells were checked to ensure that they did not activate the GAL4 reporter in isolation before the library screen was performed. Transformed yeast cells were tested on media lacking histidine with differing concentrations of 3-aminotriazole (3-AT). 3AT is a competitive inhibitor of the yeast HIS3 protein and is used in this system to inhibit low levels of HIS3 that may be expressed in a leaky manner (Durfee et al., 1993). Table 1 shows that the DBD bait vectors with or without *unc-53* sequences do not survive on media lacking histidine. DBD bait vectors with or without *unc-53* sequences were also tested for their activity towards the beta-galactosidase reporter gene. Y190 cells only and Y190 cells transformed with the DBD bait vectors did not show any blue color change when assayed for beta-galactosidase activity as shown in Table 1.

A positive control test was also performed to ensure that reporter gene activation could be detected when known interacting DBD bait and AD prey fusions were brought together. The known interactors, SNF4 (encoded by plasmid pSE1111) and SNF1 (encoded by pSE1112) were grown on DO media lacking tryptophan and leucine to select for the growth markers on the bait and prey plasmids respectively, and on triple DO media (lacking tryptophan, leucine and histidine) to verify interaction as determined by

Table 1: Yeast Two-Hybrid Control Tests.

The empty bait and prey vectors as well as single fusion proteins were tested to ensure that there was no self-activation of the reporter genes. Other combinations with UNC-53N were also tested.

PLASMIDS USED FOR TRANSFORMATION OF Y190	GROWTH ON -TRP	GROWTH ON -LEU	GROWTH ON -TRP -LEU	GROWTH ON -TRP-LEU -HIS + 25 MM 3AT	POSITIVE X-GAL ASSAY[1]
Y190	–	–	–	–	–
pAS2	+	–	–	–	–
pACT	–	+	–	–	–
pAS2 + pACT	n.d.	n.d.	+	–	–
UNC-53N	+	–	–	–	–
UNC-53(1–762)	+	–	–	–	–
UNC-53	+	–	–	–	–
pSE1111	–	+	–	–	–
pSE1112	+	–	–	–	–
pSE1111 + pSE1112	n.d	n.d.	+	+	+
UNC-53N + pSE1111	n.d	n.d.	+	–	–
UNC-53(1-762) + pSE1111	n.d	n.d.	+	–	–
UNC-53+ pSE1111	n.d	n.d.	+	–	–

+ = growth / positive color change

- = no growth / no color change

pSE1111 = pACT::*snf4*

pSE1112 = pAS1::*snf1*

n.d. = no data

[1]x-gal assays were performed on single (Sc-Trp or Sc-Leu) or double (Sc–Trp-Leu) drop out media plates where growth was optimal.

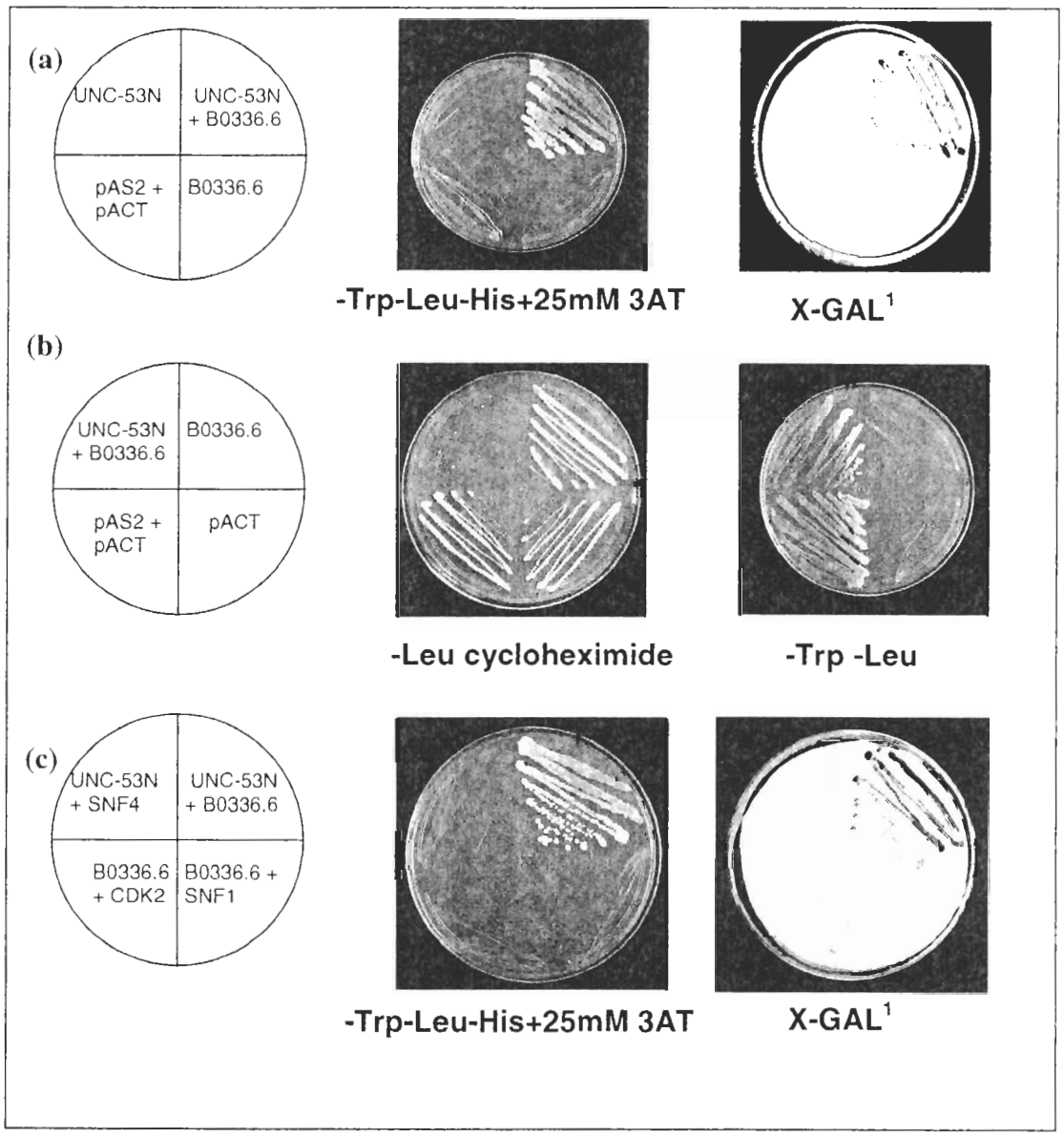
the ability to survive on histidine and turn blue in the presence of X-gal. Y190 cells, co-transformed with SNF4 and SNF1, were shown to grow on both double DO media (lacking tryptophan and leucine) and triple DO media (lacking tryptophan, leucine and histidine) containing 3AT as shown in Table 1. Likewise as a negative control, the empty bait (pAS2) and prey (pACT) plasmids (without *unc-53* sequences) were co-transformed into Y190 cells and tested for beta-galactosidase activity. Doubly transformed cells with the empty bait and prey plasmids survive on double drop out media by virtue of their selection markers but they fail to grow on triple drop out and do not turn blue in the presence of X-GAL (Table 1 and Figure 5, panel a, middle and right). The bait plasmids UNC-53N, UNC-53(1-762) and UNC-53 were also tested with the SNF4 control prey plasmid to ensure that there was no activation of the GAL4 reporter genes. The results obtained showed that the co-transformed cells grew on double DO media but not on triple DO media containing 3AT (Table 1). This verifies that the bait fusion proteins do not interact non-specifically with unrelated preys.

3.5 Yeast two-hybrid screen

Using the amino terminal portion of the *unc-53* cDNA sequence, UNC-53N as bait, and the *C. elegans* cDNA library as prey (the cDNA Library was a gift from Bob Barstead, Madison, USA and a plasmid library preparation was performed by Don Jones, UBC, Vancouver, Canada), approximately 778,802 double transformed clones were screened. Bait transformed Y190 cells were grown in liquid media lacking tryptophan and then transformed with the *C. elegans* cDNA library. Colonies that grew within 7 days on triple DO media, containing a limiting amount of 3-Aminotriazole (25 mM) were replica plated and assayed for X-gal activity (Figure 5a). Growth of colonies on double DO plates lacking tryptophan and leucine was also assayed to determine the efficiency of the transformations. Colonies that turned blue were re-streaked on triple DO media and retested for the beta-galactosidase activity in the presence of X-gal. This screen yielded 190 putative interactors based on their ability to grow on triple drop out plates (Sc-Trp-Leu-His) containing a limiting amount of 3AT (25 mM) and their ability to activate the beta-galactosidase reporter gene, which was observed by a blue color change when colonies where permeabilized in the presence of X-gal.

Figure 5: Yeast two hybrid assays.

Growth of transformed Y190 cells on various drop out media and X-gal assay[1]. **(a)** Only the bait, UNC-53N and interacting prey, B0336.6, survive on triple drop out media containing 3AT and also turn blue with X-GAL. Bait and prey plasmids singly do not grow on triple drop out media and they do not activate beta-galactosidase. Empty bait and prey vector, (pAS2+pACT) survive on double drop out media by virtue of the growth markers but do not survive on triple drop out media and do not activate beta-galactosidase. **(b)** -Leu cycloheximide media selects for the loss of the bait plasmid, pAS2 or UNC-53N. Colonies still containing both the bait and prey plasmid (i.e. leu^+trp^+) should not survive on –leu cycloheximide. Colonies that have lost the bait plasmid (leu^+trp^-), do not grow on double drop out media (-trp-leu). Colonies that have not lost the bait plasmid still survive on double drop out media. Panel (b) shows that yeast cells containing the empty bait and prey plasmids (pAS2+pACT) still survive on –leu cycloheximide, sometimes colonies that have not lost the bait plasmid survive on this media, and loss of bait plasmid must be verified by restreaking on –trp-leu drop out plates. **(c)** Identified interactors do not interact with bait fusions of unrelated function. In this case, B0336.6 is tested with SNF1 and CDK2. The bait plasmid, UNC-53N was also tested with a prey fusion of unrelated function, SNF4. Growth is observed on double drop out plates but not on triple drop out plates and these colonies do not activate beta-galactosidase.



[1] For the X-GAL assay, a replica was taken from cells growing on double drop out media (-Trp-Leu), permeabilized and then tested with X-GAL.

Only 24 however reached the final stage of sequencing to determine the molecular identity of the candidate. Sequence reads were not obtained for 6 of these 24 candidates leaving only 18 for which sequence information was determined. 1 candidate was sequenced with no significant similarity to information in the *C. elegans* database. Other colonies that initially met the selection criteria were frozen down but a significantly high proportion failed to grow when re-streaked on triple drop out media containing 25 mM 3AT. The reasons why the colonies failed to grow remain varied and unclear. Three individual colonies were picked each time for re-streaking; it is possible that a mutation had occurred in the individual yeast cells that prevented growth on the triple drop out plates. It is also plausible that taking a swab of cells would have ensured survival of some cells. Colonies were first seeded onto YPD to verify growth and then subsequently streaked on triple drop out media containing 25 mM 3AT. Yeast strains containing only the prey plasmid that still had the ability to grow on media lacking leucine and histidine (containing 25 mM 3AT), after loosing the bait plasmids were eliminated as false positives. Mating assays also revealed that the interaction between the bait fusion and prey fusion proteins was not reconstituted for some of the candidates.

Table 2: Summary of Yeast two-hybrid screen results.

DESCRIPTION	Numbers
Candidate determined	18
Could not be determined	6
Sequenced but no similarity to information in database	1
Inability to loose bait plasmid	3
Eliminated - No growth on 3AT after mating	8
Eliminated - Weak growth on 3AT after mating	5
Eliminated (False Positive) – Activates reporter gene alone strongly	8
Eliminated (False Positive) – Activates reporter gene alone weakly	9
Eliminated - No growth on 3AT on second attempt	75
Eliminated - Weak growth on 3AT on second attempt	55
Total	190

3.6 Criteria for detecting and eliminating false positives

The authenticity of candidates was determined using four control tests. In order to ensure that the prey putative candidates were unique in their interaction to UNC-53N, the prey plasmid was isolated and tested for its ability to activate the His reporter gene alone and for its affinity towards other unrelated bait constructs. After losing the bait plasmid (UNC-53N), Y190 cells containing the single prey plasmid were streaked onto double DO media lacking leucine, histidine and containing 25 mM 3AT. The lack of leucine was used to select for the GAL4 pACT AD vector and a lack of histidine to confirm that the prey plasmid did not activate the His reporter alone. Y190 cells harboring plasmids that were able to survive on this media were eliminated as false positives (Table 2).

The prey plasmid (pACT::cDNA), was extracted from Y190 yeast cells that had lost the bait plasmid (i.e. UNC-53N). Y190 cells were cured of the bait plasmid by selecting for $leu^{+}trp^{-}$ colonies on media lacking leucine and containing cycloheximide (Figure 5, panel b, middle and right). The prey plasmid was extracted from the $leu^{+}trp^{-}$ population of yeast cells. This prey plasmid was then transformed into *E. coli* DH5α cells by electroporation. This plasmid containing the cDNA sequence could now be easily obtained by performing plasmid preparations. The bait (UNC-53N) and prey (pACT::cDNA) plasmid DNAs were then retransformed together into fresh Y190 cells to confirm that activation of the reporter genes still occurred (Table 3). Combinations of bait fusion protein (i.e. UNC-53N) and the putative interactors (pACT::cDNA) in Y190 cells for which interactions were not reconstituted were eliminated from further analysis. Y190 cells expressing interacting proteins were selected on triple drop out media supplemented with 25 mM 3AT and tested for beta-galactosidase activity. As shown in Table 3, interactions were reconstituted for the clones tested. Opposite mating strains containing either a bait (i.e. UNC-53N, UNC-53(1-762) or UNC-53) or prey fusion protein (pACT::cDNA) were mated to confirm that interaction was reconstituted as shown in Table 4 (also Figure 5, panel c, middle and right). All mating tests using UNC-53N as the bait plasmid showed interaction with the prey (Table 4).

Table 3: Yeast co-transformation controls and tests for interactions with UNC-53(1-762) and UNC-53.

UNC-53N interacting proteins interact with UNC-53N but not with UNC-53(1-762) and UNC-53.

Table 4: Yeast mating interactions.

UNC-53N interacting proteins interact with UNC-53N but not with UNC-53(1-762) and UNC-53.

cDNAs in pACT plasmid	Growth on –TRP-LEU			Growth on –TRP-LEU-HIS + 25 mM 3AT			Positive X-gal assay[1]
	UNC-53N	UNC-53(1-762)	UNC-53	UNC-53N	UNC-53(1-762)	UNC-53	
B0336.6	+	+	+	+	—	—	+
T12G3.1	+	+	+	+	—	—	+
ZC404.8	+	+	+	+	—	—	+
B0546.2	n.d.	n.d	n.d	n.d	n.d	n.d	n.d
F58A4.7a	n.d	n.d	n.d	n.d	n.d	n.d	n.d
F22E12.4	n.d	n.d	n.d	n.d	n.d	n.d	n.d

[Y190(MATa) + pACT::cDNA] X [Y187(MATα) + pAS2::*unc-53*(64 - 2350)]
or pAS2::*unc-53*(64 - 5060)]

cDNAs in pACT plasmid	Growth on –TRP-LEU			Growth on –TRP-LEU-HIS + 25 mM 3AT			Positive X-gal assay[1]
	UNC-53N	UNC-53(1-762)	UNC-53	UNC-53N	UNC-53(1-762)	UNC-53	
B0336.6	+	+	+	+	—	—	+
T12G3.1	+	+	+	+	—	—	+
ZC404.8	+	+	+	+	—	—	+
B0546.2	+	+	+	+	—	—	+
F58A4.7a	+	+	+	+	—	—	+
F22E12.4	+	+	+	+	—	—	+

[1]Positive X-gal results were only obtained with UNC-53N. X-gal assays were performed on double drop out plates where growth was evident.

To ensure that the UNC-53 fusion protein was necessary for the interaction, the empty bait vector, pAS2 (i.e. devoid of *unc-53* sequence) was retransformed together with one positive pACT::cDNA fusion protein into fresh Y190 cells. Transformed Y190 cells were selected by seeding onto double and 25 mM 3AT triple drop out media. Co-transformed cells grew on the double drop out (-leu-trp) media by virtue of the selection growth markers on the bait and prey plasmids. Doubly transformed cells did not grow on the triple drop out media (-leu-trp-his +25 mM 3AT) indicating that the His reporter gene was not activated. However some activation of the beta-galactosidase gene was observed probably due to the leakiness of the lacZ reporter. It has been reported that the bait plasmid (pAS2) alone can activate lacZ weakly, but this weak activation appears to be eliminated when genes are cloned into it (Elledge 1993). It has also been suggested that this activation may be due to sequences beyond the polylinker region which are of no consequence once cDNAs are cloned into it (Elledge 1993).

Since the remaining bait constructs [UNC-53(1-762) and UNC-53] contain the amino terminus like UNC-53N, it was reasoned that the proteins which interacted with UNC-53N would also was interact with UNC-53(1-762) and UNC-53. This was tested by co-transforming the bait plasmids, UNC-53(1-762) or UNC-53 with each of the putative interacting candidates. This analysis showed that these candidates did not interact with UNC-53(1-762) or UNC-53 (see Table 3). Interactions were also verified by mating Y190 (MATa) cells containing the prey plasmid only with yeast cells of the opposite mating strain Y187 (MATα), containing only the bait construct [i.e. UNC-53(1-762) or UNC-53]. In agreement with the co-transformation results, this analysis also showed that the identified candidates did not interact with UNC-53(1-762) or UNC-53 (Table 4). Mating tests were also performed with UNC-53N to confirm that interactions were reconstituted with the putative candidates. All interactions with UNC-53N were reconstituted (Table 4).

Y190 (MATa) cells containing only the prey plasmid (pACT::cDNA) were mated with yeast cells of the opposite mating strain Y187 (MATα), containing a bait fusion protein of unrelated function (pAS2::CDK2/SNF1) (Table 5). Diploid cells were selected on double DO media lacking tryptophan and leucine to ensure mating had occurred. Cells

from the double drop out plate were assayed for X-gal activity. Mated cells were then re-streaked onto triple DO media (lacking tryptophan, leucine and histidine) containing 25 mM 3AT to assess the His reporter activity (Table 5). Figure 5 (panel c, middle and right) shows the prey plasmid, B0336.6, mated with bait fusion proteins of unrelated function (pAS2::CDK2 or pAS2::SNF1). The results show that the diploid cells containing the bait and prey plasmids survive on double DO media (by virtue of the markers on the plasmids), but fail to grow on triple DO media, indicating an absence of interaction. Also a negative test was obtained for the beta-galactosidase assay.

Table 5: Yeast mating interactions with bait fusion proteins of unrelated function.

UNC-53N interacting proteins (pACT::cDNA) do not interact with GAL4 DBD::SNF1(or CDK2) fusion protein.

[Y190(MATa) + pACT::cDNA] X [Y187(MATα) + pAS2::*snf1/cdk2*]

cDNAs in pACT plasmid	Growth on –TRP-LEU		Growth on –TRP-LEU-HIS + 25 mM 3AT		Positive X-gal assay
	SNF1	CDK2	SNF1	CDK2	
B0336.6	+	+	–	–	–
T12G3.1	+	+	–	–	–
ZC404.8	+	+	–	–	–
B0546.2	+	+	–	–	–
F58A4.7a	+	+	–	–	–
F22E12.4	+	+	–	–	–

3.7 Molecular identity of Yeast two hybrid candidates

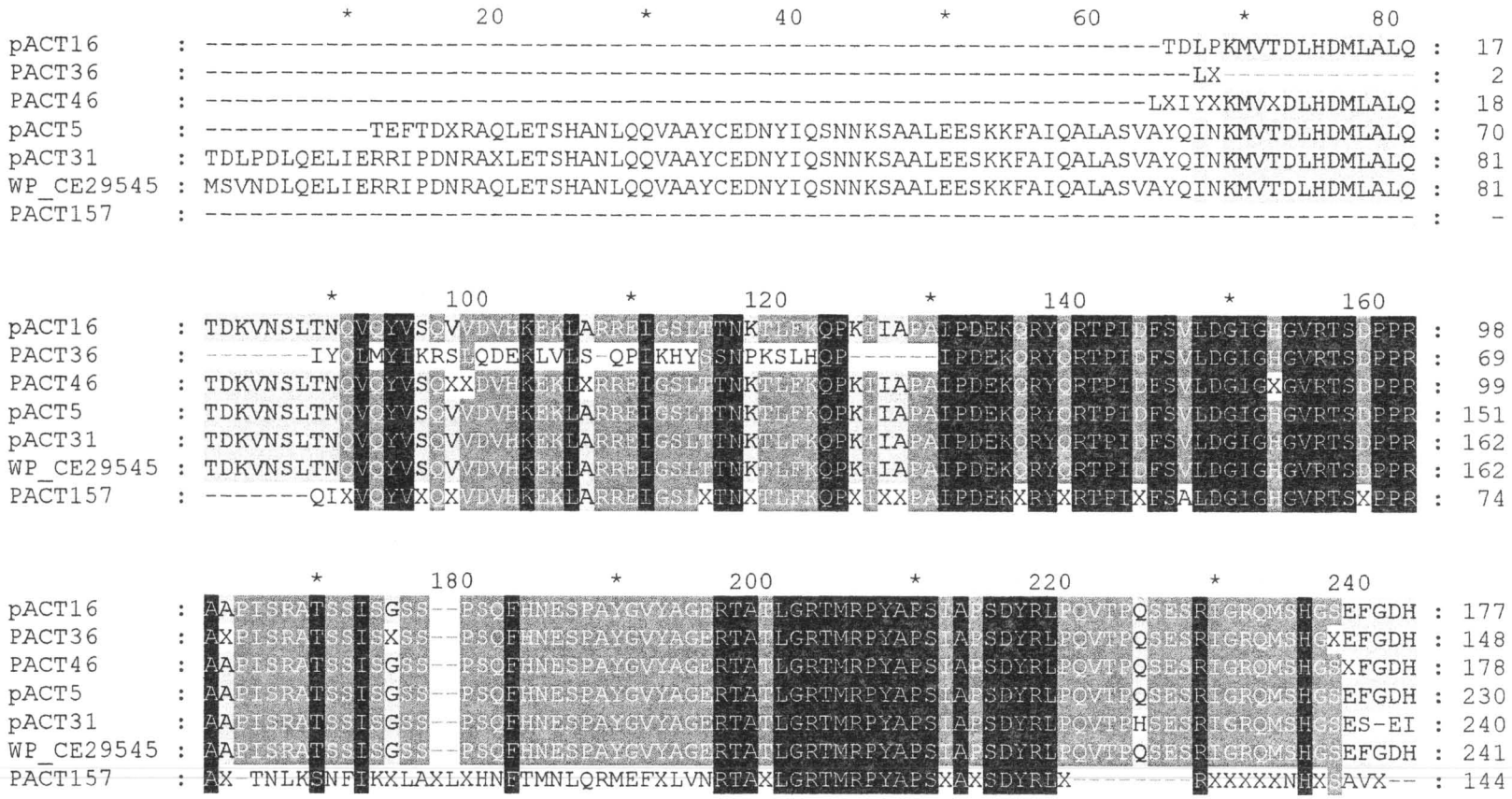
The 18 candidates whose sequences were determined divide into 6 classes of cDNAs which are shown in appendix 1. The molecular sequence in 6 of the prey plasmids (PACT::cDNA) corresponds to the *C. elegans* genomic locus B0336.6 and encodes a polypeptide of 469 amino acids for which there is no known function in the database. An alignment comparing the translated cDNA sequences of the seven clones from the yeast two hybrid screen with the *C. elegans* protein predicted by wormbase (WP_CE29545) is shown in Figure 6. While all of the cDNA clones were roughly 1.8 kb in size they clearly do not correspond to the identical cDNA clone as they differ in their amino termini. In addition, pACT157 is more highly divergent.

