# A proteomic analysis of lipid raft and GPI anchored proteins in *Caenorhabditis elegans*

Wei Rao

Submitted in accordance with the requirements for the degree of Doctor of Philosophy

The University of Leeds

Institute of Membrane and Systems Biology

Faculty of Biological Sciences

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

# Acknowledgements

I would like to thank first and foremost my supervisors Professor Elwyn Isaac and Dr. Jeff Keen, for their help and encouragements throughout this Ph.D. Special tribute must also be paid to Professor Ian Hope, Dr. Michael Walker, Dr. Hannah Craig, Dr. Sophie Bamps, Dr. Jane Shingles, Dr. Edward Parkin, Dr. Shipa Ghosh, Ms. Julia Wirtz and Mr. Paul Bedingfield, whose help and patience in teaching me the ins and outs of biological research had allowed me to finish this project. Thanks to everyone on Miall level 8 for making the place a wonderful place to work, and my other friends for bearing with me when I complained for the umpteenth time about how hard it is to do sucrose gradients.

I would like to dedicate this thesis to my family. These include my lovely wife Yawen Guo, my cousins, aunties and uncles in China, my Grandparents in Beijing and Baotou, and most of all my ever thoughtful parents, who provided me with strength and encouragement through the most difficult times. With their love and support I was able to finally finish this Ph.D. Now that I have time I will be visiting you all and boring you with the contents of this thesis for years to come.

This thesis is dedicated to my grandmother Zhang Nai Chi.

**献给我尊敬的外祖母张乃痴**。

# Abstract

Glycosylphosphatidylinositol (GPI) anchored proteins are a unique group of membrane proteins found on the surface and certain intracellular compartments of eukaryotic cells. They are bound to the membrane by a GPI moiety and have a number of important functions, including digestion, endocytosis and signal transduction. GPI anchored proteins also reside within lipid rafts, which are microdomains on the phospholipid bilayer composed of sphingolipids and cholesterol. Rafts are thought to be capable of forming semi-stable "islands" of lipids and proteins that act as a platform for a number of important cellular processes, such as T-cell activation, caveolin mediated endocytosis and protein compartmentalisation. The majority of research into rafts has been carried out in single cellular organisms or cell cultures, and their importance within development has been poorly understood.

In this project a proteomic analysis of lipid raft and GPI anchored proteins was made for the proteome of the model organism *Caenorhabditis elegans*. We found a total of 327 predicted GPI anchored proteins from the *C. elegans* genome via a novel four-program prediction method and validated three of those proteins with mass spectrometric (MS) identification. The GPI biosynthesis pathway genes of *C. elegans* were also elucidated via a bioinformatics search. 41 lipid raft proteins were identified using MS, which accounts for the largest number of such proteins found in the worm. This project will hopefully become a starting point for the research of GPI anchored proteins and lipid rafts within the nematode, and shine a light on the properties of these important classes of proteins within the context of a developmentally complex organism.

# Table of contents

# Figures

# Tables

# Abbreviations

| | |
|---|---|
| **Å** | Angstrom |
| **aa** | Amino acid |
| **AceDB** | A *C. elegans* data base |
| **Ala** | Alanine |
| **AMP** | Ampicillin |
| **APP** | Amyloid precursor protein |
| **Asn** | Asparagine |
| **Asp** | Aspartic acid |
| **BCA** | Bicinchoninic acid |
| **BLAST** | Basic local alignment search tool |
| **BLAT** | BLAST –like alignment tool |
| **BME** | β-mercaptoethanol |
| **bp** | Base pair |
| **CAV** | Caveolin |
| **CCD** | Charge coupled device |
| **CCP** | Cambridge centre for proteomics |
| **cDNA** | Complimentary deoxyribonucleic acid |
| **CHO** | Chinese hamster ovarian |
| **CICR** | Calcium induced Calcium release |
| **CID** | Collision induced dissociation |
| **Cy** | Cyanine |
| **Cys** | Cystine |
| **Da** | Dalton |
| **DHE** | Dehydroergsterol |
| **DIGE** | Differential gel electrophoresis |
| **DNA** | Deoxyribonucleic acid |
| **dNTP** | Deoxynucleotide triphosphate |
| **Dol-P-Man** | Dolichol phosphate mannose |
| **DRM** | Detergent resistant membrane |
| **dsRNA** | Double stranded ribonucleic acid |

| | |
|---|---|
| **DTT** | Dithiothreitol |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **EGFR** | Epidermal growth factor receptor |
| **ER** | Endoplasmic reticulum |
| **ESCRT** | Endosomal sorting complex required for transport |
| **ESI** | Electrospray ionisation |
| **EST** | Expressed sequence tag |
| **EtBr** | Ethidium bromide |
| **EtNP** | Phosphoethanolamine |
| **FASTA** | Fast all |
| **GEEC** | GPI-anchored protein enriched endosomal compartment |
| **GFP** | Green fluorescent protein |
| **GlcN** | Glucosamine |
| **Gly** | Glycine |
| **GO** | Gene ontology |
| **GPCR** | G-protein coupled receptor |
| **GPI** | Glycosylphosphatidylinositol |
| **HeLa cell** | Henrietta Lacks cell |
| **HEPES** | 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid |
| **HexNac** | N-acetyl hexosamine |
| **HMM** | Hidden Markov model |
| **HPLC** | High performance liquid chromatography |
| **HRP** | Horse radish peroxidase |
| **ICAT** | Isotope coded affinity tag |
| **IgSF** | Immunoglobin superfamily |
| **IEF** | Isoelectric focusing |
| **IPG** | Immobilised pH gradient |
| **iTRAQ** | Isobaric tags for relative and absolute quantification |
| **KAN** | Kanamycin |
| **kb** | Kilo bases |
| **kDa** | Kilo dalton |
| **LB** | Luria-Bertani |
| **LC** | Liquid chromatography |

| | |
|---|---|
| **MALDI** | Matrix assisted laser desorption/ionisation |
| **Man** | Mannose |
| **mb** | Mega bases |
| **MBS** | MES buffered saline |
| **MDCK** | Madin-Darby canine kidney |
| **MDLC** | Multi-dimensional liquid chromatography |
| **MES** | Morpholineethanesulfonic acid |
| **mRNA** | Messenger ribonucleic acid |
| **MS** | Mass spectrometry |
| **MS/MS** | Tandem mass spectrometry |
| **MudPIT** | Multidimensional protein identification technology |
| **m/z** | Mass to charge ratio |
| **NEB** | New England Biolabs |
| **NEP** | Neprilysin |
| **NGM** | Nematode growth media |
| **NN** | Neural network |
| **OHSt** | Overhydrated hereditary stomatocytosis |
| **o/n** | Overnight |
| **PBST** | Phosphate buffered saline Tween-20 |
| **PCR** | Polymerase chain reaction |
| **PGAP** | Post-GPI attachment to proteins |
| **p*I*** | Isoelectric point |
| **PI** | Phospatidylinositol |
| **PIG** | Phosphatidylinositol glycan |
| **PIPLC** | Phospatidylinositol-specific phospholipase C |
| **PIPLD** | Phospatidylinositol-specific phospholipase D |
| **PMF** | Peptide mass finger printing |
| **PNH** | Paroxysmal nocturnal hemoglobinuria |
| **ppm** | Parts per million |
| **Pro** | Proline |
| **$PrP^{C}$** | Prion protein cellular |
| **$PrP^{Sc}$** | Prion protein scrapie |
| **RNA** | Ribonucleic acid |

| | |
|---|---|
| **RNAi** | Ribonucleic acid interference |
| **RPLC** | Reverse phase liquid chromatography |
| **RTK** | Receptor tyrosine kinase |
| **SAGE** | Serial analysis of gene expression |
| **SCX** | Strong anion exchange |
| **SDS** | Sodium dodecyl sulfate |
| **SDS PAGE** | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| **SEC** | Size exclusion chromatography |
| **Ser** | Serine |
| **SNARE** | SNAP (soluble NSF attachment protein) receptors |
| **SOM** | Self organising map |
| **$S_{ppt}$** | Score physical property pattern |
| **$S_{profile}$** | Score profile |
| **SR** | Sarcoplasmic reticulum |
| **SRP** | **S**ignal recognition particle |
| **SV40** | Simian virus 40 |
| **SVM** | Support vector machine |
| **TAE** | Tris acetate EDTA |
| **TCA** | Trichloroacetic acid |
| **TE** | Tris EDTA |
| **TEMED** | Tetramethylethylenediamine |
| **T-cell** | Thymus cell |
| **TM** | Transmembrane |
| **Tris** | Tris(hydroxymethyl)aminomethane |
| **Trp** | Tryptophan |
| **TX-100** | Triton ™ X-100, $(C_{14}H_{22}O(C_2H_4O)_n)$ |
| **UDP** | Uridine diphosphate |
| **UNC** | Uncoordinated |
| **UV** | Ultraviolet |
| **V-ATPase** | Vacuolar adenosine triphosphatase |
| **Vps** | Vacuolar protein sorting |
| **VSG** | Variant surface glycoproteins |
| **v/v** | Volume/volume |

**w/v** Weight/volume

# *Chapter 1*

# General introduction

## 1.1 Membrane proteins and protein lipid modifications

Since the post-genomic era it has become increasingly apparent that, despite the great strides made in the elucidation of the genome of many organisms it is still not enough for a full understanding of how a cell works. Proteins are responsible for all of the processes which allow a cell to function- from energy production to gene regulation, structural integrity, environmental interface, communication with other cells, and they may even carry hereditary information via the mechanism of epigenetics (Alberts *et al.*, 2008). Therefore, the study of proteins is a subject of fundamental importance within biology, and focus has shifted greatly to their research in recent years, with a view to elucidate all of their functions within the cell and solve one of the greatest challenges within science.

The life of a protein starts from the DNA sequence of its respective gene; the primary sequence is transcribed into mRNA in the nucleus of eukaryotic organisms, which is then transported out of the nucleus where it is translated into proteins via ribosomes in the cytosol. Certain proteins carry sequences which target them to particular cellular compartments, such as the Endoplasmic Reticulum (ER), where additional processing occurs before they become functional. Many proteins undergo some form of post-translational modification, including enzymatic processing, glycosylation, phosphorylation, and various lipid modifications such as myristoylation, palmitoylation, prenylation and C-terminal anchorage via glycosylphosphatidylinositol (GPI) moieties (Hooper and Turner, 1992). Lipid modifications greatly alter the characteristics of proteins by increasing their hydrophobicity, allowing interaction with membranes and facilitate their role in many cellular processes such as signalling and antibody recognition (Carcy *et al.*, 2006; Resh, 2006).

### 1.1.1 The plasma membrane

The plasma membrane is the outermost membrane of the cell and separates its contents from the extracellular environment. It is also the only point of exchange between the intracellular and the extracellular environments, and performs a number of crucial functions for the cell, such as the absorption of nutrients, excretion of waste, communication with extracellular stimuli, protection from the environment and to ensure the correct concentrations of ions and proteins are kept within the cell. Proteins on the plasma membrane perform these vital roles and are therefore the subject of intense interest within biology.

### 1.1.2 Lipid raft microdomains

Plasma membrane proteins are able to move more or less freely within the lipid bilayer (Singer and Nicolson, 1972), and are organised according to interactions with other membrane proteins or association with parts of the cytoskeleton. In addition, distinct lipid domains have also been postulated to have a role in protein organisation within the plasma membrane. This hypothesis first began with the observation that glycosphingolipids, cholesterol and a variety of proteins were resistant to solubilisation in cold non-ionic detergents such as Triton X-100. They were hypothesised to reside within lipid rafts, which are defined as a dynamic clustering of glycosphingolipids and cholesterol in a liquid ordered phase within the outer leaflet of the plasma membrane (Simons and Ikonen, 1997). The membrane is separated into "island" like domains due to the aggregation of the glycosphingolipids and cholesterol, and this arrangement of molecules is thought to create a more thermodynamically stable lipid bilayer than a random arrangement of lipid molecules (Harder and Simons, 1997). It is this property of lipid rafts that is postulated to have a profound effect on the dynamics of proteins within the membrane.

#### 1.1.2.1 General functions of lipid rafts

The unique properties of lipid rafts allows the aggregation of specific proteins within lipid domains, such as caveolin, stomatin, GPI anchored proteins, proteins modified with a variety of lipid modifications, and raft associated cytosolic proteins such as galectins, kinases, and parts of the cytoskeleton. These proteins facilitate a large number of functions within the membrane. Lipid rafts are able to direct cell polarity by domain specific protein segregation and recruitment of cytoskeletal proteins such as actin and microtubules, as has been shown in epithelial cell polarisation (Hoekstra *et al.*, 2003), axonal growth in neurons (Kamiguchi, 2006), and fission yeast cell division and mating (Wachtler and Balasubramanian, 2006). Lipid raft association of certain ligands can be switched on and off depending on modifications such as glycosylation, phosphorylation, acylation, palmitoylation, N-myristoylation and prenylation (Alfalah *et al.*, 1999; Kabouridis and Jury, 2008; Resh, 2004; Waheed and Jones, 2002), which affects their localisation and interactions with target proteins. Rafts are also involved with other diverse cellular processes such as cell adhesion (Harris and Siu, 2002) and membrane fusion through SNARE proteins (Lang, 2007).

### 1.1.2.2 Endocytosis with caveolin

One other major function involving lipid rafts is endocytosis, and this is brought about by caveolae (Nichols, 2003), which are smooth, non-clathrin coated invaginations on the plasma membrane. Caveolae were first observed over 50 years ago (Yamada, 1955) and are formed by the 22 kDa protein caveolin (Rothberg *et al.*, 1992). This protein has 3 homologues in humans (*CAV1*, *CAV2*, *CAV3*), with *CAV1* being the most important in the creation of caveolae and has two splice variants, CAV1α and CAV1β (Schlegel *et al.*, 1998). Caveolin has one 33 amino acid transmembrane domain in the centre of the protein, and its N and C-termini are exposed to the cytosolic side of the membrane (Kurzchalia *et al.*, 1994). The structure is assembled in the Golgi apparatus before transportation to other cellular compartments, in contrast to clathrin mediated transport where the vesicles are formed *de-novo* on the plasma membrane (Schmid, 1997). Caveolae are maintained by an association of caveolin, sphingolipids, cholesterol, and various raft associated proteins such as GPI anchored proteins (Anderson, 1998). Caveolin, however, may exist in non-caveolae lipid raft environments, where they have a different set of interacting proteins and exhibit different properties (Lajoie *et al.*, 2009). Caveolae are also extensively involved in several signalling pathways including receptor tyrosine kinase (RTK) (Mukherjee *et al.*, 2006), G-protein coupled receptors (GPCR) (Patel *et al.*, 2008), and T-cell antigen receptor in the immune response (Kabouridis and Jury, 2008). Signalling proteins are sequestered within the caveolae structure, which is used as a mechanism to partition receptors from their ligands; caveolae also helps in the maintenance of the signal giving greater stability to receptor- ligand interactions once they are formed.

With involvement in so many cellular processes (especially those in cell signalling and endocytosis) it comes as no surprise that lipid rafts are thought to play a major role in a variety of disease processes. Rafts are postulated to have a role in cancer proliferation, where they contain a number of signalling pathways that cause either proliferation or apoptosis (Patra, 2008). The prion protein also reside within lipid rafts, which causes Creutzfeldt-Jakob disease in humans (Taylor and Hooper, 2006). The processing of the amyloid precursor protein (APP) in Alzheimer's disease is raft associated, with recent evidence pointing to the cholesterol synthesis inhibitor statin as a possible drug target in treatment of the disease (Reid *et al.*, 2007; Whitfield, 2006). Caveolae have an important role as an entry point for viruses and their toxins, and is involved in the infectivity of simian virus 40 (SV40) (Anderson *et al.*, 1996) and used as one of the routes of entry for the cholera toxin (Parton, 1994). Lastly, vascular diseases such as hypertension are affected by caveolae, due to the large number of signalling pathways present within this lipid domain (Callera *et al.*, 2007; Insel and Patel, 2009).

#### 1.1.2.3 The controversial nature of lipid rafts

Lipid raft research has made immense strides in the past 20 years, with the discovery of many new mechanisms of membrane biochemistry in important areas such as signalling, transport, and protein-protein interactions. However the concept of the raft is still not fully understood, with properties for the domain hotly debated within the field. Much of the controversy comes from the exact definition of what a raft is, with many researchers finding the traditional definition of extraction by cold non-ionic detergents to be arbitrary and devoid of biological meaning (Shaw, 2006); moreover, different methods of extraction can produce rafts with different lipid content and associated proteins (Gallegos *et al.*, 2006). There is a wealth of evidence in favour of the formation of thermodynamically stable, tightly packed associations of glycosphingolipids and cholesterol (Boggs, 1987; Sankaram and Thompson, 1990; Smaby *et al.*, 1996), and lipid rafts have been visualised *in vitro* using model membranes containing physiological ratios of phospholipids, sphingolipids and cholesterol (Prenner *et al.*, 2007). In recent years lipid rafts have also been visualised *in vivo* (Ishitsuka *et al.*, 2005) but the raft structures found are much more transient and smaller than the ones obtained with model membranes, prompting questions as to just how big a role lipid rafts play within the various cellular mechanisms they take part in (Shaw, 2006). The importance of lipid rafts within cell physiology is also a subject of debate, with some studies giving the conclusion that rafts are necessary for cellular function while others found them to be redundant for certain processes ascribed to them (Nichols, 2005). The study of lipid rafts is a very active field with implications in a number of diverse fields, and what can be found out in the future can only improve our understanding of many important disease processes, and our understanding of biology in general.

### 1.1.3 Glycosylphosphatidylinositol (GPI) anchored proteins

Certain proteins within the cell can become attached to the outer plasma membrane via a GPI anchor. These proteins do not have a transmembrane domain, but are covalently bonded to a glycolipid called GPI at the C-terminus of the protein that allows the structure to be stably associated with the membrane. The attachment of the anchor occurs in the ER lumen and the protein is transported to the outer membrane via the secretory pathway. GPI anchored proteins have a wide variety of functions, with the only common feature among them being a secretion signal at the N-terminal end of the protein and a GPI anchor attachment sequence at the C-terminus (Paulick and Bertozzi, 2008). Although in theory any protein may become GPI anchored as long as they contain the signal sequences present at their termini, there exist a number of proteins that possess this form of anchoring as an evolutionarily conserved feature (Udenfriend and Kodukula, 1995a). GPI anchored proteins were first found in the intracellular parasite *Trypanosoma brucei*, where they are called variant surface glycoproteins (VSG), and subsequent experiments have shown them to be crucial in the biology of the organism, in which abolition of the GPI anchor destroys the infectivity of the parasite (Lillico *et al.*, 2003). They are also important in mice, where their absence causes embryonic lethality and is postulated to be responsible for sperm/egg fusion during fertilisation (Alfieri *et al.*, 2003). The absence of two GPI anchored proteins also cause the X-linked hereditary haemophilic disease paroxysmal nocturnal hemoglobinuria in humans (Brodsky and Hu, 2006). Maturation of the malaria parasite *Plasmodium falciparum* depend on GPI anchored proteins, which are suggested as a target for drugs against the organism (Naik *et al.*, 2003). GPI anchored proteins have been shown to have roles in cell adhesion, catalysis, viral budding and antibody recognition (Karagogeos, 2003; Metzner *et al.*, 2008; Sly and Hu, 1995; Tarleton, 2007). GPI anchored proteins are associated with lipid rafts and can constitute a significant proportion of proteins found within the microdomain (Paulick and Bertozzi, 2008). Lipid raft association also allows certain GPI anchored proteins to interact with signalling pathways, including GPCRs (Landry *et al.*, 2006), T-cell activation (Wollscheid *et al.*, 2004), and the insulin signalling pathway (Sharom and Radeva, 2004). GPI anchored proteins are present in all eukaryotic organisms and can represent a significant subset of plasma membrane proteins in some species, such as in *T. brucei* and *Leishmania major* (Ferguson, 1999).

#### 1.1.3.1 The structure of the GPI anchor

There is variation between species in the exact structure and makeup of the GPI anchor. The core backbone of the anchor for a number of species is phosphoethanolamine- mannose(α1-2)mannose(α1-6)mannose(α1-4)gulcosamine(α1-6)*myo*-inositol, and can be found in organisms as diverse as *T. brucei*, *P. falciparum*, *Saccharomyces cerevisiae* and mammals (Ferguson *et al.*, 1999; Ikezawa, 2002; Pittet and Conzelmann, 2007). GPI anchored proteins in mammals have an additional phosphoethanolamine linked to the 2-position of the first mannose (adjacent to glucosamine) (Orlean and Menon, 2007). The structure is flexible and can have differences between cell types, where additional modifications occur, such as N-acetyl hexosamine (HexNAc) modification of the first mannose in rat brain Thy-1 (Homans *et al.*, 1988). Modification of the fatty acid chain in the GPI anchor takes place in the ER after transport to the Golgi, which is essential for its association with lipid rafts (Maeda *et al.*, 2007). There are 12 steps overall for the synthesis of a complete GPI anchor, with the attachment of the protein occurring via a transamidase complex in the ER (Meyer *et al.*, 2000). The complete GPI anchored protein is then transported to the Golgi, where additional modifications to the fatty acid tail occur (Fujita and Jigami, 2008), before it finally ends up on the surface of the cell.

## 1.2 Different proteomics techniques and their uses

With the production of a vast number of EST libraries and genome sequences in the last 20 years it has become increasingly clear that transcription level and gene annotation data alone are unable to explain the vast complexities of the cellular machinery that give rise to life. It was realised that in order to properly study the internal workings of living organisms a global method of protein analysis must be performed. This, in conjunction with previous organism wide studies based on mRNA and DNA, is thought to be able to give a more complete picture of the intricacies of metabolism, regulation, development and heredity, putting us one step closer to a more complete understanding of cellular biology.

Traditional techniques for the analysis of proteins generally involve the intensive characterisation of a small subset of individual proteins with respect to their expression, post translational modifications, sequence, interactions, and 3D structure. Work with large protein mixtures did occur but have mostly been confined to relatively simple analysis, as methods for the global analysis of proteins were either non-specific or time consuming (Giddings, 1984). Protein research became revolutionised with the sequencing of the genomes of various organisms in the 90's, which led directly to the invention of the field of proteomics. Coined from the words PROTEin and genOME, the study of the proteome is defined as the total analysis of all proteins within a biological system or process. The presence of well annotated genomes with EST data allowed the production of predicted protein sequence databases, which when combined with proper resolution and the use of mass spectrometry allow high-throughput identification of thousands of proteins from complex biological samples (Shevchenko *et al.*, 1996). Improvements in mass spectrometry (MS) technology have also contributed to the speed and ease with which complex mixtures of proteins are identified (Han *et al.*, 2008). This global analysis has been used as a powerful tool in many aspects of biological research, such as identification of diseased cell biomarkers, screening for interacting partners, organelle protein organisation, and global protein network analysis (Dunkley *et al.*, 2004; Motoyama and Yates, 2008; Rogers and Foster, 2009; Zhao *et al.*, 2009).

### 1.2.1 Separation techniques for proteomics

Proteomic studies require the use of multiple separation techniques that allow the resolution of individual proteins from complex mixtures. The first standard procedures involve the use of 2D gels, which is still one of the great workhorses of the proteomics field (Lopez, 2007). However, this technique has been increasingly superseded by the use of multi-dimensional liquid chromatography (MDLC), which is thought to have greater reproducibility, but lack the quantitative analysis that is available with gel based systems (Delahunty and Yates III, 2005). Hybrid techniques in which the different dimensions are separated by gel and liquid chromatography have also recently become popular, especially with the advent of "shotgun" sequencing from improved mass spectrometric analysis (Motoyama and Yates, 2008). A general workflow for 2D electrophoresis and MDLC proteomes is given in Figure 1.1; the relative merits and weakness of these different techniques will be discussed below.