Four of the candidates map to the *C. elegans* gene T12G3.1, which codes for the 753 amino acids long *C. elegans* REF (2) P like protein. The REF(2) P (refractory to sigma P) of *Drosophila* was originally identified by its ability to restrict virus multiplication (Wyers et al., 1995). An alignment comparing the translated cDNA sequences of the four clones from the yeast two hybrid screen with the *C. elegans* protein predicted by wormbase (WP_CE06438) is shown in Figure 7. They also appear to represent independent clones since they differ in their amino termini.

The sequence in 3 of the prey plasmids mapped to the *C. elegans* gene B0546.2 for which there is no known function in the database. 1 cDNA clone mapped to the *C. elegans* genes F584A.7a and another to ZC404.8. F584.7a encodes a transcription factor, while ZC404.8 encodes the GEI-20 cDNA whose protein product is considered to have a role in morphogenesis. F22E12.4a, which encodes the *C. elegans* gene egl-9, was also identified as a candidate. Egl-9 encodes a dioxygenase that negatively regulates hypoxia inducible factor (hif-1) by hydroxylating prolyl residues in HIF-1 (Epstein et al., 2001).

Figure 6: Alignment of predicted amino acid sequences encoded by B0336.6 clones identified in the yeast two hybrid screen.

cDNAs were sequenced (pACT series), translated, and compared to the full length protein predicted by wormbase (WP_CE29545). Dark dashes in the N-terminus for pACT translations indicate regions which were not contained in the identified pACT clones. Red dashes follow the end of the region sequenced where the actual 3' end of the cDNA is undetermined. Green dashes represent gaps generated by the alignment program. The complete cDNA sequences obtained as well as the sequence of the 5' primer are shown in Appendix 1.



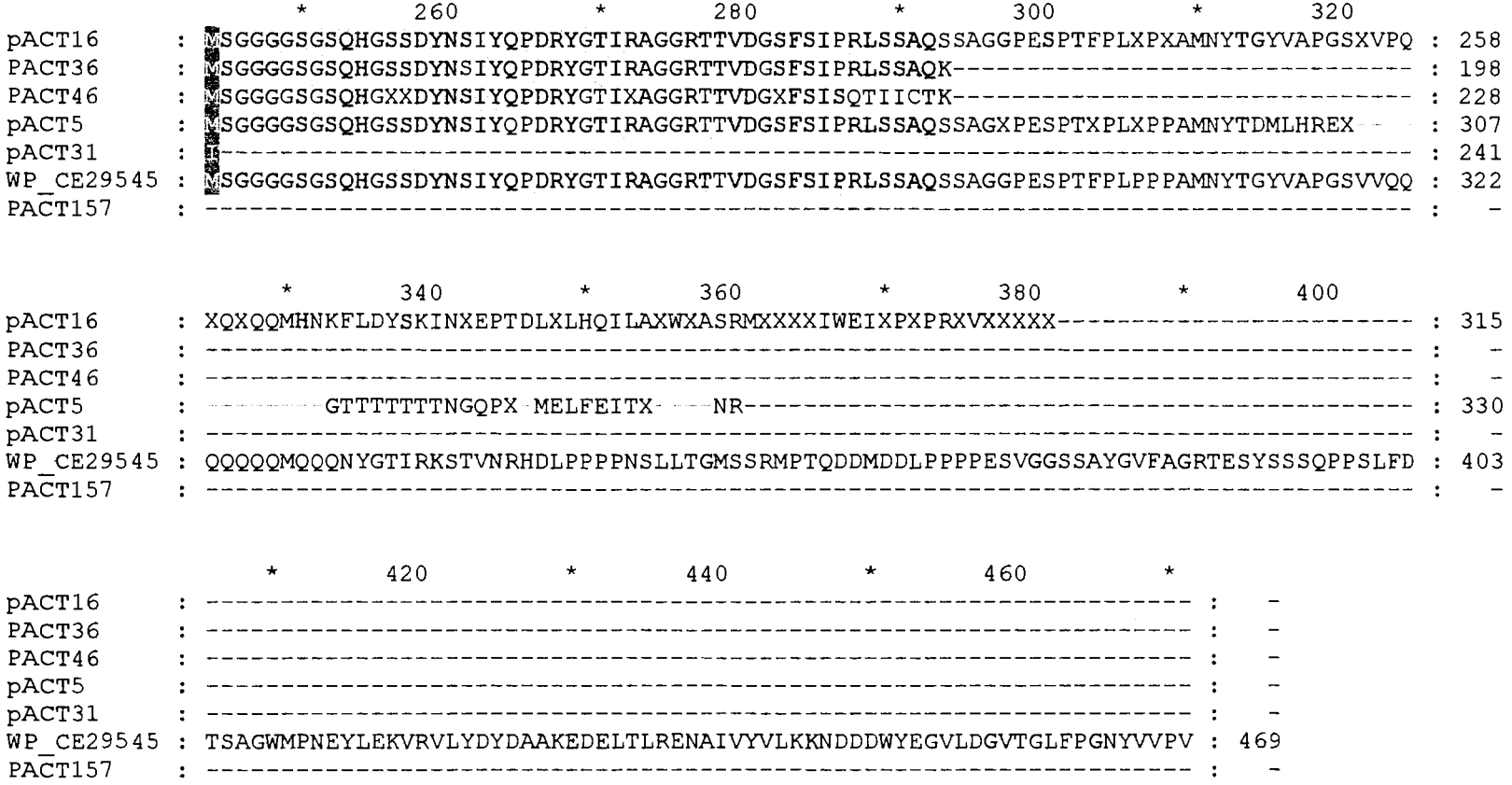
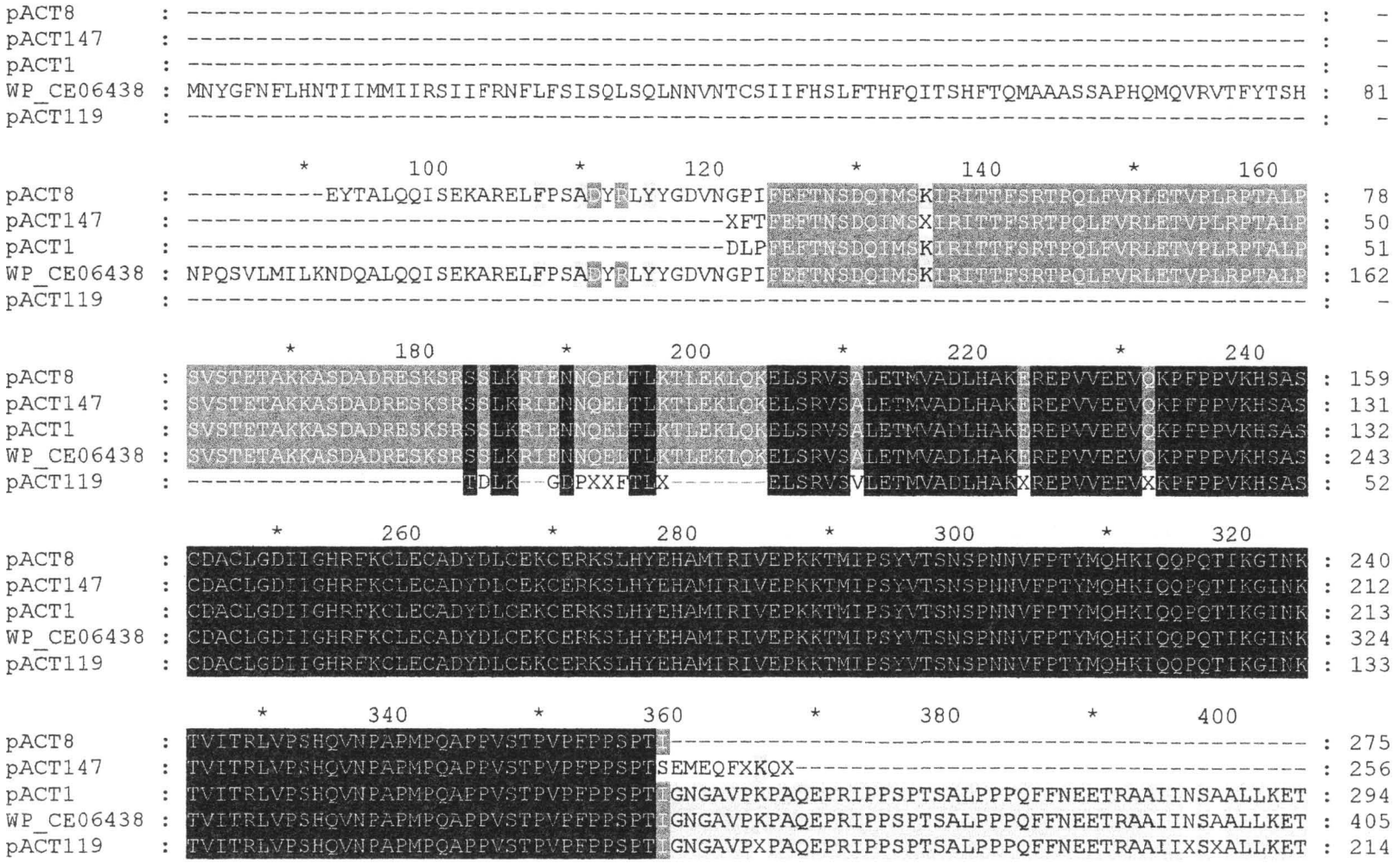
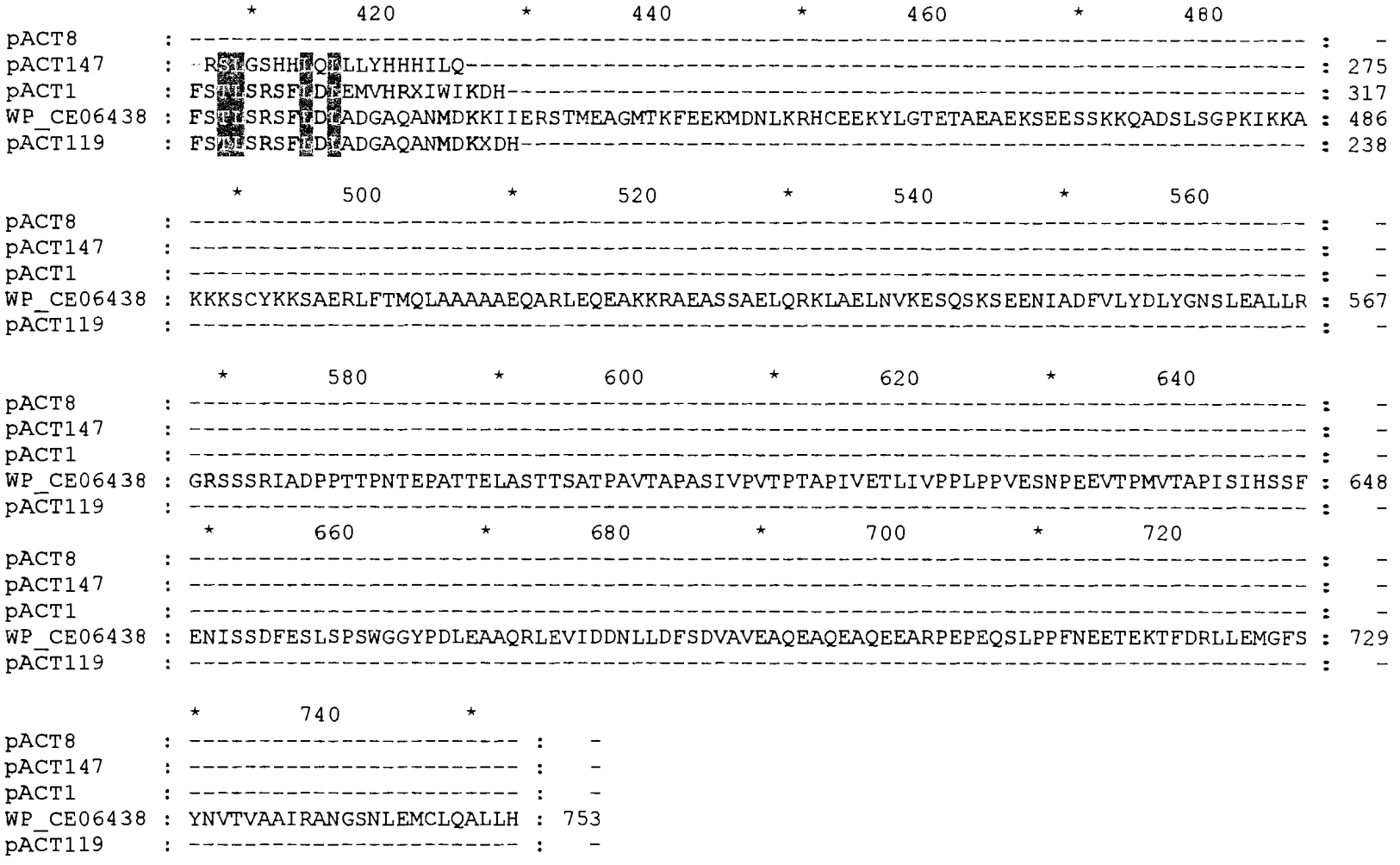

* 260 * 280 * 300 * 320
pACT16 : MSGGGGSGSQHGSSDYNSIYQPDRYGTIRAGGRTTVDGSFSIPRLSSAQSSAGGPESPTFPLXPXAMNYTGYVAPGSXVPQ : 258
PACT36 : MSGGGGSGSQHGSSDYNSIYQPDRYGTIRAGGRTTVDGSFSIPRLSSAQK---------- : 198
PACT46 : MSGGGGSGSQHGXXDYNSIYQPDRYGTIXAGGRTTVDGXFSISQTIICTK---------- : 228
pACT5 : MSGGGGSGSQHGSSDYNSIYQPDRYGTIRAGGRTTVDGSFSIPRLSSAQSSAGXPESPTXPLXPPAMNYTDMLHREX : 307
pACT31 : M---------- : 241
WP_CE29545 : MSGGGGSGSQHGSSDYNSIYQPDRYGTIRAGGRTTVDGSFSIPRLSSAQSSAGGPESPTFPLPPPAMNYTGYVAPGSVVQQ : 322
PACT157 : ---------- : -
* 340 * 360 * 380 * 400
pACT16 : XQXQQMHNKFLDYSKINXEPTDLXLHQILAXWXASRMXXXXIWEIXPXPRXVXXXXX---------- : 315
PACT36 : ---------- : -
PACT46 : ---------- : -
pACT5 : GTTTTTTTNGQPX MELFEITX NR---------- : 330
pACT31 : ---------- : -
WP_CE29545 : QQQQQMQQQNYGTIRKSTVNRHDLPPPPNSLLTGMSSRMPTQDDMDDLPPPPESVGGSSAYGVFAGRTESYSSSQPPSLFD : 403
PACT157 : ---------- : -
* 420 * 440 * 460 *
pACT16 : ---------- : -
PACT36 : ---------- : -
PACT46 : ---------- : -
pACT5 : ---------- : -
pACT31 : ---------- : -
WP_CE29545 : TSAGWMPNEYLEKVRVLYDYDAAKEDELTLRENAIVYVLKKNDDDWYEGVLDGVTGLFPGNYVVPV : 469
PACT157 : ---------- : -

Figure 7: Alignment of predicted amino acid sequences encoded by T12G3.1 clones identified in the yeast two hybrid screen.

cDNAs were sequenced (pACT series), translated, and compared to the full length protein predicted by wormbase (WP_CE06438). Dark dashes in the N-terminus for pACT translations indicate regions which were not contained in the identified pACT clones. Red dashes follow the end of the region sequenced where the actual 3' end of the cDNA is undetermined. Green dashes represent gaps generated by the alignment program. The complete cDNA sequences obtained as well as the sequence of the 5' primer are shown in Appendix 1.


pACT8 : ---------- : -
pACT147 : ---------- : -
pACT1 : ---------- : -
WP_CE06438 : MNYGFNFLHNTIIMMIIRSIIFRNFLFSISQLSQLNNVNTCSIIFHSLFTHFQITSHFTQMAAASSAPHQMQVRVTFYTSH : 81
pACT119 : ---------- : -
* 100 * 120 * 140 * 160
pACT8 : ----------EYTALQQISEKARELFPSADYRLYYGDVNGPIFEFTNSDQIMSKIRITTFSRTPQLFVRLETVPLRPTALP : 78
pACT147 : ----------XFTFEFTNSDQIMSXIRITTFSRTPQLFVRLETVPLRPTALP : 50
pACT1 : ----------DLPFEFTNSDQIMSKIRITTFSRTPQLFVRLETVPLRPTALP : 51
WP_CE06438 : NPQSVLMILKNDQALQQISEKARELFPSADYRLYYGDVNGPIFEFTNSDQIMSKIRITTFSRTPQLFVRLETVPLRPTALP : 162
pACT119 : ---------- : -
* 180 * 200 * 220 * 240
pACT8 : SVSTETAKKASDADRESKSRSSLKRIENNQELTLKTLEKLQKELSRVSALETMVADLHAKEREPVVEEVQKPFPPVKHSAS : 159
pACT147 : SVSTETAKKASDADRESKSRSSLKRIENNQELTLKTLEKLQKELSRVSALETMVADLHAKEREPVVEEVQKPFPPVKHSAS : 131
pACT1 : SVSTETAKKASDADRESKSRSSLKRIENNQELTLKTLEKLQKELSRVSALETMVADLHAKEREPVVEEVQKPFPPVKHSAS : 132
WP_CE06438 : SVSTETAKKASDADRESKSRSSLKRIENNQELTLKTLEKLQKELSRVSALETMVADLHAKEREPVVEEVQKPFPPVKHSAS : 243
pACT119 : ----------TDLK--GDPXXFTLX-------ELSRVSVLETMVADLHAKXREPVVEEVXKPFPPVKHSAS : 52
* 260 * 280 * 300 * 320
pACT8 : CDACLGDIIGHRFKCLECADYDLCEKCERKSLHYEHAMIRIVEPKKTMIPSYVTSNSPNNVFPTYMQHKIQQPQTIKGINK : 240
pACT147 : CDACLGDIIGHRFKCLECADYDLCEKCERKSLHYEHAMIRIVEPKKTMIPSYVTSNSPNNVFPTYMQHKIQQPQTIKGINK : 212
pACT1 : CDACLGDIIGHRFKCLECADYDLCEKCERKSLHYEHAMIRIVEPKKTMIPSYVTSNSPNNVFPTYMQHKIQQPQTIKGINK : 213
WP_CE06438 : CDACLGDIIGHRFKCLECADYDLCEKCERKSLHYEHAMIRIVEPKKTMIPSYVTSNSPNNVFPTYMQHKIQQPQTIKGINK : 324
pACT119 : CDACLGDIIGHRFKCLECADYDLCEKCERKSLHYEHAMIRIVEPKKTMIPSYVTSNSPNNVFPTYMQHKIQQPQTIKGINK : 133
* 340 * 360 * 380 * 400
pACT8 : TVITRLVPSHQVNPAPMPQAPPVSTPVPFPPSPTI---------- : 275
pACT147 : TVITRLVPSHQVNPAPMPQAPPVSTPVPFPPSPTSEMEQFXKQX---------- : 256
pACT1 : TVITRLVPSHQVNPAPMPQAPPVSTPVPFPPSPTIGNGAVPKPAQEPRIPPSPTSALPPPQFFNEETRAAIINSAALLKET : 294
WP_CE06438 : TVITRLVPSHQVNPAPMPQAPPVSTPVPFPPSPTIGNGAVPKPAQEPRIPPSPTSALPPPQFFNEETRAAIINSAALLKET : 405
pACT119 : TVITRLVPSHQVNPAPMPQAPPVSTPVPFPPSPTIGNGAVPXPAQEPRIPPSPTSALPPPQFFNEETRAAIIXSXALLKET : 214

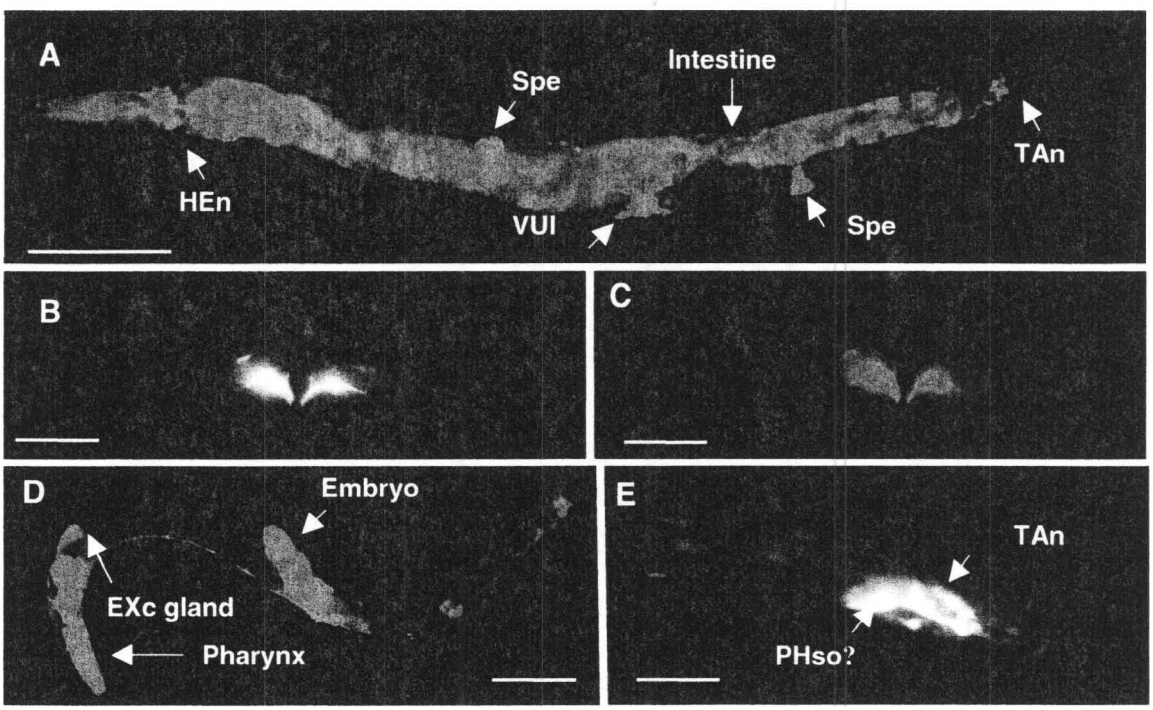

* 420 * 440 * 460 * 480
pACT8 : ---------- : -
pACT147 : RSTGSHHTQLLYHHHILQ---------- : 275
pACT1 : FSTSRSFTDEMVHRXIWIKDH---------- : 317
WP_CE06438 : FSTSRSFTDADGAQANMDKKIIERSTMEAGMTKFEEKMDNLKRHCEEKYLGTETAEAEKSEESSKKQADSLSGPKIKKA : 486
pACT119 : FSTSRSFTDADGAQANMDKXDH---------- : 238
* 500 * 520 * 540 * 560
pACT8 : ---------- : -
pACT147 : ---------- : -
pACT1 : ---------- : -
WP_CE06438 : KKKSCYKKSAERLFTMQLAAAAAEQARLEQEAKKRAEASSAELQRKLAELNVKESQSKSEENIADFVLYDLYGNSLEALLR : 567
pACT119 : ---------- : -
* 580 * 600 * 620 * 640
pACT8 : ---------- : -
pACT147 : ---------- : -
pACT1 : ---------- : -
WP_CE06438 : GRSSSRIADPPTTPNTEPATTELASTTSATPAVTAPASIVPVTPTAPIVETLIVPPLPPVESNPEEVTPMVTAPISIHSSF : 648
pACT119 : ---------- : -
* 660 * 680 * 700 * 720
pACT8 : ---------- : -
pACT147 : ---------- : -
pACT1 : ---------- : -
WP_CE06438 : ENISSDFESLSPSWGGYPDLEAAQRLEVIDDNLLDFSDVAVEAQEAQEAQEEARPEPEQSLPPFNEETEKTFDRLLEMGFS : 729
pACT119 : ---------- : -
* 740 *
pACT8 : ---------- : -
pACT147 : ---------- : -
pACT1 : ---------- : -
WP_CE06438 : YNVTVAAIRANGSNLEMCLQALLH : 753
pACT119 : ---------- : -

3.8 Expression pattern of T12G3.1

A nematode line expressing a GFP fusion protein of the GFP expression vector and the promoter region of the T12G3.1 gene was obtained from David Baillie (Alan Mah performed injections) SFU, Vancouver, Canada (Genome wide transgenic project). Expression was observed in several head and tail neurons, the vulval muscles, the intestine, coelomocytes and other unidentified cells (Figure 8). GFP fusion products were also made for the B0336.6 and ZC404.8 *C. elegans* genes, but no expression patterns were obtained. The lack of signal for some of the constructs may be indicative of an incomplete promoter region used in the fusion construct or a region lacking key regulatory elements.