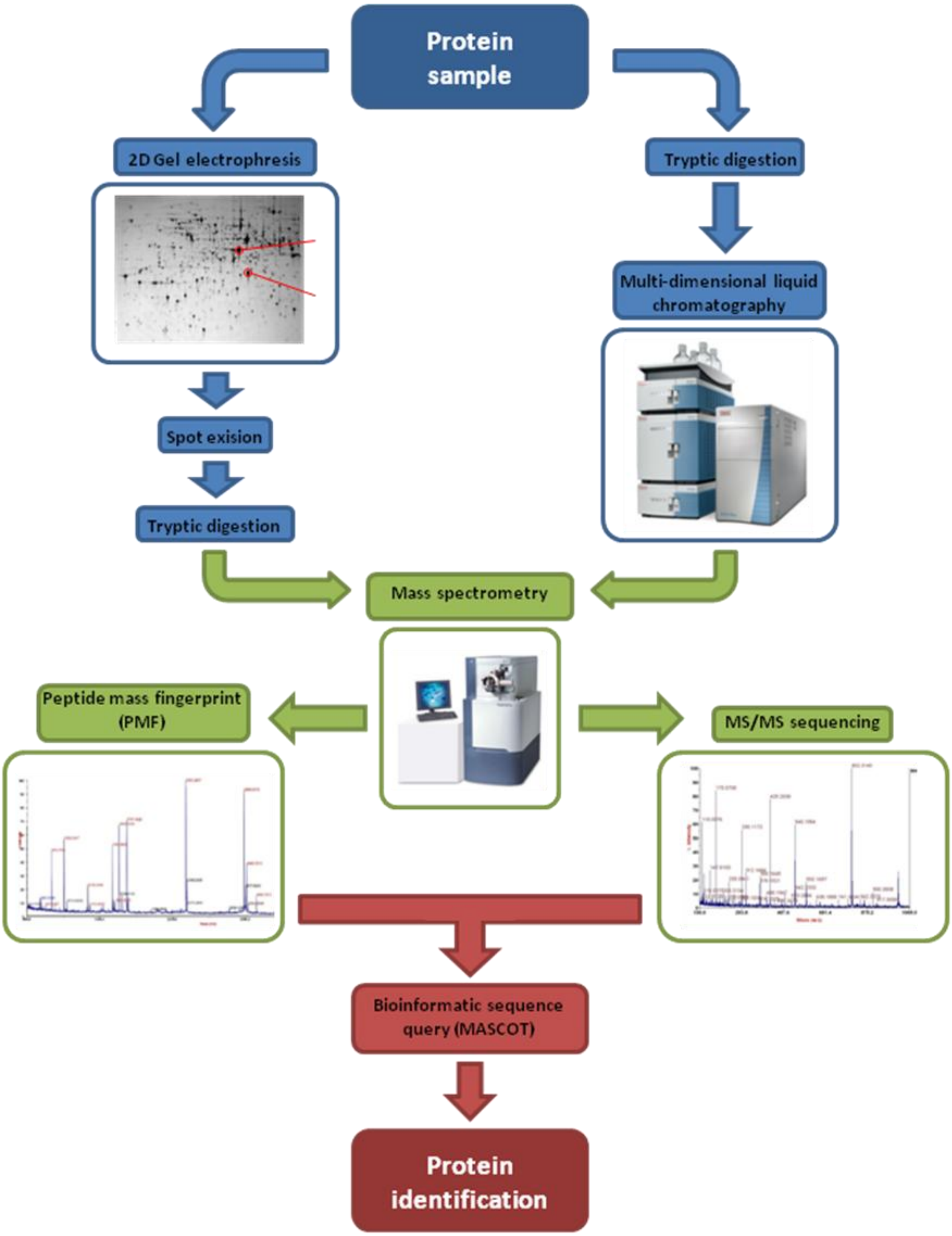



**Figure 1.1. Workflow of 2D gel electrophoresis and MDLC in proteomic studies**. The process is shown from the initial protein sample stage to the final identification of individual proteins. The 2D gel image is of an *S. cerevisiae* sample adapted from http://abdn.ac.uk/ims/proteomics/2dgelsmaps.shtml. The liquid chromatography equipment is a Thermo Fisher Scientific Accela system. The mass spectrometer is a Waters MALDI Synapt HDMS system. The PMF spectrum was adapted from http://www.york.ac.uk/res/schisto/peptide_mass_fingerprint.htm, and the MS/MS spectrum was adapted from http://www.umdnj.edu/proweb/services.htm.

#### 1.2.1.1 2D Gel electrophoresis

2D electrophoresis has been used as a technique for protein analysis long before the advent of modern proteomics. It was invented in 1956 and was first used for the separation of human serum proteins (Smithies and Poulik, 1956). Several advancements followed, culminating in the techniques developed by O'Farrell in the mid 70's (O'Farrell, 1975), which became the standard procedure for 2D analysis today.

##### 1.2.1.1.1 Principles of 2D gel electrophoresis

2D gels separate proteins in the first dimension according to their isoelectric point (p*I*) and in the second dimension by their molecular mass. The p*I* of a protein is determined by its overall charge, and the proteins are resolved via isoelectric focusing (IEF), in which a charge is placed along a pH gradient produced by carrier ampholytes- small molecules that can act as both an acid and a base- that facilitate the migration of each protein to their correct location. The invention of immobilised pH gradients (IPG) (Bjellqvist *et al.*, 1982) allowed further improvements for the resolution of proteins in the first dimension. Modern proteomic analysis tend to use commercially available precast IPG strips, with different companies offering a large selection of pH ranges for different sensitivity requirements (Taylor and Coorssen, 2006).

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) is used for the separation of proteins by molecular mass in the second dimension. This method has changed relatively little compared to the advancements made in IEF. For protein samples of high complexity a gradient gel may be used to improve resolution, and larger gels have greater resolving power than smaller gels (Lopez, 2007). The properties of the protein sample and the specific resolution required for the experiment dictates what kind of gel is best used for analysis.

#### 1.2.1.1.2 Visualisation of 2D gels

Proteins resolve into spots on the gel, which are visualised before analysis can begin. The most commonly used stains for 2D gels are Coomassie Brilliant Blue, silver stain, and fluorescent dyes such as Sypro Ruby. Coomassie has a generally linear response to protein concentration and is used when quantitation of the spots is required. Silver stain is generally non-linear for protein concentration, but can have up to 5 times the sensitivity of Coomassie Blue. It is mainly used for confirmation of the presence of proteins on the gel. Fluorescent dyes have high sensitivity and can be used to quantify proteins, and requires the use of a fluorescent scanner for visualisation. Coomassie and fluorescent stains are fully compatible with MS identification due to their ability to be destained; silver staining can be modified to become compatible with MS (Shevchenko *et al.*, 1996). The choice of staining technique in a proteomic experiment is dictated by the needs of the experimental design and the sample analysed.

#### 1.2.1.1.3 Computer analysis of 2D gels

After visualisation the gel is scanned into a computer, where the image is manipulated to align different gels together and perform spot matching, with the intensities of the spots calculated to give quantitative analysis when required. A number of commercial programs are available for this, including PDQuest, Phoretix 2D Advanced, Melanie, and others. Each of these programs has their own strengths and weaknesses, but are generally competent when used to examine most protein samples (Marengo *et al.*, 2005). After annotation of the gel, spots of interest are picked, destained and subsequently analysed by MS.

#### 1.2.1.1.4 Advantages of 2D gel electrophoresis

One of the major advantages of 2D electrophoresis is its ability to analyse individual proteins in a quantitative manner. This allows comparisons of global expression patterns between different biologically significant samples. This approach has been used to find potential biomarkers (Wong *et al.*, 2009), proteins interacting partners (Choi *et al.*, 2004), and the changes in expression profile brought about by specific conditions.

2D electrophoresis can also be used to study post-translational modifications. Modifications such as glycosylation that alter the charge of a protein can be easily visualised as a horizontal shift within the gel, and the degree of modification worked out by its isoelectric point relative to the unmodified protein (Sickmann *et al.*, 2002). Immuno-blotting of a protein subfamily or modification can also be performed, allowing very accurate analysis of important sub-families of proteins en masse (Balen *et al.*, 2006).

#### 1.2.1.1.5 Limitations of 2D gel electrophoresis

2D electrophoresis has a number of weaknesses that limits its uses when analysing certain proteomes. The technique has a limited dynamic range due to the inherit properties of polyacrylamide gels, which means proteins of low abundance such as transcription factors cannot be analysed effectively. The limited pH range (3-10 pH) that can be achieved by ampholytes also exclude the analysis of very basic and very acidic proteins (Issaq and Veenstra, 2008). Lastly, proteins with extensive hydrophobic regions, such as membrane proteins, and proteins with low solubility are poorly resolved by IEF. This is due to the need for non-ionic detergents during isoelectric focusing so that proteins can migrate to their proper p*I*. Lipids present within the sample also hamper IEF, giving a streaking effect on the gel and poorly focused spots (Issaq and Veenstra, 2008).

One other fundamental problem of 2D electrophoresis has been the lack of reproducibility between experiments. In the early days of the technique different laboratories had very different protocols for performing 2D electrophoresis, and this resulted in different looking gels for the same protein sample. Even gels within the same laboratory will run to slightly different dimensions, as the large number of variables per run (pipetting errors, gel casting, staining and destaining time, etc) makes each gel unique and non super-imposable. In recent years there has been a great many advances designed to alleviate this problem (Issaq and Veenstra, 2008). Software analysis has improved drastically in its ability to match different gels together (Marengo *et al.*, 2005), and the development of 2D differential gel electrophoresis (DIGE) allows different protein samples to be visualised on the same gel via florescent labelling, which alleviates the problem of variability between different gels.

#### 1.2.1.1.6 2D Differential Gel Electrophoresis (DIGE)

2D DIGE was developed as a technique to reduce inter-gel variation and improve reproducibility of 2D electrophoresis experiments (Unlu *et al.*, 1997). Different protein samples can be labelled with up to three different fluorescent probes (Cy2, Cy3, and Cy5) that have the same mass, charge, and different absorbance wavelengths (488nm, 532 nm and 633 nm, respectively). This enables different protein samples to be run on the same gel and eliminates variation induced by multiple gels. The inclusion of an internal standard can also aid the comparison of many different samples, improve protein quantification and reduce the number of gels needed to be run (Alban *et al.*, 2003). The dyes used have very high sensitivity, so that proteins not normally seen with conventional 2D electrophoresis can be detected (Marouga *et al.*, 2005). 2D DIGE relies on the covalent attachment of the dye to unmodified lysine residues within a protein, and in order for quantitative analysis to be performed the sample is minimally labelled at on average one dye per protein. This means that effectively only 3-5% of the total protein of any sample is labelled, and proteins that do not contain lysine will never be detected (Marouga *et al.*, 2005). The proteins must also be imaged with a specialised fluorescent scanner, with proprietary software (DeCyder) that increases running costs. Despite these shortcomings, 2D DIGE has become one of the most important techniques in proteomics today and has been used in the analysis of biomarkers (Wong *et al.*, 2009), *Arabidopsis thaliana* proteins (Borner *et al.*, 2005), human liver (Brizard *et al.*, 2009), cancer cells (Schaaij-Visser *et al.*, 2009), mitochondria (Mathy and Sluse, 2008), stem cells (Evans *et al.*, 2004), and other proteomes.

#### 1.2.1.2 Multi-dimensional liquid chromatography (MDLC) and "shotgun" sequencing

The field of proteomics experienced a mini revolution with the adoption of tandem mass spectrometry (MS/MS). This technique involves fragmenting peptides into their component amino acids, which allows the elucidation of the amino acid sequence of the peptide, which increases the accuracy of protein identification over the older peptide mass finger printing (PMF) method. MS/MS is also capable of analysing the tryptic digests of protein mixtures directly (Link *et al.*, 1999), without the need for the resolution of individual proteins prior to digestion. This new way of analysing proteomes was termed “shotgun” sequencing (Motoyama and Yates, 2008), after the well known DNA sequencing method of the same name (Wilson *et al.*, 1994).

High performance liquid chromatography (HPLC) was first explored in the 80’s as a technique for the separation of proteins in a 2D plane (Giddings, 1984). Although its application to proteomics was initially slow the use of the technique has gained momentum in recent years, and has become an advanced method of protein separation for proteomic projects today. HPLC is suited to shotgun proteomics due to its high resolving power, especially since the number of tryptically digested peptides generated from a complex protein sample can be as high as 600,000 (Motoyama and Yates, 2008). A milestone for this technique was achieved in 2001 with the invention of Multidimensional protein identification technology (MudPIT) (Washburn *et al.*, 2001), which has shaped the course of MDLC analysis in proteomics.

#### 1.2.1.2.1 Principles of MDLC analysis

While it is theoretically possible for any combination of different techniques to be used for the two (or more) dimensions of separation, a set of common practices have started to become established, in accordance to the specific requirements of the experiment. The analysis of proteomes with MDLC can be partially (offline) or fully (online) automated. For the first dimension of peptide separation a variety of techniques can be used, which includes LC methods such as size exclusion chromatography (SEC) (Peuravuori *et al.*, 2007), strong cation exchange (SCX) (Washburn *et al.*, 2001), strong anion exchange (Motoyama *et al.*, 2007), as well as non- LC methods such as SDS PAGE (Trelle *et al.*, 2009) and IEF (Cargile *et al.*, 2005). The second dimension separation can in theory be achieved with any technique that is orthogonal to the one used in the first dimension; however this part of MDLC analysis is almost always performed with reverse phase LC (RPLC), as this method has a high resolving power and has the advantage that the column can be linked directly to certain mass spectrometers for coupled peptide elution and analysis (Motoyama and Yates, 2008). Offline 2D can be performed with any of the techniques in the first dimension, with LC methods for full shotgun experiments involving the tryptic digestion of protein samples at the start, and SDS PAGE/IEF used for partial shotgun experiments where intact proteins are resolved before being digested for the second dimension, allowing for the inclusion of additional information such as proteins mass and fraction pI range. There is also scope for optimisation of each fraction to achieve the highest number of protein identifications for the sample. Online methods require the use of an LC method in the first dimension, with computer controlled automated valves that feed the fractions from the first dimension to be separated in an orthogonal technique in the second dimension. SCX is usually used for the first dimension, though others have also been used for the analysis of different protein samples (Nägele *et al.*, 2004). Online LC has less resolving power than offline due to the lack of optimisation of each fraction in between each dimension of analysis. It is however the preferred method for large scale proteomic projects, as its high degree of automation allows a high turnover of protein analysis and the uniformity of conditions also allow better comparisons between different samples. This method is also preferred when the protein sample size is small, as the amount of sample wastage is minimised during handling between the different dimensions (Motoyama and Yates, 2008).

#### 1.2.1.2.2 Advantages of MDLC analysis

MDLC has many advantages over 2D electrophoretic techniques as a method of proteomic analysis. The technique has a high dynamic range and may detect proteins of low abundance, due to a lack of need for protein detection before being identified by MS/MS. It is capable of a much higher throughput than 2D electrophoresis, since the second dimension can be directly attached to the mass spectrometer for extremely rapid analysis. Lastly liquid chromatography allows the analysis of proteins that are unsuited for 2D electrophoresis, such as membrane proteins, highly acidic and highly basic proteins, as the digestion of proteins prior to analysis reduce problems with solubilisation. These advantages have lead to the technique becoming widely adopted for proteomics projects in recent years, including post translational modifications (Trelle *et al.*, 2009), sub proteomes (Feuk-Lagerstedt *et al.*, 2007), model organisms (Baggerman *et al.*, 2005; Husson *et al.*, 2009; Washburn *et al.*, 2001) and biomarker discovery (Whelan *et al.*, 2009).

#### 1.2.1.2.3 Limitations of MDLC analysis

With recent trends in LC technology becoming increasingly sophisticated some of the earlier limitations with the technique, such as an inability to analyse post-translational modifications, have been steadily resolved (Rogers and Foster, 2009). The technique still has a few weaknesses, such as when peptides of highly abundant proteins are preferentially sampled, leading to the peptides of low abundance proteins becoming swamped out and unidentified in the mass spectrometer (Han *et al.*, 2008). This situation can be avoided by better pre-digestion fractionation of protein samples, and by careful optimisation of eluted fractions from each dimension. The biggest limitation of the technology is its difficulty in the analysis of proteins in a quantitative manner. Many technologies have been developed to alleviate this problem in recent years, and most involve the isotopic tagging of proteins to quantify them in the mass spectrometer, such as isotope coded affinity tag (ICAT) (Gygi *et al.*, 1999) and isobaric tags for relative and absolute quantification (iTRAQ) (Ross *et al.*, 2004). This involves subjecting the protein sample to chemical reactions with isotopically labelled tags, which are then detected in the mass spectrometer as a series of peaks with stereotyped differences in detected mass. Different protein samples (or an internal standard for one sample) may be tagged with different isotopes and the relative heights of the isotopic peaks can then be used as a measurement of relative protein abundance. Recently massive strides have also been made in non-labelled protein quantification, where spectrometrical peaks from ordinary runs of LC MS/MS are analysed with computer programs that allow quantitative comparison between different experiments (America and Cordewener, 2008).

### 1.2.2 The use of mass spectrometry in proteomic studies

Mass spectrometry is one of the oldest techniques for the analysis of compounds in organic chemistry (Borman *et al.*, 2003). It works by first converting the sample to be analysed into gas phase ions with an ion source, which are the placed into a mass analyser that separates them based on their mass to charge ratio (m/z), which is recorded by a detector at the end of the instrument. The electron bombardment in the first stage of the mass spectrometer fragments the compound into a distinct set of ion peaks, and this unique pattern is used to elucidate the structure of the sample under test. Proteomic analysis however requires whole peptides to be analysed in a relatively intact manner, as extensive fragmentation will produce too much noise in the ion peaks, which would hinder the identification of the peptide. Proteomic samples therefore need to be subjected to "soft" ionisation, where the peptides are ionised in the mass spectrometry instrument for detection, but are otherwise left relatively unchanged (Canas *et al.*, 2006).

#### 1.2.2.1 Development of the ion source

Several soft ionisation techniques such as fast atom bombardment (Morris *et al.*, 1981) and plasma desorption (Macfarlane and Torgerson, 1976) were developed in the 70's and 80's when interest grew in the use of mass spectrometry for the study of proteins. The techniques offered unique perspectives on peptide analysis, but were generally less sensitive than other widely used peptide sequencing methods such as Edman sequencing, requiring much higher amounts of sample and thus were not routinely adopted for the analysis of proteins. It was not until the late 80's that protein mass spectrometry came of age with the invention of electrospray ionisation (ESI) (Fenn *et al.*, 1989) and matrix assisted laser desorption/ionisation (MALDI) (Karas and Hillenkamp, 1988; Tanaka *et al.*, 1988). These techniques allowed accurate and speedy analysis of peptides, which paved the way for the advent of the field of proteomics today. Both John Bennett Fenn and Koichi Tanaka, who were the first people to develop ESI and laser desorption techniques respectively, each received the

2002 Nobel Prize in Chemistry for their pioneering work in the field of protein analysis and their overall contributions to biological research.

#### 1.2.2.2 ESI

ESI works by forming small charged micro droplets of soluble peptides by passing them through a narrow capillary under high voltage, which can be done under atmospheric conditions. As the droplets fragment and evaporate ionised peptides are formed, which is then analysed in the rest of the instrument. Salts and detergents need to be removed from the sample to prevent adduct formation, and this is usually done by reverse phase chromatography. Ions produced in ESI tend to be multiple charged, which can give a range of m/z ratios for each peptide and aid in the accurate mass analysis of the peptide. The multiple charge also allows easier fragmentation of the peptide for further analysis with MS/MS (Canas *et al.*, 2006). One disadvantage of ESI is its inability to retain the sample once it has been sprayed into the mass spectrometer, which allows less scope for optimisation of the sample within the instrument.

#### 1.2.2.3 MALDI

The principles of MALDI mass spectrometry involve the ionisation of the sample by the transfer of energy from a matrix compound via ultraviolet (UV) excitation. Peptides are co-dissolved with the matrix compound at a molar ratio of 1 to 10,000, which are subsequently plated onto a sample probe. This creates a crystal structure of matrix compound with embedded peptides within. After being hit by a pulse UV laser under vacuum the matrix absorbs the energy and becomes partially vaporised along with some of the embedded peptides (Hillenkamp *et al.*, 1991). The matrix causes the ionisation of the peptides in the gas phase, which is then passed onto the rest of the MS instrument for analysis. MALDI is relatively tolerant of sample contaminants such as buffers and salts, and has the advantage that the proteins analysed can be re-examined many times before they are depleted. The technique however has different ionisation properties for peptides of different amino acid sequences, and is less amenable to automation than ESI due to the need for the plating of the sample before analysis in a mass spectrometer (Canas *et al.*, 2006).

#### 1.2.2.4 Fragmentation and identification of proteins with mass spectrometry

Proteins samples are commonly examined with mass spectrometry after digestion with an endopeptidase such as trypsin. This procedure produces defined peptides that result in less complex fragmentation patterns within the instrument and allows clearer interpretation of the results. The first level of peptide identification comes from the total mass of the peptide, which is produced by a single MS run. The mass of one peptide gives little information about the amino acid constitution of the peptide in question; however, the masses of all of the peptide fragments from the tryptic digest can be pooled together to form a "fingerprint" of peptide masses for the protein of interest. This fingerprint can then be searched with an algorithm (such as MASCOT (Perkins *et al.*, 1999)) against theoretical tryptic digestions of protein sequences *in silico*, which results in the identification of the sample protein. This method is called peptide mass fingerprinting (PMF) and was one of the first methods adopted in proteomic studies for the identification of proteins (Pappin *et al.*, 1993).

With the advent of tandem MS/MS instruments it has become possible to produce further fragmentations of the peptides produced during tryptic digestion, which allows the sequencing of those peptides from the resulting MS/MS spectra (Hunt *et al.*, 1986). MS/MS is a more sensitive method of protein identification than PMF and is also compatible with the analysis of peptide mixtures, paving the way for “shotgun” MDLC based methods for proteomic studies (Motoyama and Yates, 2008). Recent advances in mass spectrometry technology include $MS^n$ fragmentation, which is used on samples such as phosphorylated peptides that have proven difficult to fragment using MS/MS alone (Rogers and Foster, 2009). New instruments such as Orbitrap mass spectrometers are able to analyse intact proteins, and have show great promise in improving the analysis of post-translational modifications with even greater coverage than before (Yates *et al.*, 2009).

### 1.2.3 The contributions of proteomics to biology

Proteomics has become one of the most widely used techniques in biology today. Proteomic projects have been used on model systems as diverse as viruses, bacteria, eukaryotes and whole organisms such as *drosophila*, *Arabidopsis* and humans, which has contributed greatly to the understanding of the biology of those organisms (Engstrom *et al.*, 2004; Han and Lee, 2006; Komatsu *et al.*, 2007; Mathy and Sluse, 2008). Studies on subcellular locales and post-translational modifications have improved our understanding of important processes such as signal transduction in a global manner (Mathy and Sluse, 2008; Rogers and Foster, 2009). Quantitative proteomic techniques have been used in the study of disease biomarkers, especially for a variety of cancers that has yielded many novel potential therapeutic targets and new methods for treating the disease (Conrad *et al.*, 2008; Ikonomou *et al.*, 2009;

Zhao *et al.*, 2009). Proteomics projects have also been used extensively in the emerging field of systems biology, where the technique has been used in the creation and validation of models for complex regulatory networks (Ivakhno and Kornelyuk, 2006; Kreeger and Lauffenburger, 2010; Maurya *et al.*, 2007). The field of proteomic research has enjoyed an explosive growth in the past decade and will likely become one of the most import techniques in biology for the post genomic era.

## 1.3 Progress in proteomics for lipid raft and GPI anchored proteins

### 1.3.1 Proteomic analysis of lipid rafts

Ever since the explosive growth in proteomic analysis of the past ten years there has also been a large amount of interest in using these techniques for the study of proteins in lipid rafts. Membrane proteins are notoriously difficult to analyse using 2D electrophoresis, due to their alkaline nature and poor insolubility in the non-ionic detergents required for IEF (Santoni *et al.*, 2000). Shotgun techniques using MDLC MS/MS do not have these disadvantages and are used more frequently for the analysis of these proteins (Wu and Yates, 2003). Gel based methods however may reveal different sets of proteins to shotgun techniques when used on the same sample (Li *et al.*, 2003; Li *et al.*, 2004a). Past studies of lipid raft proteomes include Jurkat T-cells (von Haller *et al.*, 2001), bovine Neutrophils (Nebl *et al.*, 2002), HeLa cells (Foster *et al.*, 2003), Human smooth muscle (MacLellan *et al.*, 2005), adipocytes (Kim *et al.*, 2009), the fungus *Candida albicans* (Insenser *et al.*, 2006) and others. Commonly identified proteins were those involved in the make-up of the cytoskeleton, signalling molecules such as heterotrimeric G-proteins, stomatin, flotillin, caveolin, lectins, heat shock proteins such as hsp90, and endosomal proteins such as components of the proton pump V-ATPase. There are two major contaminants within almost all proteomic studies of lipid rafts, namely mitochondria proteins and ER associated proteins. In fact, these contaminants are so ubiquitous that some have questioned whether they might indeed have raft association in some way or other (Bae *et al.*, 2004); other studies, however, seem to refute such an idea, based on more traditional methods of raft determination such as sensitivity to cholesterol depletion (Foster, 2008; Zheng *et al.*, 2009).

### 1.3.2 Proteomic analysis of GPI anchored proteins

Progress with GPI anchored proteins using proteomics has been relatively slow compared to the analysis of lipid rafts. Most of the proteomic work on this class of proteins has been done in the plant model *A. thaliana*, with the first such study performed on 2D gels using antibody staining and N-terminal sequencing for protein identification (Sherrier *et al.*, 1999). This was followed up later with a large scale 2D DIGE analysis in which 30 GPI anchored proteins were identified with LC-MS/MS (Borner *et al.*, 2003). Other proteomic projects include the identification of GPI anchored proteins from pollen (Lalanne *et al.*, 2004), myelin sheath (Lalanne *et al.*, 2004), the parasite *P. falciparum* (Gilson *et al.*, 2006) and human HeLa cells (Elortza *et al.*, 2003). All of these projects follow the same procedure of "Shave and conquer" (Elortza *et al.*, 2003), in which membrane proteins were subjected to phosphoinositol-specific phospholipase C (PIPLC) digestion and the released GPI anchored proteins were extracted via Triton X-114 phase partitioning (Bordier, 1981). Phosphoinositol-specific phospholipase D (PIPLD), an alternative phospholipase with specificity also for GPI anchored proteins, was used by Elortza *et al.* in a study of GPI anchored proteins in HeLa cells and *A. thaliana*, and was found produce a different set of proteins when compared with digestion with PIPLC (Elortza *et al.*, 2006; Elortza *et al.*, 2003).

### 1.3.3 Prediction of GPI anchoring using bioinformatics programs

There has been a large amount of progress made in the past 10 years on the prediction of GPI anchored proteins from protein databases. Most prediction programs focus on the C-terminal anchor motif, using the sequences from a learning set of experimentally determined GPI anchored proteins to predict protein anchorage and the GPI attachment site. The first program came from Eisenhaber *et al.* and is called the BIG PI prediction program, available online at http://mendel.imp.ac.at/sat/gpi/gpi_server.html (Eisenhaber *et al.*, 1999). Since then the program has been updated (Eisenhaber *et al.*, 2003d), with several other programs GPI-SOM http://gpi.unibe.ch/ (Fankhauser and Maser, 2005), DGPI (http://129.194.185.165/dgpi/), FragAnchor http://navet.ics.hawaii.edu/%E2%88%BCfraganchor/NNHMM/NNHMM.html (Poisson *et al.*, 2007) and PredGPI http://gpcr2.biocomp.unibo.it/predgpi/ (Pierleoni *et al.*, 2008) also available on the web. The wealth of prediction programs allows an in-depth bioinformatic analysis of potential GPI anchored proteins within a genome, which paves the way for a comprehensive proteomic study of these proteins within the desired organism.

## 1.4 *Caenorhabditis elegans* and its contributions to biological research

*C. elegans* is a small soil living nematode that was first analysed by Sydney Brenner over 30 years ago (Brenner, 1974) and has since become one of the most intensely studied model organisms in the world. Brenner wanted to find a model organism to bridge the gap between simple unicellular organisms such as yeast and developmentally complex organisms such as *Drosophila melanogaster*, and *C. elegans* was chosen for this purpose after much consideration. The multicellular nature, ease of genetic manipulation and invariant lineage of *C. elegans* made the nematode an ideal organism for the study of development, growth, and aging.

### 1.4.1 The biology of *C. elegans*

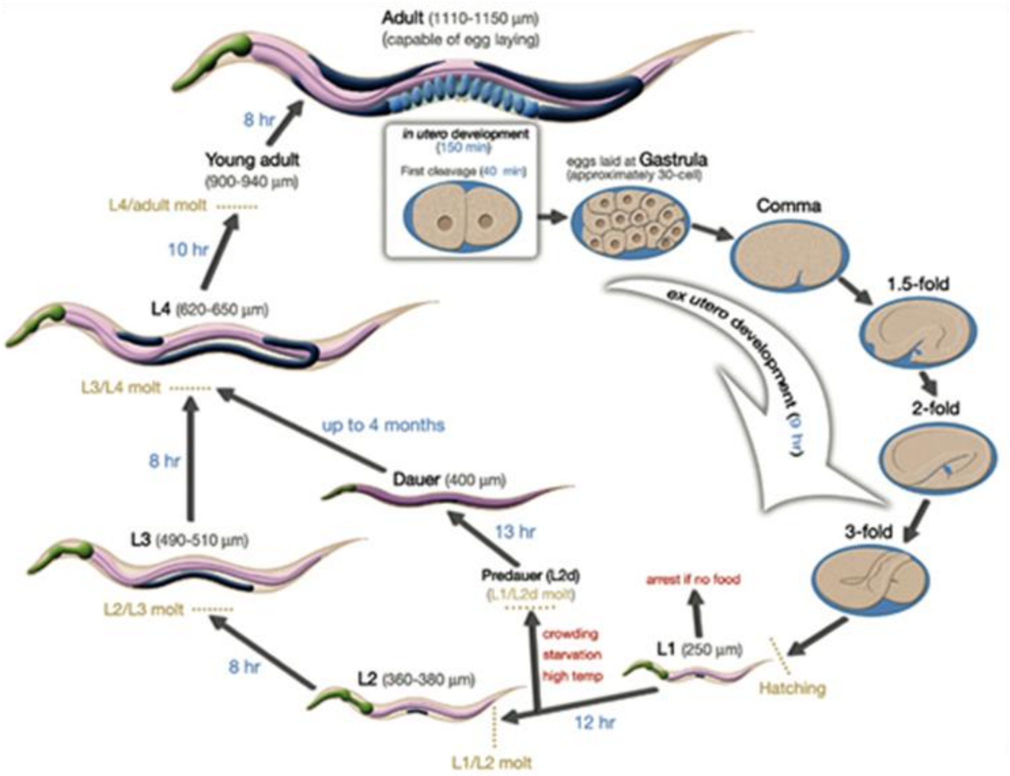
*C. elegans* worms are easily cultivated, have a short generation time of 3-5 days and can be maintained on agar plates with *E. coli* as its sole food source. Each worm develops from an egg and goes through four larval molts (stages L1-L4) before the final molt into the adult form (Figure 1.2a). *C. elegans* has an invariant lineage which ends with 959 cells in the adult hermaphrodite and 1,031 cells in the male. Under stressed conditions such as a lack of food or overcrowding the nematode can enter into a dauer stage after L1, where the animal becomes thin and elongated. Life expectancy of dauer stage worms can last for months and is thought to be a mechanism for stress resistance in the wild. Upon favourable conditions the worm exits this dauer stage and develops straight into the L4 stage of the life cycle.

*C. elegans* mostly mate as a hermaphrodite by self fertilisation. Occasionally males are produced as the result of a rare loss of the X chromosome, which occurs with a frequency of around 0.05%. Males are more motile than hermaphrodites and have special appendages around their tails for mating. Self fertilisation of hermaphrodites produce typically 300 offspring, while male-hermaphrodite matings can produce more than 1,000 young, and gives an equal ratio of males and hermaphrodites in the offspring.

The invariant lineage of the worm has allowed the characterisation of all of the developmental stages of each cell as a lineage map. *C. elegans* organs include a mouth, pharynx, gonad, intestine, cuticle and nerve cells (Figure 1.2b). Hermaphrodites have two ovary arms that move away from the middle of the worm towards both ends before turning back towards the middle, where they pass through a spermatheca before joining into a common uterus. Males are characterised by their thin shape, smaller size and modified tail structure which is used in attaching the worm to the hermaphrodite during mating.

Additional background information regarding *C. elegans* morphology and development can be found in the online resources Wormatlas (http://www.wormatlas.org/) and Wormbook (http://www.wormbook.org/).

a)




b)

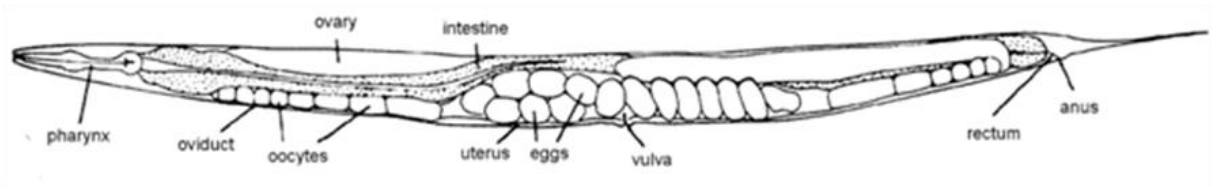



**Figure 1.2. Overview of development and morphology of *C. elegans*.**

a) The life cycle of *C. elegans*, showing the development of a hermaphrodite nematode from the egg stage through to the L1-L4 larval stages, before the final molt into the adult. The entire life cycle takes around 3-5 days to complete. Worms may enter into a long lived dauer stage after L1 due to stress and may live for months in this form. The diagram was adapted from Wormatlas at http://www.wormatlas.org/ver1/handbook/anatomyintro/anatomyintro.htm.

b) Anatomical features of a *C. elegans* adult hermaphrodite. Adapted from http://avery.rutgers.edu/WSSP/StudentScholars/project/introduction/worms.html.

### 1.4.2 *C. elegans* genetics and genomics

*C. elegans* was the first multicelullar organism to have its entire genome sequenced (Consortium, 1998). Since then the amount of annotation and manipulation of its genome has been steadily increasing. Originally the data was analysed with A *C. elegans* Data Base (AceDB) (Kelley, 2000), an open-source software developed in 1992 by Richard Durbin and Jean Thierry-Miegas as a tool for data management for the *C. elegans* genome project, and is maintained today at the Sanger Institute at http://www.acedb.org/. The program has since then evolved into a web based repository called Wormbase (http://wormbase.org/, (Stein *et al.*, 2001)), which holds information for all the current sequence data, splice models, protein sequences, expression profile, RNA interference (RNAi) experiments, phenotypes, ESTs, gene ontology (GO) terms, homologies to other species, and the literature references available for every *C. elegans* gene. Wormbase is updated frequently (1-2 months between updates) and contains 97 mb of DNA information and more than 20,000 protein sequences in the latest version.

*C. elegans* is a model that is very amenable to genetic manipulation. Many knockout strains of worms are available, with new strains being generated continuously for researchers from two major knockout consortiums, the Mitani lab at the Women's Medical University School of Medicine in Japan (http://www.shigen.nig.ac.jp/c.elegans/index.jsp) and the *C. elegans* Gene Knockout Consortium at the University of Minnesota, USA (http://biosci.umn.edu/CGC/). The mechanism of RNAi, which allows post-transcriptional gene silencing via the breakdown of mRNA by the action of double stranded RNA (dsRNA) interference, was first observed in *C. elegans* (Fire *et al.*, 1998). Since then great advances within the field has been made, with the discovery that RNAi can be mediated by simple feeding of dsRNA to the worms (Timmons and Fire, 1998). This has lead to the establishment of a number of RNAi libraries (Kamath and Ahringer, 2003; Rual *et al.*, 2004), which helped to produce several genome wide RNAi screens for a diverse number of processes such as general metabolism, embryogenesis, cell migration, neurotransmission and others (Cram *et al.*, 2006; Gottschalk *et al.*, 2005; Simmer *et al.*, 2003; Sonnichsen *et al.*, 2005). There is also extensive data for *C. elegans* in the form of expressed sequence tags (ESTs) (Kohara, 1996; McCombie *et al.*, 1992; Waterston *et al.*, 1992), the ORFeome (Reboul *et al.*, 2003), yeast 2-hybrid interactome (Li *et al.*, 2004b), and the Promoterome of GFP tagged genes for the analysis of expression patterns (Dupuy *et al.*, 2004), which all make the worm an excellent model system for large scale genomics studies.

### 1.4.3 Transcriptomics and proteomics studies in *C. elegans*

Many projects have attempted to analyse *C. elegans* expression profiles on a global scale, using transcriptomic techniques such as microarrays (Schena *et al.*, 1995) and Serial analysis of gene expression (SAGE, (Velculescu *et al.*, 1995)). Microarrays in the worm have been used to elucidate the total expression profiles of its germline, heat shock response, aging, dauer formation, non-coding RNA, alternative splicing, and many other processes (Barberan-Soler and Zahler, 2008; Blumenthal *et al.*, 2002; GuhaThakurta *et al.*, 2002; He *et al.*, 2006; Murphy *et al.*, 2003; Reinke *et al.*, 2000; Wang and Kim, 2003). SAGE analysis has been used to study the changes in expressions associated with aging within the nematode (Halaschek-Wiener *et al.*, 2005; Jones *et al.*, 2001).

*C. elegans* research is dominated by genetic studies, but in recent years there has also been an increased interest in the use of proteomic techniques to study of the worm (Audhya and Desai, 2008). The first proteomic study for the nematode was performed on whole worm lysate with 2D gels and MALDI MS peptide mass fingerprinting, which identified 12 proteins within the worm (Kaji *et al.*, 2000). A subsequent analysis of the same sample was able to produce 152 identified proteins (Schrimpf *et al.*, 2001). Both of these studies used relatively harsh techniques such as sonication and freeze-thawing for the extraction of proteins from the worms, as *C. elegans* has a tough cuticle that has proven to be problematic for biochemical studies in the past. Other 2D electrophoresis analysis have been used on the *C. elegans* proteome for quantitative assessment using $^{15}N$ labelling (Krijgsveld *et al.*, 2003), and to study the effects of cholesterol depletion (Choi *et al.*, 2003), heat sensitivity (Madi *et al.*, 2003) and the apoptotic signalling pathway (Greetham *et al.*, 2004). More recently studies have focused on the use of newer techniques such as 2D DIGE (Tabuse *et al.*, 2005), and LC MS/MS analysis for the elucidation of neuropeptides and mitochondria proteins (Husson *et al.*, 2007; Li *et al.*, 2009). Proteomics is a growing field within *C. elegans* research, with an increasing integration of proteomic results with the genetic data in Wormbase (Rogers *et al.*, 2008). This represents a significant step towards a more systematic understanding of *C. elegans* biology, which will help us gain a greater insight into complex processes such as development, signal transduction, organelle function and aging.

### 1.4.4 *C. elegans* as a potential model for lipid raft and GPI anchored proteins

Lipid rafts and GPI anchored proteins have been relatively poorly studied in *C. elegans*. A lipid raft fraction was extracted from the worm by a previous study, which found the presence of the stomatin homologues UNC-1 and UNC-24, as well as an interacting partner of UNC-1 named UNC-8 (Sedensky *et al.*, 2004). One GPI anchored protein, PHG-1 (also known as PHAS-1), was found to be sensitive to PIPLC digestion when it was expressed in a mammalian cell line (Agostoni *et al.*, 2002). There has been a relatively large body of work on the *C. elegans* caveolin homologues *cav-1* and *cav-2* within recent years. Studies on *cav-1* had shown that the protein is expressed strongly throughout embryonic development, and becomes localised in the nervous system and body-wall muscles from L1 to adult stages (Scheel *et al.*, 1999). CAV-1 has been shown to be involved in the meiotic cell cycle and acetylcholine signalling of nematodes, and interacts with dynamin within the worm to affect locomotion (Parker *et al.*, 2007). CAV-2 was found to be localised to the apical membrane of the *C. elegans* intestinal cells, where it was shown to be required for lipid trafficking (Parker *et al.*, 2009). Predictions of GPI anchored proteins with bioinformatics tools has been popular with *C. elegans* due to the presence of a well annotated genome for the nematode (Eisenhaber *et al.*, 2000; Fankhauser and Maser, 2005; Poisson *et al.*, 2007), but no experimental work have been attempted to follow up on these studies.

*C. elegans* has the potential to become an excellent model organism for the study of lipid rafts and GPI anchored proteins. Lipid raft and GPI anchored protein research tends to be confined to single cellular organisms and cell lines, which do not give a good overview of their effect in the complex processes of development and growth.

This is even more important when both of these classes of proteins are found to be involved in a number of signalling processes, such as GPCR signalling, T-cell activation, the insulin signalling pathway, and others (Bickel, 2002; Kabouridis and Jury, 2008; Landry *et al.*, 2006). The well annotated genome, wealth of developmental knowledge and the ease of genetic manipulation of *C. elegans* makes the worm an attractive target for the study of lipid rafts and GPI anchored proteins, within the context of a developmentally complex organism. This will also help us gain a greater insight into *C. elegans* membrane biology, and help us better understand the intricate biological processes within this model organism.

## 1.5 Outline for this thesis

A report on the study of lipid rafts and GPI anchored proteins in *C. elegans* is presented in this thesis. Chapter 2 contains an in-depth analysis of predicted GPI anchored proteins from the *C. elegans* genome with four bioinformatic programs, BIG PI, GPI SOM, FragAnchor and PredGPI. In Chapter 3 an analysis of the GPI anchor synthesis pathway is presented for *C. elegans*, as well as other processes, such as dolichyl phosphate mannose synthesis and lipid modifications of the anchor tail, which are essential for the production of a fully functional GPI anchored protein. Chapter 4 presents an account of the extraction of a lipid raft fraction from nematode membranes with Triton X-100 sucrose density gradient centrifugation, and the extraction of GPI anchored proteins with PIPLC digestion. In Chapter 5 a proteomic analysis of the lipid raft and GPI anchored proteins found in the worm is presented. Chapter 6 is the general discussion and the last chapter of this thesis, which summarises the findings of this project and their relevance to nematode membrane studies. The Chapter also outlines the wider field of research concerning lipid raft and GPI anchored proteins, as well as the potential for their further study within the *C. elegans* model organism.

***Chapter 2***

# Bioinformatics study of GPI anchored proteins in *Caenorhabditis elegans*

## 2.1 Introduction

The attachment of a GPI anchor to a protein is a highly conserved and important post-translational modification in eukaryotic organisms (Paulick and Bertozzi, 2008). They were first discovered when researchers found that the Thy-1 antigen in mice and the Variant Surface Glycoproteins (VSG) in *Trypanosoma brucei* behaved like typical membrane proteins but contained no transmembrane domains, and were released from the cell surface by bacterial phospatidylinositol-specific phospholipase C (PIPLC). Similar results with other proteins such as acetylcholine esterase and alkaline phosphatase came together in the 80's, and a novel mode of attachment of proteins onto the cell surface via GPI moieties was proposed (Ferguson and Williams, 1988). Data from cDNA of VSGs in *Trypanosoma brucei* has shown a need for an N-terminal secretion signal sequence and a C-terminal hydrophobic region, which are both cleaved off in the mature protein found on the cell surface (Boothroyd, 1985).

### 2.1.1 Expression of GPI anchored proteins within the cell

The life of a GPI anchored protein begins with binding of its N-terminal signal peptide sequence with the signal recognition particle (SRP) in the cytoplasm, which directs the ribosome onto the translocon where the protein is co-translocated into the ER lumen before the cleavage of the signal (Walter and Johnson, 1994). Once inside the ER lumen the C-terminal propeptide sequence is proteolytically cleaved by a transamidase complex and a GPI moiety becomes attached to the residue at the carboxyl terminus of the protein called the ω site (Figure 2.1) (Udenfriend and Kodukula, 1995a). Mature GPI anchored proteins mostly contain no stretches of hydrophobic sequences and are transported via the secretory pathway though the Golgi apparatus, until they are finally expressed on the outer surface of the plasma membrane (Figure 2.1c).

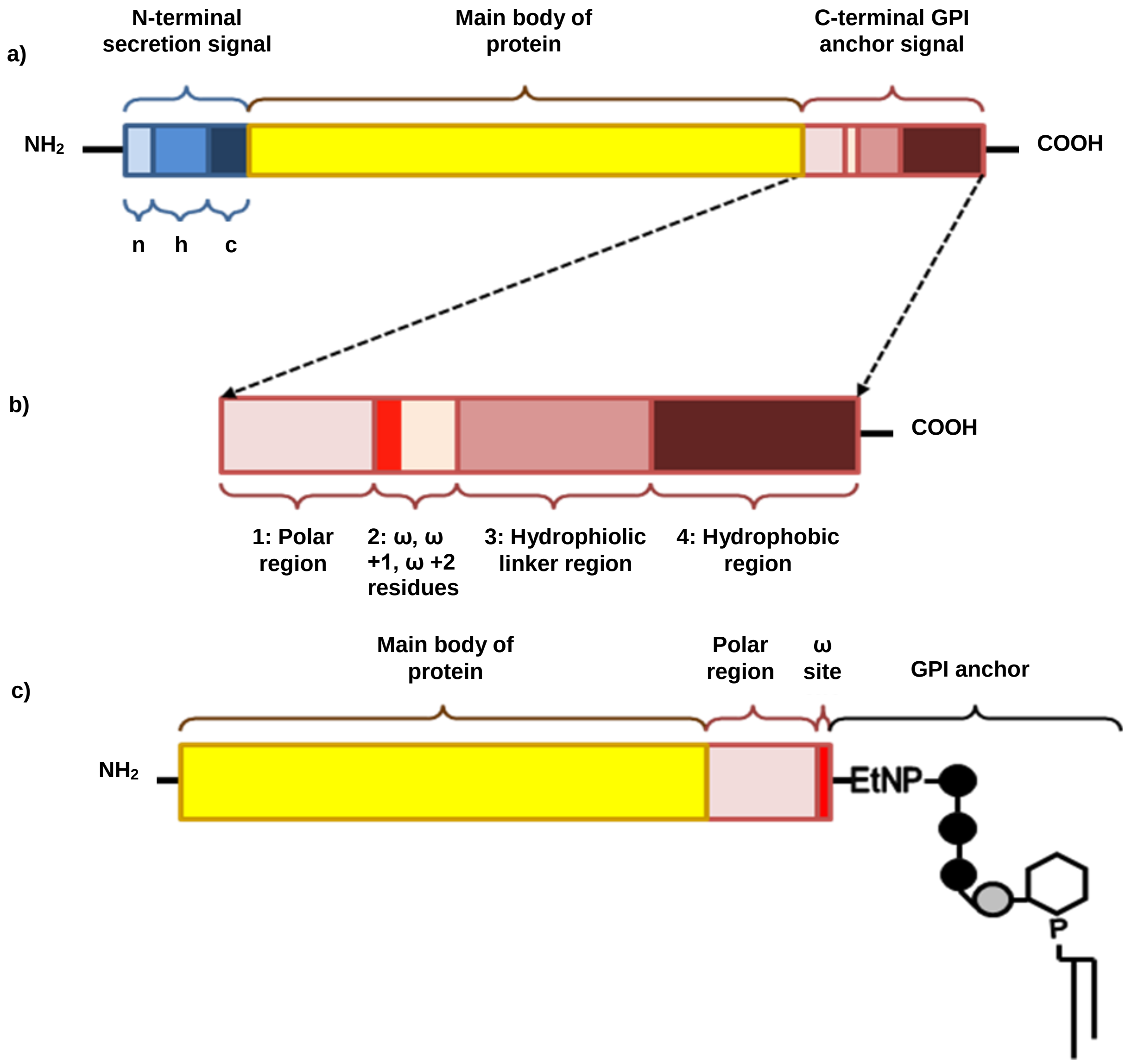



**Figure 2.1. Diagrammatical representation of the structure of a GPI anchored protein.**

a) The transcript of the GPI anchored protein codes for an N-terminal secretion signal (with n, h, and c regions) and a C-terminal GPI anchor signal. The main body of the protein does not contain any Transmembrane domains.

b) Structure of the C-terminal GPI anchor signal, which contains four sections. Section 1: polar residues, section 2: ω, ω +1 and ω +2 residues, section 3: hydrophobic residues, section 4: hydrophobic tail.

c) Final protein structure after attachment to GPI anchor. For the features of the GPI anchor moiety refer to Figure 3.1.

### 2.1.2 Sequence features of a GPI anchored protein

#### 2.1.2.1 Property and bioinformatic prediction of the N-terminal secretion signal

GPI anchored proteins found in nature typically contain a secretion signal (Gerber *et al.*, 1992), although some synthetic proteins made without the sequence have been observed to be capable of GPI anchor attachment (Howell *et al.*, 1994). The signal sequence has been very well studied and contains a set of consensus features, which include an N-terminal (n)- region of 1-5 charged residues, followed by a central hydrophobic (h)- region of 7-15 residues, and finally 3-7 polar uncharged residues at the C-terminal (c)- region, with some sequence conservation around the cleavage site. The secretion signal has a final size of around 15-30 residues (Figure 2.1a) (von Heijne, 1990). Numerous attempts have been made since the 1980's on a prediction program for the secretion sequence, initially based on simple weight matrix approaches (von Heijne, 1986), which was later diversified into other more sophisticated machine learning methods. In 1997 Niesel *et al.* produced a prediction program based on a neural network (NN) (Nielsen *et al.*, 1997), which was followed up with the addition of a separate predictor based on a hidden Markov model (HMM) (Nielsen and Krogh, 1998). These two approaches became integrated into the web based program SignalP 3.0 with significant improvements in the quality of sequences in the training set (Bendtsen *et al.*, 2004). The prediction program is thought to be very robust and has become a *de-facto* standard in the prediction of signal peptides for researchers (Emanuelsson *et al.*, 2007).

#### 2.1.2.2 The C-terminal GPI anchor signal

##### 2.1.2.2.1 Sequence properties of the C-terminal GPI anchor signal

Characterisation of the C-terminal sequence began with the discovery that the terminal signal sequence was not simply proteolytically cleaved for VSGs from *Trypanosome brucei*, but occurred with the addition of ethanolamine to the terminal amino acid in an amide linkage (Holder, 1983). Subsequently it was found that this reaction happens soon after the translation of the protein in the ER (Bangs *et al.*, 1985; Conzelmann *et al.*, 1987). Analysis of the sequences of known GPI anchored proteins and their cDNA produced a list of putative rules for this attachment, such as the requirement for a small amino acid residue at the ω site, more stringent requirements for the two amino acid positions immediately after the ω site, and a run of hydrophobic residues at the end of the protein (Ferguson and Williams, 1988). Further experiments revealed that the sequence at the C-terminus was sufficient for GPI anchoring when expressed at the end of the secreted human growth hormone (hGH) (Moran and Caras, 1991). Single amino acid changes within the sequence can abolish or rescue GPI anchoring (Moran *et al.*, 1991), and certain features of the C-terminal sequence such as the hydrophobic tail were necessary, but insufficient to direct GPI anchoring (Caras *et al.*, 1989). Synthetic peptide sequences following the rules of the signal motif allowed the attachment of GPI anchors to proteins (Coyne *et al.*, 1993), suggesting that it was the combined features of the signal, not the specific residues in the sequence *per se*, that allowed a protein to become GPI anchored.