Figure 8: Expression profile of T12G3.1.

A *T12G3.1::gfp* fusion construct was injected into hermaphrodites and transgenic animals were viewed using a Zeiss compound microscope with FITC filters for immunofluorescence. Expression was observed in (**a**) the vulva (VUl) muscles, spermatheca (SPe), head neurons (HEn) and tail neurons (TAn), intestine (**b** & **c**) vulva, (**d**) tail neurons, embryos, pharynx, excretory gland cell (**e**) tail neurons, phasmid sockets (PHso). Expression was also observed in coelomocytes (not shown). Scale bars: 100 μm.

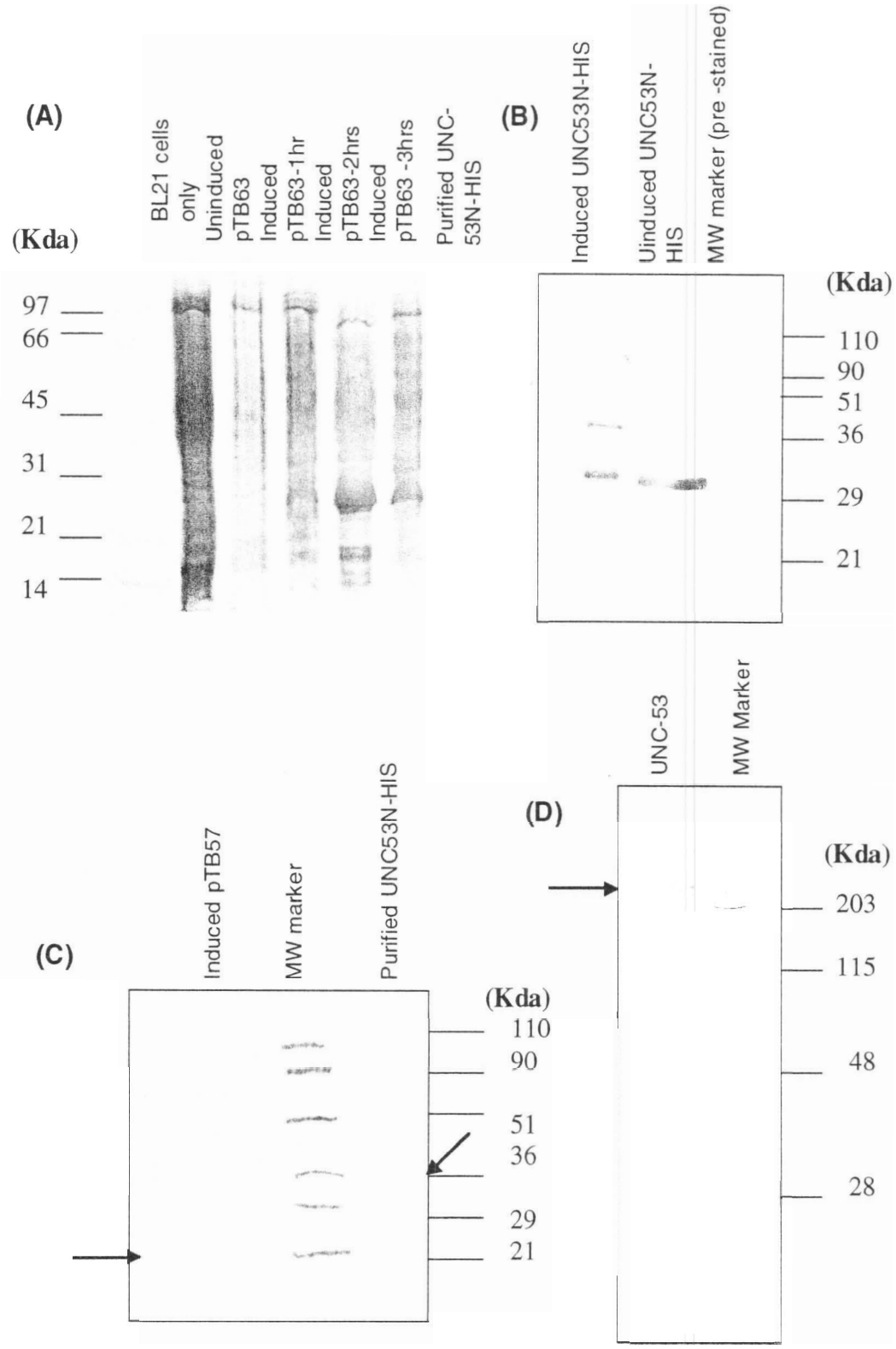

A
Spe
Intestine
TAn
HEn
VUl
Spe
B
C
D
Embryo
EXc gland
Pharynx
E
TAn
PHso?

3.9 Analysis of UNC-53 protein isoforms using PAb UNC-53N

The molecular complexity of the *unc-53* locus (multiple promoters, alternative splicing) suggests that there are multiple protein isoforms *in vivo*. In order to better understand the nature of these isoforms, their patterns of tissue and temporal expression, polyclonal antisera was generated against the N-terminus of UNC-53; a region identical to that contained in the bait used in the yeast two-hybrid screen, UNC-53N. A His-tag fusion containing the N-terminus region of UNC-53 was purified (Figure 9a) and inoculated into a female New Zealand white rabbit to raise antibodies, which will be referred to as PAb UNC-53N (polyclonal antibody against the N-terminus of UNC-53) henceforth. The antibody was determined to be effective at 1:30,000 on western blots of purified protein (see Figure 9b). The polyclonal antibody was also shown to detect non-epitope tagged portions of UNC-53 containing the N-terminus. pTB57 encodes the same region as pTB63 excluding the histidine tag while pTB74 encodes the full-length protein devoid of the histidine tag as well. The fact that PAB UNC-53N identifies these other UNC-53 recombinant proteins confirms that antibodies were made against UNC-53N and not just against the histidine tag (Figure 9c and 9d). Attempts to determine the specificity and effectiveness of PAb UNC-53N on yeast and worm lysates have been inconclusive. However, embryos in which UNC-53 was ectopically expressed in the intestine under control of a heat shock promoter, or in body muscle under control of the *unc-54* promoter, stained in these tissues with PAb UNC-53N while staining was absent in wild type embryos (results not shown).

Figure 9: Protein expression and purification analysis.

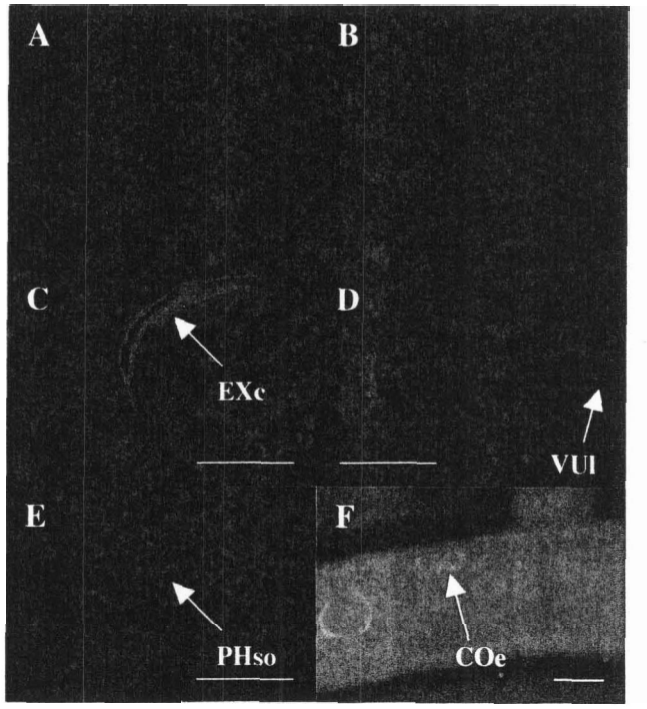
(**a**) Commassie stain of a 12% SDS-PAGE gel. Induction at 1, 2, and 3 hours is shown. BL21 cells only and uninduced lane shows that the protein is not expressed. Expression is obtained however after induction with IPTG. Histidine tagged purified sample is also shown. (**b**) Lysates from uninduced and induced samples were transferred to a nylon membrane and UNC53 proteins detected with polyclonal antisera at 1:5,000. All antisera used in this analysis were pre-absorped with bacteria acetone preparation (see materials and methods) Western analysis reveals that protein expression may be leaky as a signal is obtained before induction.(**c**) Purified pTB63 was also transferred to a western blot and detected. The polyclonal antibody was also shown to detect the truncated UNC-53 protein encoded by pTB57, which contains the same region as UNC-53N excluding the histidine tag. Detection at 1:10,000. (**d**) The polyclonal antisera also detected the full-length UNC-53 protein encoded by pTB74. pTB74 was expressed in bacteria, the cell lysate run on a polyacrylamide gel and transferred to a blot for western analysis.

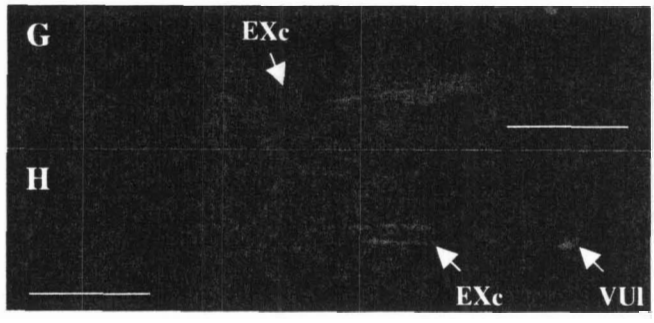

(A)
(Kda)
BL21 cells only
Uninduced pTB63
Induced pTB63-1hr
Induced pTB63-2hrs
Induced pTB63 -3hrs
Purified UNC-53N-HIS
97
66
45
31
21
14
(B)
Induced UNC53N-HIS
Uinduced UNC53N-HIS
MW marker (pre -stained)
(Kda)
110
90
51
36
29
21
(C)
Induced pTB57
MW marker
Purified UNC53N-HIS
(Kda)
110
90
51
36
29
21
(D)
UNC-53
MW Marker
(Kda)
203
115
48
28

Immunofluorescence was performed on adult hermaphrodites and staining was observed in the excretory canals, coelomocytes, vulva and staining was observed in the posterior end of the worm, possibly in the anal sphincter as well as other unidentified cells (Figure 10). Interestingly this pattern of expression was retained in both *unc-53(n152)* and *unc-53(n166)* mutant alleles; however the canals in these mutants were truncated (Figure 10g and h). No staining was observed in control experiments using the secondary antibody alone (Figure 10) or in combination with the pre-immune sera (Figure 10b).

Figure 10: PAB UNC-53N staining patterns.

Immunofluorescence on N2 wild type hermaphrodites using the polyclonal antibody (PAB UNC-53N). (**a**) Control experiments using the secondary antibody alone or (**b**) in combination with the pre-immune sera did not give any expression. (**c**) Staining in wild type N2 worms was observed in the excretory canals (EXc) (**d**) vulva (VUl) (**e**) phasmid socket (PHso) (**f**) coleomocyte (Coe). Truncated canals in the vicinity of the vulva are also shown for (**g**) *unc-53(n152)* and (**h**) *unc-53(n166)*. Scale bars: 100 μm.


A
B
C
EXc
D
VUl
E
PHso
F
COe


G
EXc
H
EXc
VUl

4. DISCUSSION:

4.1 UNC-53 is expressed in the cells in which it is required.

In this study, polyclonal antisera specific to the N-terminus of UNC-53 was generated and used to detect the full length UNC-53 isoform in whole animals by immunohistochemistry. The expression profile of UNC-53, as determined by staining with PAb UNC53N, generally complemented the data obtained in previous promoter-GFP analysis and rescue experiments. UNC-53 is made in the excretory cell and the vulval myoblasts, cells which are known to require UNC-53 activity during outgrowth (Stringham et al., 2002). These results are consistent with a cell autonomous role for UNC-53 in cell guidance. Interestingly, staining was observed in adult tissues after the completion of cell migration and outgrowth suggesting that UNC-53 may play a role in the maintenance or stabilization of the cellular cytoskeleton after outgrowth.

Expression was also retained in the *unc-53* mutants *n152* and *n166*, suggesting that both *n152* and *n166* animals express truncated UNC-53 proteins. The molecular lesion in *n152* is known and predicted to encode a gene product truncated at the C-terminus (Stringham et al., 2002). While the molecular lesion of *n166* remains undetermined, the results of this study suggests that a truncated UNC-53 protein containing at least part of the CH domain is produced, since it is recognized by PAb UNC-53N.

4.2 UNC-53N interacts with the T12G3.1 *C. elegans* gene product.

T12G3.1 encodes the *C. elegans* Ref(2)P *like* protein that contains two intriguing motifs suggestive of function: a zinc finger motif (predicted by smart, pfam and LOAD) of the ZZ type found in the cytoskeletal protein dystrophin, and a ubiquitin associated domain (UBA) (predicted by smart) at the carboxy terminus of the protein (Schultz J et al., 1998) (Figure 11). Blastp searches with the *C. elegans* Ref(2)P revealed homologous similar proteins in other organisms. Verprolin, the *S. cerevisiae* counterpart is 29.3%

similar to the *C. elegans* REF(2) P like protein, while *Drosophila's* REF(2) P protein is 42.9% similar to the *C. elegans* protein. An alignment comparing the amino acid sequence of the predicted *C. elegans* protein against the *D. melanogaster* and *S. cerevisiae* similar proteins are depicted in Figure 12. The alignment shows regions of homology distributed weakly all along the length of the amino acid sequence. Proline stretches are also found interspersed along the amino acid sequence of the yeast protein.

While the candidate cDNAs that mapped to the T12G3.1 locus are not identical along their length, alignment of the predicted protein sequences reveal an almost complete region of identity between amino acids 205 and 359 with respect to the predicted *C. elegans* protein encoded by T12G3.1 (WP_CE06438) from wormbase. This suggests that the upstream limit to the region of interaction with UNC-53 is after amino acid 205. While the actual 3' ends of cDNA clones were not determined, the size of all cDNAs was approximately 1.8 kb, indicating that none appear to encode the entire full length protein of 753 amino acids encoded in T12G3.1. Interestingly, the zinc finger domain is located between amino acids 238 and 281 of the WP_CE06438 amino acid sequence (SMART), and is contained in all clones (Figure 7).

Figure 11: Predicted domains in candidates that interact with UNC-53N

Predictions were made using SMART (Schultz J et al., 1998).

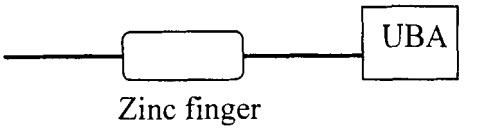
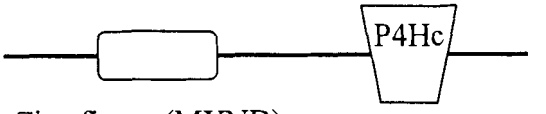
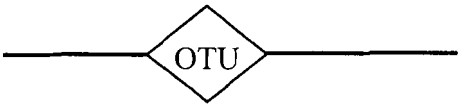
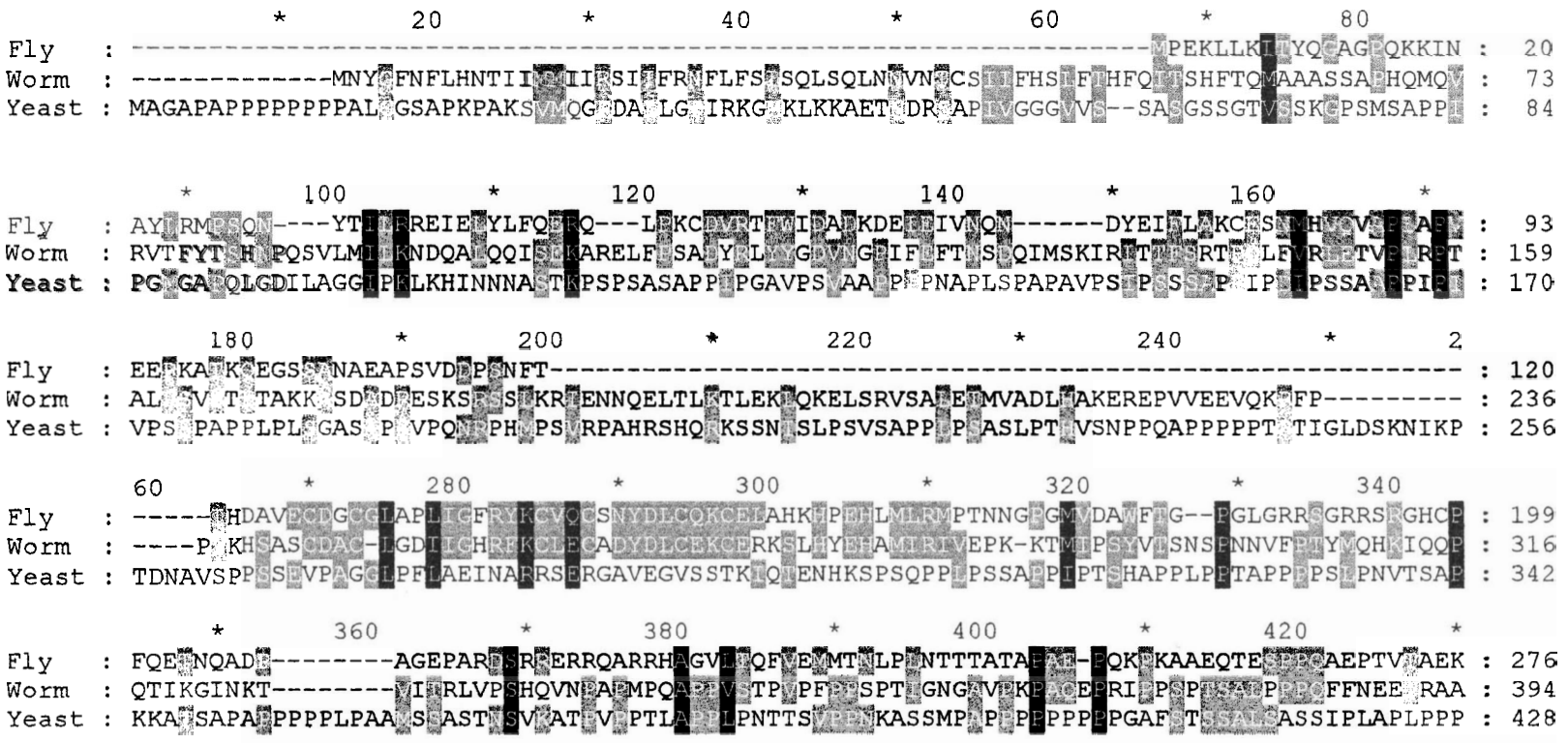
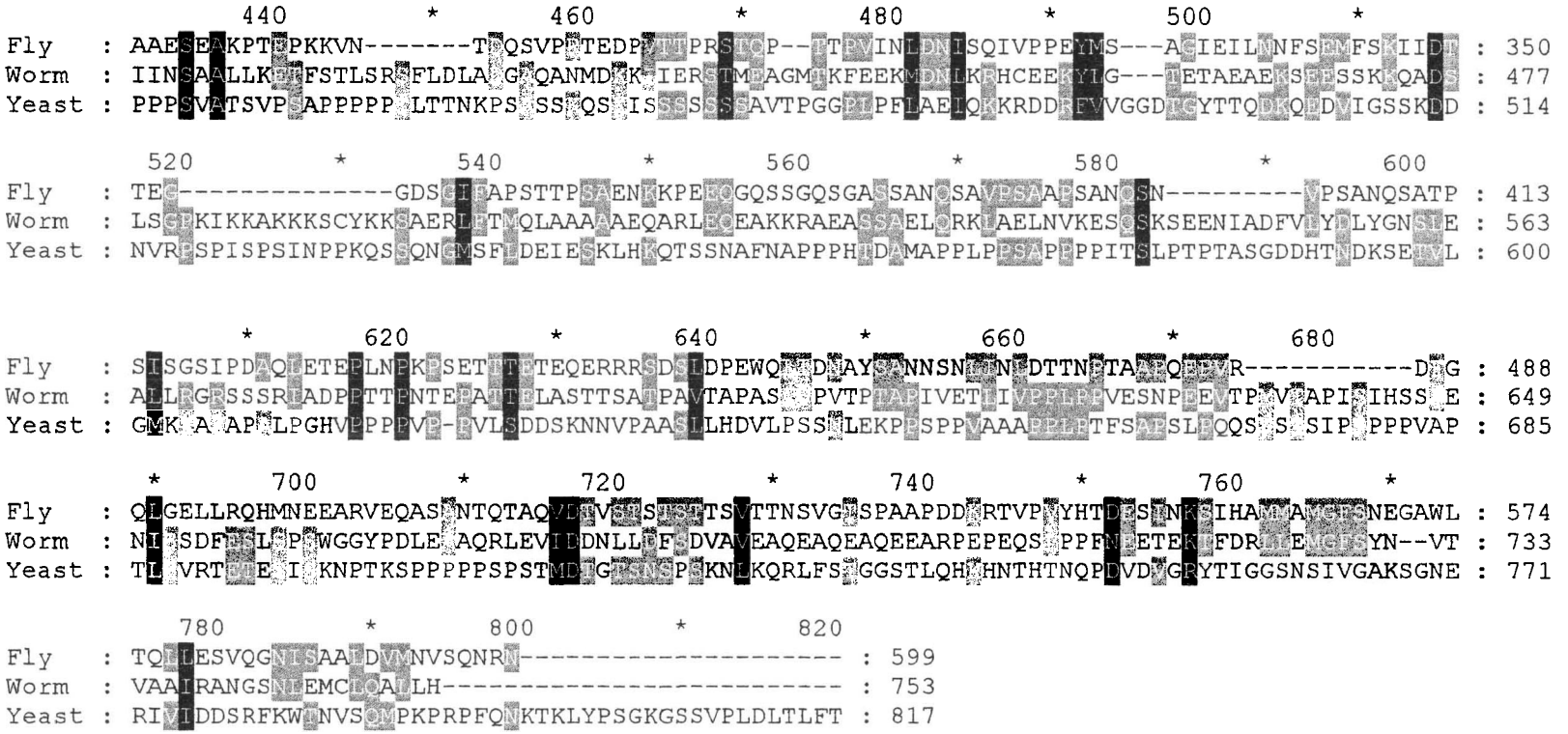
Gene/Size/	Predicted domains
T12G3.1 (REF(2)P *like* protein) 83.7 KDa	 Zinc finger
B0336.6 / 51.3 KDa	
ZC404.8 (Spn-4)/ 39.2 Kda	 RNA recognition motif
F22E12.4 (Egl-9) / 79.7KDa	 Zinc finger (MYND) Prolyl 4-hydroxylase
B0546.2 / 48.3 KDa	 OTU-like cysteine protease
F58A4.7a / 45.9 KDa	 Helix-loop-helix

Figure 12: Alignment of the *C. elegans*, *D. melanogaster* and *S. cerevisiae* predicted amino acid sequences encoded by T12G3.1.

Regions of complete conservation are shown in black while areas of partial homology are depicted in grey.


Fly
Worm
Yeast


440 * 460 * 480 * 500 *
Fly : AAESEAKPTEPKKVN-------TEQSVPETEDPVTTPRSTQP--TTPVINLDNISQIVPPEYMS---AGIEILNNFSEMFSKIIDT : 350
Worm : IINSAALLKEFFSTLSRSFLDLAGGQANMDKKIERSTMEAGMIKFEEKMDNLKRHCEEKYLG---TETAEAEKSEESSKKQADS : 477
Yeast : PPPSVATSVPSAPPPPPLTTNKPSSSSQSISSSSSSSAVTPGGPLPFLAELQKKRDDRFVVGGDTGYTTQDKQLDVIGSSKDD : 514
520 * 540 * 560 * 580 * 600
Fly : TEG--------------GDSGIFAPSTTPSAENKKPEEQGQSSGQSGASSANQSAVPSAAESANQSN---------VPSANQSATP : 413
Worm : LSGPKIKKAKKKSCYKKSAERLPTMQLAAAAAEQARLEQEAKKRAEASSAELQRKLAELNVKESQSKSEENIADFVLYDLYGNSLE : 563
Yeast : NVRESPISPSINPPKQSSQNGMSFLDEIESKLHKQTSSNAFNAPPPHTDAMAPPLPPSAPEPPITSLPTPTASGDDHTNDKSEIVL : 600
* 620 * 640 * 660 * 680
Fly : SISGSIPDAQLETEPLNPKPSETTTETEQERRRSDSLDPEWQNMDNAYSANNSNTNNTDTTNPTAAEQFEVR----------DNG : 488
Worm : ALLPGRSSSRLADPPTTPNTEPATTPLASTTSATPAVTAPASSPVTPTAPIVETLIVPPLPPVESNPEEVTPVVAPIPIHSSSE : 649
Yeast : GMKLAAPSLPGHVPPPPVP-RVLSDDSKNNVPAASLLHDVLPSSSLEKPPSPPVAAAPPLPTFSAESLPQQSSSSIPPPPVAP : 685
* 700 * 720 * 740 * 760 *
Fly : QLGELLRQHMNEEARVEQASSNTQTAQVDTVSTSTSTTSVTTNSVGTSPAAPDDSRTVPNYHTDESTNKSIHAMMAMGTSNEGAWL : 574
Worm : NISSDFTSLSPTWGGYPDLEAQRLEVIDDNLLCFSDVAVEAQEAQEAQEEARPEPEQSSPPFNPETETKTFDRLLEMGTSYN--VT : 733
Yeast : TLSVRTEEEISKNPTKSPPPPPSPSTMDIGTSNSPSKNLKQRLFSSGGSTLQHSHNTHTNQPDVDVGRYTIGGSNSIVGAKSGNE : 771
780 * 800 * 820
Fly : TQLLESVQGNLSAALDVMNVSQNRN---------- : 599
Worm : VAALRANGSNLEMCLQALLH---------- : 753
Yeast : RIVIDDSRFKWTNVSQMPKPRPFQNKTKLYPSGKGSSVPLDLTLFT : 817

Deviations between the amino acid sequences predicted by cDNA clones corresponding to a single gene may be the result of :the following: (a)DNA sequencing errors, (b) incomplete or aberrant cDNA products which are the result of mistakes in cDNA synthesis by reverse transcriptase and (c) clones may represent alternative protein isoforms that result from alternative transcriptional starts or RNA splicing.

The REF(2)P protein was first identified in *Drosophila* for its ability to interfere with the multiplication of the sigma rhabdovirus (Wyers et al., 1995). Viral replication involves a number of stepwise processes that ensures survival of the virus. These include attachment and entry of the virus particle, decoding of genome information, translation of viral mRNA by host ribosomes, genome replication, and assembly and release of viral particles containing the genome. The *ref(2)P* gene is highly polymorphic in natural *Drosophila* populations and the several alleles are classified either as permissive or restrictive by comparing their effects on sigma virus replication (Wyers et al., 1995).

Analysis of the T12G3.1 gene product by the SMART protein database revealed the presence of a zinc finger domain of the ZZ type found in a variety of dystrophin-like proteins, and CREB-binding protein/p300 homologues (Schultz et al., 1998). It has long being purported that dystrophin is a cytoskeletal protein (Koenig et al., 1988, Ponting et al., 1996) whose function in skeletal muscle is thought to involve linking the actin cytoskeleton to the extracellular matrix (Rentschler et al., 1999).In dystrophin, the zinc finger motif is thought to bind calmodulin (Ponting et al., 1996), an intracellular protein that mediates several calcium regulated processes by binding calcium. This association with calcium results in a conformational change that allows calmodulin to bind various proteins and thereby modulate their activity making calmodulin a crucial player in the transmission of signals in several signaling cascades. The second region of interest revealed an Ubiquitin Associated (UBA) domain. UBA domains are seen in molecules involved in diverse cellular activities including DNA repair, proteolysis and signal transduction (Madura, 2002).

The corresponding yeast candidate, verprolin, is an actin binding protein that has been implicated in several crucial developmental processes such as cell growth,

cytoskeletal organization, migration, endocytosis and mitochondrial protein differentiation. Yeast verprolin (VRP1) encodes a proline rich protein (Donnelly et al., 1993), which contains the highly conserved LKKAET signature motif present in other actin-binding proteins like WASP (Wiskott-Aldrich syndrome protein) and N-WASP (*C. elegans* homologue of WASP) (Vaduva, G. et al., 1997). There has been an increasing body of evidence suggesting a connection in the processes that modulate the actin cytoskeleton and control of endocytosis (Munn, et al., 1995, Vaduva et al., 1997, Naqvi et al., 1998). Vaduva et al, provided evidence, which suggest that verprolin, is a polarity development protein that modulates the actin cytoskeleton possibly by serving as an anchor for actin. Concurrently, they showed in a yeast two-hybrid assay that verprolin binds actin via its LKKAET domain (Vaduva et al., 1997). Yeast cells with mutations in VRP1 have also been shown to be defective in endocytic processes (Munn et al., 1995). Interestingly the UNC-53 bait fusion used in this screen contains a LKKDQK motif and a calponin homology domain, both of which are seen in actin binding proteins (Stringham et al., 2002). In addition, preliminary results suggest that UNC-53 binds F-actin in a co-sedimentation assay (Stringham, unpublished results). Thus it is possible that both verprolin and UNC-53 are localized to the actin cytoskeleton by virtue of their binding abilities to actin but then interact in a secondary fashion with each other. Verprolin also interacts with other proteins that are important in maintaining the fluid endocytic processes as well as regulating actin dynamics such as Las17p (encodes the yeast homologue of the human Wiskott-Aldrich Syndrome protein, WASP) and Sla1p (Dewar et al., 2002, Roumanie et al., 2002).

4.3 T12G3.1 expression in vulva muscles and neurons

The expression pattern of T12G3.1 was characterized using a *T12G3.1::GFP* expression construct (Baillie, unpublished results). Strong expression was observed in the vulva muscles, several head and tail neurons, the spermatheca, and in the intestine (Figure 8, Table 6). *unc-53* gene expression in the vulva muscles has been observed and a defect in the longitudinal migration of the sex myoblast required for proper egg laying has been reported (Stringham et al., 2002).

Table 6: RNAi data and GFP expression analysis obtained from current literature.

All GFP expression results were obtained from David Baillie, Vancouver, Canada, from the transgenic genome project.

Gene	ORF	RNAi Phenotype	GFP Expression	Reference
B0336.6	469aa	embryonic lethal	no signal	(Kamath et al., 2003) (Baillie, unpublished)
T12G3.1	753aa	n.d.	intestine, pharynx, head neurons, vulva, excretory gland, tail neurons spermatheca, coelomocytes	(Baillie, unpublished)
ZC404.8	351aa	embryonic lethal	no signal	(Tsuboi et al., 2002) (Baillie, unpublished)
F22E12.4a	723aa	n.d.	n.d.	
F58A4.7a	429aa	n.d.	n.d.	
B0546.2	441aa	wild type	n.d.	(Kamath et al., 2003)

Expression of *unc-53* has also been noted in the DA motorneurons, and some neurons in the head and tail ganglia. The ALN and PLN neurons, which send axons anteriorly as far as the head along the sublateral cord, stop prematurely in *unc-53* mutants (Stringham et al., 2002).

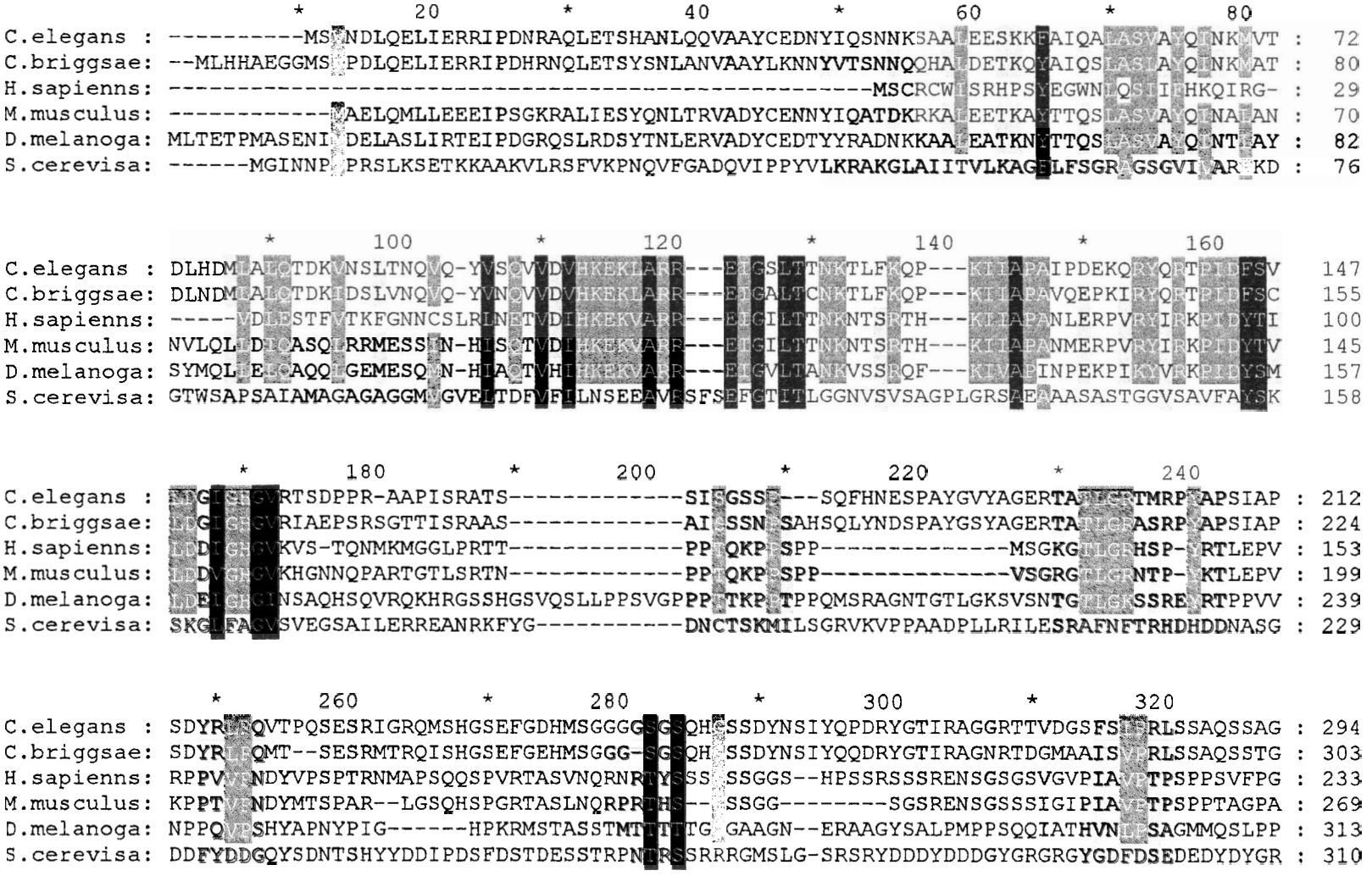
Taken together, expression of T12G3.1 in the vulva muscles, the head and tail neurons place T12G3.1 in the right location to interact with UNC-53 in directing cell migration. It should be noted however that the reported *unc-53* GFP data was based on expression driven from two internal promoters (Stringham et al., 2002) which may not reflect the expression pattern of a longer transcript. *unc-53* mutants also display a truncated canal phenotype and *unc-53* is expressed in the excretory canals (Figure 10c). However, T12G3.1 did not show any expression in the excretory canals. T12G3.1 maps to the T12G3 cosmid, which rescues the *let-52* gene on chromosome IV, and homozygous *let-52* nematodes arrest at the early larval stage (Schein, J.E., MSc Thesis. SFU).

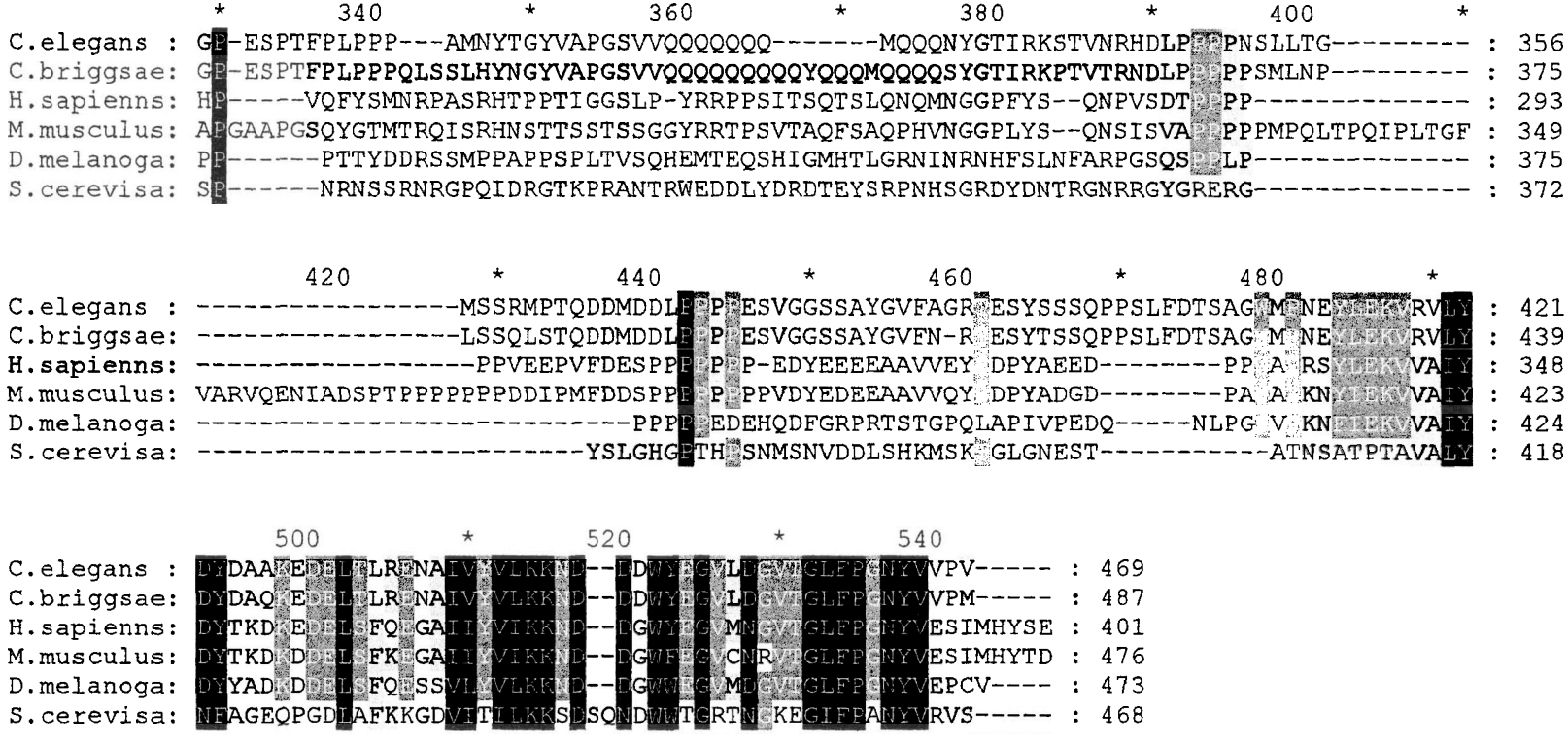
4.4 UNC-53N interacts with the B0336.6 *C. elegans* gene product

The highest numbers of sequenced candidates interacting with the UNC-53N bait construct were found to correspond to the *C. elegans* gene encoded by B0336.6. BLAST searches of the complete B0336.6 protein sequence identified an SH3 domain at the carboxy terminus (Figure 11) (predicted by smart and pfam), suggesting that the B0336.6 protein may be involved in signal transduction and is a plausible interactor for UNC-53. Interestingly, the truncated UNC-53N bait construct used in this screen does not contain the polyproline repeats suspected of being an SH3 binding domain, suggesting that the interaction with B0336.6 is through a region separate and distinct from the SH3 domain.

Blastp searches revealed proteins of significant similarities to the *C. elegans* protein in *D. melanogaster, H. sapiens, M. musculus, S. cerevisiae* and *C. briggsae*. An alignment comparing the amino acid sequence of the predicted *C. elegans* protein encoded by B0336.6 with the predicted proteins in *D. melanogaster, H. sapiens, M. musculus, S. cerevisiae* and *C. briggsae* is shown in Figure 13.

Figure 13: Alignment of the *C. elegans*, *D. melanogaster*, *C. briggsae*, *H. sapiens*, *M. musculus* and *S. cerevisiae* predicted amino acid sequences encoded by B0336.6. Regions of complete conservation are shown in black while areas of partial homology are depicted in grey.


C.elegans :
C.briggsae:
H.sapienns:
M.musculus:
D.melanoga:
S.cerevisa:


340
360
380
400
C.elegans : GP-ESPTFPLPPP---AMNYTGYVAPGSVVQQQQQQQ-------MQQQNYGTIRKSTVNRHDLPPPPNSLLTG--------- : 356
C.briggsae: GP-ESPTFPLPPPQLSSLHYNGYVAPGSVVQQQQQQQQQQYQQQMQQQQSYGTIRKPTVTRNDLPPPPPSMLNP--------- : 375
H.sapienns: HP-----VQFYSMNRPASRHTPPTIGGSLP-YRRPPSITSQTSLQNQMNGGPFYS--QNPVSDTPPPP------------- : 293
M.musculus: APGAAPGSQYGTMTRQISRHNSTTSSTSSGGYRRTPSVTAQFSAQPHVNGGPLYS--QNSISVAPPPPMPQLTPQIPLTGF : 349
D.melanoga: PP------PTTYDDRSSMPPAPPSPLTVSQHEMTEQSHIGMHTLGRNINRNHFSLNFARPGSQSPPLP------------- : 375
S.cerevisa: SP------NRNSSRNRGPQIDRGTKPRANTRWEDDLYDRDTEYSRPNHSGRDYDNTRGNRRGYGRERG------------- : 372
420
440
460
480
C.elegans : ----------------MSSRMPTQDDMDDLPPPESVGGSSAYGVFAGRESYSSSQPPSLFDTSAGMMNEVLPKVRVLY : 421
C.briggsae: ----------------LSSQLSTQDDMDDLPPPESVGGSSAYGVFN-RESYTSSQPPSLFDTSAGMMNEVLPKVRVLY : 439
H.sapienns: ----------------PPVEEPVFDESPPPPPP-EDYEEEEAAVVEYDPYAEED--------PPAARSVLPKVVAIY : 348
M.musculus: VARVQENIADSPTPPPPPPPDDIPMFDDSPPPPPPVDYEDEEAAVVQYDPYADGD--------PAAAKNVLPKVVAIY : 423
D.melanoga: ------------------------PPPPPEDEHQDFGRPRTSTGPQLAPIVPEDQ-----NLPGVVKNPLPKVVAIY : 424
S.cerevisa: ----------------------YSLGHGPTHPSNMSNVDDLSHKMSKPGLGNEST----------ATNSATPTAVALY : 418
500
520
540
C.elegans : DYDAAVEDELTLRRNAIVYVLKKNL--PDWYPGVLDGVTGLFPGNYVVPV----- : 469
C.briggsae: DYDAQVEDELTLRRNAIVYVLKKND--DDWYHGVLDGVTGLFPGNYVVPM----- : 487
H.sapienns: DYTKDVEDELSFQRGAIIYVIKKND--DGWYEGVMNGVTGLFPGNYVESIMHYSE : 401
M.musculus: DYTKDVDDELSFKRGAIIYVIKKND--DGWFEGVCNRVTGLFPGNYVESIMHYTD : 476
D.melanoga: DYYADVDDELSFQRSSVLYVLKKND--DGWWEGVMDGVTGLFPGNYVEPCV---- : 473
S.cerevisa: NFAGEQPGDLAFKKGDVITILKKSDSQNDWWTGRTNGKEGIFPANYVRVS----- : 468

Of particular note, there is a strong region of conservation all across the species in the SH3 domain (SMART) (amino acids 415-469 with respect to the *C. elegans* amino acid sequence). Regions of proline conservation are also seen along the amino acid sequence (at amino acids 209, 319, 330, 393, 394, 443, 444, 445). In addition, a tract of polyglutamines are present in the *C. elegans* and *C. briggsae* proteins (amino acids 306 – 326 with respect to the *C. elegans* amino acid sequence).

While there is a high degree of conservation in the C-terminal area of the sequences, the yeast protein appears to have diverged more in the N-terminal region. Alignment of the predicted amino acid sequences derived from translations of candidate cDNAs that mapped to the B0336.6 locus show an almost complete region of identity between amino acids 67 and 287 with respect to the translation of the complete B0336.3 locus (WP_CE29545) from wormbase (Figure 6). Since the cDNA clones differ in their N-termini, the region of interaction with UNC-53 must be carboxyl to amino acid 67. While the cDNAs ranged from 1.8 to 2.0 kb in size, incomplete sequencing of the 3' ends rendered it impossible to determine the C-terminal border in protein B0336.6 for the UNC-53 interaction. The SH3 domain is located between amino acids 415 and 469 in the C-terminal region of the B0336.6 protein (SMART).

The *Drosophila* Abelson interacting protein (dAbi), along with the human and mouse homologues are SH3 domain containing proteins that bind to the proline-rich motifs of the Abelson protein kinase (Abl) (Juang, J. et al., 1999). The Abl-interactor (Abi) proteins are recognized as targets of the Abl-family of non-receptor tyrosine kinases (NRTKs) and have been implicated in Rac-dependent cytoskeletal reorganization in response to growth factor stimulation (Dai and Pendergast 1995). Interestingly, *unc-53* has also been grouped with *unc-73*, a GEF (guanine nucleotide exchange factor) molecule which activates Rac signaling, in a gonad independent pathway for migration of the sex myoblasts (Chen et al., 1997). UNC-53 also contains a putative nucleotide-binding domain in the C-terminus (Stringham et al., 2002).