The various features of the C-terminal signal sequence were gathered together in the 90’s to establish a set of criteria necessary for the GPI anchoring of proteins. Several studies involving sequencing of the C-terminus of GPI anchored proteins have established the ω sites for a number of them, which were brought together by Sidney Udenfriend for one of the first attempts at a prediction of protein GPI anchoring through the amino acid sequence at the C-terminus (Udenfriend and Kodukula, 1995b). It was reported that the ω site permitted only use of the amino acids Gly, Ala, Ser, Cys, Asp and Asn, with different affinities for GPI anchoring for each residue; the amino acid type for the ω +1 site was not found to be important (any except for Pro and Trp), and the ω +2 site was found to be the most stringent of all of the residues at the C-terminus (Gly, Ala and Ser). A simple probability based on multiplication of the proportional occurrence of an amino acid at the ω and ω+2 sites was produced to determine the likelihood of GPI anchoring for an unknown protein. The paper also acknowledged the importance of the flanking sequences of the ω site with respect to anchor attachment, noting the need for hydrophobic residues to be present at the C-terminus of the protein for efficient GPI anchor attachment.

Further advances in the field indicated a need for a hydrophilic spacer region of 6-14 amino acids between the ω site and the hydrophobic tail, with 8 being the optimal number (Furukawa *et al.*, 1997). Point mutational analysis in the hydrophobic tail suggested that this sequence is also subject to certain rules regarding its amino acid preference, such as different requirements for hydrophobicity along the sequence and a possible tendency for the tail to form an alpha helix (Yan *et al.*, 1998). It was thought that the hydrophobic C-terminal tail may be inserted into the ER membrane to assist the protein’s reaction with the transamidase complex. In 1998 Eisenhaber *et al.* took the available information regarding the properties of the C-terminal peptide sequence and produced a comprehensive bioinformatic analysis of the GPI anchor attachment signal (Eisenhaber *et al.*, 1998). In the paper, Eisenhaber *et al.* used all protein sequences found in SwissProt with annotations for GPI anchoring (155 at time of writing, with various degrees of confidence) as a reference for the amino acid composition of the C-terminal signal, and established a set of criteria for the attachment of the anchor. The C-terminal sequence is split up into four sections with distinct properties (Figure 2.1b). Section one, which begins at the position around 11 amino acids in the N-terminal direction of the ω site (through residues ω -11 to ω -2), has a generally polar profile that is flexible and unstructured. This is thought to help the reaction of the transamidase complex by the minimisation of steric effects around the active site. Section two concerns the amino acid residues around the ω site. The requirements for the positions ω, ω +1 and ω +2 were found to be similar to previous suggestions (Udenfriend and Kodukula, 1995b). In addition, amino acids from positions ω -1 to ω +2 were found to occupy a restricted volume, due to the size constraints of the active site within the transamidase complex, and the residue makeup of the region showed mutual compensatory effects with respect to this restricted volume; the volume of the active site was estimated to be around 540 $Å^3$. Section three covers residues ω +3 to ω +8, and is essentially another linker region with no sequence specificity, but a general property of being hydrophilic. The author noted some specific properties for a number of amino acids in this section, with was thought to allow better interaction of the signal sequence with the transamidase complex. Section four runs from ω +9 to the end of the protein, and constitutes a run of hydrophobic residues that extends to at least ω +21. The authors also detected differences in the features of the C-terminal signal in metazoans and protozoans, which prompted them to divide the GPI anchor attachment sequences into these categories.

### 2.1.2.2.2 Bioinformatic prediction programs for the C-terminal GPI anchor signal

The analysis produced by Eisenhaber *et al.* was later used to produce the first GPI anchor predictor program called BIG PI (Eisenhaber *et al.*, 1999). The program is capable of producing an output of the likelihood of GPI anchoring of a protein from its amino acid sequence. Eisenhaber *et al.* used 177 proteins as a learning set and the program produces a score function (S) based on the addition of two scores, the amino acid preference profile at the C-terminus ($S_{profile}$) and the level of conservation of the physical properties of the sequence with relation to the four sections previously described ($S_{ppt}$, physical property pattern). The predictor is available in two formats for the analysis of metazoan and protozoan protein sequences. This program was subsequently used on an early version of the *C. elegans* genome in which 86 proteins were predicted to be GPI anchored (Eisenhaber *et al.*, 2000). The program was later refined and used on a variety of genomes, and found strong predictions for eukaryotic and some archaea bacteria species, but none for eubacteria (Eisenhaber *et al.*, 2001). A plant specific Big PI predictor was made in 2003, with data from various *Arabidopsis thaliana* projects as the source of the learning set (Eisenhaber *et al.*, 2003c).

In subsequent years other researchers have also attempted to produce GPI anchor prediction programs. Kronegg and Buloz created a program called DGPI in 1999, using the amino acid composition around the ω site as the basis for prediction (http://129.194.185.165/dgpi/). Borner *et al.* produced a list of predicted *Arabidopsis* GPI anchored proteins with an in-house developed program based on the detection of hydrophobic stretches and co-confirmation with SignalP 2.0 (Borner *et al.*, 2002).

Fankhauser and Maser produced a GPI prediction program based on a Kohonen self-organising map called GPI-SOM, which used a set of GPI anchored protein training set to product a neural network with which to find the pattern of amino acids for the signal sequence (Fankhauser and Maser, 2005). FragAnchor was produced by Poisson *et al.*, who used a two stage process involving a neural network coupled with HMM to identify a protein and its ω site (Poisson *et al.*, 2007). Lastly, Pierleoni *et al.* made PredGPI using a combination of HMM and support vector machine (SVM) to predict GPI anchored proteins and their ω sites (Pierleoni *et al.*, 2008). There have also been other general membrane protein prediction programs with GPI prediction functions but do not perform detailed analysis on the ω site (Chou and Shen, 2007). The general consensus is that a combination of prediction programs will most likely produce a more accurate set of predictions (Elortza *et al.*, 2006). In tests with annotated GPI anchored and non-anchored proteins it was found that the programs performed equally well, with generally small false positive rates and prediction rates somewhere in the 80% range; BIG PI, the original prediction program, was still found to be the most stringent predictor, but the program was also found to have the highest number of false negatives in its output (Pierleoni *et al.*, 2008; Poisson *et al.*, 2007).

### 2.1.3 Outline for this Chapter

In this chapter I will give details of the use of these prediction programs for the elucidation of predicted GPI anchored proteins in *C. elegans*. *C. elegans* has a very well annotated genome which was first published in 1998 (Consortium, 1998). The repository for this information is available on the web at www.wormbase.org and is updated frequently. The programs BIG PI, GPI-SOM, FragAnchor and PredGPI were used to produce a list of potential GPI anchored proteins, and SignalP 3.0 was used to verify the N-terminal secretion peptide for these sequences. In order to further verify these predictions the *Caenorhabditis briggsae* orthologues of the predicted proteins from *C. elegans* were also subjected to the four GPI anchoring programs. *C. briggsae* is a closely related nematode to *C. elegans* and a great degree of genetic conservation has been shown between the two nematodes (Stein *et al.*, 2003); this may result in a greater degree of accuracy in the prediction for a particular protein, if it's predicted to be GPI anchored in both of the nematode species. The list of predicted proteins will be used for analysis of the GPI anchoring process in *C. elegans*, and as a starting point for the proteomic analysis of this class of proteins in this model organism.

## 2.2 Methods

### 2.2.1 Sequences for *C. elegans* and *C. briggsae*

All annotated and predicted *C. elegans* protein sequences were downloaded from the Wormbase website at ftp://ftp.wormbase.org/pub/wormbase/genomes/c_elegans/sequences/protein/ and were from release WS183 version of the genome as of November 2007. A total of 23,541 protein sequences were presented in FASTA format and saved as a Microsoft .txt file. *C. elegans* gene descriptions were retrieved from Wormbase using the Batch Genes function at http://wormbase.org/db/searches/batch_genes. *C. briggsae* genes orthologous to *C. elegans* genes of interest and their protein sequences were found via basic local alignment search tool (BLAST) search using the Batch Genes website.

### 2.2.2 Prediction of the N-terminal secretion signal

Prediction of N-terminal secretion signal was made with SignaP 3.0 program available at http://www.cbs.dtu.dk/services/SignalP/. Protein sequences were uploaded in FASTA format and analysed using the Eukaryotic parameter group with the Neuronal network model. All protein sequences were truncated to the first 70 amino acids before analysis as recommended by the program.

### 2.2.3 Prediction of the C-terminal GPI anchor signal

#### 2.2.3.1 Big PI predictor program

All *C. elegans* protein sequences from release version WS183 were loaded onto the Big PI prediction program at http://mendel.imp.ac.at/gpi/gpi_server.html, with the metazoan learning set used as the criteria for prediction. Prediction results contain a score for the level of confidence and a putative site of cleavage. The sequences for proteins with predicted GPI anchoring were tested for N-terminal secretion signal with SignalP 3.0. Proteins with positive prediction from both programs were considered to be acceptable. *C. briggsae* orthologues of the predicted genes were taken from Wormbase and also subjected to the Big PI and SignalP 3.0 predictor programs to determine their GPI anchoring status.

#### 2.2.3.2 GPI SOM

*C. elegans* release WS183 protein sequences were uploaded to the GPI SOM predictor program at http://gpi.unibe.ch/. GPI SOM carries out a tandem prediction with SignalP 2.0 built into the program, and the final result is verified for both C-terminal and N-terminal signal sequences. GPI SOM does not generate a score for the protein but does give a putative cleavage site for the C-terminus. The *C. briggsae* orthologues of these genes were taken from Wormbase and GPI anchor prediction was also made for them in GPI SOM.

#### 2.2.3.3 FragAnchor

The FragAnchor prediction program can be found at http://navet.ics.hawaii.edu/~fraganchor/NNHMM/NNHMM.html. *C. elegans* protein sequences were uploaded as a file to the prediction program. The program automatically discards protein sequences with 50 or less amino acids and non-standard amino acid letters. FragAnchor uses a two stage prediction process in which the sequence is analysed with a NN algorithm and then passed through a HMM program. Positively identified predictions are placed under four categories based on the score of the identification, which are highly probable (HMM score $\geq$ 5.4), probable (5.4 > HMM score $\geq$ 2.2), weakly probable (2.2 > HMM score $\geq$ 0.2) and potentially false positive (0.2 > HMM score), and generates a putative cleavage site for GPI anchoring. Predicted genes from the highly probable, probable and weakly probable were put through SignalP 3.0 to generate a final list of predicted GPI anchored proteins. *C. briggsae* orthologues of these proteins were also put through the FragAnchor and SignalP 3.0 predictor programs to determine their GPI anchor status.

### 2.2.3.4 PredGPI

PredGPI can be found at http://gpcr.biocomp.unibo.it/predgpi/. *C. elegans* proteins were analysed with the program in batches of 500 or less sequences due to a restriction placed by the website. The outcomes are presented with a putative cleavage site and a score for the protein identification as highly probable (p $\geq$ 99.9), probable (99.9 > p $\geq$ 99.5), or lowly probable (99.5 > p $\geq$ 99.0). Sequences from all three categories were subjected to SignalP 3.0 to test for the presence of an N-terminal secretion motif. Proteins with both predictions were considered to be putative GPI anchored proteins. Sequences for the *C. briggsae* orthologues were taken from Wormbase and subjected to both PredGPI and SignalP 3.0 prediction programs.

### 2.2.4 Gene Ontology (GO) terms for the predicted terms

The GO term for each prediction were taken from the Wormbase website using the batch genes webpage (http://www.wormbase.org/db/searches/batch_genes). GO terms were presented in three categories, which are Molecular Function, Biological Process and Cellular Component. Where multiple GO terms were present for a gene the most representative term was chosen for its description. Finally GO terms with similar overall description were placed in broad groups for clarity, such as placing GO:0008237 (metallopeptidase activity) and GO:0008236 (serine-type peptidase activity) into the Catalytic group for the Molecular Function category.

## 2.3 Results

23,541 proteins were present in release WS183 of the *C. elegans* genome. In order to find the number of proteins that are GPI anchored from this list the sequences for each protein were subjected to 4 different C-terminal sequence GPI prediction programs Big-PI, GPI SOM, FragAnchor and PredGPI. BIG PI, FragAnchor and PredGPI give a score for the likelihood of their prediction, with FragAnchor and PredGPI presenting three different levels of confidence for their predictions. All of the programs give a prediction for a putative transamidase cleavage site.

The presence of the N-terminal secretory sequence is necessary for GPI anchored proteins and is predicted by SignalP 3.0. Protein sequences with both N and C termini hits were considered to contain true GPI anchor predictions. GPI SOM has SignalP 3.0 search as a part of its function.

*C. briggsae* orthologues of the predicted proteins from each program were taken from Wormbase and subjected to GPI anchor prediction with the same program. This is used to test the fidelity of the prediction as *C. briggsae* is a close evolutionary relative of *C. elegans* and has been shown to be a good complimentary organism with regard to genetics research (Stein *et al.*, 2003).

### 2.3.1 Individual results from all prediction programs

#### 2.3.1.1 Big PI

The BigPI program produced a list of 125 GPI anchored proteins with N-terminal secretion signal, which is the smallest number of proteins for the prediction programs tested (Figure 2.2). 52 of these proteins also have GPI anchored orthologues from *C. briggsae* predicted with the same program.

#### 2.3.1.2 GPI SOM

GPI SOM produced the longest list of predicted GPI anchored proteins with 657 sequences predicted. 348 of these proteins have orthologues in *C. briggsae* that are also predicted to be GPI anchored with this program. GPI SOM produced the largest list of predicted proteins of all of the programs (Figure 2.2).

#### 2.3.1.3 FragAnchor

FragAnchor produced 237 proteins as potential GPI anchored proteins. Of these sequences 109 are predicted to be highly probable, with 71 predicted to be probable and 57 weakly probable by the criteria of the program. *C. briggsae* orthologues with predicted GPI anchoring is present for 146 of these proteins (Figure 2.2).

#### 2.3.1.4 PredGPI

362 proteins were predicted from PredGPI to be GPI anchored from the *C. elegans* genome, with 157 classified by the program as highly probable, 111 probable and 94 as lowly probable. Of these 186 proteins had predicted GPI anchored proteins for their *C. briggsae* orthologues (Figure 2.2).

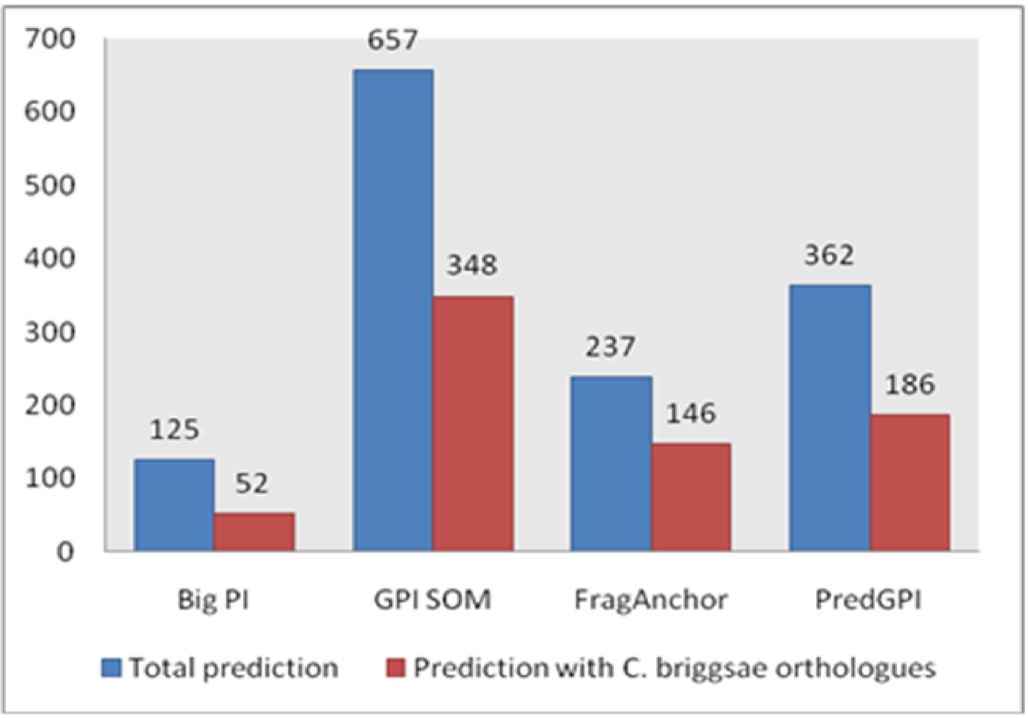



**Figure 2.2. Total number of predictions for the four prediction programs.** The numbers above each bar indicates the number of predicted proteins. Blue bars represent the total number of positive outputs for each program and the red bars represent the number of proteins that also have GPI prediction in their *C. briggsae* orthologue.

### 2.3.2 Prediction across all programs

As expected a large number of proteins are predicted to be GPI anchored with more than one program. There are a total of 778 unique proteins overall predicted from all four programs (Table 2.1, for a full list of proteins refer to Appendix 1). Of these 81 protein sequences were found to be GPI anchored from all four programs, 112 proteins were found with three programs, 134 sequences were scored with two programs, and 451 were predicted to be GPI anchored from only one program (Figure 2.3).

| | Total number of unique *C. elegans* proteins predicted to be GPI anchored | *C. elegans* predictions with *C. briggsae* orthologues that also have GPI anchor prediction |
|---|---|---|
| Number of predicted proteins from all four programs | **778** | **382** |
| Number of proteins with predictions in two or more programs | **327** | **201** |

**Table 2.1. Total numbers of GPI anchored proteins predicted for *C.elegans*.** Presented here are the total numbers of unique *C. elegans* GPI anchoring predictions across all four prediction programs, the number of those proteins that also have orthologues in *C. briggsae* that are also positive for GPI prediction, and the number of proteins in both of the categories that have the more stringent criteria of being predicted by two or more programs.

382 unique *C. elegans* sequences were found with prediction of GPI anchoring also in their *C. briggsae* orthologues (Table 2.1). 38 of those proteins were found with all four programs, 73 predicted from three programs, 90 were predicted with two programs, and 181 were found with one program (Figure 2.3).

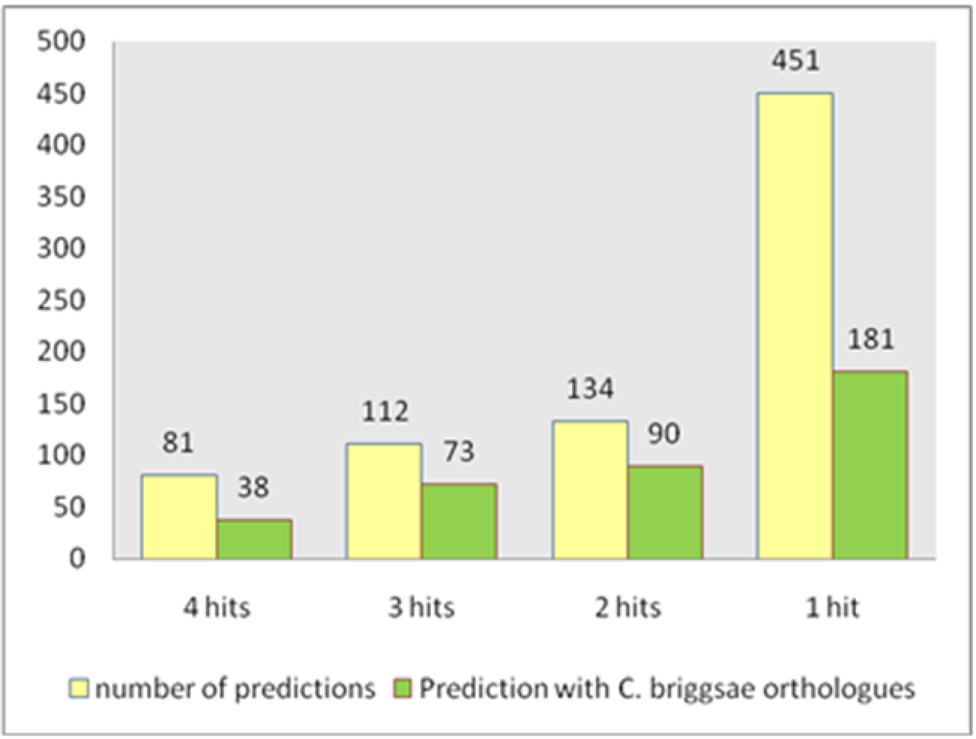



**Figure 2.3. The criteria used to determine of GPI anchor prediction.** The graph shows the number of proteins with independent hits from 4, 3, 2 or 1 of the prediction programs tested. The yellow bars represent the total number of proteins for each category and the green bars represent those proteins that also have orthologues in *C. briggsae*. Proteins with 2 hits or more are considered to have positive predictions for GPI anchoring.

Out of the total 778 proteins 451 were found with GPI prediction from just a single prediction program. Of these GPI SOM account for the highest proportion of predictions (Figure 2.4). GPI prediction was found to be more accurate for humans and *Arabidopsis* when multiple prediction programs were used (Elortza *et al.*, 2006). A list of proteins with prediction from two or more programs was made. This reduced the total number of predicted proteins to 327, with 201 sequences that also have GPI anchored *C. briggsae* orthologue (Table 2.1).

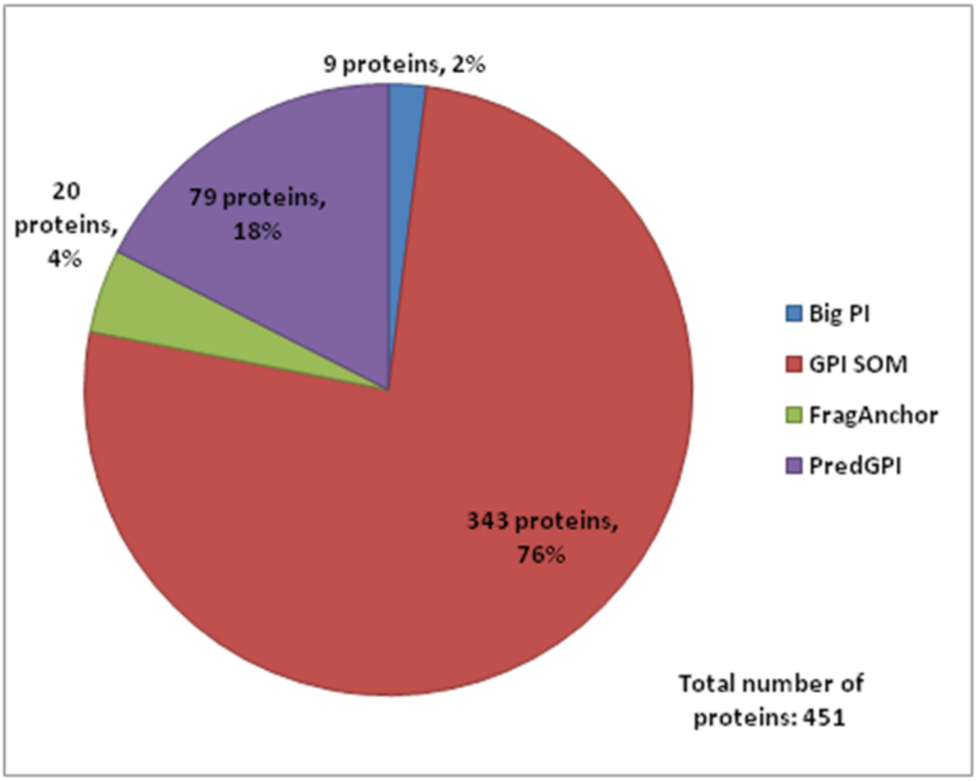



**Figure 2.4. Percentage of proteins with only a single prediction from a program.** The total number of proteins with only one hit from a predictor is 451. Of these GPI SOM accounts for the highest proportion of predictions.