This family of NRTKs includes c-Abl and Arg, both of which are required for proper neurulation in mice (Koleske et al., 1998). Both Abi and the Abl-family proteins

are expressed in the developing nervous system, in the neuron cell body, synapses and growth cones (Courtney et al., 2000). While evidence suggest a role for Abl and Arg in neuronal development and axonogenesis, the targets for these proteins in neuronal development remain obscure (Dai and Pendegast 1995). Thus the finding of Courtney et al that c-Abl, Arg and Abi are expressed in the same neuronal tissues is note worthy, suggesting that Abi may act to relay signaling events downstream of the Abl kinases or contribute to the regulation of c-Abl and Arg kinase (Courtney et al., 2000). Likewise, genetic studies in Drosophila have pointed to a role for the Abl family of kinases in axonogenesis and growth cone pathfinding (Hu and Reichardt 1999). Interestingly, GFP reporter analysis to characterize the expression pattern of UNC-53 in *C. elegans* showed expression in several neurons in the head, pioneering neurons of the nerve ring and in the DA motorneurones pioneering the dorsal cord. Further, UNC-53 mutants show premature terminations in the ALN and PLN neurons (Stringham et al., 2002). It is possible that both Abi and UNC-53 exhibit overlapping expression patterns that link them together in a pathway that guides cell migration in the nervous system.

Abi has also been linked to cytoskeletal reorganization by virtue of its interaction with Eps8, which is a substrate of several receptor tyrosine kinases, including epidermal growth factor (EGFR) (Scita et al., 1999). Scita et al demonstrated that a tri-complex of Eps8, E3bl (the mouse homologue of Abi) and Sos-1 exhibit Rac-specific guanine nucleotide exchange factor activities which are characterized by membrane ruffling and lamellipodia formation (Scita et al., 1999). Together these results link the Abi/E3b1/B0336.6 protein to both receptor and non-receptor tyrosine kinases as well as GTPase mediated signaling (Courtney et al., 2000).

The corresponding yeast protein, Ysc84p, revealed by the blastp search, is a protein that couples the activities of the actin cytoskeleton with that of the endocytotic machinery. It was identified in a yeast two-hybrid screen for its interaction with Sla1p, a signaling adaptor molecule in yeast. Like the other Abi proteins, Ysc84p, possesses a SH3 domain in its C-terminal region, which is essential for its interaction with Sla1p. Ysc84p localizes to the actin cortex and an intact actin cytoskeleton is necessary for this localization (Dewar, H. et al., 2002). Dewar et al further explain that while Ysc84p and

Sla1p have been shown to physically interact, localization of these proteins is not dependent on each other. Ysc84p localizes to the cortical actin structures at the cell membrane requiring the presence of F-actin for this process whereas Sla1p localizes to the cortex independently of F-actin and Ysc84p (Dewar et al., 2000). This raises the question of how Ysc84p localizes to F-actin as analysis of its sequence does not reveal any known actin-binding motifs. Dewar et al suggest that Ysc84p achieves this feat probably by binding to another molecule that binds actin. In line with this view, they show that WASP/N-WASP/Las17p and the actin-binding protein 1(Abp1p) are required for proper localization of B0336.6/Abi/Ysc84p to the cell cortex (Dewar et al., 2000). The UNC-53 bait construct that was employed in this screen possesses a calponin homology domain that is seen in several actin binding proteins. It is possible that UNC-53 may help link B0336.6/Abi/Ysc84p to the cell cortex through actin binding.

4.5 Sla1p links Verprolin/T12G3.1 to Ysc84p/Abi/B0336.6

It is important to note that two interactors of UNC-53 defined so far, T12G3.1/verprolin and B0336.6/Abi/Ysc84p, both interact with the Sla1p protein in yeast. The yeast interactions DB (GRID) also identified both verprolin and Ysc84p as interactors with Sla1p when it (Sla1p) was used as a bait to search for interacting prey proteins. Verprolin which interacts with Sla1p through a region that has not yet described was discovered in a large-scale analysis in which tandem-affinity purification (TAP) and mass spectrometry were used to characterize multi-protein complexes (Gavin et al., 2002). The SH3 domain in the terminal carboxyl domain of Ysc84p/Abi has been shown to be necessary for interaction with Sla1p (Dewar et al., 2002). At this point, investigating the *in vivo* role of Sla1p may be important in understanding how UNC-53 may be functioning. A blast search of the *C. elegans* protein database using the amino acid sequence of the yeast Sla1p protein revealed a *C. elegans* protein defined by the genomic locus ZK470.5. ZK470.5 is defined as a SH2-SH3 containing protein in *C. elegans*.

The *SLA1* gene was originally identified in a synthetic lethal screen to identify mutants that required the expression of the *Abp1* (actin-binding protein 1) gene in yeast

(Holtzman et al., 1993). Abp1p belongs to a class of actin binding proteins that alter actin dynamics by binding directly to it. In addition, Abp1 has been shown to activate Arp2/3 in yeast (Warren et al., 2002), a complex that appears to be a critical component of the actin cytoskeleton in all eukaryotes and plays a central role in the assembly of actin networks (Goode et al., 2001). Recently, Sawa et al. (2003) have reported the presence of the Arp2/3 complex in *C. elegans* which also consists of seven subunits as follows; actin-related protein 2/3 complex *(arx)-1*, *arx-2*, *arx-3*, *arx-4*, *arx-5*, *arx-6* and *arx-7* (Sawa et al., 2003). Sla1p is the only yeast protein known to date to interact with all three known Arp2/3-activating proteins (Abp1, Las17p/WASP/N-WASP and Pan1p) in yeast (Tang et al., 2000).

Sla1p is a multifunctional protein that has been shown to perform overlapping functions; namely that of coupling the process of endocytosis to factors regulating actin dynamics. Credence for this comes from its ability to interact both with proteins regulating actin dynamics and with proteins required for endocytosis (Warren et al., 2002). In line with the above finding, yeast cells lacking Sla1p possess an aberrant cytoskeleton, exhibit temperature sensitive growth defects (Holtzman et al., 1993) and exhibit a significant although not complete, defect in both fluid-phase and receptor mediated endocytosis (Warren et al., 2002). Sla1p contains three SH3 domains in its N-terminal third and a C-terminal domain made up of multiple repeats rich in proline, glutamine, glycine and threonine. This region of multiple repeats has been shown to interact with the N-terminal EH-domain of End3p and the LR1 domain of Pan1p. Both End3p and Pan1p are EH-domain containing proteins that have been shown to be necessary for endocytosis and normal actin organization in yeast (Tang et al., 2000). What then is Sla1p doing in vivo? Using a combination of immunofluorescence microscopy and biochemical approaches, Warren et al showed localization of Sla1p to the cell cortex through an interaction with End3p, a component of the endocytic machinery. In the absence of End3p, Sla1p shows no significant cortical localization indicating that End3p is required for this localization (Warren et al., 2002). Sla1p however was also shown to bind the actin regulating proteins, Abp1p and the yeast homologue of WASP, Las17p. In addition to earlier reports that Las17p/WASP/N-WASP can immunoprecipitate with Sla1p (Li, 1997), Warren et al., demonstrated in GST pull down

assays that both Las17p and Abp1p bind to Sla1p. Interestingly, other proteins such as actin, Arp2p and cofilin did not bind the Sla1p beads (Warren et al., 2002). These findings led Warren et al to postulate that Sla1p is situated at a relatively static cortical complex that contains proteins known to be required for endocytosis (End3p and Pan1p). Sla1p is then able to couple an association between this complex at the endocytic phase and the actin patches through interactions with both Abp1p and Las17p (which in turn activates the Arp2/3 complex which is important for organizing actin networks).

Possibly, verprolin/T12G3.1 and Ysc84p/Abi/B0336.6 are part of a protein complex that leads to the activation of Arp2/3. What this also means is that the actin polymerization that ensues at the cell surface is specific and required for the endocytic processes occurring at this location as well (Warren et al., 2002).

4.6 UNC-53N interacts with the *spn-4 C. elegans* gene product.

Another gene identified in this yeast two-hybrid screen was the *C. elegans* gene *spn-4*. Also known as *gei-20*, it is situated at the genomic locus ZC404.8. Blastp searches revealed a RNA recognition motif (predicted by PFAM) in Spn-4 (Figure 9) and homologues in other organisms. The corresponding *C. briggsae* homologue, CBP11200, encoded by the *C. briggsae* gene *CBG19017*, is 99.4% similar to *C. elegans* Spn-4 protein. The yeast counterpart, YCL011C, encodes a G-strand binding protein that is 27.4% similar to the *C. elegans* protein. The *Drosophila* homologue, CG32062-PB, whose function is also unknown, is 39.6% similar to the *C. elegans* protein. Both the human and mouse homologues of this protein belong to the Ataxin-2 binding protein family and are 28.2% and 23.9% similar respectively to the *C. elegans* Spn-4.

Spn-4 belongs to a small subfamily of proteins that contain a RNA recognition motif (RRM) and has been identified in several screens linking it to proteins with functions in mitotic spindle orientation and embryogenesis. It was first identified in a screen for mutants defective in mitotic spindle orientation in the posterior cell of the two-cell stage embryo (Gomes et al., 2001). Huang et al. (2002) identified *spn-4* in a yeast two hybrid while looking for proteins that interact with MEX-3, a RNA-binding protein that

localizes the PAL-1 protein to the posterior blastomeres. In yet another study, Ogura and colleagues identified *spn-4* in a screen for proteins that interact with POS-1, a CCCH-type zinc-finger protein that is involved in transducing polarity cues at the posterior end of the embryo. They discovered that SPN-4 and POS-1 work together to regulate *glp-1* mRNA translation (Huang et al., 2002).

In a more recent study, SPN-4 was identified as an interactor with GEX-3 in a yeast two hybrid screen and named GEI-20 for GEXs-interacting molecules (Tsuboi et al., 2002). GEX-3 is vital for the initial processes of body morphogenesis in *C. elegans*. Cells lacking *gex-3* differentiate properly but fail to become organized. The initial ventral migrations of the epidermal cells are absent; consequently dorsal intercalation of the hypodermis that seals the developing embryo does not take place which results in 100% embryonic lethality (Soto et al., 2002). *gex* mutants also show disorganization of other embryonic tissues. The body muscles, which require the hypodermis for proper positioning, are highly disorganized. The pharynx and intestine fail to elongate into the appropriate tubular structures and homozygous postembryonic *gex-3* hermaphrodites show defects in egg-laying (*Egl*) and in gonadal morphogenesis (Soto et al., 2002). Interestingly, similar gonadal morphogenesis defects were observed in the *C. elegans ced-10* mutants. *ced-10* is the *C. elegans* Rac homolog that is required for the engulfment of apoptotic cell corpses and has also been shown to be necessary maternally for embryogenesis (Lindquist et al., 2001). Further to this, Soto et al have showed that arresting embryos produced by *ced-10* null homozygous mothers are severely defective in body morphogenesis including defects in some but not all of the morphogenetic events that require *gex-3* function (Soto et al., 2002). Soto et al give stronger evidence for a possible association between Gex-3 with Ced-10 by examining *gex-3* mutants for phenotypes present in *ced-10* mutants. In addition to its role of engulfing apoptotic cell corpses that arise during the normal course of embryogenesis, *ced-10* is also required for the migration of the distal tip cells (DTCs), which are necessary for the formation of the symmetric U-shaped gonad (Reddien and Horvitz, 2000). Soto et al examined the gonadal morphology in *gex-3* mutants and found them to be defective at levels similar to *ced-10/rac-1* mutants. The gonads failed to move to the dorsal side of the animal, presence of premature bends in the gonad arm and meandering motions, back and forth between the

dorsal and ventral sides (Soto et al., 2002). Interestingly, *unc-53* mutants also show defects in migration of the distal gonad cells (Stringham, personal communication). Together, these findings with those that point to *gex-3* like phenotypes in *ced-10/rac-1* mutants provide support for the involvement of Rac signaling in *gex-3* function. Interestingly homologues of GEX-3 have been implicated in Rac signaling. In fact the mammalian GEX-3 homologue, HEM2 was initially identified as an interactor of Rac1 GTPase (Kobayashi et al., 1998). In *Drosophila*, there is convincing evidence for a link between the GEX-3 homologue HEM2/NAPI/KETTE and Rac signaling. Genetic interactions in *Drosophila*, point to the *Drosophila* homologue KETTE, as being vital for controlling axonal pathfinding and proper actin cytoarchitecture (Hummel et al. 2000).

How do these findings support the plausibility of an interaction between GEI-20, the GEX-3 interactor, and UNC-53N? The findings clearly point to a role for GEX-3 in the regulation of tissue morphogenesis and cell migrations. The RNAi phenotype of *gei-20*, like that of *gex-3* is also embryonic lethal, suggesting that Gei-20/Spn-4 cooperates with GEX-3 in its activities. Like the GEX-3;SPN-4 complex, UNC-53 has also been shown to be important for the migration and outgrowth of a subset of cells including neuronal cells, in which GEX-3 may have an important function. GEX-3 has been linked to Rac signaling which suggests it may modulate cytoskeletal rearrangements indirectly. UNC-53N, the domain that was used in this yeast two-hybrid screen contains a CH and LKK domains, both of which are characteristic of actin binding proteins. It is possible that UNC-53 may serve as a downstream component of a Gei-20;GEX-3 signaling cascade.

On the other hand GEI-20/SPN-4 is classified as a RNA binding protein (Kamath et al., 2003), but UNC-53 does not contain any domains that suggest direct interactions with a RNA binding protein. The possibility that this protein may be situated in the nucleus would make it difficult to reconcile how it would function with UNC-53, which is a cytoplasmic protein (Stringham et al., 2002). It is possible, however, that the RNA binding motif is not required for interaction. GEX-3 for instance was shown to localize predominantly to cell boundaries although it has been suggested that four hydrophobic regions contained within this protein could act as transmembrane regions (Baumgartner et

al., 1995), but there is no evidence for this position. Further more, Kobayashi et al have shown that mammalian GEX-3 can be isolated from brain cytosol (Kobayashi et al., 1998) suggesting that GEX-3 may actually be in the cytoplasm, which in turn would infer that GEI-20/SPN-4 might also localize to the cytoplasm of cells. Finally, Labbe and Goldstein's (2002) question and position on the seeming contradictory roles of Spn-4 is insightful. "How can a single protein do so much?" They suggest that the diversity of SPN-4 functions probably results not only from its predicted ability to bind diverse RNAs, including mRNAs which might allow it to bind and regulate the translation of several maternal mRNAs that are present in the early embryo, but also from its involvement with various protein complexes (Labbe and Goldstein 2002).

4.7 UNC-53N interacts with the F22E12.4 *C. elegans* gene products.

F22E12.4 defines an egg-laying abnormal (*egl*) gene in *C. elegans*. This gene, *egl-9*, was initially identified in a screen for egg-laying defective animals (Trent et al., 1983). Blastp matches to the longest protein product in other animals revealed some similarities, but not with any of known function. The *C. briggsae* protein encoded by *CBGO9602* is 95.7% similar to the *C. elegans* Egl-9, while *YMR317W*, which encodes the *S. cerevisae* homolog, is 58.2% similar to the *C. elegans* Egl-9. The mammalian EGLN1 and the mouse Smooth muscle protein are 55.3% and 35.5% respectively similar to the *C. elegans* Egl-9 protein. The *C. elegans* homologue to this protein is the polycistic kidney disease protein 1, which is 46.5% similar to the Egl-9 protein. SMART protein searches revealed a Zn-finger domain of the MYND type and a prolyl 4-hydroxylase domain of the alpha subunit, while INTERPRO revealed an oxygenase domain (Figure 9). In addition to its egg-laying defect, a loss-of-function mutation in *egl-9* confers a strong resistance to paralysis by *Pseudomonas aeruginosa* (Darby et al., 1999). *egl-9* is expressed in both muscle and neuronal tissues and Darby et al hypothesize that the bacterial toxin may aberrantly activate muscle contraction either by acting directly on Egl-9 or on a pathway that includes it (Darby et al., 1999). The specific function of the protein however remains elusive. Using sequence profile analysis to predict protein function, Aravind and Koonin have placed Egl-9 together with AlkB and leprecan in a novel family of enzymes known

as 2-oxoglutarate (2OG) and Fe(II)-dependent oxygenases (2OG-Fe(II) oxygenase). These enzymes are well represented in eukaryotes and bacteria and they catalyze a variety of interactions typically involving the oxidation of an organic substrate using a dioxygen molecule (Aravind and Koonin, 2001). They predict that EGL-9 is a novel hydroxylase that is capable of eliciting its action through the modification of sidechains of intracellular proteins (Aravind and Koonin, 2001).

Some alleles of *unc-53* are also *egl* (Stringham et al., 2002), raising the possibility that UNC-53 may interact with EGL-9 in some fashion but the mechanisms by which this would occur is purely speculative. It is also possible that EGL-9 does not interact via the prescribed domains.

4.8 UNC-53N interacts with the B0546.2 C. elegans gene product

BO546.2 encodes a *C. elegans* protein of unknown function, which contains an OTU-like (for Ovarian Tumour) cysteine protease domain predicted by PFAM (Figure 9). In order to understand the function of the Ovarian Tumor (OTU) gene, which is involved in oocyte morphogenesis in Drosophila, Makarova et al used the OTU sequence as a query to search the protein database at NCBI using the gapped BLAST program. This analysis revealed the *C. elegans* gene and some highly conserved OTU homologs in humans, HIN1 (for HIV-induced protein) (Makarova et al., 2000). None of the detected homologs of OTU has a known biochemical function. BLASTp matches to other proteins are as follows: the *C. briggsae* protein BP:CBP24133 is 98.4% similar to the *C. elegans* protein while the *M. musculus* protein, TR:Q9JIG6 is 59.4% similar to the *C. elegans* protein. The *Drosophila* protein is encoded by CG6091 and it is 27% similar to the *C. elegans* protein.

4.9 UNC-53N interacts with the F58A4.7a *C. elegans* gene product

The *C. elegans* gene product F58A4.7a, which encodes a protein product of unknown function, was identified as an interactor in the screen with UNC-53N. A

BLASTp search of the amino acid sequence revealed a helix-loop-helix domain. BLASTp searches to the longest protein product also revealed similarities with proteins in other organisms, however none of the homologues has a known biochemical function. The corresponding protein in *C. briggsae* is 88.4% similar to the *C. elegans* protein. In *Drosophila*, *crp* which encodes CG7664-PA is 68.4% similar to the *C. elegans* protein. The yeast protein, YIR019C, which is presumed to be required for invasion and pseudohyphae formation in response to nitrogen starvation, is 46.9% similar to the *C. elegans* protein. Both the *H. sapiens* and *M. musculus* proteins describe a transcription factor AP-4, and are 37.1% and 37.4% respectively similar to the *C. elegans* protein.

4.10 Models for UNC-53 activity

The cell has been defined as "a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines" (Alberts, 1998). Even though proteins constitute a major proportion of the dry mass in a cell, the temptation to think that these proteins are constantly colliding in a random fashion falls far short of the current understanding of intracellular activities. Instead, it is envisaged that almost all major processes are effected by an assembly of 10 or more protein molecules. As each of these assemblies carries out its biological function, it interacts with other complex protein assemblies (Albert, 1998). UNC-53 with its many intriguing domains and pleiotropic array of phenotypes probably functions in a protein complex similar to those described above. This screen for interacting proteins has revealed interactors that may be components of a protein complex in which UNC-53 may function.

Cell migration plays a crucial role in various aspects of development in both vertebrates and invertebrates. Consequently, understanding the factors that initiate, promote extension and outgrowth continue to be highly studied. It is known that the process of migration usually includes a complex interplay between extracellular cues, cell surface receptors that internalizes the external signal and a series of protein complexes that relay this information to the cytoskeleton, which in turn orients the cell in a specific

direction. Bentley and O'Connor hypothesize that when the signal reaches the cytoskeleton, there is a localized accumulation of polymerized actin at the growth cone that results in extension in a preferred direction and in like manner a repulsive extracellular cue would result in a reduction or destabilization of polymerized actin at the growth cone (Bentley and O'Connor, 1994). Currently two models exist for UNC-53 function within the cell. Stringham et al. (2002) suggest a model in which UNC-53 activity is activated by SEM5, resulting in the localization of UNC-53 to the actin cytoskeleton. UNC-53 is in turn activated perhaps through nucleotide binding and promotes the crosslinking of actin molecules to orient, guide and promote growth cone extension in a specific direction. Alternatively, Stringham and colleagues propose that UNC-53 may not be directly involved in modulating the actin cytoskeleton, but may serve as an intermediary in a relay of events whose ultimate function is extension of the growth cone (Stringham et. al. 2002).

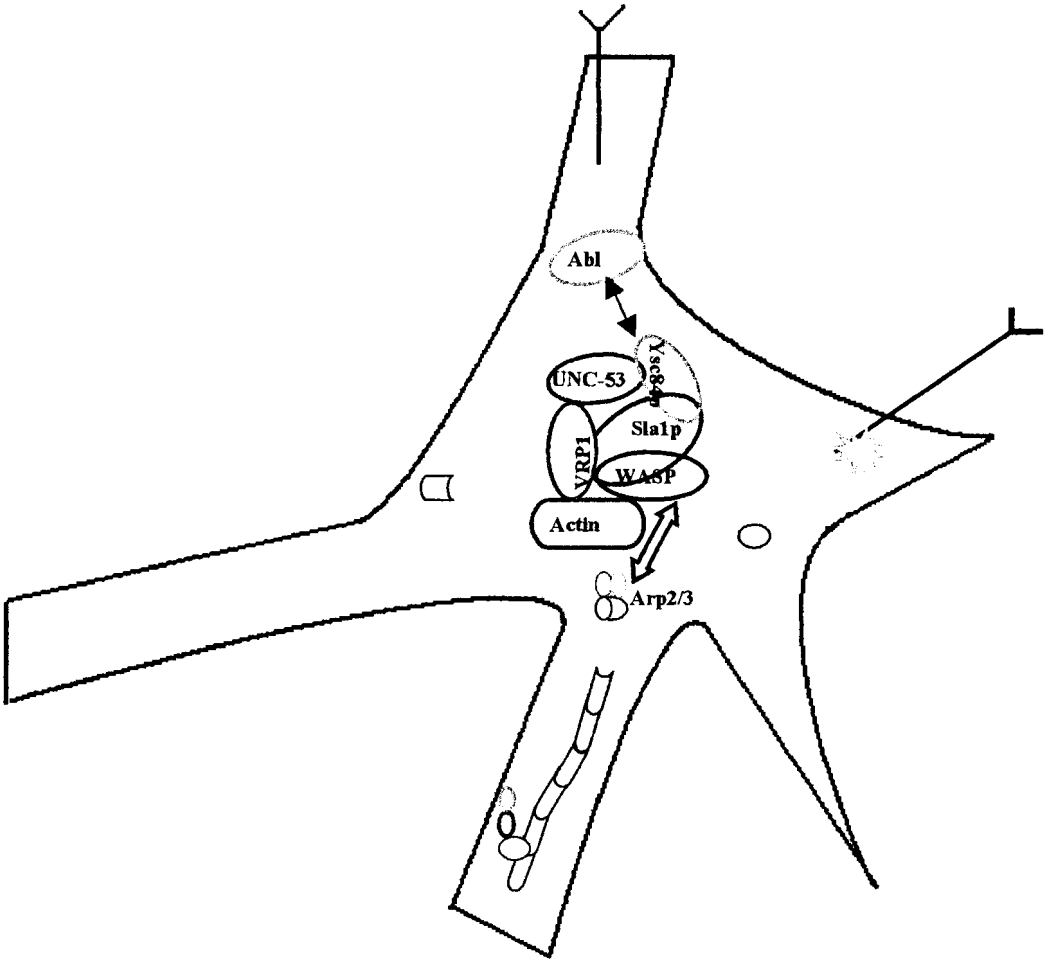
The novel interactors that have been identified clearly link UNC-53 to cytoskeletal events. It has been postulated that verprolin modulates the actin cytoskeleton possibly by serving as an anchor for actin. This idea of an anchor for actin concurs well with the view that sustained actin polymerization occurs to stabilize the growth cone in a specific direction. Perhaps verprolin binds and anchors actin and this location then serves as a nucleation site for further actin polymerization. Preliminary results have suggested that UNC-53 binds F-actin in a co-sedimentation assay (Stringham, unpublished results). Interestingly, this screen did not identify interactions with actin. It is possible that UNC-53 also localizes to the actin filaments and then as part of the protein complex at the actin filament, binds verprolin on a secondary level or verprolin/T12G3.1 and Abi/Ysc84p/B0336.6 may act to localize UNC-53 to actin. Importantly, both verprolin and Abi interact with other actin binding proteins and verprolin interacts with actin directly. Verprolin via its carboxyl terminus binds the amino terminus of Las17p/WASP/N-WASP (Naqvi et al., 1998), and in addition to binding actin, Las17p recruits through its carboxyl terminal the Arp2/3 complex to initiate and promote nucleation of actin filaments (Naqvi et al., 1998, Hu and Reichardt, 1999).

Both verprolin/T12G3.1 and Ysc84p/Abi/B0336.6 display a cortical localization that is reminiscent of actin staining. In mammalian cells Ysc84p/Abi/B0336.6 localize to protruding lamellipodia and filopodia tips, where new actin subunits are added to growing filaments. Stradal et al. (2001) reported a very interesting finding; that in some cells, rosette-like arrays were observed that harbored Abi proteins as well as components of the Arp2/3 complex (Stradal et al., 2001).

Together these results point to the formation of a protein complex around Las17p/WASP/N-WASP and the Arp2/3p complex. How are all these proteins localized to this site? Are they recruited together, do they have other functions such that may be found at other sites as well? These are all questions that would need to be addressed to understand how UNC-53 may be interacting *in vivo*. Possibly, a complex consisting of UNC53/NAV, verprolin/T12G3.1, Abi/Ysc84p/B0336.6, Sla1p/K06A9.1a and actin is formed at a growth cone. Establishment of this complex may activate Las17P/WASP/N-WASP, which in turn activates the Arp2/3 complex (Figure 14). It has been suggested that Abi associates with the Arp2/3 complex, which indicates that both UNC-53 and Abi/Ysc84p/B0336.6 may be situated at the nucleating site, perhaps stabilizing actin polymerization. This model, like the first proposed by Stringham et. al. (2002) would infer that UNC-53 is directly involved with actin polymerization in contrast to serving as an intermediary that relays information which ultimately modulates actin dynamics.

Figure 14: Model for UNC-53 activity.

Abl kinases may mediate their activity through Abi/Ysc84p/B0336.6 which act as adaptors for the Abl family of NRTKs. UNC-53 binds both verprolin/T12G3.1 and Ysc84p/Abi/B0336.6. Sla1p/ZK470.5 links verprolin/T12G3.1 to Ysc84p/Abi/B0336.6, while binding to WASP as well. Verprolin also binds WASP, which activates the Arp2/3 complex. Alternatively the protein complex may form at the protruding end of the filaments where verprolin anchors actin and the other molecules may also facilitate stabilization.


Abl
UNC-53
Sla1p
VRP1
WASP
Actin
Arp2/3

5. FUTURE DIRECTIONS:

The data obtained from this yeast two hybrid screen suggest that through its amino terminus, UNC-53 may interact with novel molecules not identified previously through genetic screens. It is possible that in genetic screens, mutants in which the interaction complex is faulty are lethal at various stages in development, consequently making it difficult to characterize these mutations. Genetics may also require extensive screens of worms for mutations that may be elusive or easy to overlook.

Secondly, these candidates have not been identified in large-scale yeast two hybrid screens and this may be due to the large size (5kb) of the full-length *unc-53* transcript, which probably is under represented in any cDNA library. In addition the prevalence of smaller transcripts from internal promoters may mean most transcripts if present are devoid of the N-terminal region. However in this study, by using portions of UNC-53 as bait, interactions which otherwise may have been missed are revealed. This point is important because these interactions may be only relevant to the full length isoform since the N-terminal region is absent in all other isoforms driven by internal promoters or resulting from alternative splicing events (Stringham et al., 2002). In addition, large proteins may not fold properly in yeast, which can prevent the detection of some interactions. Interestingly, preliminary attempts in this study to use the full-length *unc-53* gene product as bait did not yield results.

Of course the two-hybrid screen has its own caveats and as rightly suggested by Fields and Sternglanz (1994), an interaction between a target and library-encoded proteins does not necessarily mirror the *in vivo* relationship. Interactions detected in a two-hybrid screen still need to be confirmed by other biological or biochemical experiments (Fields and Sternglanz, 1994). It is also possible that the two-hybrid system may detect an interaction between domains that may not normally be accessible or exposed in the native protein, especially when short sequences are used to test an interaction (Fields and Sternglanz, 1994). The yeast two hybrid screen may also fail to detect some interactions for the following reasons: (a) the failure of the bait and prey fusion proteins to localize to the nucleus (b) interaction between proteins may depend on post-translational

modifications that are absent in yeast cells. For instance, proteins that contain a SH2 domain require phosphotyrosine in the proteins to which they bind, and this phosphorylation may not occur in yeast (c) accessibility to certain domains may be blocked. For example if the protein-protein interaction is mediated through the amino-terminus of either of the proteins, the conventional orientation of the hybrid constructs may obscure an interaction because the transcription factor domain blocks accessibility (Fields et al., 1994; Walhout et al., 2000). It is possible that accessibility to the N-terminus may be diminished in a fusion protein between GAL4 and the full length UNC-53 protein. This may explain why interactions were not detected when the longer UNC-53 constructs were used as bait.

It is worth noting, however, that some of the domains found in the positive candidates analyzed are consistent with roles in signal transduction and/or cytoskeleton dynamics. Two of the six classes of cDNAs point to actin binding proteins, as well as a significant signaling molecule that may be important in linking the activation of a receptor tyrosine kinase with the organization of the actin cytoskeleton, all roles that concord with UNC-53 as a cytoplasmic signaling molecule. The findings suggest that UNC-53 may be modulating cytoskeletal organization via actin binding proteins as opposed to binding actin directly, even though *in-vitro* evidence suggests that UNC-53 co-sediments with actin (Stringham, E. et al., 2002). In addition, this study did not reveal any direct interactions between actin and the N-terminus of UNC-53. Interestingly there is evidence that suggests that the single CH domain of calponin is neither sufficient nor necessary for binding F-actin (Gimona, M., et al., 1998). It is possible that a direct interaction with actin may be detected if a longer cDNA consisting of both LKK domains is used as bait in a two-hybrid screen. The significant proportion of actin binding homology proteins in this screen (33%) is probably due to the calponin domain used in screening; a domain found usually in actin binding proteins. It is possible that we have designed a screen that would identify predominantly downstream components of this pathway. To further characterize this pathway, a screen using the mid-portion of the UNC-53 protein would probably identify other upstream components because of the presence of the SH3 binding motif that typifies several intermediate signaling molecules. Intriguingly Abi/Ysc84p/B0336.6, SH3 signaling molecule was identified even though

the bait did not contain the putative SH3 binding region. Presumably, UNC-53 interacts with the candidate via a region distinct from the SH3 domain. It would be interesting to see what biological candidates are identified from screens using the other bait constructs.

RNAi is also a useful and rapid means for ascertaining gene function. Unfortunately, no definitive RNAi phenotypes were obtained for four classes of cDNAs based on results from the RNAi genome wide *C. elegans* project. The two classes of cDNAs that had RNAi phenotypes were defined as embryonic lethals (Kamath et al., 2003). Neuronal genes are often resistant to RNAi procedures and UNC-53 happens to be involved in neuronal guidance and as such obtaining a phenotype may be difficult via traditional RNAi methods. Secondly, because these analyses are genome wide, they may omit features that are specific to certain genes that are not immediately obvious. It may be necessary to examine the canals, neurons and sex muscles to see if these structures are truncated or misguided in animals that have been subjected to RNAi (Table 5).

Overlapping pattern of gene expression between two genes can be a useful determinant to confirm *in-vivo* interaction, as both genes would have to be expressed in the same region of the cell for physical interaction to occur. The transgenic genome wide project (Baillie, unpublished results) has provided some preliminary data as observed by the expression pattern of T12G3.1, which shows some overlap with the pattern of *unc-53* expression (Figure 8). Two other genes were analyzed, B0336.6 and ZC404.8, but no expression data was obtained (Table 6), which may be indicative of a short promoter region used in the fusion construct.

Two-hybrid protein interaction data can be verified using gluthatione-*s*-transferase (GST) pull-down assays. This involves expressing the protein of interest in bacteria as a GST fusion protein and bound to glutathione beads. A cell extract is passed over the beads and proteins that interact with the protein of interest bind to the beads while non-interacting proteins do not. The protein complex can then be analyzed by SDS-PAGE. Similarly UNC-53 has been conveniently fused to a histidine tag, which can be bound to a nickel column and used to pull down interactors. Prey plasmids in the two-hybrid screen contain cDNAs of the interactors and these DNAs can also be transferred into appropriate protein expression vectors for further analysis.

Since a non-lethal allele of *let-52* is available (Baillie, personal communication) it would be beneficial to determine if T12G3.1 corresponds to the *let-52* gene. Northern blot analysis or RT-PCR could be used to ascertain the amount of T12G3.1 RNA transcript made in N2 wild type animals compared to the amount made in mutant *let-52* worms. A reduction in the T12G3.1 transcript in the mutants would confirm that T12G3.1 was indeed *let-52*. Sequencing the *let-52* gene to determine the position of the molecular lesion would also help in clarifying its identity.

Another biochemical approach that could be used to identify other components of this pathway is co-immunoprecipitation. In co-immunoprecipitation, endogenously interacting proteins can be detected using specific antibodies against the protein of interest (or bait protein) or an epitope tag to which the protein of interest has been fused. The protein of interest is precipitated from cell extracts together with other cellular proteins for which it has affinity, the entire complex is collected by binding it to Sepharose beads, and the complex is analyzed by SDS-PAGE. Proteins may then be further characterized and purified and their identities determined by peptide sequencing or mass spectrometry. UNC-53N has been conveniently fused to a histidine tag and polyclonal antibodies have been raised against this portion. These polyclonal sera could therefore be used for future co-immunoprecipitation experiments.

In summary, there are several tools available that can be used both for the verification of the data already obtained, and for extending the UNC-53 pathway by the identification of additional molecules. Finally, this study has revealed new molecular partners for UNC-53 that should help in elucidating the role this complex protein plays in directing cell migration.

6. REFERENCES:

Alberts, B. (1998). The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291-294.

Aravind, L. and Koonin, E. (2001). The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. *Genome Biol.* **2**, 1-8.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidmand, J. G., Smith, J. A. et al. (1993). *Current Protocols in Molecular Biology*. New York: J. Wiley & Sons.

Baumgartner, S., Martin, D., Chiquet-Ehrismann, R., Sutton, J., Desai, A., Huang, I., et al. (1995). The HEM proteins: A novel family of tissue-specific transmembrane proteins expressed from invertebrates through mammals with an essential function in oogenesis. *J. Mol. Biol.* **251**, 41-49.

Bentley, D. and O'Connor, T. P. (1994). Cytoskeletal events in growth cone steering. *Curr. Opin. Neurobiol.* **4**, 43-48.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acid Res.* **7**, 1513-1523.

Branda, C. S. and Stern, M. J. (2000). Mechanisms controlling sex myoblast migration in *Caenorhabditis elegans* hermaphrodites. *Dev. Biol.* **226**, 137-151.

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

Chan, S. S., Zheng, H., Su, M. W., Wilk, R., Killeen, M. T., Hedgecock, E. M., Culotti, J. G. (1996). UNC-40, a *C. elegans* homolog of DCC (Deleted in Colorectal Cancer) is required in motile cells responding to UNC-6 netrin cues. *Cell* **87**, 187-195.

Chen, E. B., Branda, C. S. and Stern, M. J. (1997). Genetic enhancers of *sem-5* define components of the gonad-independent guidance mechanism controlling sex myoblast migration in *Caenorhabditis elegans* hermaphrodites. *Dev. Biol.* **182**, 88-100.

Clark, S. G., Stern, M. J. and Horvitz, H. R. (1992). *C. elegans* cell-signalling gene *sem-5* encodes a protein with SH2 and SH3 domains. *Nature* **356**, 340-344.

Confalonieri, F. and Duguet, M. (1995). A 200-amino acid ATPase module in search of a basic fuction. *BioEssays* **17**, 639-650.

Courtney, K. D., Grove, M. Vandongen, H., Vandongen, A., LaMantia, A. and Pendergast, A. M. (2000). Localization and Phosphorylation of Abl-Interactor Proteins, Abi-1 and Abi-2, in the Developing Nervous System. *Molecular and Cellular Neuroscience* **16**, 244-257.

Dai, Z., and Pendergast, A. M. (1995). Abi-2, a novelSH3-containing protein interacts with the c-Abl tyrosine kinase and modulates c-Abl transforming activity. *Genes Dev.* **9**, 2569-2582.

Darby, C., Cosma, C. L. Thomas, J. H. and Manoil, C. (1999). Lethal paralysis of *Caenorhabditis elegans* by *Pseudomonas aeruginosa. Proc. Natl. Acad. Sci. USA* **96**, 15202-15207.

Dewar, H., Warren, D. T., Gardiner, F. C., Gourlay, C. G., Satish, N., Richardson, M. R. et al. (2002). Novel Proteins Linking the Actin Cytoskeleton to the Endocytotic Machinery in *Saccharomyces cerevisiae. Mol. Biol. Of the Cell* **13,** 3646-3661.

Donnelly, S. F., Pocklington, M. J., Pallotta, D. and Orr, E. (1996). A proline-rich protein, verprolin, involved in cytoskeletal organization and cellular growth in the yeast *Saccharomyces cerevisiae. Mol Microbiol.* **10**, 585-596.

Durfee, T., Bechere, K., Chen, P. L., Yeh, S. H., Yang, Y., Kilburn, A. E. et al. (1993). The retinoblastoma protein associates with the protein phosphotase type 1 catalytic subunit. *Genes and Dev.* **7,** 555-569.

Elledge, S. J. (1993). Protein Interaction Cloning Using the Two Hybrid System: Library transformation using Y190. Baylor College of Medicine, Houston Texas, U.S.A.

Epstein, A. C. R., Gleadle, J. M., McNeill, L. A., Hewitson, K. S., O'Rourke, J., Mole, D. R. et al (2001). *C. elegans* EGL-9 and Mammalian Homologs Define a Family of Dioxygenases that Regulate HIF by Prolyl Hydroxylation. *Cell* **107**, 43-54.

Epstein, H. F. and Shakes, D. C. (1995). *Caenorhabditis elegans Modern Biological Analysis of an Organism.* New York: Academic Press.

Fields, S. and Song, O. (1989). A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245-246.

Fields, S. and Sternglanz, R. (1994). The two-hybrid system: an assay for protein-protein interactions. *TIG* **10**, 286-292.

Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M.,Sohrmann, M. and Ahringer, J. (2000). Functional genomic analysis of C. elegans chromosome 1 by systematic RNA interference. *Nature* **408**, 325-330.

Gavin, A., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A. et al. (2002). Functional organization of yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147.

Gimona, M. and Mital, R. (1998). The single CH domain of calponin is neither sufficient nor necessary for F-actin binding. *Jour. of Cell Sci.* **111**, 1813-1821.

Gomes, J., Encalada, S. E., Swan, K. A., Shelton, C. A. Carter, J. C. and Bowerman, B. (2001). The maternal gene *spn-4* encodes a predicted RRM protein required for mitotic spindle orientation and cell fate patterning in early *C. elegans* embryos. *Dev.* **128**, 4301-4314.

Goode, B. L., Rodal, A. A., Barnes, G. and Drubin, D. G. (2001). Activation of the Arp2/3 complex by the actin filament binding protein Abp1p. *J. Cell Biol.* **153**, 627-634.

Hall, A. (1998). Rho GTPases and the Actin Cytoskeleton. *Science* **279**, 509-514.

Hao, J. C., Yu, T. W., Fujisawa, K., Culotti, J. G., Gengyo-Ando, K., Mitani, S. et al. (2001). *C. elegans* Slit acts in midline, dorsal-ventral, and anterior-posterior guidance via the SAX-3/Robo Receptor. *Neuron* **32**, 25-38.

Hedgecock, E. M., Culotti, J. G., Hall, D. H. (1990). The *unc-5*, *unc-6*, and *unc-40* genes guide circumferential migrations of pioneer axons and mesodermal cells on the epidermis in *C. elegans*. *Neuron* **4**, 61-85.

Hedgecock, E. M., Culotti, J. G., Hall, D. H. and Stern, B. D. (1987). Genetics of cell and axon migrations in *Caenorhabditis elegans*. *Dev.* **100**, 365-382.

Hoffman, C. S. and Winston F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* **57**, 267-72.

Holtzman, D. A., Yang, S. and Drubin, D. G. (1993). Synthetic-lethal interactions identify two novel genes, *SLA1* and *SLA2*, that control membrane cytoskeleton assembly in *Saccharomyces cerevisiae*. *J. Cell Biol.* **122**, 635-644.

Hu, S. and Reichardt, L. F. (1999). From Membrane to Cytoskeleton: Enabling a Connection. *Neuron* **22**, 419-422.

Huang, N. N., Mootz, D. E., Walhout, A. J. M., Vidal, M. and Hunter, C. P. (2002). MEX-3 interacting proteins link cell polarity to asymmetric gene expression in *Caenorhabditis elegans*. *Dev.* **129**, 747-759.

Huang, X., Huang, P., Robinson, M. K., Stern, M. J., and Jin, Y. (2003). UNC-71, a disintegrin and metalloprotease (ADAM) protein, regulates motor axon guidance and sex myoblast migration in *C. elegans*. *Dev.* **130**, 3147-3161.

Hummel, T., Leifker, K. and Klambt, C. (2000). The *Drosophila* HEM-2/NAP1 homolog KETTE controls axonal pathfinding and cytoskeletal organization. *Genes & Dev.* **14,** 863-873.

Hynes, R. O. and Lander, A. D. (1992). Contact and Adhesive Specificities in the Associations, Migrations, and Targeting of Cells and Axons. *Cell* **68**, 303-322.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S. (2001). Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* **98**, 218-223.

Juang, J. and Hoffmann, M. F. (1999). *Drosophila* Abelson interacting protein (dAbi) is a positive regulator of Abelson tyrosine kinase activity. *Oncogene* **18**, 5138-5147.

Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M. et al. (2003). Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421,** 231-237.

Keegan, L., Gill, G., Ptashne, M. (1986). Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231,** 699-704.

Keynes, R. and Cook, G. M. W. (1995). Axon Guidance Molecules. *Cell* **83,** 161-169.

Kobayashi, K., Kuroda, S., Fukata, M., Nakamura, T., Nagase, T., Nomura, N. et al. (1998). p140Sra-1 (Specifically Rac1- associated protein) is a novel specific target for Rac1 small GTPase. *J. Biol. Chem.* **273,** 291-295

Kokel, M., Borland, C. Z., DeLong, L., Horvitz, H. R. and Stern, M. (1998). *clr-1* encodes a receptor tyrosine phosphatase that negatively regulates an FGF receptor signaling pathway in *Caenorhabditis elegans*. *Genes & Dev.* **12**, 1425-1437.

Kuwabara, P. E., and O'Neil, N. (2001). The use of functional genomics in *C. elegans* for studying human development and disease. *J. Inherit. Metab. Dis.* **24**, 1-11.

Labbe, J. and Goldstein, B. (2002). Embryonic Development: A New SPN on Cell Fate Specification. *Curr. Biol.* **12,** R396-R398.

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

Lauffenburger, D. A. and Horwitz, A. F. (1996). Cell Migration: A Physically Integrated Molecular Process. *Cell* **84**, 359-369.

Li, R. (1997). Bee1, a yeast protein with homology to Wiscott-Aldrich Syndrome Protein, is critical for the assembly of cortical actin cytoskeleton. *J. Cell Biol.* **136**, 649-658.

Lundquist, E. A., Reddien, P. W., Hartwieg, E., Horvitz, H. R., and Bargmann, C. I. (2001). Three *C. elegans* Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. *Dev.* **128,** 4475-4488

Madura, K. (2002). The ubiquitin-associated (UBA) domain: on the path from prudence to prurience. *Cell Cycle* **1**, 235-244.

Maes, T., Barcelo, A. and Buesa, C. (2002). Neuron navigator: a human gene family with homology to *unc-53*, a cell guidance gene from *Caenorhabditis elegans. Genomics* **80**, 21-30.

Makarova, K. S., Aravind, L., Koonin, E. V. (2000). A novel superfamily of predicted cysteine proteases from eukaryotes, viruses and Chlamydia pneumoniae. *Trends Biochem Sci* **25**, 50-52.

Manser, J. Roonprapunt, C. and Margolis, B. (1997). *C. elegans* cell migration gene *mig-10* shares similarities with a family of SH2 Domain proteins and acts cell nonautonomously in excretory canal development. *Dev. Biol.* **184**, 150-164.

Merrill, R. A., Plum, L. A., Kaiser, M. E. and Clagett-Dame, M. (2002). A mammalian homolog of unc-53 is regulated by all-trans retinoic acid in neuroblastoma cells and embryos. *Proc. Natl. Acad. Sci. USA* **99**, 3422-3427.

Munn, A. L., Stevenson, B. J., Geli, M. I. and Riezman, H. (1995). *end5*, *end6*, and *end7*: mutations that cause actin delocalization and block the internalization step of endocytosis in *Saccharomyces cerevisiae. Mol. Biol. Cell* **6**, 1721-1742.

Naqvi, S. N., Zahn, R., Mitchell, D. A., Stevenson, B. J. and Munn, A. L. (1998). The WASp homologue Las17p functions with the WIP homologue End5p/verprolin and is essential for endocytosis in yeast. *Curr. Biol.* **8**, 959-962.

Ogura, K., Shirakawa, M., Barnes, T. M., Hekimi, S. and Ohshima, Y. (1997). The UNC-14 protein required for axonal elongation and guidance in *Caenorhabditis elegans* interacts with the serine/threonine kinase UNC- 51. *Genes Dev.* **11**, 1801-1811.

Ponting, C. P., Blake, D. J., Davies, K. E., Kendrick-Jones, J. and Winder, S. J. (1996). ZZ and TAZ: new putative zinc fingers in dystrophin and other proteins. *Trends Biochem. Sci.* **21**, 11-13.

Reddien, P. W. and Horvitz, H. R. (2000). CED-2/CrkII and CED-10/Rac control phagocytosis and cell migration in *Caenorhabditis elegans. Nature Cell Biol.* **2**, 131-136.

Rentschler, S., Linn, H., Deininger, K., Bedford, M. T., Espanel, X., Sudol, M. (1999). The WW domain of dystrophin requires EF-hands region to interact with beta-dystroglycan. *Biol. Chem.* **380**, 431-42.

Riddle, D. L., Blumenthal, T., Meyer, B. J. and Priess, J. R. (1997). *C. elegans* II. New York: Cold Spring Harbor Press.

Roumanie, O., Peypouquet, M., Thoraval, D., Doignon, F. and Crouzet, M. (2002). Functional interactions between the *VRP1-LAS17* and *RH03-RHO4* genes involved in actin cytoskeleton organization in *Saccharomyces cerevisiae*. *Curr. Genet.* **40**, 317-325.

Rubin G. M., Yandell M. D., Wortman J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K. et al. (2000). Comparative genomics of the eukaryotes. *Science* **287**, 2204-2215.

Sawa, M., Suetsugu, S., Sugimoto, A., Miki, H., Yamamoto, M. and Takenawa, T. (2003). Essential role of the *C. elegans* Arp2/3 complex in cell migration during ventral closure. *J. of Cell Science* **116**, 1505-1518.

Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P. and Bork, P. (2000). SMART: A Web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28,** 231-234.

Schultz, J., Milpetz, F., Bork, P. and Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 5857-5864.

Scita, G., Nordstom, J., Carbone, R., Tenca, P., Giardina, G., Gutkind, S. et al. (1999). EPS8 and E3B1 transduce signals from Ras to Rac. *Nature* **401**, 290-293.