### 2.3.3 GO terms for predicted GPI anchored proteins

A list of all available GO terms was analyzed for genes found with two or more prediction programs alongside those for genes with *C. briggsae* orthologues. The GO terms fall into three categories, which are Molecular Function, Biological Processes and Cellular Component. The proportion of genes with GO terms for each of the categories is shown in Table 2.2. A comparison of each GO term was made for all predicted proteins versus those predicted with 2 or more prediction programs (Figure 2.5).

| Categories | Proteins with 2 or more hits | Proportion of total | Proteins with 2 or more hits that also have *C. briggsae* orthologues | Proportion of total |
|---|---|---|---|---|
| **Molecular function** | 88 | 27% | 61 | 30% |
| **Biological process** | 93 | 28% | 63 | 31% |
| **Cellular component** | 149 | 46% | 93 | 46% |

**Table 2.2. Proportion of proteins with GO terms.** The three broad categories of GO terms are Molecular Function, Biological Processes, and Cellular Component. Presented here are the number of proteins with GO terms in each of the categories and the proportion they represent within the total number of predicted proteins (327 for the number of proteins with 2 or more hits from the four prediction programs, and 201 for those proteins that also have a *C. briggsae* orthologue that is also predicted to be GPI anchored).

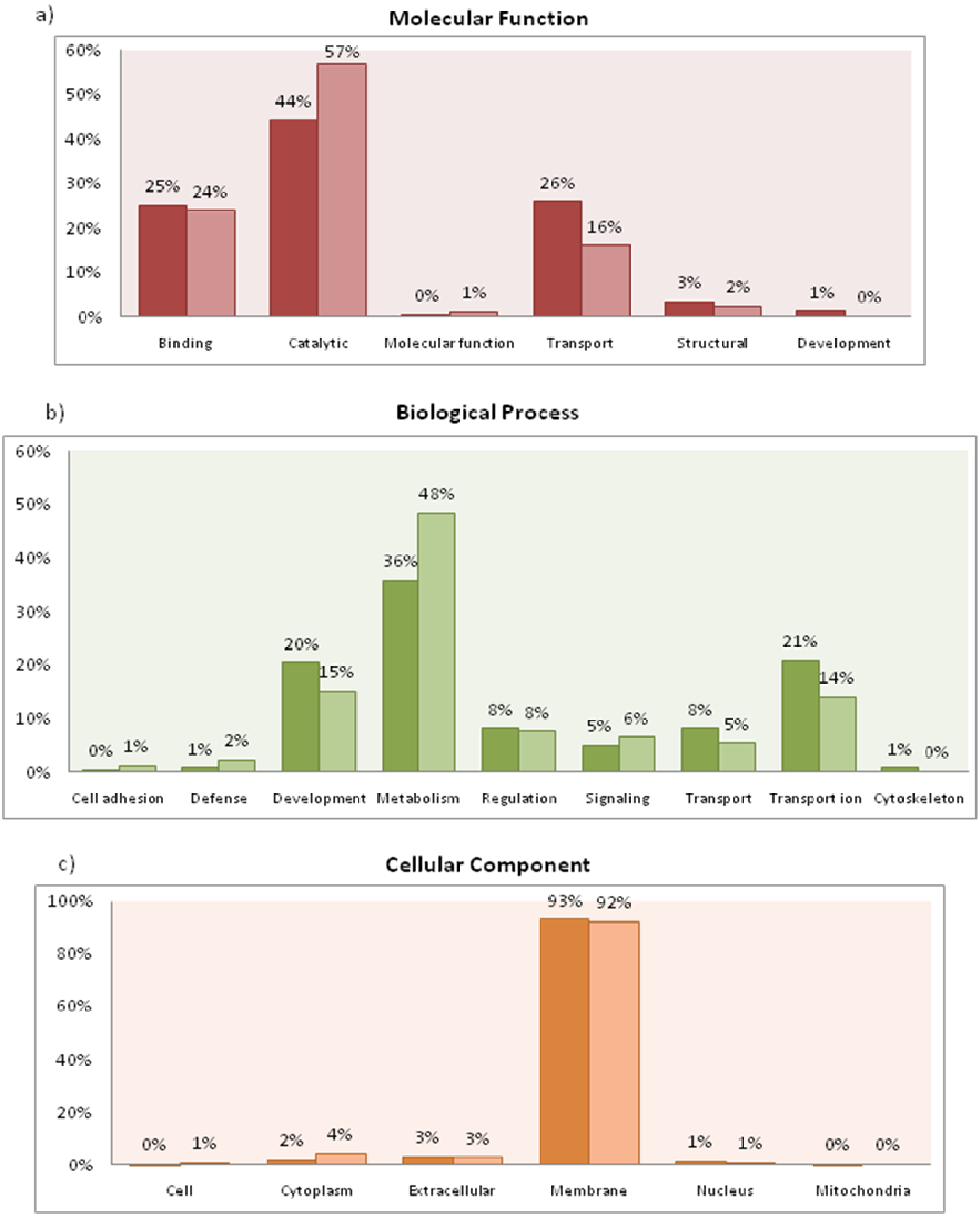



**Figure 2.5. Comparison of GO term categories for predicted GPI anchored proteins.** The bars on the left for each term correspond to all proteins with prediction hits from one or more programs while the numbers on the right correspond to proteins with hits from 2 or more prediction programs. Graph a) represents GO terms in Molecular Function, b) Biological Process, c) Cellular Component.

#### 2.3.3.1 Molecular Function

In the Molecular Function category the majority of the genes are involved in Catalytic processes, with 56.8% of proteins having GO terms with this particular function. There are also large numbers of genes for binding (21.6%) and transport (15.9%). 2 genes each were assigned as having receptor function and structural, and one gene had a generic "molecular function" term from the database. The proportion for each of the terms is similar in genes with *C. briggsae* orthologues, with no genes present within the structural and molecular function GO groups (Figure 2.6).

#### 2.3.3.2 Biological Process

In the Biological Processes category the group with the most genes are involved with metabolism (48.4%), followed by transport (19.4%) and development (15.1 %). There are also a small number of genes involved in regulation, signalling, defence and cell adhesion. The number of genes with *C. briggsae* orthologues also have similar percentages to the overall GO groups (Figure 2.6).

#### 2.3.3.3 Cellular Component

The majority of GO terms for the Cellular Component category belong to the membrane group with 91.9% of the total. There are a small number genes belonging to the extracellular, cytoplasmic, nuclear, and cell group. Genes with *C. briggsae* orthologues also have similar proportions of entries within each of these groups, with the one nuclear localised gene absent (Figure 2.6).

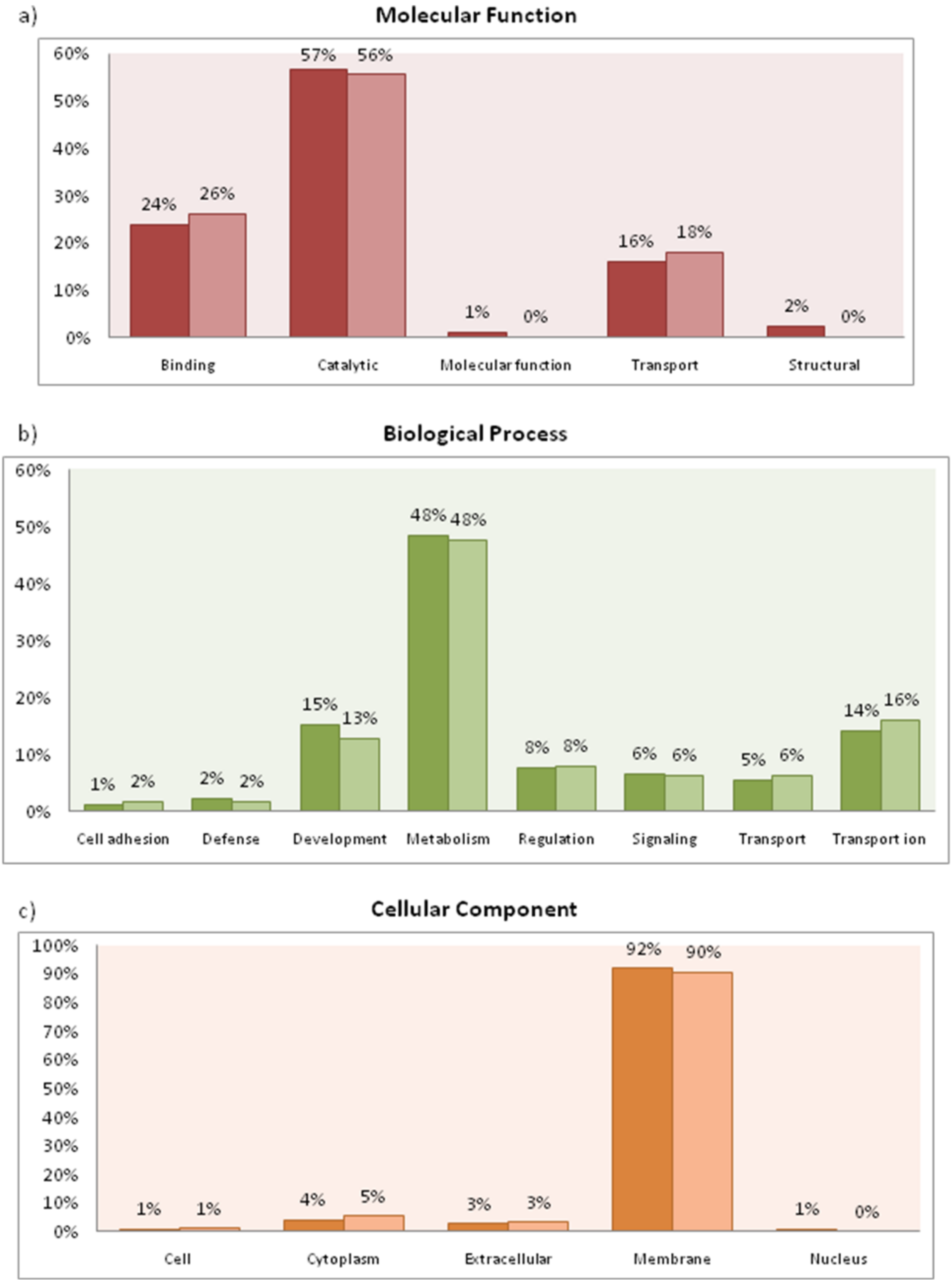



**Figure 2.6. Comparison of GO term categories for GPI anchored proteins predicted with 2 or more programs.** The bars on the left for each term correspond to all *C. elegans* proteins that fit this criterion while the numbers on the right correspond to these proteins that also have a *C. briggsae* orthologue with predicted GPI anchoring. Graph a) represents GO terms in Molecular Function, b) Biological Process, c) Cellular Component.

## 2.4 Discussion

A comprehensive report on the number of potential *C. elegans* GPI anchored proteins is presented in this chapter. SignalP 3.0 was used to predict the N-terminal secretory sequence, while four prediction programs using different algorithms for the C-terminal motif were used to produce a list of proteins that are potentially GPI anchored. There was a large amount of overlap with the output of predicted proteins from each of the four programs, and in the end 778 unique GPI anchored protein predictions were produced in this study. Of these proteins 81 were predicted by all four of the prediction programs, 112 by a combination of three prediction programs, 134 by two, and 451 by only one of the prediction programs (Figure 2.3).

### 2.4.1 Analysis of the different prediction programs

#### 2.4.1.1 Big PI

Big PI is the oldest of these programs and uses a weight matrix approach to produce a list of potential GPI anchored proteins. The parameters of the weight matrix is determined by Eisenhaber *et al.*'s study of C-terminal residue positions of experimentally determined GPI anchored proteins (Eisenhaber *et al.*, 1998). Big PI produced the lowest number of predictions of all of the programs (Figure 2.2). The majority of the Big PI predicted proteins are predicted by two or more programs (93%), with a high proportion of those predictions in the 4 hits category (65%) (Table 2.3a).

#### 2.4.1.2 GPI SOM

GPI SOM uses a Kohonen Self-Organizing Map method for assigning GPI anchoring to potential sequences (Fankhauser and Maser, 2005). GPI SOM produced the largest number of potential GPI anchored proteins with 657 predictions. While the program does have the largest number of proteins with only a single validation (52%) (Table 2.3a), GPI SOM is also present with most of the proteins with three (98%) or two (92%) prediction program validations (Table 2.4a).

#### 2.4.1.3 FragAnchor

FragAnchor uses a two stage process of Neural Network and Hidden Markov Model to validate potential GPI anchored proteins (Poisson *et al.*, 2007). FragAnchor produced 237 proteins and a large proportion of the proteins are validated by two or more prediction programs (92%) (Table 2.3a).

#### 2.4.1.4 PredGPI

PredGPI is the latest program available to researchers of GPI anchoring and also uses a two stage process involving a Support Vector Machine and Hidden Markov Model for determination of GPI anchoring (Pierleoni *et al.*, 2008). The number of predictions produced by the program was 362. The proportions of proteins with 3 and 2 hits that also have prediction with PredGPI is high (94% and 72%, respectively, Table 2.4a).

a)

| Prediction program | number in 4 hits | proportion of total | number in 3 hits | proportion of total | number in 2 hits | proportion of total | number in 1 hit | proportion of total | total |
|---|---|---|---|---|---|---|---|---|---|
| Big PI | 81 | 65% | 25 | 20% | 10 | 8% | 9 | 7% | 125 |
| GPI SOM | 81 | 12% | 110 | 17% | 123 | 19% | 343 | 52% | 657 |
| FragAnchor | 81 | 34% | 96 | 41% | 40 | 17% | 20 | 8% | 237 |
| PredGPI | 81 | 22% | 105 | 29% | 97 | 27% | 79 | 22% | 362 |

b)

| Prediction program | number in 4 hits | proportion of total | number in 3 hits | proportion of total | number in 2 hits | proportion of total | number in 1 hit | proportion of total | total |
|---|---|---|---|---|---|---|---|---|---|
| Big PI | 38 | 73% | 6 | 12% | 7 | 13% | 1 | 2% | 52 |
| GPI SOM | 38 | 11% | 73 | 21% | 82 | 24% | 155 | 45% | 348 |
| FragAnchor | 38 | 26% | 68 | 47% | 34 | 23% | 6 | 4% | 146 |
| PredGPI | 38 | 20% | 72 | 39% | 57 | 31% | 19 | 10% | 186 |

**Table 2.3. The number and proportion of outputs from each prediction program.** Number and percentages of proteins with a total of 4, 3, 2 and 1 hits are shown for each program. For example the Big PI predictor program occurs in 81 of the proteins with 4 hits in the four prediction programs, which represents 65% of the total number of Big PI predictions from the *C. elegans* genome (125). Table a) data from all predicted *C. elegans* proteins. Table b) data from *C. elegans* proteins that also have GPI anchor predicted *C. briggsae* orthologues.

a)

| Prediction program | proportion of proteins with 3 hits | proportion of proteins with 2 hits | proportion of proteins with only one hit |
|---|---|---|---|
| Big PI | 22% | 7% | 2% |
| GPI SOM | 98% | 92% | 76% |
| FragAnchor | 86% | 30% | 4% |
| PredGPI | 94% | 72% | 18% |

b)

| Prediction program | proportion of proteins with 3 hits | proportion of proteins with 2 hits | proportion of proteins with only one hit |
|---|---|---|---|
| Big PI | 8% | 8% | 1% |
| GPI SOM | 100% | 91% | 86% |
| FragAnchor | 93% | 38% | 3% |
| PredGPI | 99% | 63% | 10% |

**Table 2.4. Analysis of the fidelity of each prediction program from protein predictions.** The percentage contribution of each of the four prediction programs to predicted proteins with 3, 2, and 1 total hits is shown. For example of the proteins with 3 hits from prediction programs 98% have one of their predictions in GPI SOM, while 86% have one of their predictions in FragAnchor. Table a) data from all predicted *C. elegans* proteins. Table b) data from *C. elegans* proteins that also have GPI anchor predicted *C. briggsae* orthologues.

#### 2.4.1.5 Comparison between the four prediction programs

The metric of a good prediction program comes from maximising the number of real positive predictions while minimising both false positive and false negative results, so that the program is stringent enough to include real potential sequences and at the same time generalised enough to not exclude other genuine GPI anchored proteins. Out of all of the prediction programs Big PI has emerged with the highest stringency, with the highest number of its predictions validated by the other programs (Table 2.3a). However Big PI predictions are not represented in a large number of the proteins that have 3 or 2 hits, suggesting that the program has a high false negative rate, which may be due to the relatively strict weight matrix approach used in its algorithm. Both FragAnchor and PredGPI performed well with a large proportion of genes also validated with three prediction programs. PredGPI has a large percentage of proteins validated by two programs (72%) while FragAnchor has a small proportion (30%) (Table 2.4a), which indicates that FragAnchor is less general and more stringent than PredGPI. GPI SOM has the largest number of predictions which makes the program the most generalised of the four tested predictors. The large proportion of proteins that are predicted by just GPI SOM (76%, Figure 2.4, Table 2.4a) suggests that the program also has a high false positive rate. However, the proportion of proteins with three and two hits that also have GPI SOM prediction is also high (Table 2.4a), suggesting that the program is capable of producing good quality predictions. All of the prediction programs show a steady reduction in the proportion of proteins within the data that have three hits, two hits and one hit, which is to be expected from a data set with various total numbers of predictions. Taken together all four programs are capable of producing good GPI predictions that are validated in the expected pattern with other programs, with BigPI being the most stringent and least generalised, followed by FragAnchor, PredGPI, and lastly GPI SOM as the most generalised and least stringent.

### 2.4.2 Total GPI anchored protein prediction from the *C.elegans* genome

The total number of proteins predicted to be GPI anchored from all programs was 778. Of these 327 were validated by at least two prediction programs (Table 2.1, for the full list of proteins see Appendix 1). For the proteins with only one validation there is a disproportionate number from GPI SOM (76%, Figure 2.4). Analysis of the GO terms from the predicted proteins revealed a large proportion of proteins with the label of “transport ion” for their Biological Process description (Figure 2.5). Since ion transportation involves the formation of transmembrane pores it would be unlikely for these proteins to be designated as GPI anchored. Proteins with two or more prediction program validations showed a decrease in the proportion of proteins designated with transport ion. Previous proteomic studies in human cell lines and *Arabidopsis* found the use of multiple prediction programs improves the fidelity of validation of experimentally derived GPI anchored proteins (Elortza *et al.*, 2006). The final number of predicted GPI anchored proteins from this analysis is designated to be 327 sequences as predicted by two or more prediction programs (Appendix 1).

### 2.4.3 Validation of predictions with *C. briggsae* orthologues

Out of the total number of *C. elegans* predicted GPI anchored proteins 382 were also found with orthologues in *C. briggsae* that are also predicted to be GPI anchored, with 201 of those proteins predicted with two or more programs (Table 2.1, for the full list of proteins see Appendix 2). *C. briggsae* is a well known companion model organism for *C. elegans* and there is close conservation between their genomes (Stein *et al.*, 2003). It was therefore postulated that conserved genes for GPI anchoring in both organisms would lead to better validation of the prediction. Of the 201 genes the proportion for the GO terms were similar in all the three categories recorded, indicating that there is no marked difference of predictive power by the use of *C. briggsae* orthologue for validation (Figure 2.5). *C. briggsae* orthologue validated proteins may represent a core list of proteins with potential GPI anchor modifications. The list of proteins may also be used as a starting point for the study of potential GPI anchored proteins in *C. briggsae*.

### 2.4.4 Functions of GPI anchored proteins in *C. elegans*

#### 2.4.4.1 GO terms of likely functions for the predicted proteins

GO terms are a set of curated annotations which describe the characteristic of genes in a non-species dependent manner. GO terms are split into three broad categories based on the gene's Molecular function, Biological process and Cellular component. Of the 327 GPI anchored proteins there were 88 proteins with entries for Molecular Function, 80 entries for Biological Process, and 149 with entries for Cellular Component (Table

2.2). For Molecular Function the majority of the proteins were involved in catalysis, with many of the proteins present having carboxypeptidase activity. This is in line with the finding that carboxypeptidase M is GPI anchored in mammalian cells (Skidgel *et al.*, 1996). A large proportion of GPI anchored proteins also appear to be involved in the binding of substrates and transport, with a relatively small number involved in receptor binding and structural roles (Figure 2.6a).

For Biological Processes the majority of GPI anchored proteins appear to be involved in metabolic processes. A large percentage of genes are also involved in regulation, development and signalling (Figure 2.6b), which is consistent with the roles of GPI anchored proteins in other organisms (Ikezawa, 2002). There are a large percentage of proteins with the description of transport ion in the prediction, which may represent transmembrane proteins that have been identified as false positives. Most of these proteins are however validated with three or more prediction programs, and so may be genuine GPI anchored proteins with miss-annotations for their GO terms. One protein (C05D9.3) is involved in cell adhesion, which is also documented to occur in the adhesion of neural cells (Karagogeos, 2003).

For the Cellular Component part of the prediction programs the vast majority of the proteins were annotated as membrane, which supports the presence of GPI anchoring (Figure 2.6c). The proteins annotated as extracellular may still possess a GPI anchor as certain anchored proteins can be released from the cell surface as a part of their function (Yoon *et al.*, 2007). There are 6 proteins designated as cytoplasmic, which on further analysis were all curated with predicted GO terms and do not have experimental data to verify the annotation. The gene with the cell annotation is acetylcholine esterase 2 (*ace-2*) and is a well known GPI anchored protein in other systems. The one nuclear gene is called *bli-4* and has multiple splice variants with different C-terminal sequences, one of which could potentially be GPI anchored.

### 2.4.4.2 Genes of interest in *C. elegans* with prediction for GPI anchoring

Many interesting genes were present within the list of potential GPI anchored proteins found in this analysis (Table 2.5). Five genes have descriptions as lysosomal carboxylpeptidases, and this sub cellular compartment has been shown to be involved in GPI anchored protein sorting and have associated GPI anchored proteins (Grunfelder *et al.*, 2002). 20 peptidases, including the *acn-1* gene that has lost its metallopeptidase active site but is still important for larval development and moulting, are also predicted to be GPI anchored. *C. elegans* contains four acetylcholine esterase genes (*ace-1, 2, 3* and *4*) and three of them, *ace-2*, *ace-3* and *ace-4* are present within the predicted results. Acetylcholine esterase is a involved in neural transmission at the synaptic cleft and has a highly conserved GPI anchored form (Nalivaeva and Turner, 2001). The genes *tre-3* and *tre-5* encode trehalases which are also commonly found to be GPI anchored in mammalian cells (Netzer and Gstraunthaler, 1993); they account for two out of the five putative trehalases in *C. elegans*. The *C. elegans* gene *odr-2*, an olfactory neuron gene with homology to *Ly-6* (leucocyte antigen-6) (Chou *et al.*, 2001), was found to have validation in three of the prediction programs tested. Related to this are *hot-3*, *4*, and *7*, genes of unknown function that are homologues to *odr-2* are also present on the list of potential proteins, with *hot-5* predicted to be GPI anchored with GPI SOM only. *wrk-1* encodes a widely expressed homologue of a GPI-anchored immunoglobin superfamily (IgSF) protein and has five potential isoforms, three of which are found here. Two forms of the apical gut protein *tag-10* were found to be GPI anchored. *Tag-10* is orthologous to the GA1 apical gut protein of *Haemonchus contortus* that was demonstrated to be a GPI anchored protein in immunisation studies of sheep (Jasmer *et al.*, 1996). Lastly *phg-1* (also known as *phas-1*) was predicted to be GPI anchored by two programs in this analysis, in which the gene has also been demonstrated to be GPI anchored when expressed in a mammalian cell line (Agostoni *et al.*, 2002).

### 2.4.5 Conclusion

The proportion of genes with potential GPI anchoring found in this study accounts for 1.39% of the *C. elegans* genome. Previous estimates of *C. elegans* GPI anchored protein amount have all been attempted with only one prediction program, with the more conservative Big PI estimating the number to be 0.45% (Eisenhaber *et al.*, 2001), 0.66% for FragAnchor when only the Highly probable category of proteins was considered (Poisson *et al.*, 2007), and around 2.8% by GPI SOM (Fankhauser and Maser, 2005). This chapter presents the most comprehensive analysis of potential GPI anchored proteins in *C. elegans*, with the stringency of validation from multiple programs to reduce potential false positive and false negative predictions of GPI anchoring. The results presented here can help *C. elegans* researchers interested in GPI anchored proteins to look at their gene of interest in a different way, and may also aid researchers in the field of GPI anchored proteins by offering them another resource for the analysis of these proteins.

# *Chapter 3*

# Analysis of GPI biosynthesis genes in *Caenorhabditis elegans*

## 3.1 Introduction

The glycosylphospatidylinositol (GPI) anchor is a branched glycolipid that requires a complex biosynthetic pathway for its production. The use for this molecule as a protein anchor is widespread within living organisms, and GPI anchored proteins have been ubiquitously found in eukaryotes, including vertebrates, plants, insects, fungi and protozoa (Ferguson *et al.*, 1985b; Ferguson and Williams, 1988; Hortsch and Goodman, 1990; Morita *et al.*, 1996). The presence of GPI anchoring is less certain within the Eubacteria and Archaeobacteria kingdoms, with no evidence found so far for any eubacterial species that possess this post-translational modification. There is however tentative suggestion that certain Archaeobacteria also possess this protein anchor, as postulated by bioinformatic searches (Eisenhaber *et al.*, 2001) and experimentally verified in the archaea species *Sulfolobus acidocaldarius* (Kobayashi *et al.*, 1997). The Sulfolobus genus has been considered to be a close relative of eukaryotes (Iwabe *et al.*, 1989; Lake *et al.*, 1984; Woese *et al.*, 1990), which raises the possibility that this form of membrane attachment had its evolutionary origin in the Archaea.