Selfors, L., Schutzman, J. L., Borland, C. Z. and Stern, M. (1998). *soc-2* encodes a leucine-rich repeat protein implicated in fibroblast growth factor receptor signaling. *Proc. Natl. Acad. Sci. USA* **95**, 6903-6908.

Serafini, T. and Kennedy, T. E. (1994). The netrins define a family of axon outgrowth-promoting proteins homologous to *C. elegans* UNC-6. *Cell* **78,** 409-424.

Soto, M. C., Qadota, H., Kasuya, K., Inuoue, M., Tsuboi, D., Mello, C. C. et al (2002). The GEX-2 and GEX-3 proteins are required for tissue morphogenesis and cell migrations in *C. elegans*. *Genes & Dev.* **16**, 620-632.

Stern, M. J., Marengere, L. E., 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*. *Mol. Biol. Cell* **4**, 1175-1188.

Steven, R., Kubiseski, T. J., Zheng, H., Kulkarni, S., Mancillas, J., Ruiz Morales, A. et al. (1998). UNC-73 activates the Rac GTPase and is required for cell and growth cone migrations in *C. elegans*. *Cell* **92**, 785-795.

Stradal, T., Courtney, K. D., Rottner, K., Hahne, P., Small, J. V. and Pendergast, A. M. (2001). The Abl interactor proteins localize to sites of actin polymerization at the tips of lamellipodia and filopodia. *Curr. Biol.* **11**, 891-895.

Stradal, T., Kranewitter, W., Winder, S. J. and Gimona, M. (1998). CH domains revisited. *FEBS Lett.* **431**, 134-137.

Stringham, E., Pujol, N., Vandekerckhove, J., Bogaert, T. (2002). *unc-53* controls longitudinal migrations in *C. elegans*. *Dev.* **129**, 3367-3379.

Sulston, J. E., Schierenberg, E., White, J. G. and Thompson, J. N. (1983). The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64-119.

Tang, H. Y., Xu, J. and Cai, M. J. (2000). Pan1p, End3p, and Sla1p, three yeast proteins required for normal cortical actin cytoskeleton organization, associate with each other and play essential roles in cell wall morphogenesis. *Mol. Cell. Biol.* **20,** 12-25.

Tessier-Lavigne, M. and Goodman, C. S. (1996). The molecular biology of axon guidance. *Science* **274**, 1123-1133.

The *C. elegans* Sequencing Consortium. (1998). Genome sequence of the Nematode *C. elegans*: A Platform for Investigating Biology. *Science* **282**, 2012-2018.

Trent, C., Tsung, N. and Horvitz, H. R. (1983). Egg-laying defective mutants of the nematode of *C. elegans*. *Genetics* **104**, 619-647.

Tsuboi, D., Qadota, H., Kasuya, K., Amano, M., and Kaibuchi, K. (2002). Isolation of the Interacting Molecules with GEX-3 by a Novel Functioning Screening. *Biochem. and Biophys. Res. Comm.* **292**, 697-701.

Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R. et al. (2000). A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-627.

Vaduva, G., Martin, N. C. and Hopper, A. K. (1997). Actin-binding Verpolin Is a Polarity Development Protein Required for the Morphogenesis and Function of the Yeast Actin Cytoskeleton. *Jour. of Cell Biol* **139**, 1821-1833.

Van Troys, M., Vandekerckhove, J. and Ampe, C. (1999). Structural modules in actin-binding proteins: towards a new classification. *Biochim. Biophys. Acta* **1448**, 323-348.

Venter, J. C., Adam,s M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G. et al. (2001). The sequence of the human genome. *Science* **291**, 1304-13051.

Walhout, A. J. M., Sordella, R., Xiaowei, L., Hartley, J. L., Temple, G. F., Brasch, M. A., Thierry-Mieg, N., Vidal, M. (2000). Protein Interaction Mapping in C. elegans Using Proteins Involved in Vulval Development. *Science* **287,** 116-122.

Walhout, A. J. M., Boulton, S. J. and Vidal, M. (2000). Yeast two-hybrid systems and protein interaction mapping projects for yeast and worm. *Yeast* **17**, 88-94.

Warren, D. T., Andrews, P. D. Gourlay, C. W. and Ayscough, K. R. (2002). Sla1p couples the yeast endocytic machinery to proteins regulating actin dynamics. *Jour. of Cell Science* **115**, 1703-1715.

Wightman, B., Baran, R. and Garriga, G. (1997). Genes that guide growth cones along the *C. elegans* ventral nerve cord. *Dev.* **124**, 2571-2580.

Wolf, F. W., Hung, M. S., Wightman, B., Way, J., Garriga, G. (1998). *vab-8* is a key regulator of posteriorly directed migrations in *C. elegans* and encodes a novel protein with kinesin motor similarity. *Neuron* **20**, 655-666.

Wyers, F., Petitjean, A. M., Dru, P., Gay, P. and Contamine, D. (1995). Localization of Domains within the *Drosophila* Ref(2)P Protein Involved in the Intracellular Control of Sigma Rhabdovirus Multiplication. *Jour. of Virology* **69**, 4463-4470.

Yu, H., Chen, J. K., Feng, S., Dalgarno, D. C., Brauer, A. W. and Schreiber, S. L. (1994). Structural basis for the binding of proline-rich peptides to SH3 domains. *Cell* **76,** 933-945.

7. APPENDIX 1

A. Sequencing primers for cDNA inserts in the GALAD pACT plasmid

5' pACT sequencing primer	5' TTCGATGATGAAGATACCCC 3'
3' pACT sequencing primer	3' AGGCAAAACGATGTATAA 5'

B. Partial sequences of cDNAs obtained from the yeast two hybrid screen (courtesy of the NAPS Unit at UBC, Vancouver). Sequencing was performed using the 5' pACT sequencing primer.

T12G3.1 (pACT1)

5'AGATCTGGAATTCGGATCCTCGAGATTTACCCTTTGAGTTCACCAATTCGGATCAGAT
TATGTCTAAGATTCGAATCACTACATTCTCTCGAACTCCTCAACTTTTTGTTCGCCTGGA
AACTGTTCCCCTTCGTCCAACAGCCCTACCATCAGTGAGCACTGAAACTGCCAAAAAAGC
TTCCGATGCTGATCGTGAATCAAAGAGCAGATCCTCGCTGAAGCGTATCGAGAACAACCA
GGAGCTCACATTGAAGACGCTGGAGAAGTTGCAAAAGGAGCTCTCAAGAGTTAGCGCTTT
GGAGACGATGGTTGCTGATCTTCATGCAAAGGAAAGAGAACCTGTTGTCGAGGAAGTTCA
GAAGCCATTCCCTCCAGTGAAGCACTCGGCGTCTTGTGATGCTTGCCTCGGTGACATCAT
CGGACATAGGTTCAAGTGTCTCGAATGTGCCGATTATGATTTGTGCGAGAAGTGTGAGAG
AAAATCTTTACATTATGAGCATGCAATGATCAGAATTGTGGAGCCAAAGAAAACAATGAT
CCCATCATATGTGACATCAAACTCTCCGAACAATGTCTTCCCAACTTACATGCAGCACAA
GATCCAACAACCACAAACTATCAAGGGGATCAACAAGACTGTTATCACGAGACTCGTTCC
TTCACATCAAGTTAATCCAGCTCCAATGCCTCAAGCTCCACCAGTCTCCACTCCAGTTCC
ATTCCCTCCATCTCCAACCATCGGAAATGGAGCAGTTCCAAAACCAGCCCAGGAGCCTCG
GATCCCACCATCTCCAACTTCTGCTCTACCACCACCACAATTCTTCAATGAGGAGACCCG
TGCAGCCATTATTAATAGTGCAGCGTTGCTCAAGGAAACTTTCAGTACTTTGAGCAGAAG
CTTCTTGGATTTGGAGATGGTGCACAGGNCAATATGGATAAAAGATCATTGAACGCTCGA
CCATGGAAGCAGGAATGACCAATTCGAAGAGAAAATGATANTTGAAAGACACTGTGAGGA
GAATATCTTGGACGGAACTGCTGAACTGANAAAGCGAGGAATCTTCGAAAAGCAGCAGAT
GCTGNCTGGACAAGATCAGAAGCTAAANAATCTGCTCANAAACTGCGAAGATATCCATGC
ACTTGTGNGNGNTGNTGACAGCCGCTGACAAAACCANAAAACTGAGCNCTCTCTGATANN
AGACTTGTNGNTATGN-----3'

T12G3.1 (pACT8)

5′AGATCTGGAATTCGGATCCTCGAGAGTACACGGCTCTTCAACAGATCTCCGAGAAGGC
AAGAGAACTGTTCCCATCGGCTGATTATCGTCTCTACTACGGTGATGTCAACGGACCAAT
CTTTGAGTTCACCAATTCGGATCAGATTATGTCTAAGATTCGAATCACTACATTCTCTCG
AACTCCTCAACTTTTTGTTCGCCTGGAAACTGTTCCCCTTCGTCCAACAGCCCTACCATC
AGTGAGCACTGAAACTGCCAAAAAAGCTTCCGATGCTGATCGTGAATCAAAGAGCAGATC
CTCGCTGAAGCGTATCGAGAACAACCAGGAGCTCACATTGAAGACGCTGGAGAAGTTGCA
AAAGGAGCTCTCAAGAGTTAGCGCTTTGGAGACGATGGTTGCTGATCTTCATGCAAAGGA
AAGAGAACCTGTTGTCGAGGAAGTTCAGAAGCCATTCCCTCCAGTGAAGCACTCGGCGTC
TTGTGATGCTTGCCTCGGTGACATCATCGGACATAGGTTCAAGTGTCTCGAATGTGCCGA
TTATGATTTGTGCGAGAAGTGTGAGAGAAAATCTTTACATTATGAGCATGCAATGATCAG
AATTGTGGAGCCAAAGAAAACAATGATCCCATCATATGTGACATCAAACTCTCCGAACAA
TGTCTTCCCAACTTACATGCAGCACAAGATCCAACAACCACAAACTATCAAGGGGATCAA
CAAGACTGTTATCACGAGACTCGTTCCTTCACATCAAGTTAATCCAGCTCCAATGCCTCA
AGCTCCACCAGTCTCCACTCCAGTTCCATTCCCTCCATCTCCAACCATCTGAAATGAGCA
GTTCCCAAACCAGCCCAGGAGCCTCGGATNCCACCATCTCCCACTTCTGCTCTACCACCA
CCACAATTCTTCAATGANGAGACCCGTGCAGCCNTTATTAATAANGCAGCGTTGCTCAAG
GAACTTTCATACTTTGAGCANAANCTTCTTGGANTTGGAGATGGNGCACAGGNCATATGG
ATAAAAGATCTTGAACGCTCGACNTGGAGNAGGATGACCAATTCGANAAAAATGATATTG
AAANACCTGNGGGAAAAATCTGGACGAACTGTGAACTGAAANACGGGANCTCCAAAACAA
CAATACTGTTTGNCAAATANAAGGTAAAAANCCGGTNAAAATTGCANAAAATTCAGGAAT
GTGGGGG-----3′

T12G3.1 (pACT119)

5′GACCTAAAGGGGGATCCTCNANAGTTCACGCTAAANGAGCTCTCAAGAGTTAGCGTTT
TGGAGACGATGGTTGCTGATCTTCATGCAAAGGANAGAGAACCTGTTGTCGAGGAAGTTC
ANAAGCCATTCCCTCCAGTGAAGCACTCGGCGTCTTGTGATGCTTGCCTCGGTGACATCA
TCGGACATAGGTTCAAGTGTCTCGAATGTGCCGATTATGATTTGTGCGAGAAGTGTGAGA
GAAAATCTTTACATTATGAGCATGCAATGATCAGAATTGTGGAGCCAAAGAAAACAATGA
TCCCATCATATGTGACATCAAACTCTCCGAACAATGTCTTCCCAACTTACATGCAGCACA
AGATCCAACAACCACAAACTATCAAGGGGATCAACAAGACTGTTATCACGAGACTCGTTC
CTTCACATCAAGTTAATCCAGCTCCAATGCCTCAAGCTCCACCAGTCTCCACTCCAGTTC
CATTGTGATTTGCTNTCNANTTACGAANTCCTGGANTTGNTNNAGGCGTCNGNNNAAACC
CAACCCCTGCTCTACCACCACCACAATTCTTCAATGAGGAGACCCGTGCAGCCATTATTN
ATAGTGNAGCGTTGCTCAAGGAAACTTTCAGTACTTTGAGCAGAAGCTTCTTGGATTTGG
CAGATGGTGCACAGGCCAATATGGATAAANAAGATCATTGAACGCTCGACCNTGGAAGCT
GGAATGACCAAATTCNAANAGAAAATGGATAATTTGAAAGACACTGTGAGGAAAAANATC
TTTGGAACGGAAACTGCTGAACTGANAANAGCGAGGAATTTTCAAGAAACAAGCAGATAG
CTGTCTGGACCAAGATCAANAAGGCTAAGAAGAAGCCTGCTACAGAAATNTGCCGAGAGA
TNTTCCNATGCACTGNTGCGCGNTGCTGANAAGCCCGTCTGAAAAAAAACAAAAANANCT
GAAGCCCGTCTGCTGATTANNNAGAANTTGNTNACTTANGGAAGNANNNAATCAAAACCN
ANGAANATTCCCTCCATCTCCAACCATCGGAAATGGAGCAGTTCCAAANCCAGCCCAGGA
GCCTCGGATCCCACCATCTCCAACTTC-----3′

T12G3.1 (pACT147)

5′ACCTAAANGCGGATCCTCGANAGTTCACCTTTGAGTTCACCAATTCGGATCAGATTAT
GTCTAANATTCGAATCACTACATTCTCTCGAACTCCTCAACTTTTTGTTCGCCTGGAAAC
TGTTCCCCTTCGTCCAACAGCCCTACCATCAGTGAGCACTGAAACTGCCAAAAAAGCTTC
CGATGCTGATCGTGAATCAAAGAGCAGATCCTCGCTGAAGCGTATCGAGAACAACCAGGA
GCTCACATTGAAGACGCTGGAGAAGTTGCAAAAGGAGCTCTCAAGAGTTAGCGCTTTGGA
GACGATGGTTGCTGATCTTCATGCAAAGGAAAGAGAACCTGTTGTCGAGGAAGTTCAGAA
GCCATTCCCTCCAGTGAAGCACTCGGCGTCTTGTGATGCTTGCCTCGGTGACATCATCGG
ACATAGGTTCAAGTGTCTCGAATGTGCCGATTATGATTTGTGCGAGAAGTGTGAGAGAAA
ATCTTTACATTATGAGCATGCAATGATCAGAATTGTGGAGCCAAAGAAAACAATGATCCC
ATCATATGTGACATCAAACTCTCCGAACAATGTCTTCCCAACTTACATGCAGCACAAGAT
CCAACAACCACAAACTATCAAGGGGATCAACAAGACTGTTATCACGAGACTCGTTCCTTC
ACATCAAGTTAATCCAGCTCCAATGCCTCAAGCTCCACCAGTCTCCACTCCAGTTCCATT
CCCTCCATCTCCAACATCGGAAATGGAGCAGTTCCNAAAACAGNCCAGGAGCCTCGGATC
CCACCATCTCCAACTTCTGCTCTACCACCACCACATTCTTCAATGAGGAGACCCGGNNGC
CNTTTTAATAGTGCANCGTTGNTCAAGAAACTTCANTACTTTGAGCANAAGNTCTTGGAT
TTGCAATGGGCACAGGCCATNTGANAAAAANATCNTGANCNCTCNACNTGNAGCCGGATG
ACCAATTCCANAAAAATGGGNAAATTNNAACANCTGNGAGNNAANTTNTGGACGAAACTG
TGANCTNAAAAANNNGGAANTCTAAANCANCAATANTTGCNGGNCAAAAAANAGNTAAAA
AACCTGTNNAAATNNCNAN-----3′

B0336.6 (pACT5)

5′GAGATCTGGAATTCGGATCCTCGAGAGTTCACCGATNATCGAGCTCAACTGGAAACGA
GTCATGCGAATCTTCAACAAGTTGCCGCGTATTGTGAGGATAATTATATACAATCAAACA
ATAAATCTGCTGCGCTAGAGGAATCCAAGAAATTCGCGATCCAGGCACTCGCCAGCGTAG
CCTACCAGATTAACAAGATGGTTACAGATTTACACGATATGCTTGCTCTACAAACCGATA
AAGTGAACTCTTTAACAAATCAAGTTCAATATGTTAGCCAAGTAGTTGATGTACATAAAG
AGAAGCTTGCAAGACGAGAAATTGGTTCTCTCACAACCAATAAAACATTATTCAAGCAAC
CCAAAATCATTGCACCAGCAATCCCAGATGAAAAGCAGAGATATCAACGAACGCCCATCG
ATTTTTCTGTTCTTGACGGAATAGGGCATGGTGTCAGAACATCGGATCCACCGAGAGCAG
CACCAATCTCAAGAGCAACTTCATCAATTTCTGGCAGTTCTCCATCACAATTTCACAATG
AATCTCCAGCGTATGGAGTTTATGCTGGTGAACGAACGGCTACGTTAGGAAGAACAATGA
GACCGTATGCTCCATCAATTGCTCCATCGGATTATCGGTTGCCACAGGTCACACCACAAT
CAGAATCACGAATCGGCCGTCAAATGAGCCACGGATCAGAGTTCGGAGATCATATGAGCG
GTGGTGGTGGAAGCGGAAGTCAACACGGATCATCAGACTATAATTCCATTTATCAACCTG
ATCGTTACGGAACTATTCGAGCTGGTGGTCGGACTACAGTGGATGGTAGCTTTTCTATTC
CCAGACTATCATCTGCACAAAGTAGTGCTGGAGGNCCAGAATCACCAACATTNCCACTTN
CACCACCAGCTATGAATTATACTGATATGTTGCACCGGGAGTGNGGTACAACAACCACCA
CCACCACCAATGGACAACCAANTATGGAACTATTCGAAATCACGGNGAACCGCTGATCTT
CNCCTCNCCAATTCTTGCTCCTGGATGNCAGTCGAAGGCCACACAGAATGNTGNTGATCT
ACNCTCCCCNAAATNGTGGGGGGCTCCCCGAGAGTGTGGTGNAGANANANCGNNAGTCAG
TCGCNCAGTNTNTTGNACAANGTGNNGNTGCCACAGANTTGGAAAGACGGGCTGACACAT
NTGTNCAAGGGC-----3′

B0336.6 (pACT16)

5′AAGGCGGATCCTCGAGATTTACCCAAGATGGTTACAGATTTACACGATATGCTTGCTC
TACAAACCGATAAAGTGAACTCTTTAACAAATCAAGTTCAATATGTTAGCCAAGTAGTTG
ATGTACATAAAGAGAAGCTTGCAAGACGAGAAATTGGTTCTCTCACAACCAATAAAACAT
TATTCAAGCAACCCAAAATCATTGCACCAGCAATCCCAGATGAAAAGCAGAGATATCAAC
GAACGCCCATCGATTTTTCTGTTCTTGACGGAATAGGGCATGGTGTCAGAACATCGGATC
CACCGAGAGCAGCACCAATCTCAAGAGCAACTTCATCAATTTCTGGCAGTTCTCCATCAC
AATTTCACAATGAATCTCCAGCGTATGGAGTTTATGCTGGTGAACGAACGGCTACGTTAG
GAAGAACAATGAGACCGTATGCTCCATCAATTGCTCCATCGGATTATCGGTTGCCACAGG
TCACACCACAATCAGAATCACGAATCGGCCGTCAAATGAGCCACGGATCAGAGTTCGGAG
ATCATATGAGCGGTGGTGGTGGAAGCGGAAGTCAACACGGATCATCAGACTATAATTCCA
TTTATCAACCTGATCGTTACGGAACTATTCGAGCTGGTGGTCGGACTACAGTGGATGGTA
GCTTTTCTATTCCCAGACTATCATCTGCACAAAGTAGTGCTGGAGGTCCAGAATCACCAA
CATTCCCACTTCCNCCANCAGCTATGAATTATACTGGATATGTTGCCCCGGGAAGTGNGG
TACCACAACNACAACNACAACAAATGCACAACAAATTTTTGGACTATTCCAAAATCAACG
GNGAACCGACTGATCTCCNNCTCCACCAAATTCTTGCTCNCTGGANGGCAAGTCGAATGN
CANCCAANANGATTTGGGAGATCNCCCCCNCCCCCAGAANNGTTGNGGGNNTCNNCGTNT
GAGNGTTGCTGGAAANAAAACCNCAGTTCAANNCNCCCCAGCTNTTTGATCNANNGNTGG
AGGAGCCCNNGANNNTTGGAAAANNNGGCCCGNCCAATTTNAGNTGCAAAAAACAANTGN
ACTT-----3′

B0336.6 (pACT31)

5′AAAGGCGGATCCTCGAGATTTACCTGATCTTCAAGAGCTCATCGAGCGACGGATACCC
GATAATCGAGCTCANCTGGAAACGAGTCATGCGAATCTTCAACAAGTTGCCGCGTATTGT
GAGGATAATTATATACAATCAAACAATAAATCTGCTGCGCTAGAGGAATCCAAGAAATTC
GCGATCCAGGCACTCGCCAGCGTAGCCTACCAGATTAACAAGATGGTTACAGATTTACAC
GATATGCTTGCTCTACAAACCGATAAAGTGAACTCTTTAACAAATCAAGTTCAATATGTT
AGCCAAGTAGTTGATGTACATAAAGAGAAGCTTGCAAGACGAGAAATTGGTTCTCTCACA
ACCAATAAAACATTATTCAAGCAACCCAAAATCATTGCACCAGCAATCCCAGATGAAAAG
CAGAGATATCAACGAACGCCCATCGATTTTTCTGTTCTTGACGGAATAGGGCATGGTGTC
AGAACATCGGATCCACCGAGAGCAGCACCAATCTCAAGAGCAACTTCATCAATTTCTGGC
AGTTCTCCATCACAATTTCACAATGAATCTCCAGCGTATGGAGTTTATGCTGGTGAACGA
ACGGCTACGTTAGGAAGAACAATGAGACCGTATGCTCCATCAATTGCTCCATCGGATTAT
CGGTTGCCACAGGTCACACCACATTCAGAATCACGAATCGGCCGTCAAATGAGCCACGGA
TCAGAGTCGGAGATCATATGAGCGGTGGTGGTGGAAGCGGAAGTCAACACCGAACATCAN
ACTATATTCCATTTATCNACCTGACGTTACGGAACTATTCCAGCTGGTGGTCGGACTNCA
GTGGATGGTAGCTTTTCTATTCCCAACTATCATCTGCNCAAAGTANNGCTGGAGNCCAGA
ATCACCAACATTCCCNCTNCCCACCNGCTNTGANTNTACTGGNNTGTTGNCCNGGAAGNG
GGNCCACACTCCACNNCNNAATGCACNCCAAATTTGGACTTTCCGAANCANNNGGACCGA
CTGTTCTCCCCCCCCTATTTTTTGNCCNGNAAGCANCCAANGGCNACNCAATNNNTGGAT
NA-----3′

B0336.6 (pACT36)