### 3.1.1 The GPI anchor core structure

The structural determination of the GPI anchor began in the 1980's with the work of Fergurson *et al.* producing a partial structure for the variant surface glycoprotein anchor of *Trypanosoma brucei* (Ferguson *et al.*, 1985b), which led later to its determination by a combination of techniques involving nuclear magnetic resonance spectroscopy, mass spectrometry, chemical modification, and exoglycosidase digestion (Ferguson *et al.*, 1988). Since then more than 20 different GPI anchor structures have been solved from a variety of different organisms, which provided much insight into the properties of the anchor within the cell (Ferguson, 1999). All GPI anchors contain a highly conserved backbone, which begins with the C-terminal residue of the protein (ω site, see below) attached via an amide bond to phosphoethanolamine. This in turn is linked to a glycan core with the structure mannose(α1-2)mannose(α1-6)mannose(α1-4)glucosamine(α1-6)*myo*-inositol. Finally, the molecule ends with a phospholipid tail that anchors the structure within the membrane (Figure 3.1).

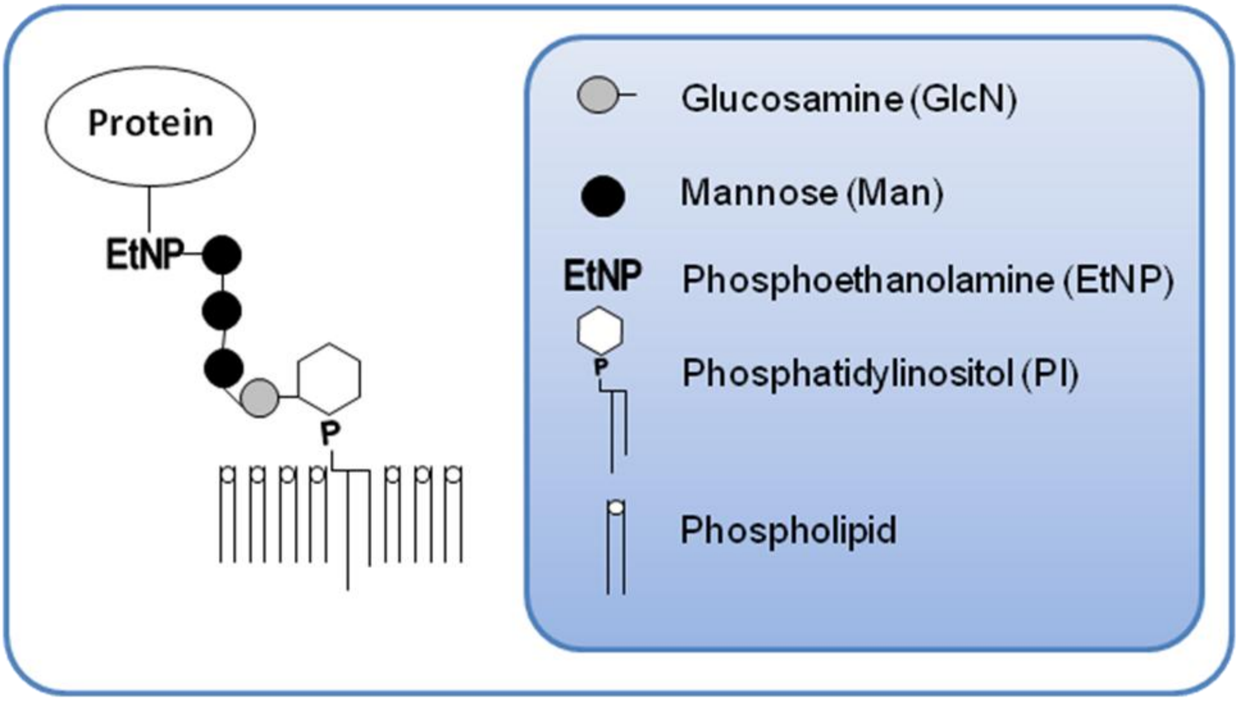



**Figure 3.1. The conserved core structure of GPI anchors.** The molecule has the structure EtNP-Man(α1-2)Man(α1-6)Man(α1-4)GlcN(α1-6)*myo*-PI, with the protein attached to the EtNP moiety.

### 3.1.2 Modifications to the core structure

All GPI backbones have a variety of species and cell type specific side chain additions in their glycan core. Most of these modifications involve the mannose subunits, with the addition of complex arrays of mannose (Man), galactose, N-acetylgalactosamine, sialic acid, N-acetylhexosamine and phosphoethanolamine observed in several organisms, including several species of protozoan parasites, *Saccharomyces cerevisiae*, plants, rat, human, and others (Brewis *et al.*, 1995; Deeg *et al.*, 1992; Ferguson *et al.*, 1988; Fontaine *et al.*, 2003; Homans *et al.*, 1988; Nakano *et al.*, 1994). Mannose is the most common side chain addition to the mannose closest to the protein in the glycan core. The addition of phosphoethanolamine to either the middle or glucosamine attached mannose occurs only in higher eukaryotes and is not found in protozoa. All known mammalian GPI anchors are found with this modification on the mannose adjacent to glucosamine. Complex side-chains of polysaccharides are found mainly on this mannose as well (Paulick and Bertozzi, 2008). The glucosamine of the core glycan has been found to be modified in *Trypanosoma cruzi* with 2-aminoethylphosphonate (Almeida *et al.*, 2000), but is otherwise unmodified in most other cases. It is thought that these side chain modifications occur for the specific needs of the anchor in different conditions, such as dense packing in VSGs and other steric effects in relation to the lipid bi-layer (Ferguson, 1999; Homans *et al.*, 1989). The inositol moiety may become palmitoylated at the 2 position in certain GPI anchors (Treumann *et al.*, 1995). This modification makes the anchor resistant to cleavage with PI specific phospholipase C (PIPLC), but not resistant to the action of PI specific phospholipase D (PIPLD) (Deeg and Davitz, 1995). Lastly, fatty acid remodelling in the phosphoinositol tail may occur, which involves replacement of the unsaturated fatty acid chains of phosphotidylinositol to diacylglycerol, alkylacylglycerol, myristate or ceramide (Kerwin *et al.*, 1994; McConville and Ferguson, 1993; Morita *et al.*, 2000; Sipos *et al.*, 1997). The replacement to a saturated chain is essential for the localization of the GPI anchor within lipid raft subdomains within the membrane (Maeda *et al.*, 2007), which may be due to the tight packing requirements within the environment.

### 3.1.3 GPI anchor synthesis and modification

GPI anchor synthesis is a multistage biochemical process and takes place within the Endoplasmic Reticulum (ER). The biosynthetic pathway is different with regard to the specific organism, with the most notable difference between the protozoan pathway and that of higher eukaryotes (Ferguson, 1999). Most research on the biosynthetic pathway comes from studies of two organisms, human and *S. cerevisiae* (here on referred to as yeast), in which 23 genes have been found so far to be involved in the process (Orlean and Menon, 2007). The making of a GPI anchor starts off on the cytoplasmic surface of the ER membrane and finishes with the attachment of the GPI anchored protein in the lumen of the ER and takes 12 steps, with one of the steps being tissue specific in humans. After the synthesis and attachment of the protein to the anchor the GPI structure is further modified in the ER and Golgi before final transport to the cell surface. A detailed description of all known processes involved in the human and yeast GPI anchor modification is given in figure 3.2. All genes referred to in this section are human/yeast unless otherwise specified.

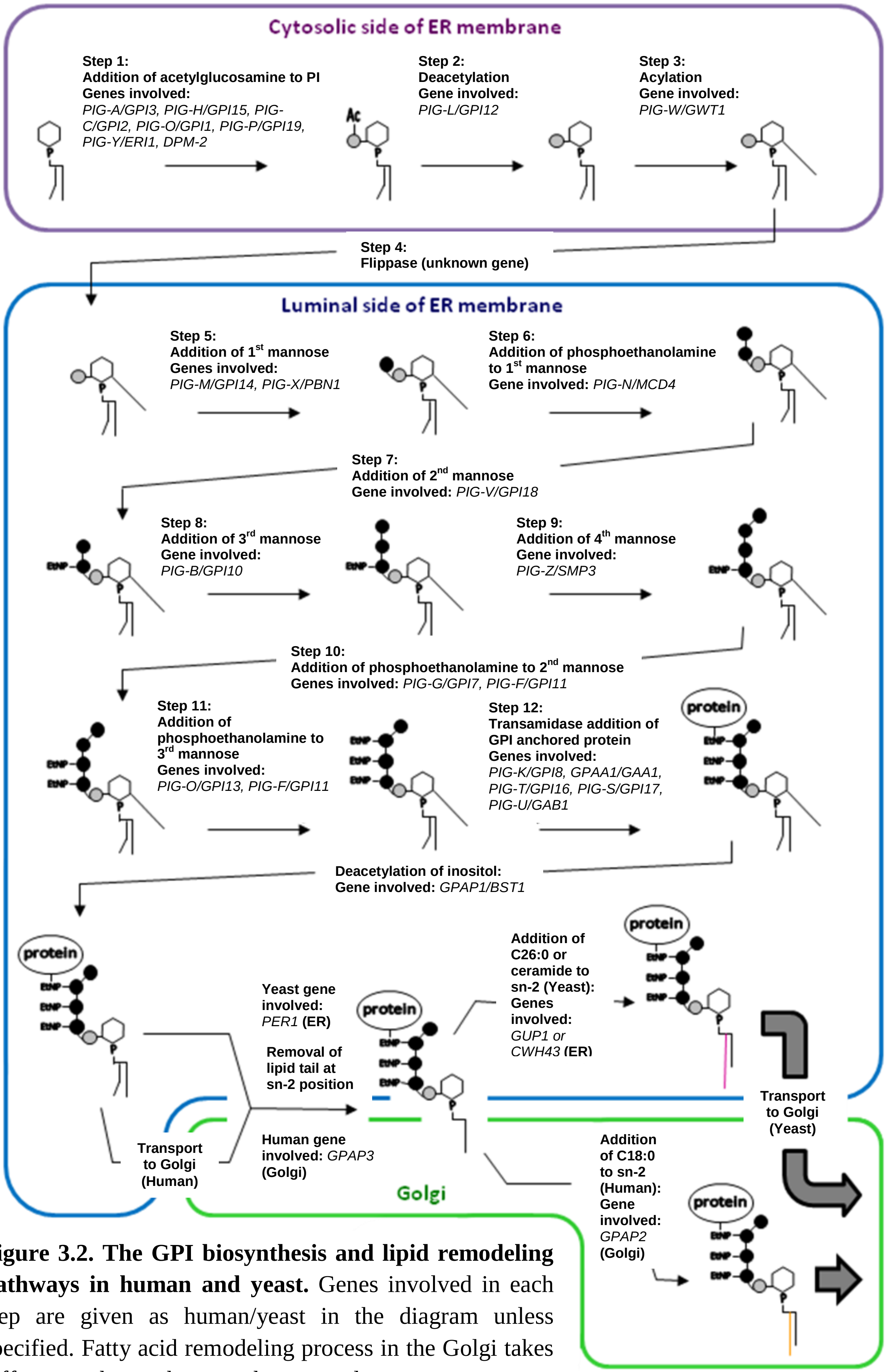



**Figure 3.2. The GPI biosynthesis and lipid remodeling pathways in human and yeast.** Genes involved in each step are given as human/yeast in the diagram unless specified. Fatty acid remodeling process in the Golgi takes different pathways between human and yeast.

### 3.1.3.1 The GPI anchor synthesis pathway

#### 3.1.3.1.1 Step 1: Transfer of α-1-6-N-acetyglucosamine (GlcNAc) to phosphoinositol (PI) to form GlcNAc-PI

The first reaction of GPI anchor synthesis is the formation of GlcNAc-PI from uridine diphosphate (UDP)-GlcNAc and PI (Eisenhaber *et al.*, 2003a). This reaction is catalysed on the cytoplasmic leaflet of the ER membrane by the UDP-GlcNAc transferase complex (Vidugiriene and Menon, 1993), which so far has seven components implicated for its function (Murakami *et al.*, 2005). The enzyme is also negatively regulated by Ras in yeast (Sobering *et al.*, 2004), but such regulation is not detected in mammalian systems (Murakami *et al.*, 2005).

*PIG-A/GPI3*

The *PIG-A/GPI3* component of the GlcNAc transferase is the catalytically active part of the enzyme in humans and yeast. The human PIG-A is 484 amino acids long and the yeast protein is 452 amino acids in size. PIG-A has a single transmembrane domain near the C terminus of the protein with its catalytic subunit exposed to the cytosolic side of the ER membrane, with its short C terminal ER luminal domain implicated as a signal for its orientation within the ER membrane (Watanabe *et al.*, 1996). PIG-A is vital for GPI anchor production, and the lack of this protein causes the onset of the haemophilic disease, Paroxysmal nocturnal haemonglobineria, in humans through the loss of the regulatory proteins CD55 and CD59 (Parker, 1996), and female infertility in mice (Alfieri *et al.*, 2003).

*PIG-H/GPI15*

PIG-H forms a complex with PIG-A and is essential to achieving physiological levels of GPI anchors in humans, but a measurable amount (<1% normal) can be detected in its absence (Watanabe *et al.*, 1996). The protein is 188 amino acids in humans and 229 amino acids in yeast, and forms a tight hairpin loop with both the N and C termini pointing into the cytoplasm of the cell.

*PIG-C/GPI2*

*GPI2* was found in a yeast temperature sensitive lethal strain that had interactions with *GPI1* (Leidich *et al.*, 1995). PIG-C in humans has a hydropathy plot typical of a transmembrane protein and is predicted to have 8 transmembrane regions, with both its N and C termini on the cytoplasmic side of the ER (Inoue *et al.*, 1996; Tiede *et al.*, 2000). It is speculated that PIG-C/Gpi2p acts as a scaffolding protein for the enzyme complex, so that the transferase is secured to the cytosolic side of the ER membrane. PIG-C has a size of 297 amino acids and GPI2 is 280 amino acids long.

*PIG-Q/GPI1*

*GPI1* was found in a conditionally lethal strain of yeast in a screen for GPI synthesis genes (Leidich *et al.*, 1994). PIG-Q/Gpi1p are predicted to have 6 transmembrane domains with both of its amino acid termini emerging onto the cytoplasmic side of the ER (Tiede *et al.*, 2000). *PIG-Q* loss of function in humans results in a significant reduction of transferase activity in humans. The loss of PIG-Q leads to reduced cellular levels of PIG-C and PIG-H, and causes inhibition of association between PIG-C, PIG-A and PIG-H. PIG-Q is thought to have the role of stabilizing the transferase by protecting it from cellular degradation (Hong *et al.*, 1999b). The human PIG-Q protein is 581 amino acids and the yeast protein is 609 amino acids long.

*PIG-P/GPI19*

Human *PIG-P* produces a small protein (158 amino acids) that interacts with PIG-A and PIG-Q. It is found to be essential for the first step of GPI biosynthesis, but its exact mode of function has not yet been elucidated (Watanabe *et al.*, 2000a). The yeast homologue has recently been found to have a similar phenotype with a size of 140 amino acids, and is predicted to form a hairpin loop within the ER membrane with both ends pointing into the cytoplasm (Newman *et al.*, 2005).

*PIG-Y/ERI1*

*PIG-Y* was found in a human cell line with a severe defect in surface GPI anchor protein expression (Murakami *et al.*, 2005). *PIG-Y* encodes a 71 amino acid protein that directly binds to PIG-A, although a 6 member UDP-GlcNAc transferase complex can be formed in its absence. The protein bears some sequence similarities to yeast Eri1p, which has also been shown to be involved in the first step of GPI anchor synthesis (Sobering *et al.*, 2004).

*DPM2*

*DPM2* exist in mammals as a cytoplasmic protein of 88 amino acids and is the regulatory subunit of dolichol phosphate mannose (Dol-P-Man) synthase enzyme complex (Maeda *et al.*, 2000). The protein weakly interacts with PIG-A, PIG-C and PIG-Q and has been shown to enhance the transferase activity by 3 fold (Watanabe *et al.*, 2000a). No ortholog has been found so far in yeast.

#### 3.1.3.1.2 Step 2: De N-acetylation of GlcNAc-PI to form glucosaminyl (GlcN)-PI

*PIG-L/GPI12*

The second reaction within the GPI anchor synthesis also occurs on the cytosolic side of the ER membrane (Vidugiriene and Menon, 1993). The reaction involves the deacetylation of the GLcNAc-PI by PIG-L/GPI12 into glucosaminylphosphatidylinositol (GlcN-PI) (Nakamura *et al.*, 1997a) and was shown to be essential in yeast (Watanabe *et al.*, 1999). PIG-L is a type I membrane protein of 252 amino acids with a single transmembrane domain and has two independent ER retention signals (Pottekat and Menon, 2004). Further analysis identified the protein to be a zinc metalloenzyme and a possible target for an antiprotozoan drug (Urbaniak *et al.*, 2005).

#### 3.1.3.1.3 Step 3: Acylation of inositol ring on GlcN-PI to form GlcN-acyl-PI

*PIG-W/GWT1*

Step 3 of GPI anchor synthesis involves the addition of an acyl group (usually palmitate) to the inositol ring of GlcN-PI at position 2 to produce GlcN-acyl-PI. This process is carried out on the cytosolic side of the ER membrane and is carried out by the *PIG-W/GWT1* gene. The protein consists of 504 amino acids in humans and 498 amino acids in yeast. *GWT1* deletion confers lethality in yeast (Umemura *et al.*, 2003), and a study on *PIG-W* implicates a role for the acyl group in the addition of phosphoethanolamine to the third mannose (Murakami *et al.*, 2003).

### 3.1.3.1.4 Step 4: Flipping of GlcN-acyl-PI into the lumen

Since the rest of the reactions of the GPI biosynthetic pathway occur within the ER lumen the GlcN-PI molecule needs to be "flipped" across the membrane bilayer before it can be further processed into a functional anchor. Flipping of glycerophospholipids is an energetically expensive process that rarely occurs spontaneously, and requires the action of special "flippase" enzymes for their efficient transfer (Pomorski and Menon, 2006). No GPI specific flippase has been found so far, but research has discovered that flipping of GlcN-PI occurs in model membranes in the presence of a number of ER phospholipid flippases, indicating the possibility that this process is shared with the general phospholipid flipping pathways within the ER (Vishwakarma and Menon, 2005).

### 3.1.3.1.5 Step 5: Addition of 1st mannose subunit to GlcN-acyl-PI to form Man-GlcN-acyl-PI

*GPI-manosyltransferase-I (GPI-MT-I, PIG-M/GPI14)*

The main catalytic subunit of this enzyme is called *PIG-M* in humans and *GPI14* in yeast (Maeda *et al.*, 2001c). The human and yeast proteins are 423 and 403 amino acids in length, respectively. *GPI14* loss of function alleles causes cell wall instability in yeast and an increase in transcription of cell wall related genes (Davydenko *et al.*, 2005).

*PIG-X/PBN1*

*PIG-X/PBN1* is an essential interaction partner of *PIG-M* with a size of 252 amino acids in human and 416 amino acids in yeast (Ashida *et al.*, 2005b). This protein forms an association with PIG-M and stabilises it in the ER. Pbn1p is also required for folding and stability of a number of other proteins in yeast and act as an essential chaperone-like protein within the ER (Subramanian *et al.*, 2006).

#### 3.1.3.1.6 Step 6: Modification of Man-GlcN-acyl-PI with ethanolphosphoamine (EtnP) at the 1st mannose to form (EtnP)-Man-GlcN-acyl-PI

*PIG-N/MCD4*

*PIG-N* catalyses the addition of EtnP to the 1st mannose in humans and has a size of 931 amino acids (Hong *et al.*, 1999a), with the yeast gene *MCD4* as its homolog with a size of 919 amino acids (Gaynor *et al.*, 1999). This modification is important for the addition of the third mannose in yeast, and has been shown to be important for subsequent remodelling of the lipid anchor in the Golgi (Wiedman *et al.*, 2007; Zhu *et al.*, 2006). In humans the gene is not essential but significantly affects surface expression of GPI anchored proteins by the recognition of this moiety by the transamidase complex (Vainauskas and Menon, 2006).

#### 3.1.3.1.7 Step 7: Addition of 2nd mannose to Man-GlcN-acyl-PI to form Man-(EtnP)-Man-GlcN-acyl-PI

*GPI-MT-II (PIG-V/GPI18)*

*PIG-V* was recently found to be the gene responsible for *GPI-MT-II* activity in humans (Kang *et al.*, 2005). The gene codes for a protein of 493 amino acids, and has the ortholog gene *GPI18* in yeast (433 amino acids), which shows a weakened cell wall phenotype (Fabre *et al.*, 2005). Both proteins are predicted to have 8 transmembrane domains and functionally conserved regions in their ER luminal sequences. Human cells mutated in *PIG-V* accumulated EtnP modified Man-GlcN-acyl-PI, while yeast mutants have both modified and unmodified Man-GlcN-acyl-PI, which indicates alternative routes within the biosynthetic pathway in yeast.

#### 3.1.3.1.8 Step 8: Addition of 3rd mannose to Man-(EtnP)-Man-GlcN-acyl-PI to form Man-Man-(EtnP)-Man-GlcN-acyl-PI

*GPI-MT-III (PIG-B/GPI10)*

The addition of the 3rd mannose mediated by *PIG-B* in humans (Takahashi *et al.*, 1996) and *GPI10* in yeast (Sutterlin *et al.*, 1998). The human gene encodes a protein that is 554 amino acids long and the yeast protein length is 616 amino acids. PIG-B was found to have 12 transmembrane domains in a bioinformatic comparison of related mannosyltransferases (Oriol *et al.*, 2002).

#### 3.1.3.1.9 Step 9: Addition of 4th mannose to Man-Man-(EtnP)-Man-GlcN-acyl-PI to form (Man)-Man-Man-(EtnP)-Man-GlcN-acyl-PI

*GPI-MT-IV (PIG-Z/SMP3)*

The addition of the 4th mannose is essential in yeast but appears to be tissue specific in humans, where it occurs in the brain (Orlean and Menon, 2007; Stahl *et al.*, 1992; Taron *et al.*, 2004a). The fourth mannose transferase for humans is named *PIG-Z* and has a size of 579 amino acids. The yeast homologue of the gene is called *SMP3* and has a size of 516 amino acids (Grimme *et al.*, 2001).

#### 3.1.3.1.10 Step 10: Addition of EtnP to 2nd mannose of (Man)-Man-Man-(EtnP)-Man-GlcN-acyl-PI to form (Man)-Man-(EtnP)-Man-(EtnP)-Man-GlcN-acyl-PI

*PIG-G/GPI7*

*PIG-G* encodes a protein of 975 amino acids in humans and is responsible for the addition of EtnP to the $2^{nd}$ mannose in the core glycan (Shishioh *et al.*, 2005). The yeast gene, *GPI7* is 831 amino acids and disruption of the gene causes cell wall defects, such as protein anchoring and cell wall separation (Benachour *et al.*, 1999; Fujita *et al.*, 2004; Richard *et al.*, 2002). In humans, however, the modification has little effect on GPI anchor attachment, and produces a minor type of GPI anchor that may also be present on the cell membrane without protein attachment (Shishioh *et al.*, 2005).

#### 3.1.3.1.11 Step 11: Addition of EtnP to 3rd mannose of (Man)-Man-(EtnP)-Man-(EtnP)-Man-GlcN-acyl-PI to form EtnP-(Man)-Man-(EtnP)-Man-(EtnP)-Man-GlcN-acyl-PI

*PIG-O/GPI13*

*PIG-O/GPI13* is responsible for the addition of the EtnP to the glycan backbone at the $3^{rd}$ mannose, which is the final structure needed for the completion of the core GPI anchor (Hong *et al.*, 2000; Taron *et al.*, 2000). The human *PIG-O* gene produces a protein of 1089 amino acids and the yeast *GPI13* gene encodes a protein of 1017 amino acids, with both essential for GPI anchor synthesis in each organism.

#### 3.1.3.1.11.1 Additional gene involved in steps 10 and 11

*PIG-F/GPI11*

*PIG-F/GPI11* both encode proteins of 219 amino acids in humans and yeast (Inoue *et al.*, 1993; Taron *et al.*, 2000). They are involved in the EtnP modification of the 2$^{nd}$ and 3$^{rd}$ mannose and interact directly with *PIG-G/GPI7* and *PIG-O/GPI13*. *PIG-F* in human is essential for the action of *PIG-O* in the addition of EtnP to the third mannose (Hirose *et al.*, 1992; Puoti and Conzelmann, 1993; Sugiyama *et al.*, 1991), with the *PIG-G* gene implicated in the regulation of *PIG-O* via competition for PIG-F proteins (Hong *et al.*, 2000). *GPI11* was found to be an essential gene in yeast but was shown not to be a requirement for EtnP addition by *GPI13*, implicating it in other cellular processes (Taron *et al.*, 2000).

#### 3.1.3.1.12 Step 12: attachment of GPI anchor via the GPI transamidase complex

The attachment of the GPI anchor to a protein is catalysed by the GPI transamidase (GPIT) complex. This enzyme consists of 5 confirmed subunits, *PIG-K, GPAA1, PIG-T, PIG-S* and *PIG-U*, which co-immunoprecipitate to form the functional transamidase (Hong *et al.*, 2003). GPIT does not have any sequence specificity but recognises a conserved C-terminal sequence motif, with the amino acid residue of attachment on the protein called the ω site. The motif can be split into 4 regions; the first contain 11 mostly polar residues acting as a linker to the main protein, the second contain small residues including the ω site, the third region is a spacer region of around 7 moderately polar residues, and the last section consists a sequence of hydrophobic amino acids up to the C-terminus (fig.3.3) (Eisenhaber *et al.*, 1998). It was recently found that the GPIT subunit *PIG-U* was upregulated in bladder cancer

(Guo *et al.*, 2004) and that *GPAA1* and *PIG-T* over-expression causes invasiveness in breast cancer (Wu *et al.*, 2006). A study of all 5 GPIT subunits in 19 different cancers implicated these genes in a variety of oncogenic roles, including upregulation in cancers of the breast, ovarian, uterus, lymphoma, lung, and deregulation in a number of other cancer types (Nagpal *et al.*, 2008). Taken together, it seems that GPIT subunits are of immense interest to medical science, and the importance of GPI anchoring is just beginning to be explored within human biology.

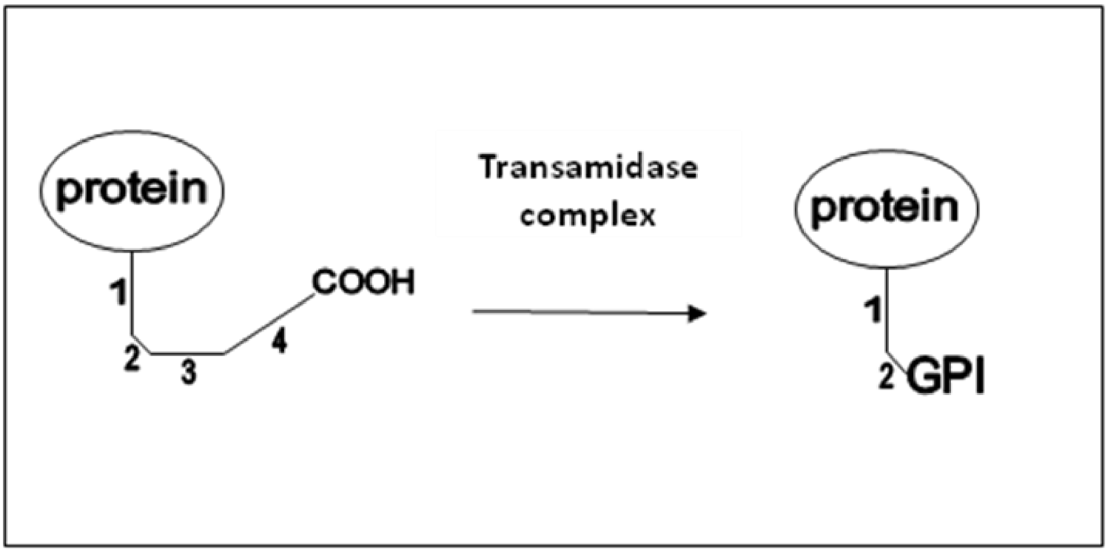



**Figure 3.3. Reaction of the transamidase complex.** GPI anchored proteins contain a C-terminal consensus motif with 4 characteristic regions. The second region (2) contains the ω site, which is the residue of attachment to the GPI anchor.

*PIG-K/GPI8*

*PIG-K* is the human gene that encodes the catalytic subunit of the GPI transamidase. The protein product for this gene is 395 amino acids, with its yeast ortholog at 411 amino acids long. This protein functions as a cysteine endopeptidase with a pair of conserved active sites at His157 and Cys199, and has a segment of TM region around 30 amino acids at the C-terminus (Meyer *et al.*, 2000). The TM domain is not essential for the function of the protein (Ohishi *et al.*, 2000). Gpi8p was found to form a prolonged association with the C-terminal signal sequence of unanchored proteins and catalyses the reaction at the ω site by forming a thioester intermediate with the proprotein (Chen *et al.*, 2003a; Spurway *et al.*, 2001). Knockout of the *PIG-K* ortholog in African trypanosomes (gpi8) abolished the attachment of GPI anchored proteins (Lillico *et al.*, 2002). PIG-K is enzymatically active when expressed as a recombinant protein in *E. coli* (Kang *et al.*, 2002), but it's activity is greatly attenuated *in vivo* by the subunits associated with it (Chen *et al.*, 2003a; Ohishi *et al.*, 2003).

*GPAA1/GAA1*

The human gene *GPAA1* encodes a protein of 621 amino acids with 7 transmembrane domains. The yeast ortholog of the gene is 614 amino acids long. The protein interacts with the other GPIT subunits through a large ER lumenal domain in between the first and second transmembrane domains (Vainauskas *et al.*, 2002). GPAA1 forms a complex with PIG-K where it is required for the recognition of the proprotein substrate (Chen *et al.*, 2003a). GPAA1 also has a proline residue in the C-terminal TM region found to be essential for GPI anchor recognition (Vainauskas and Menon, 2004b), suggesting a role in the recognition of both of the substrates of transamidase.

*PIG-T/GPI16*

*PIG-T* encodes a protein of 578 amino acids in human. The yeast ortholog *GPI16* is 610 amino acids long and exists as an integral membrane protein with a single transmembrane domain (Fraering *et al.*, 2001). This protein has structural similarities to prolyl oligopeptidase, a porcine protein with a novel beta-propeller structure which may be able to confer specificity to the PIG-K cysteine protease (Eisenhaber *et al.*, 2003a). An intermolecular disulfide bridge forms between Cys92 on PIG-K and Cys182 of PIG-T, and this covalent modification is essential for normal levels of transamidase activity within the cell (Ohishi *et al.*, 2003). Affinity purification of GPIT in yeast resulted in a complex of Gpi16p, Gpi8p and Gaa1p, suggesting that these three proteins form a core structure within which transamidase activity occurs (Fraering *et al.*, 2001).

*PIG-S/GPI17*

*PIG-S* in humans encodes a protein of 555 amino acids with two putative transmembrane domains at each ends of the protein. The yeast ortholog is called *GPI17* and is 534 amino acids long. *PIG-S* is an essential gene for transamidase activity and has been implicated in a structural role for the complex, and may confer species specific selectivity for protein targets (Eisenhaber *et al.*, 2003a; Ohishi *et al.*, 2001). Unlike Gpi16p and Gaa1p, Gpi17p associates transiently with the GPIT complex in yeast (Zhu *et al.*, 2005).

*PIG-U/GAB1*

*PIG-U* is a recently found subunit of human transamidase and encodes a protein of 435 amino acids. Its ortholog in yeast is called *GAB1* (394 amino acids) and the protein is predicted to have 8 to 10 transmembrane domains, which partially rescues *PIG-U* knockout in human (Hong *et al.*, 2003). The function of *PIG-U* has been speculated to be recognition of either the GPI attachment signal or the lipid portion of GPI. Gab1p was found to form a complex with Gpi17p in yeast, suggesting the GPIT complex functions as two multi-subunit components (Grimme *et al.*, 2004). Gab1p may also have other functions in yeast, as depletion of the protein causes actin bar formation, suggesting the protein has functions in actin organization.

**3.1.3.2 Synthesis of Dol-P-Man, the mannose donor**

The mannose donor Dol-P-Man required by GPI synthesis are produced in human and yeast by the gene *DPM1/DPM1*. This involves the reaction between Dol-P and GDP-Man, which occurs at the cytosolic side of the ER membrane and is transported into the luminal side of the ER via a flippase (Eisenhaber *et al.*, 2003a). Dol-P-Man is used extensively within the cell to modify various structures with mannose, including O-mannosylation and N-glycosylation of proteins (Orlean, 1990a). In yeast, only Dpm1p is required for this reaction, and the enzyme has a membrane transmembrane domain at the C-terminus which tethers it to the ER membrane. DPM1 in humans lack this domain, and needs to be stabilised by DPM2 and DPM3 in order to function (Ashida *et al.*, 2005c). DPM3 has been shown to have the domain required for anchoring to the ER membrane, and interacts with DPM1 to stabilise it for the reaction. DPM3 interaction also prevents DPM1 from becoming degraded by the cell machinery, possibly by blocking its ubiquitination. DPM2 acts to stabilise DPM1 within the complex (Maeda *et al.*, 1998b), and is also directly implicated in the complex used in the first step of GPI biosynthesis.

#### 3.1.3.3 Lipid remodelling

##### 3.1.3.3.1 Inositol deacylation

*PGAP1/BST1*

Inositol deacylation occurs after anchor attachment of the protein and is important for transport of the GPI anchored protein to the Golgi (Orlean and Menon, 2007). *PGAP1/BST1* encodes membrane proteins of 922 and 1029 amino acids in human and yeast and performs inositol deacylation within the ER (Tanaka *et al.*, 2004). *BST1* is also involved in quality control of GPI anchored proteins, where a delay in the deacylation process reduces the efficiency of degradation of misfolded GPI anchored proteins (Fujita *et al.*, 2006b).

##### 3.1.3.3.2 Fatty acid remodelling

The relatively short unsaturated lipid tail of the GPI anchor is subjected to modification in both human and yeast before the structure is transported to the surface of the cell. The yeast lipid tails can either be replaced by longer saturated fatty acids or ceramide (Sipos *et al.*, 1997), while in humans the modification involves replacement with a saturated lipid tail (Ikezawa, 2002). The process starts with the removal of the acyl group on the sn-2 position of the glycerol backbone of the GPI anchor, which is catalysed by PGAP3 (320 amino acids) in humans in the Golgi, and in yeast by its ortholog Per1p (357 amino acids) in the ER (Fujita *et al.*, 2006a).

Subsequently a saturated (C18:0) fatty acid is added to the anchor by PGAP2 (315 amino acids) in the Golgi in human (Tashima *et al.*, 2006), while in yeast the sn-2 position is first filled in the ER with a long saturated C26:0 chain catalysed by Gup1p (560 amino acids) (Bosson *et al.*, 2006), and may subsequently be modified with a ceramide in a multistep pathway within the ER and Golgi by as yet unidentified genes (Reggiori *et al.*, 1997). Yeast does contain a homologue to human *PGAP2*, which is called *CWH43* (953 amino acids) and adds a ceramide moiety to the GPI anchor tail (Ghugtyal *et al.*, 2007). Fatty acid remodelling is important for protein transport to the surface of the cell, where it is also required for association of the protein within lipid rafts (Maeda *et al.*, 2007).

### 3.1.4 The *C.elegans* model system and contributions to genetics research

The nematode *C. elegans* has a reputation as an excellent model system for elucidating the role of individual genes within a developmental context. *C. elegans* has a transparent appearance and has an invariant lineage from the first meiotic division to the adult (Brenner, 1974), which allows detailed analysis of temporal and spatial gene expression under a light microscope. Transformation of *C. elegans* with knock in of genes is relatively straightforward compared with other developmentally complex models. A common technique involves the injection of the DNA of interest into the germline of the worm, which causes stable inheritance and expression of the gene, allowing a variety of developmental questions to be answered. This technique was first demonstrated with the suppression of sex transformation in an amber suppressible *tra-3* strain, following injection of tRNAs from a *sup-7* amber suppressor mutant (Kimble *et al.*, 1982). Fire demonstrated the versatility of this approach by showing that injection of a *lacZ* gene fused at the 5' end with a *Drosophila* heat shock promoter is able to produce its gene expression pattern *in vivo* (Fire, 1986). The injection procedure of Fire produced genomically integrated genes of 1-10 copies with very similar expression levels to the wildtype; it was however technically demanding due to the need for the DNA to be injected into oocytes. A more accessible protocol of injecting into the germline syncytium of the worm was developed by Stichcomb *et al.*, which forms the basis for the most popular method of transformation used today (Stinchcomb *et al.*, 1985). Stinchcomb's protocol is technically less demanding but creates large extrachromosomal arrays of 80-300 tandem repeats of the injected plasmid with varying levels of inheritance stability. The development of green florescent protein (GFP) reporter constructs (Chalfie *et al.*, 1994) paved the way for the analysis of a gene's expression pattern in real time. Selectable markers for positive injection were also developed to aid the identification of successful DNA integration, with the use of the dominant *rol-6* gene giving an easily scorable "rolling" phenotype when co-injected with the desired vector (Mello *et al.*, 1991). The technique of micro particle bombardment, which involves the introduction of DNA into the worm germline via microcarrier gold beads, was also adapted for transformation, with the rescue of the *unc-119* mutant worms strain (Maduro and Pilgrim, 1995) used as a selectable marker for successful integration (Praitis *et al.*, 2001). Transformation of worms using this technique yielded chromosomally integrated lines with low copy numbers of the injected DNA.

### 3.1.5 Expression pattern analysis in *C. elegans*

Expression patterns of *C. elegans* genes were first analysed with the introduction of promoter::reporter fusions made from the insertion of genomic fragments within a *lacZ* reporter plasmid (Hope, 1991). More precise methods for the creation of DNA fusion products followed, culminating with the highly versatile and accurate Gateway recombination approach, which uses the site specific recombination of bacteriophage lambda to create promoter::reporter fusion constructs (Hartley *et al.*, 2000). This approach was first used to produce a library of 12,000 open reading frames (ORF) from the *C. elegans* genome, which was termed the ORFeome of the worm (Reboul *et al.*, 2003). A library of promoter::reporter constructs was then created from 6,000 *C. elegans* gene promoters fused to GFP, and was named the Promoterome version 1.1 (Dupuy *et al.*, 2004). Transformation of 366 nematode lines for worm transcription factor promoters was recently performed with the Promoterome using a combination of microparticle bombardment and injection techniques, which yielded extensive information on the developmental expression pattern of a number of transcription factor gene families (Reece-Hoyes *et al.*, 2007). The promoter regions used for the Promoterome are all under 2,000 bp in length, which represents the size of 5' intergenic regions of a large proportion of genes (60%) and is likely to contain most of the *cis* regulatory elements of the gene. However, the size of the promoter regions may still be too small for some genes with large intergenic regions, and some of their crucial regulatory elements may not be present within their Promoterome construct. The Promoterome constructs also do not take into account of regulatory elements outside of the 5' region of the gene, such as in introns, 3' untranslated regions and *trans* acting elements, which may hinder its accuracy as a representation of the gene's expression pattern in vivo.

### 3.1.6 Plan for this chapter

In this chapter a detailed bioinformatic analysis of *C. elegans* and *C. briggsae* GPI anchor synthesis pathway genes was made with respect to the known human and yeast genes. *C. briggsae* is an excellent companion model organism to *C. elegans* with a completed genome (Stein *et al.*, 2003), which may shed insight into some of the homologues found in *C. elegans*. We also speculated into the nature of GPI anchor modifications within the nematode, and presented a possible structure and synthesis pathway for the anchor inside the worm. Expression profiles for important synthesis genes was also carried out via microparticle bombardment and injection analysis, with the use of the Promoterome and novel promoter::GFP constructs made with Gateway recombination. An analysis of the GPI synthesis pathway may give us a greater understanding of GPI anchoring within the worm, and an expression profile of these genes may provide insight into the role of GPI anchors within the context of tissue specific processes and development.

## 3.2 Method

### 3.2.1 Search for *C. elegans* homologues of GPI anchor synthesis pathway genes from humans

Human and yeast genes in the GPI anchored synthesis pathway were found through the literature search and their sequences were taken from the Ensembl web genome browser (http://www.ensembl.org/index.html). Sequences from the human pathway genes were searched against the *C. elegans* and *C. briggsae* genomes via BLAST at the Wormbase website (http://wormbase.org/db/searches/blast_blat). Sequence alignment was done with the ClustalX 2.0 tool (Larkin *et al.*, 2007).

### 3.2.2 Maintenance of *C. elegans* strains

Wild type *C. elegans* worms came from the N2 Bristol strain as described by Brenner (Brenner, 1974) and *unc-119* strain worms were provided courtesy of the Hope lab. Worms were kept on in 55 mm diameter agar plates made from nematode growth media (NGM, 50 mM NaCl, 1 mM $CaCl_2$, 25 mM $KH_2PO_4$, 1 mM $MgSO_4$, 5 µg/ml Cholesterol, 0.25% (w/v) peptone, 1.7% (w/v) agar) and seeded with OP50 strain *E. coli* bacteria (Brenner, 1974). Worms were kept at 20°C for 4 days or until most of the food was consumed and need renewal, which was done by moving 3-4 worms to freshly seeded plates with a platinum wire.

### 3.2.3 Liquid culture of *C. elegans*

*Unc-119* strain worms from 2 fully grown NGM plates were washed into 100 ml of S basal solution (0.1 M NaCl, 0.05 M potassium phosphate, pH 6, 5 µg/ml cholesterol) via pipetting. 100 µl of Streptomycin (50 mg/ml), 100 µl of Nystatin (50 mg/ml) and 4.5 ml of HB101 bacterial suspension were added to the S basal solution and the total mixture was incubated at 20°C shaking for 3 days, after which 1 ml of worms from the previous liquid culture was used to inoculate a new batch. The culture solution was checked daily and fresh bacteria were added as necessary.

### 3.2.4 Bacteria strains

Bacteria strains were kept at 4°C on 90 mm diameter agar plates with Luria-Bertani (LB) agar formula (8.6 mM NaCl, 1% (w/v) peptone, 0.5% (w/v) yeast extract, 1.5% (w/v) bacteriological agar). Strains requiring selection were streaked onto plates with supplied with the appropriate antibiotic at a final concentration of 100 µg/ml.

#### 3.2.4.1 OP50 *E. coli* strain

*E. coli* OP50 strain was acquired courtesy of Hope lab and kept on agar plates as described above. OP50 bacteria for NGM plates were grown in 100mL LB media (8.6 mM NaCl, 1% (w/v) peptone, 0.5% (w/v) yeast extract) at 37°C shaking overnight (o/n) and 5-6 drops were added to each NGM plate in a laminar flow hood and left to dry for 24 hours.

### 3.2.4.2 HB101 *E. coli* strain

HB101 *E. coli* strain was acquired courtesy of Hope lab. HB101 stock was kept on 140 mm diameter LB agar plates with streptomycin (50 µg/ml). Bacteria for worm liquid culture were grown in 1L LB media at 37°C shaking o/n and spun at 3,000g for 5 minutes. The supernatant was discarded and the bacterial pellet was resuspended in an equal volume of S basal and stored at 4°C. Typically 12 ml of final bacterial suspension was made per 1l of LB media.

### 3.2.5 Extraction of plasmids with miniprep

Plasmid extraction was performed using QIAprep Miniprep kit from Qiagen. A single colony of the desired strain of *E. coli* was taken from a selective plate and incubated in 2.5 ml of LB media (10 g tryptone, 5 g yeast extract and 5 g of NaCl in 1l of $dH_2O$) at 37$^o$C overnight while shaking. The bacteria were spun at 6,000 g for 3 mins and the supernatant was discarded. The pellet was resuspended in 250 µl of QIAprep buffer P1 (RNAase added, LyseBlue solution at 1:1,000) and shaken gently. 250 µl of QIAprep buffer P2 was the added to the solution and mixed thoroughly until a homogenous blue solution was visible. 350 µl of QIAprep buffer N3 was then added to the solution and mixed with inversion until the blue colour turns colourless and a cloudy precipitant is visible. The solution was then centrifuged on a benchtop centrifuge for 10 mins at approx. 10,000 *g* (13,000 rpm). The supernatant was then applied to a QIAprep spin column and centrifuged at 13,000 rpm for 60 sec. The flow-through was discarded and 0.75 µl of QIAprep buffer PE (with added EtOH) was applied to the column and spun at 13,000 rpm for 60 sec. The flow-through from this was also discarded and the column was spun again at 13,000 rpm to remove residual PE buffer. The spin column was then placed onto a 1.5 ml tube and 50 μl of buffer EB (10 mM Tris-HCL, pH 8.5) was added to the column and let stand for 60 sec, and then spun at 13,000 rpm for 60 sec. The final eluted DNA solution as checked by running in an agarose gel.

### 3.2.6 Polymerase Chain Reaction (PCR)

PCR was performed with Expand High Fidelity PCR system from Roche. Two master mixes of PCR reagents were prepared prior to loading onto the PCR machine (PCR Express, Hybaid). Master mix A consists of 0.5 μl of dNTP, 0.15 μl of upstream primer (in 10 mM Tris-HCl, pH 8.5), 0.15 μl of downstream primer (in 10 mM Tris-HCl, pH 8.5), 0.5 μl of template in (10 mM Tris-HCl, pH 8.5) and 23.7 μl of $dH_2O$ for a total volume of 25 μl per reaction; Master mix B consists of 5 μl of Expand High fidelity buffer (x10 without $MgCl_2$), 6 μl of $MgCl_2$ (25 mM stock solution, final solution 3 mM), 0.75 μl of Expand High Fidelity Enzyme mix (2.6U/reaction stored in 20 mM Tris-HCl, pH 7.5 (25°C), 100 mM KCl, 1 mM dithiothreitol (DTT), 0.1 mM EDTA, 0.5% Nonidet P40 (v/v), 0.5% Tween 20 (v/v), 50% glycerol (v/v)), and 13.25 μl $dH_2O$ for a total volume of 25 μl per reaction. 25 μl of Master mix A and 25 μl of Master mix B were added to one PCR tube and placed in the PCR machine. The program used was as follow- step 1: 94°C for 2 min, x1 repeat; step 2: 94°C for 15 sec, 59°C for 30 sec, 68°C for 5 min, x10 repeat; step 3:- 94°C for 15 sec, 59°C for 30 sec, 68°C for 5 min +5 sec per cycle, x10 repeat; step 4: 72°C for 7 min, x1 repeat.

Final hold step was at 4°C. DNA prepared from PCR were visualised with DNA agarose gel.

### 3.2.7 DNA sample running in agarose gel and visualization

DNA gels were made by mixing 0.4 g of agarose with 50 ml of Tris –acetate EDTA buffer (TAE, 40 mM acetate, 1 mM EDTA) and boiling the solution in a microwave. 3µl of ethidium bromide (EtBr) was added to the solution, which was then poured into a gel box with lane separators and left to set for 30 minutes. The gel was then placed into a gel tank and submerged in TAE buffer.  10 µl of each DNA sample was mixed with 1 µl of DNA loading buffer (10X buffer made up of 0.025 g bromophenol blue, 1.25 ml of 10% SDS, 12.5 ml of glycerol and 6.25 ml of $dH_2O$) and loaded onto into the lanes of the gel, with 6 µl of size markers (Fermentas Generuler 1KB DNA ladder) loaded into the lanes at each end. The gel was then run at 90 V for 45 mins or until the bromophenol blue front had reached the desired distance. The gel was visualised with a CCD camera under UV light.

### 3.2.8 Genomic cosmids

Genomic cosmid for D2085 was obtained from the Wellcome Trust Sanger Institute. The clone arrived as a stab culture and was plated on ampicillin-selective agar plates and stored at 4°C. The bacteria colonies were selected and subjected to Miniprep for the extraction of the cosmid.

### 3.2.9 Restriction digestion of DNA

All restriction enzymes were purchased from New England Biolabs (NEB). Reaction mixtures were made with 0.5 μl enzyme, 2 μl of desired DNA, 1 μl of buffer appropriate for the enzyme (x10 solution), and 6.5 μl of $dH_2O$. The reaction mixture was then incubated in a PCR machine at 37°C for 2 hours. Digested DNA was visualised on an agarose gel.

### 3.2.10 Gold particle bombardment of DNA constructs from the Promoterome

Promoter::GFP fusion DNA constructs from the Promoterome were supplied courtesy of Dr. Jane Shingles from the Hope lab. Promoterome strains for the gene of interest were unfrozen from -80°C and maintained on bacteria agar plates. Plasmids containing the Promoter::GFP fusion were prepared with Miniprep and linearised with restriction digestion as described. A gold particle solution was prepared by mixing 60 mg of gold particles (0.3–3 μm, Chempur, Germany) to 2ml of 70% ethanol, which was then spun briefly and the supernatant discarded; the pellet was washed 3 times with $dH_2O$ and resuspended in 1 ml of 50% glycerol. 30 μl of linearised DNA (approx. 7 μg of DNA) was added dropwise to 70 μl of gold suspension. 300μl of 2.5M $CaCl_2$ and 112 μl 0.1M spermidine were also added dropwise and the solution was centrifuged at 3,000 g for 30 sec and the supernatant discarded. The pellet was resuspended in 800 μl of 70% ethanol and centrifugated again at 3,000 g for 30 sec. The supernatant was again discarded and the pellet was resuspended in 70 μl of 100% ethanol. The DNA- gold particle solution was vortexed regularly to prevent clumping of the gold particles. 10 μl of gold particle solution was spread on microcarriers in the hepta macrocarrier holder of the gold bombardment machine (PDS-1000/He from BioRad). *Unc-119* strain of worms were taken from liquid culture and suspended in a wide test tube under gravity at 4°C and harvested as a pellet at bottom of the tube. 1 ml of worms was distributed evenly over the seven target spots of a 90mm diameter NGM plate. The bombardment procedure from the PDS-1000/He Biolistic was followed and 1 ml of M9 buffer (3 g of $KH_2PO_4$, 6 g of $Na_2HPO_4$, 5 g of NaCl, 1 ml of 1 M $MgSO_4$ in 1 l of $dH_2O$) was added to the worms and rested for 1 hour. 4 ml of M9 was then added to the plates for resuspension and 0.5 ml of the worms was added to eight NGM plates each. Each plate was incubated at 20°C under normal conditions and 8 transformed lines (into wildtype phenotype) were chosen after 3-4 weeks. 4 worms from each plate with a transformed line were transferred to individual 50 mm NGM plates and assessed for stability after 7 days. The line with the highest transmission of GFP was taken and the rest discarded.