5′AAAGGNGGATCCTCAANATTTACCAGTTGATGTACATAAAGAGAAGCTTGCAAGACGA
GAAATTGGTTCTCTCACAACCAATAAAACATTATTCAAGCAACCCAAAATCATTGCACCA
GCCAATCCCAGATGAAAAGCAGAGATATCAACGAACGCCCATCGATTTTTCTGTTCTTGA
CGGAATAGGGCATGGTGTCAGAACATCGGATCCACCGAGAGCAGCNCCAATCTCAAGAGC
AACTTCATCAATTTCTGGNAGTTCTCCATCACAATTTCACAATGAATCTCCAGCGTATGG
AGTTTATGCTGGTGAACGAACGGCTACGTTAGGAAGAACAATGAGACCGTATGCTCCATC
AATTGCTCCATCGGATTATCGGTTGCCACAGGTCACACCACAATCAGAATCACGAATCGG
CCGTCAAATGAGCCACGGATCNGAGTTCGGAGATCATATGAGCGGCGGTGGTGGAAGCGG
AAGTCAACACGGATCATCAGACTATAATTCCATTTATCAACCTGATCGTTACGGAACTAT
TCGAGCTGGTGGTCGGACTACAGTGGATGGTAGCTTTTCTATTCCCAGACTATCATCTGC
ACAAAAGTAGTGCTGGAGGTCCAGAATCACCAANATTCCCACTTCCACCACCAGCTATGA
ATTATNCTGGATATGTTGCACCCGGGAAGTGTGGTNCAACAACAAACNACCACAACAATT
GCACAACAAAATTATGGAACTATTCNGAAAATCACGGNGAACCGAATGATCTTCCACCTC
CANCCAATTCTTTGCTCNCTGGAATGTCAGTCGAATGCCANCCCAAGATGANTGGATGAT
CTNCCNCCNNCCCNAAAATCAGTTNGGGGGCANCANCGNATGGAAGNGTTGCTGGNATAA
ANAAATCGTAAAGTTCNNNCNGCACCANTTTCTTTGTNNCNANGNTGGTGGTNCCNACCA
GTNTTTGAAAAAANACCGGNCTGGNNATNNNNANNNNNAAANANAAAAATGANNNTTNCN
ANANCATTNNNTACTNTTNAAAAACCANAACNCTGGTTGANGGTCNTGAAGGNCCCNGGN
TTCCCG-----3′

B0336.6 (pACT46)

5′GANCNTAAAGGGGGATCCTCAANATTTACCNAAAGATGGTTACNGATTTACACGATAT
GCTTGCTCTACAAACCGATAAAGTGAACTCTTTAACAAATCAAGTTCAATATGTTAGCCA
AGTNNTTGATGTACATAAAGAGAAGCTTGCNAGACGAGAAATTGGTTCTCTCACAACCAA
TAAAACATTATTCAAGCAACCCAAAATCATTGCACCAGCAATCCCAGATGAAAAGCAGAG
ATATCAACGAACGCCCATCGATTTTTCTGTTCTTGACGGAATAGGGNATGGTGTCAGAAC
ATCGGATCCACCGAGAGCAGCACCAATCTCAAGAGCAACTTCATCAATTTCTGGCAGTTC
TCCATCACAATTTCACAATGAATCTCCAGCGTATGGAGTTTATGCTGGTGAACGAACGGC
TACGTTAGGAAGAACAATGAGACCGTATGCTCCATCAATTGCTCCATCGGATTATCGGTT
GCCACAGGTCACACCACAATCAGAATCACGAATCGGCCGTCAAATGAGCCACGGATCANA
GTTCGGAGATCATATGAGCGGTGGTGGTGGAAGCGGAAGTCAACACGGANCATCNGACTA
TAATTCCATTTATCAACCTGATCGTTACGGAACTATTCNAGCTGGTGGTCGGACTACAGT
GGATGGTANCTTTTCTATTTCCCAGACTATCATCTGCACAAAGTAGTGCTGGAGGTCCAN
AATCACCAACATTCCCACTTCCACCACCAGNTANGAATTATACTGGATATGTTGCACCGG
NAAGNGTGGTNCAACAACAACAACACNACNAATGCNAANANCAAANTTATGGAACTATTC
GAAAATCAACGGNGAANCNAATGATCTTCNCCTCCCCAAATTNNTTGCTCATGGAATGTC
AANCCAATGCACCCAGATGNTATGGTTGANCTCCACCNCCCCANAATCAGGTGGTGGNCA
NNACNATNGNGNGTTTGTGGNAANAAAATCGANGTTGAGNACCCCNAANTNTTTTGAACA
ANGNTGAAGGTGCCAAGNNATTGAAAANACGGTCTGNNCANTTNNTGTNAAAAAAAAATT
NNNCTCCNGACCCATTNNNNCNNTNA-----3′

B0336.6 (pACT93)

5′ACCNNAAGGNGGATNTCCAAAAGTTCTCGCTACATAATGGGGAAGGAGGAANGAGTGT
TAATGATCTTCANGAGCTCATCGAGCGACGGATANCCGATAATCGAGCTCAACTGGAAAC
TTAGTCATGCGAATCTTCAACAAGTTGCCGCGTATTGTGAGGATAATTATATACAATCAA
ACAATAAATCTGCTGNGCTAGAGGAATCCAAGAAATTCGCGATCCAGGCACTCGCCAGCG
TANNCTACCAGATTAACAAGATGGTTACNGATTTACACGATATGCTTGCTCTACAAACCG
ATAAAGTGAACTCTTTAACAAATCAAGTTCAATATGTTAGCCAAGTAGTTGATGTACATA
AGAGAGAAGCTTGCAAGACGAGAAATTGGTTCTCTCACAACCAATAAAACATTATTCAAG
CAACCCAAAATCATTGNACCAGCAATCCCAGATGAAAAGCAGAGATATCAACGAACGCCC
ATCGATTTTTCTGTTCTTGACGGAATAGGGCATGGTGTCAGAACATCGGATCCACCGAGA
GCAGCACCATCTCAAGAGCAACTTCATCAATTTCTGGCAGTTCTCCCTCACAATTTCACA
ATGTATCTCCAGCGNATGGAGTTTATGCTGGTGAACGAACGGCTACGTTAGGAAGAACAA
TGAGACCGNATGCTCCATCAATTNGCTCCATCGGATTATCGGNTGGCACAGGTCNACACC
CCAATCAGAATCNCNAATCGGGCGTCTAATGANCCNCGGATCAGAGTTCGGAGATCATNT
TGACNGGTGGTGGTGGAAGCGGAAGNCNANCCGGATCATCAGACTTTAATTCCATTTATN
ANCCTGANCGTTNCGGAACNATTCNAGCTGGTGGCGGACTAANNTGGATGGAGNTTTCTT
TNCCNGAATATNCTCNGCACAAGTAANGTGGAGGTCNAAAATCACCAACNTCCCACTTCC
CCCNCNAGTNTNATNNNCTGANAAGTNGCACCGGGAAGGGGGTACTCNNNANNNACACNN
NNATGNCCAACAATTANGGACTTTNAAANNCCGGAACCANTNNNTCTNCTCNCCCCAATT
TTTGNNCAGNANGNCAGGCAATNNNNNAAAANTNTGC-----3′

B0336.6 (pACT157)

5′ANCNAAAGGGGGAACCTCAAAAGGGTCCCTAGNTNGNGATGCCCCAATCNTGGNGAAG
GCNNAATGAGCGTNTANTGAACNNNCNNGAGCTCATCGAGCGACGGATACCCGANAATCN
AGCTCAACTGGAAACGANTTCATGACGAATCTTCTANAAGGNTGCCGCGAATTGTGAGGA
TAATTATANACAATCAAACAATAAANCTGCTGCGCTAGAGGAATCCAAGAAATTCNCGAT
CCAGGAACTCGCCAGCGNAGCCTACCAGANTAACAAGATGGTTACNGATTTACNCGATAT
GCTTGCTCTACAAACCGATAAAGTGAACTCTTTAACAAATCATNGTTCAATATGTTANCC
AAGNAGTTGATGTACATAAAGAGAAGCTTGCCAGACGAGAAATTGGTTCTCTCACNACCA
ATANAACATTATTCAAGCAACCCANAATCNTTGCNCCAGCAATCCCAGATGAAAAGNNGA
GATATCANCGAACGCCCATCNATTTTTCTGCTCTTGACGGAATAGGGCATGGTGTCAGAA
CATCGGANCCACCGAGAGCANGCACCAATCTCAAGAGCAACTTCATCAAANTTCTGGCAG
NTCTCCNTCACAATTTCACAATGAATCTCCAGCGTATGGAGTTTNTGCTGGTGAACCGAA
CGGCTNCGTTAGGAAGAACAATGAGACCGTATGCTCCATCANTTGCTNCATCGGATTATC
GGTTGNCCCGGNNANCNCNCNATCNNAATCACNAATCGGCCGTCANATGACCACGGATCN
AANTTCGNANATNATTGAGNGGGGGTGGNGGAAGCNGAAGTCAACACNGATCATCAGACN
ANATTCNTTTATNACCTGNACGTNNCGGACTNTCAAGCNGNGGNCGGACNACANTGGNTG
GNAGTTTTCNTNTCCCAGACTNNCATCTGAACAANGNANGNGGNGGGCCNAAATCCNCAA
CCTTCCCCTTCNCCCCNAGTANNTATTNCNNGGAATGTGGNCCGGGAGGGGGGAAANCNN
CCACNCAACNAAGGNCACNAATNNGGANTTTNNAAATCNCGGNGACNAANCGNTTNCCCC
CAATATTTTNCCCNGGNNGCNAGNNATNNCACAC-----3′

B0546.2 (pACT108)

5′ACNTAAAGGGNGGATCCTCNANAGNTCTCGACGATTTTACCANNCAAAAAAAAGCGNG
AANAGGATTCCCNNCACAGATCAGTTTCAANAAATTGGAAGTTCTGCTGANAAATAGCCG
TCANNTACTCTCTGGACCATCCTGCAGCTCATCAGTCTTCATCTTCNAGCTCCGCGGATT
AAAAGGATCCGTTCTCAGGAGGATGGACCTCCGAGNAAGCGATGGGTCCTATCGGCTTCA
TCTACCTCATCAACGAACAAAAGCCCAATTCCATCATCATCACCAGTCCTACCACCACCT
CAACCTGAAATTGAGCCAAATCGAAGCCAATCTCCAGGAAATGCCGACTATAATTCAGAC
GATGAATATGAAATGCAAAATATTCGAGACGATGAGANCNAAACGGAATTCNCNAATCGT
CTGGCGGCCCGNGGCCTAATAATCAAAGAAATGGTCGGCGACGGAGCCTGTATGTTCCGT
TCAATTGCCGAACAAATCTACGGTGATCAGGAAATGCACGGACAAATTCNCNGTCTTTGT
ATGGATTATATGTCGAATAACCGCGATCATTTCAAAGAATTCATTACCGGAAAACTTTGA
GAATTATATTCNACGAAAACGCGAGGAAAATGTGCACGGGAATCATGTTGAGCTTCAAGC
GATCTCTGAAATGTTTGCACGGCCTGTNNAGGNGNATCAGTACNGTGNTGAACCAATAAA
TGTGCTTCTTCCCCGTCCGGAACCANAAGCTACTGNTCTAACANANGGCTCCGCGTCCAC
ATCGAAANGCTTGCTCCGCCGCCGTCTANNGTACCAAATCAACANATTCCACCGCTTNGG
TTANNTTTCNNTCGAGCTGTCCACTATAATGCTNTTCTTGNCCCAAAATTCCAANAATCG
GAANNGGTCTTNGATTCCCGGAATGTNCCCNGGTCCANTGANAAGAGTTATGNNCCNAGC
NANGAAGACCTCGNAGTTGGNNCATNTTGAANAACNNCGNNCCGGANAAACCNANTTGNC
GATTTCCACNAACNAGGGNANTGAANNAATTNCCGGAGANCCCNAATTTATNNTTAATGN
NTTTNAAAAGNTAAAANGCCCGCAANGCNNNGGNNCCTGNCTTCNATTCC-----3′

B0546.2 (pACT142)

5′GAACNAAAAGGGGGATCCNCCAAAAGNNCNCGCNNGATAAGGATTCCCAGCNGAGATC
ANTTTCCACCATTGGAAGNTCTGCTGAAAATAGCCGTCNACAACTCTCTGGACCATCCTG
CANCTNATCANCTTCATCTTCGAGCTCCGCGGCTAAAAGGATCCGTTCTCAGGAGGATGG
ACCTCCGAGCAAGCGATGGNTCCTATCGGCTTCATCTACCTCATCAANGAACAAAAGCCC
AATTCCATCATCNTCACCAGTCCTACCACCACCTCAACCTGAAANTGAGCCANATCGAAG
CCAANCTCCAGGAAATGCCGACTATAATTCAGACGATGAATATGAAATGCAAAATATTCG
AGACGATGAGATCGAAACGGAATTCTCGAATCGTCTGGCGGCCCGTGGGCTAATAATCNA
AGAAATGGTCGGCGACGGAGCCTGTATGTTCCGTTCAATTGCCGAACAAATCTACGGAGA
TCAGGAAATGCACGGACTAATTCGCCGTCTNTGAATGGATTATATGTCGAATAACCGCGA
TCATTTCAAAGAATTCATTACGGAAAACTTTGANAATTATATTCAACGAANACGCGAGGA
AAATGTGACGGGAATCATGTTGAGCTTCAAGCGATCTTTGAAATGTTTGCACGGCCCTGT
ANAGGTGTATCAGTCAGNNGATGAACCANTAANTGTGCTTCTTCCCGTCGGAACCAAAAG
CTNCTGCTCTAAAANATGGTTCCGCGTCCNCATCGAAGCTGNTCCGCCGCCGCTACGTAC
CATTCAANNANAATCCCCGCTTNGGCTTAAGTANNNNCGANCTGTNNCTATANGCATTCT
TGCTCCAAAATTCAACAATNGAATCGGNTCNGATTACCCGGATGATNCCNCGATCAGCTG
ATAAANAGTNATGNGCCCGNNTGAAAACTCNGNTTGGAGCTANTNGANAACTANNTCCNG
GATAAANCGATATGACNATTTCNNCGACTCGGNNATATGAAAANAATCTCGNNANCCCCA
TTCTACTCAANGNNTTNAAAGGCTAAANGNTCGNCNAANCATTNGANTCCGGNCTCNATT
NCGNGNANCCCCGTNNNGGA-----3′

B0546.2 (pACT169)

5′GACNTAAAGGNGGATCCNCGAAANNTNCNANCTNCGAGCAAGCNATGGGTCCTATCCG
GATTCAATCTACCCTCATTCAANCGTAACAAAAGCCCNATTCCATCATCATCACCAGNTC
CCTACCCCCATCCTCAACCCTGAATANTGAGCCAAATCGAANTCCACATCATCCAGGAAA
TGCCGACTATAATTNAGACNATGAATATGANNTGCAAAATATTCNAGACGATGAGATCTG
NAACGGAATTCTCNAATCGTCTGGCGGCCCGTGGCCTAATAATCAAAGAAATGGACGGCG
ACGNAGCCTGNATGTTCCGTTCAATTGCCGAACAAATCTNCGGTGATCAGGAAATGCACG
GACCAATTCGNCGTCTTTGTATGGATTATATGTCGAATAACCANCGATCATTTCANAGAA
TTCNTTACGGAAAACTTTGAAAATTATATTCAACGAAAACGCNAGGAAAATGTGCACGGG
ANTCATGTTGAGCTTCAAGCNATCTCTGAAATGTTTGCACGGNCCTGNANAGGTGTATCA
GTNCAGNGANGAACCNANTAAATGTGCTTCTTCCCCGTCCGGAACCAGAAGCTACTGCTN
CAACAGATGGTTCCGCGNCNACANCGANNGCTGCTCCGCCNCCNTCTTCAGNACCAATTT
CAANNGAATCCCACCNGNTTCGGGCTTANTTNTCCNTCGAGCTGGTCNCTATAATGNTNT
TNTTGCNCCAAACCATTNCAAACANTCCGCAATCGGTCTCCGTATATCCGGAATGGAACC
CNGGNTCNCCTTGATAAANAGTTNANTGNNCTCCNGNNTGAANACCTCNNANTTGGANCN
NNTGNAAAAAACANGGCNCCNNGATAAAAANCNTNTNNANGATTTTCAACNAACCCCGGG
GGGATNTGGAAAANCCATTCCNCNNAACCCNNAATTTCCNNCTTNAANGACTTTGAAANT
NNAAANGCTCCNCNACCCANNTGGNANNCNGGNCCTCCNNATNNCNGGGNANCCCCCGTT
TNGGNCTAGNCNNTTCAANAAANGCCNCCNCCCATNTTTGAATGGNCCCCNNCCCCCCAN
GTNNAATNGCCAAGNTNTCTANCCACTGTNTTTTGNNNTNNACCTANNCANANACANGTT
GGGC-----3′

ZC404.8 (pACT10)

5′AGATCTGGAATTCGGATCCTCGAGAGTACACGCAAAACACACAGATNTTTACTAACTT
CGCTCACAGAGCACATGATGGATTACCGTTTAACAGTGCGAATCCTTCCAACAAAGATCC
AATTTTCACAATGCCAATATCGGTCAAACCGAANACTCGTGAGCCGGCTTCTTTCACAGA
TAACAGGATATATGTCAGCAATATTCCCTTCTCGTTTCGTGAACAAGATTTGGCGGCAAT
GTTCTTCGCATATGGAAGAGTCCTGAGTGTGGAAATCGTCACAAATGATCGTGGATCCAA
AGGGTTCGGGTTTGTCACACTCGATTCCATCGAATCCTGTGAGAAAGCTCGTGCTGCGCT
TCACGAATCACATGTTCAAGGAAGAATTATAGAAGTGAGAAGAGCGACACCAACCCGCAG
AAAGCTTATCAACAATCCACAAAATGAAGTTTTGCCACCACCAAAGCTGTGTGTCGATCT
TCGAGCCCCTCATAATTTATGGAGAGCTGAGCCAATGCATCAGTTGTTCAAGGAAAAGGA
GAACACAACATGTTTTCCCGAAGCTGGATTCATGATGGCACCATACCGTAGCAATGGAAT
TTTCAACACGCGTAGTCTTGTGCAGACCAAACCACCTCGATGCACCAAGCACAGCGAGCT
CAAGCTTTCTTCAGCTGGTGAATACTTCTGCAAAAACGGCGAGCCTACGACGGAAACAAG
TATTCTGATGTGCATGCACAGACCAAACTCACCATGCAGCAATAAGTGTTCTGATTCTTC
GAATCACGAGCTGTCTGATGTGGAGTTGAACTCTATATTCCCACATCATCTTCGTGACCA
GATTACTGCTCTTCTCGACACTTCAAACCATTTTGGATCAGGAAATAATAGTGCTAACAA
AGGAAAGAGAGCACCATCTGTGACATCTTCTGGATTGAGATCATCAGAGAGCGAGACAGT
TTCAGACGAANAAATCATTGGTCCCNCATAACAGCCCTGATATCTTCTCGCTGCTCTCTC
GAAGGTCACATCGTTCCCGGAAGCTGTTCTCNCCNAAGATCGCCAGCCAGNATCAGTCAC
TGAACNCCATTTATTCNNCCACTGGGGTTGATTTATCGAAACCCTACTTCCTTACAAAAG
TTTACAANTTTACGAACTCTGAANGNCCCNTGCCTGNAANANCTGAGTTTCTCCGAGNCA
AN-----3′

F22E12.4 (pACT80)

5′AAGGGTGGATCTNCAAAAGNTACCAGACAAGTAATTTNACGCCTCNGGNAGCACCTGC
TGNNATGGCTCCATATTCCACCTTACATGANNNCNCCTTGATGATNCTGTACTTACCACG
ATTCACTTCTTCATATCATCTTCAAAAAAATCCAATTTCNGAATCAAACTATTTCNAATT
TTCNGNCAACATTTTCAGATCACATCTGANGACCGAACCAGAGCCATCGATTCCNATCCA
ANTNCCACAAAGGANATCATCCACANGTACAGTACCGTTCANTANNGANNGANGTGCATT
CAAACCATACAGAAATACGCATGTGCATTAATTCAATTTCTTNTGAATCAATGTCTTCCA
TGTGCACATCACATGATNNCATCNCTTGAATACAGGTCATCAGNCTTCCCTNGCAATGTT
CCCAACACGTAGCACTGCTCAAAAGTGATATCANTAGACTCGCTCAAAGTTTTGAGNCTT
GCTGGAAATTCACCAANCNTTCNTTGGCTCTTGTCNCAAACTTCGGTATCNGTCAACTGC
TTCNACAGNANCTATTTCCCACCTCCAGAGNACAACGACAAGTTCAGCTACNAAGTCNCN
NGCTAAAGCGANACAACAACTGTNGGAAAAGAAAAAGANAATTCACACTGATGATCCCGN
ATATTCNGATCATCGAAACTANAGGTGGACANAACCCANCGGTTTCCAGAAATACGGAAA
CNACCAANTNCTTCNNAACTCCNNTGACCCAAAAATTAATTACNAGGATCACANTTAGAT
GTCGNTTTCTCTNNCNCCCTNCAANAAATTCAGAAAGCATCTTCNNAANCGTGGNCTCNN
NCTTANCANTCCCAAAACNNTGGTTNTACNGTAAAAANCNTGNCCNAACAGNGGNCCTAA
ACCCGAAAGAATTTGGNGGGCCNNTNCTNACNATTTTTTGNCTCNGATNNANACAAANTT
CCCCCGAAANGAATTGNGNCCCCTTNANCNGGNCNCTNTNCNGGGTTGCGAGGGAANNAA
AATNNNAAGTNAATTNCNCTCTAAATTTCTATNGNACAANTTTNTNGATTNNGNNTNTNN
GNNNTGTNNGNCNGNGAANNACNNCCNTNNTCCAGANTTNTN-----3′

F58A4.7a (pACT161)

5′ACNAAAGGGNGANCCNCNAAAGGNCANAAATTCATTTGCNNTCTGANTTTGCCTACCT
CTCCGTATTGGNNNATGCCCNAATGNCTCGCCTACCTGGCCNCTTATTCCCAACCGGNGN
CCAACTTTNTCCAACGTCNCTCGACCCCGNACTGTAANATNCANAATGCGTCGNCAAANT
GNCTATNGCNANNAGCGGNNNNNGATGCNAGAGCNTCANTGCGGNAATTCTTGGCCTCNA
AGAGCCTTACTTCNCAGAAAAGANGGANAAATNCTGTAATAGGCGGCTATTCNGCAGCAA
ACTGCANATATGGTTCATCANCTTTTGGGACACAAAGGNGAAGATATACCTGACGGTGNA
GAACCCAGAANNCTCNAACTNGAAAAAGATCACCACGACNNTGACCATCNTAGCGCCNCA
NTAGCNCANTCTNCCAAANCAATANTAGAAACTGATCGAGCCGCCNCGTAAGGCTTTGGA
NAGTCAAGCANTTCANCTCCGNGANCCTGNTCCAAATGACNACAAACATCATCGCAAGCA
TNNTCACCAGTGANACCANAGAACTAATGGTTCTGGCGGCTTCACATTGNCATCNNCATN
ATGCTTTCTTCGGCTCTTCCTACTCCGANNCGAGAATCACCGTAAAGGNAACCAATCNTT
NCAAGATACAACANGTACTCCGATGACTCTTTTTGACNCTGAACCGGNAGCNCAACTTCT
NCCTGAANCNTNGGCGTCNCCNCCGCATATTCNAACCGCNNCCAATTTGNCCAAGTNTTN
NAGTNAACNNGNTNTCCGNCTTATNCCAANNCTNCAANNATAAGTNGAAANNNNATTACN
CTCTNTGTNAACTNCATCAACTTTTGACTGCCGCNCCAANNATCNTTCTNNNACNCNNNT
TCNACTCANCCNAGTATTTTGCNNCCAAANTNCNGCTNCGNGGCCCCNCGGNCCTTGTCT
ATCGNAANCNCGTNNCCCNTCCTCTTNCCCNNNCNATGGAANGACANNATTTNCGNTTGN
ANCACNANACNNATCTGACCGTTNANNCCGNTAANNNTTTTNNACNTNCTTATNTNNAGN
NAAGNTATTTNCCAAANCCNCTCNNNCNACNNCAAATGTTNNCNTGANCNGANTTTGNAA
CGGNANTCCTTNNNC-----3′