### 3.2.11 Promoter::GFP fusion of D2085.6 with GATEWAY recombination

GATEWAY recombination was performed with the Invitrogen GATEWAY Cloning kit. The promoter region for the Promoter::GFP fusion of D2085.6 was chosen 5,155 bases upstream of the start codon of the gene according to sequences from Wormbase. Oligos for the promoter were designed online with Primer3 (http://frodo.wi.mit.edu/primer3/). Gateway recombination site attB4 was fused to 21 bp of the sequence at the 5' end of the promoter the to produce the forward primer (sequence- GGGGACAACTTTGTATAGAAAAGTTGTCGGTAACATCTTTCCAA

TCC) and Gateway recombination site attB1r was fused with 22 bp of the sequence at the 3’ end of the promoter (including the start methionine ATG) to produce the reverse primer (sequence- GGGGACTGCTTTTTTGTACAAACTTGTCATGCATT AAAGTGATTATTGT), which were ordered from Sigma-Genosys. Forward and reverse primers were used in a PCR reaction (Expand High Fidelity PCR system, Roche) with the D2085 cosmid as a template to produce a D2085.6 promoter sequence flanked with attB4 and attB1r sites. The Gateway BP reaction mixture was made using 1.15 μl of D2085.6 promoter PCR product (20 fmol), 0.25 μl pDON_P4_P1r vector (in TE buffer- 10mM Tris-HCl, pH 7.5, 1mM EDTA, from The Andrew Fire vector kit, courtesy of Dr. Sophie Bamps), 2.6 μl of TE buffer and 1 μl of BP Clonase II enzyme mix and incubated at 25°C overnight in a PCR machine (PCR Express, Hybaid). The BP reaction was stopped with the addition of 0.5 μl proteinase K and incubated at 37°C for 10 min and at 95°C for 5 min. BP reaction products were then transformed into *E. coli* DH5α strain cells by the addition of 5 μl of BP reaction to 50 μl of DH5α cells on ice for 30 min, which were then placed in a 42°C waterbath for 90 sec for heat shock. Induced dh5α cells were incubated in 1ml LB media at 37°C for 1 hour, then plated on KAN (kanymycin, 100 μg/ml) selective agar plates and incubated at 37°C overnight. Colonies from KAN plates were subjected to miniprep and digested with restriction enzymes HindIII (cuts twice for 2,350 bp and 5,448 bp fragments) and EcoRV (cuts thrice for 1,103 bp, 2,662 bp and 4,033 bp fragments) for validation. 1 μl of validated BP reaction products was then added to 1.5 μl of destination vector pJS02_469 (linearised with SalI restriction enzyme, contains GFP construct, courtesy of Dr. Sophie Bamps), 5.5 μl of TE buffer, and 2 μl of LR Clonase II reaction mix and incubated at 25°C overnight in the PCR machine, and then stopped with the addition of 0.5 μl proteinase K, incubated at 37°C for 10 min and at 95°C for 5 min. 5 μl of LR reaction products were added to 50 μl dh5α cells on ice for 30 min and heat shocked in a 42°C waterbath for 90 sec for induction, incubated in 1ml LB media at 37°C for 1 hour and then placed on AMP (ampicillin) section agar plates, which was incubated at 37°C overnight. Colonies from AMP plates were miniprepped and digested with restriction enzymes BamHI (cuts twice for 3,629 bp and 7,597 bp fragments) and XbaI (cuts twice for 1,696 bp and 9,530 bp fragments) for validation of the correct product.

### 3.2.12 Injection of worms

Injection of reporter constructs was performed on *C. elegans* N2 hermaphrodites by standard microinjection techniques (Mello *et al.*, 1991). Agarose pads were made by placing a drop of 2.5% agarose (w/v) in between two 22 x 50 mm coverslips for 2 min, taking them apart and leaving the coverslip with agarose to dry overnight. Needles for injection were made from a needle puller (Narishige Scientific Instruments, Japan) with borosilicate microcapillary glass tubes (Clark Electromedical Instruments, UK). D2085.6 promoter::GFP construct was diluted to 20 ng/μl in TE buffer and was mixed with 100 ng/μl plasmid DNA containing the *C. elegans rol-6* gene sequence (pRF4 plasmid in TE buffer, courtesy of Dr. Hannah Craig). The mixed DNA was then loaded into the needle with mouth pipetting from a drawn out glass tube. The needle was mounted onto the injection equipment which consists of an inverted optics microscope (Zeiss Axiovert 10), micromanipulator arm (Narishige Scientific Instruments, Japan) and a $N_2$ cylinder set at 50 Barr pressure, with the tip of the needle broken with abrasion against an agarose pad. Young adult worms were placed onto the agarose pad with a drop of mineral oil (Sigma-Aldrich Co. Ltd., UK) and injected with DNA into the syncytium of the distal arm of the gonad. After injection a drop of M9 buffer was placed on the worms and they were allowed to recover for 20 min before transfer to NGM plates. $F_1$ transformants displaying the dominant *rol-6* phenotype were transferred to fresh NGM plates for propagation and observation of the stability of transmission. After the $F_3$ generation worms still displaying the *rol-6* phenotype were visualised for GFP activity.

### 3.2.13 Visualisation of GFP tagged worms

Worms transformed with promoter::GFP constructs were subjected to visualisation with fluorescence microscopy. *C. elegans* worms were grown on NGM plates for 2-3 days until most of the bacteria food have been consumed and were washed off with 1 ml M9 solution and settled out in an Eppendorf tube for 10 min at 4°C. The worm pellet was distributed on 8 well microscope slides and 0.5 μl of 20 mM levamisole was added to each well. Slides were mounted on a Zeiss Axioplan microscope equipped with DIC optics and visualised through Chroma Technology Corp. filter set 41012. Spatial and temporal expression patterns of GFP were determined for all stages of development. Representative images of the observed expression pattern were collected with Improvision Openlab software on a Photometrics CoolSNAP camera.

## 3.3 Results

### 3.3.1 Homology search of *C. elegans* and *C. briggsae* genes

#### 3.3.1.1 GPI synthesis pathway genes

Genes involved in the synthesis of the GPI anchor in the ER were found with literature search for humans and *S. cerevisiae* (yeast). The human genes chosen for the homology search are listed in Table 3.1, with the *C. elegans* and *C. briggsae* homologues found by BLAST search from Wormbase. Of the 23 genes in the synthesis pathway 16 have homologues within *C. elegans* and *C. briggsae*, with *C. briggsae* also containing an additional 2 homologues that were absent in *C. elegans*. Homologues for most of the GPI synthesis steps are present within both nematodes. Three out of the seven genes involved in the first step of synthesis have no homologues in either nematode species, as well as the interacting partner *PIG-X* in step 5 and *PIG-Z* from step 9, which adds the fourth mannose to the structure. Of note are *PIG-L* and *PIG-F* (GPI anchoring steps 2 and 10/11, respectively) which have homologues within *C. briggsae* but did not have significant hits within the *C. elegans* genome.

| Stage | Human gene | Description | Size (aa) | Yeast gene | Size (aa) | *C. elegans* gene | Blast score | Size (aa) | *C. briggsae* gene | Blast score | Size (aa) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| step 1 | *PIG-A* | Enzymatic part of complex | 484 | *GPI3* | 452 | D2085.6 | 1.30E-112 | 444 | CBG00513 | 4.60E-112 | 393 |
| | *PIG-H* | Binds PIG-A, helps catalysis | 188 | *GPI15* | 229 | n/a | n/a | n/a | n/a | n/a | n/a |
| | *PIG-C* | Scaffolding of complex, bind PIG-Q | 297 | *GPI2* | 280 | T20D3.8 | 2.10E-32 | 282 | CBG21692 | 3.80E-28 | 267 |
| | *PIG-Q* | Stabilise complex | 581 | *GPI1* | 609 | F01G4.5 | 5.50E-30 | 269 | CBG06019 | 1.70E-31 | 248 |
| | *PIG-P* | Interact with PIG-A + Q | 158 | *GPI19* | 140 | Y48E1B.2 | 1.20E-10 | 890 | CBG20762 | 7.70E-11 | 871 |
| | *DPM2* | Regulate DPM1, enhances GlcNAc | 82 | *n/a* | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| | *PIG-Y* | Binds to PIG-A | 114 | *ERI1* | 68 | n/a | n/a | n/a | n/a | n/a | n/a |
| step2 | *PIG-L* | GlcNAc-PI deacetylase | 252 | *GPI12* | 304 | n/a | n/a | n/a | CBG07954 | 8.50E-24 | 147 |
| Step 3 | *PIG-W* | Addition of acyl group to inositol ring | 504 | *GWT1* | 490 | Y110A2AL.12 | 2.00E-33 | 480 | CBG19615 | 3.00E-31 | 827 |
| Step 5 | *PIG-M* | Add 1$^{st}$ mannose to | 423 | *GPI14* | 403 | B0491.1 | 4.90E-79 | 417 | CBG02919 | 5.80E-73 | 394 |
| | *PIG-X* | Interaction partner of PIG-M | 217 | *PBN1* | 416 | n/a | n/a | n/a | n/a | n/a | n/a |
| step 6 | *PIG-N* | Add phoshoethanolamine to 1st mannose | 931 | *MCD4* | 919 | Y54E10BR.1 | 4.80E-134 | 912 | CBG04200 | 4.40E-137 | 920 |
| | | | | | | F28C6.4 | 5.40E-16 | 745 | CBG00550 | 1.10E-09 | 721 |
| | | | | | | | | | CBG01149 | 0.01 | 483 |
| step 7 | *PIG-V* | Add 2$^{nd}$ mannose | 493 | *GPI18* | 433 | T09B4.1 | 4.60E-24 | 672 | CBG12553 | 7.20E-15 | 673 |
| Step 8 | *PIG-B* | Add 3$^{rd}$ mannose | 554 | *GPI10* | 616 | T27F7.3 | 1.00E-71 | 496 | CBG02293 | 1.20E-74 | 495 |
| Step 9 | *PIG-Z* | Add 4$^{th}$ mannose | 579 | *SMP3* | 516 | n/a | n/a | n/a | n/a | n/a | n/a |
| Step 10 | *PIG-G* | Add phoshoethanolamine to 2nd mannose | 975 | *GPI7 (LAS21)* | 830 | F28C6.4 | 2.30E-77 | 745 | CBG00550 | 1.20E-71 | 721 |
| | | | | | | C27A12.9 | 4.50E-39 | 883 | CBG20246 | 2.10E-34 | 453 |
| | | | | | | Y54E10BR.1 | 1.60E-08 | 912 | CBG04200 | 1.10E-08 | 920 |
| Step 11 | *PIG-O* | Add phoshoethanolamine to 3rd mannose | 1089 | *GPI13* | 1017 | C27A12.9 | 1.30E-92 | 883 | CBG20246 | 1.90E-59 | 453 |
| | | | | | | F28C6.4 | 8.30E-31 | 745 | CBG00550 | 1.50E-35 | 721 |
| | | | | | | Y54E10BR.1 | 2.00E-05 | 912 | CBG04200 | 1.30E-05 | 920 |
| Step 10/11 | *PIG-F* | Required for 2nd/3rd mannose modification | 219 | *GPI11* | 219 | n/a | n/a | n/a | CBG05911 | 2.40E-08 | 554 |
| step 12 | *PIG-K* | Transamidase protease | 395 | *GPI8* | 411 | T05E11.6 | 3.90E-86 | 319 | CBG06010 | 2.00E-86 | 319 |
| | | | | | | T28H10.3 | 3.60E-24 | 462 | CBG23516 | 6.10E-28 | 463 |
| | *GPAA1* | May bind free GPI lipid anchor | 621 | *GAA1* | 614 | F33D11.9b | 3.40E-21 | 676 | CBG04019 | 3.10E-16 | 508 |
| | *PIG-T* | May regulate active site of PIG-K | 578 | *GPI16* | 610 | F17C11.7 | 6.60E-40 | 531 | CBG23063 | 3.30E-39 | 531 |
| | *PIG-S* | May be structural | 555 | *GPI17* | 534 | T14G10.7 | 3.30E-15 | 544 | CBG03410 | 1.70E-15 | 695 |
| | | | | | | | | | CBG17621 | 0.0092 | 106 |
| | *PIG-U* | May be involved in substrate recognition | 435 | *GAB1* | 394 | T22C1.3 | 1.80E-33 | 421 | CBG08253 | 1.70E-32 | 419 |
| | | | | | | B0491.1 | 0.00065 | 417 | | | |
| | | | | | | srz-103 | 0.0016 | 326 | | | |

**Table 3.1. Homology search of GPI anchor synthesis pathway genes in *C. elegans* and *C. briggsae*.** All known genes of the GPI anchor synthesis pathway from humans and yeast are presented here with a brief description and their predicted size in amino acids (aa). *C. elegans* and *C. Briggsae* homologues were obtained from BLAST searches against the human pathway genes and are presented with their BLAST scores (significance at $p < 0.05$) and their size in amino acids.

#### 3.3.1.2 Genes involved in Dol-P-Man synthesis

Genes involved in the synthesis in Dol-P-Man, an essential component of GPI anchor synthesis were also searched against the *C. elegans* and *C. briggsae* genomes for homology. Three human genes are involved in this process and of these *DPM1* and *DPM3* have homologues in both nematodes (Table 3.2), with *DPM1* having multiple hits in BLAST. *DPM2* is also a component of step 1 of GPI anchor synthesis, but does not have a homologue in either *C. elegans* or *C. briggsae* (Table 3.1).

| Human gene | Description | Size (aa) | Yeast gene | Size (aa) | *C. elegans* gene | Blast score | Size (aa) | *C. briggsae* gene | Blast score | Size (aa) |
|---|---|---|---|---|---|---|---|---|---|---|
| DPM1 | Catalytic unit for Dol-P-Man synthesis | 260 | *DPM1* | 267 | Y66H1A.2 (*dpm-1*) | 1.10E-81 | 239 | CBG13497 (*Cbr-dpm-1*) | 3.40E-84 | 343 |
| | | | | | H43I07.3 | 4.80E-08 | 339 | CBG01437 | 7.70E-09 | 338 |
| | | | | | gly-8 | 0.00096 | 421 | | | |
| DPM2 | Regulate *DPM1*, enhances GlcNAc | 82 | n/a | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| DPM3 | tethers DPM1 to membrane | 122 | n/a | n/a | F28D1.11 (*dpm-3*) | 9.20E-07 | 95 | CBG03325 (*Crb-dpm-3*) | 3.30E-06 | 95 |

**Table 3.2. Homology search of Dol-P-Man synthesis genes in *C. elegans* and *C. briggsae*.** Known genes within the human and yeast pathways are presented with a description and their size in amino acids (aa). *C. elegans* and *C. briggsae* homologues were obtained with BLAST searches from Wormbase against the human genes. BLAST scores for significant results ($p < 0.05$) and their predicted size in amino acids are presented.

#### 3.3.1.3 Lipid remodelling

The fatty acid chains of GPI anchors are modified within the ER and Golgi apparatus before they are transported to the surface of the cell. Human and yeast differ slightly in the types of modifications they perform to the anchor, most notably at the sn-2 position of the lipid where the human protein PGAP2 replaces the fatty acid with a saturated C18:0 chain, while the yeast protein Gup1p adds a longer saturated C26:0

species to the position, which can be modified further by other genes such as *CWH43*. The human fatty acid remodelling genes *GPAP1*, *GPAP2* and *GPAP3* all have homologues in both *C. elegans* and *C. briggsae* (Table 3.3), with *GPAP2* having multiple significant hits by BLAST search in the two nematodes (5 in *C. elegans* and 4 in *C. briggsae*). The yeast protein Gup1p has a weak homologue in *C. elegans* (hhat-2, $p = 0.026$) which is a putative palmitoyltransferase in the hedgehog signalling pathway (Burglin and Kuwabara, 2006), while no significant homologues were found for *C. briggsae* with BLAST.

| Human gene | Description | Size (aa) | Yeast gene | Size (aa) | *C. elegans* gene | Blast score | Size (aa) | *C. briggsae* gene | Blast score | Size (aa) |
|---|---|---|---|---|---|---|---|---|---|---|
| *PGAP1* | Inositol deacylation | 922 | *BST1* | 1029 | T19B10.8 | 3.00E-25 | 733 | CBG23146 | 6.00E-25 | 1550 |
| *PGAP3* | Removes acyl group on sn-2 position | 320 | *PER1* | 357 | R01B10.4 | 7.00E-25 | 320 | CBG09260 | 6.00E-28 | 326 |
| *PGAP2* | Addition of saturated fatty acid to sn-2 | 315 | *CWH43* | 953 | T04A8.12 (tag-189) | 6.00E-36 | 263 | CBG18005 (Crb-tag-189) | 5.00E-35 | 263 |
| | | | | | Y38F1A.8 | 1.00E-08 | 303 | CBG02772 | 5.00E-09 | 299 |
| | | | | | T23B12.5 | 4.00E-04 | 224 | CBG26903 | 0.005 | 253 |
| | | | | | Y11D7A.9 | 0.010 | 297 | CBG15066 | 0.012 | 297 |
| | | | | | ZK185.4 | 0.015 | 281 | | | |
| n/a | n/a | n/a | *GUP1* | 560 | Y57G11C.17a (hhat-2) | 0.026 | 524 | n/a | n/a | n/a |

**Table 3.3. Homology search of fatty acid modification genes in *C. elegans* and *C. briggsae*.** Known genes within the human and yeast pathways are presented with a description and their size in amino acids (aa). *C. elegans* and *C. briggsae* homologues searched against the human genes with BLAST, with significant scores ($p < 0.05$) and the protein's predicted size (aa) presented.

### 3.3.2 Analysis of *C. elegans PIG-K* homologues

#### 3.3.2.1 Sequence analysis

*PIG-K* is the catalytic part of the GPI transamidase involved in the final stage of GPI anchor attachment. Mutation of *PIG-K* homologues in humans, yeast and *trypanosome brucei* have all shown a phenotype lacking in GPI anchoring, suggesting that the protein is essential for the addition of GPI to proteins (Kang *et al.*, 2002; Meyer *et al.*, 2000; Ohishi *et al.*, 2000). Both *C. elegans* and *C. briggsae* contain two homologues to the PIG-K protein after BLAST search (Table 3.1). T05E11.6 is the highest scoring *C. elegans* homologue followed by T28H10.3, and in *C. briggsae* the CBG06010 gene had the highest BLAST score followed by CBG23516. A CLUSTALX alignment was made for all the *PIG-K* homologues (Figure3.4.a). T05E11.6 and CBG06010 are homologues of each other and have 95.9% sequence identity, while T28H10.3 and CBG23516 are homologous to each other and also have high sequence identity (90.3%) (Figure 3.4.b). PIG-K contains two active site residues His157 and Cys199 and they are both present within all of the homologous sequences (Figure 3.2.a). The T05E11.6 and CBG06010 protein sequences lack the hydrophobic C-terminal domain found in PIG-K, while the T28H10.3 and CBG23516 protein sequences appear to contain the domain.

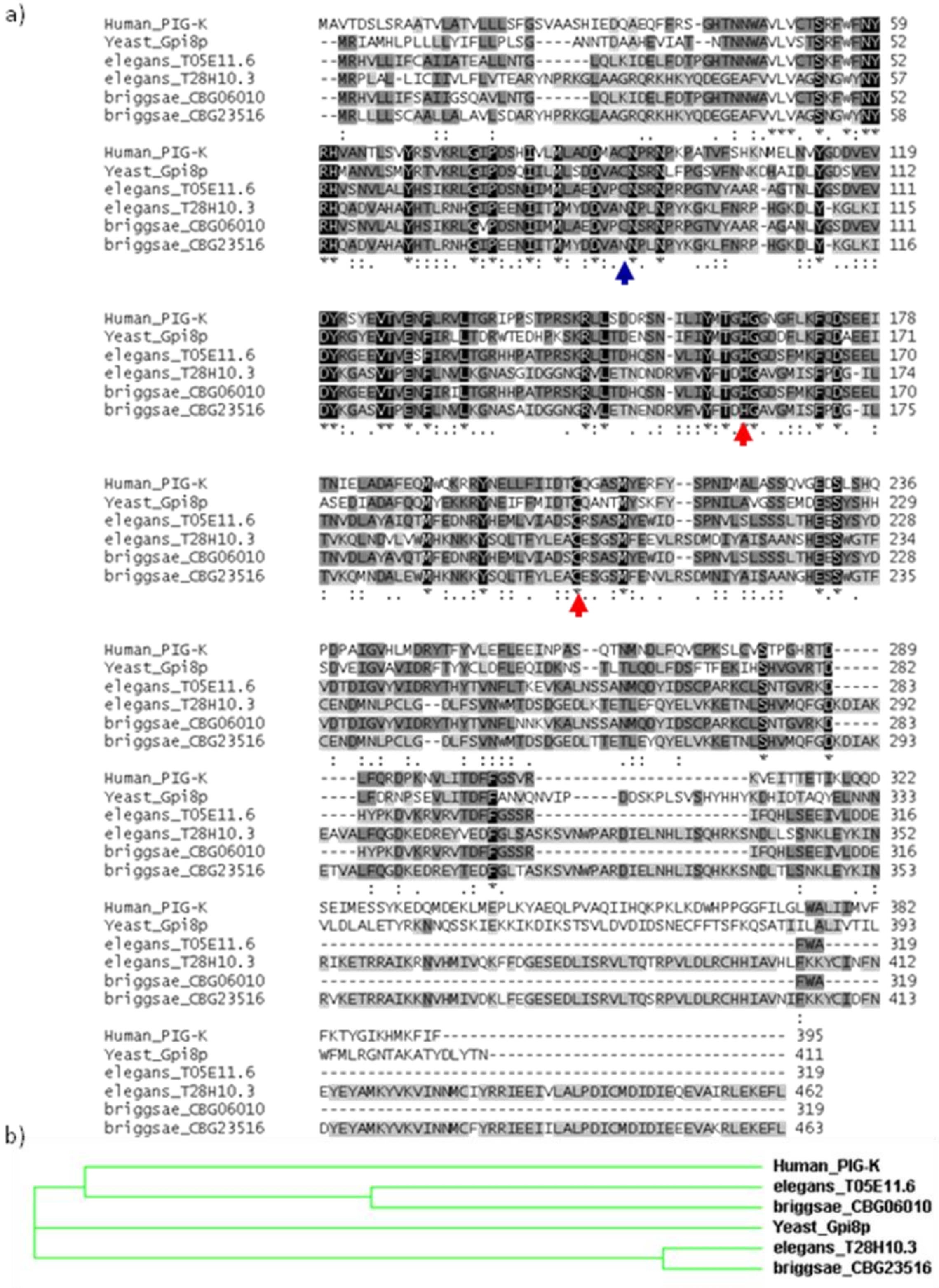



**Figure 3.4. Analysis of the protein sequences of PIG-K homologues.** Sequences for human PIG-K, yeast Gpi8p, and the *C. elegans* and *C. briggsae* homologues were analysed with CLUSTAX version 2.0.12.

a) Multiple sequence alignment of the four protein sequences. Descriptions for the symbols in the graph can be found in Figure 3.8. Red arrows represent the active site residues His157 and Cys199 (PIG-K) and the blue arrow indicates the position where the disulfide bridge forms with PIG-T (Cys92 on PIG-K).

b) Cladogram of the six protein sequences.

### 3.3.2.2 Expression analysis of *C. elegans PIG-K* homologue T28H10.3

#### 3.3.2.2.1 Properties of promoter region

*C. elegans* T28H10.3 was found to be present within the Promoterome, a repository of promoter::GFP fusions for expression analysis available from the Hope Lab (Dupuy *et al.*, 2004). T28H10.3 is present on Chromosome V on the *C. elegans* genome (Figure 3.5.c) between positions 12,512,999 and 12,514,925 and lies within a gene rich area, with eight other gene models present within the surrounding 25 kb region (Figure 3.5.b). T28H10.3 also has 28 EST sequences attributed, suggesting that the gene is highly expressed (Figure 3.5.a).

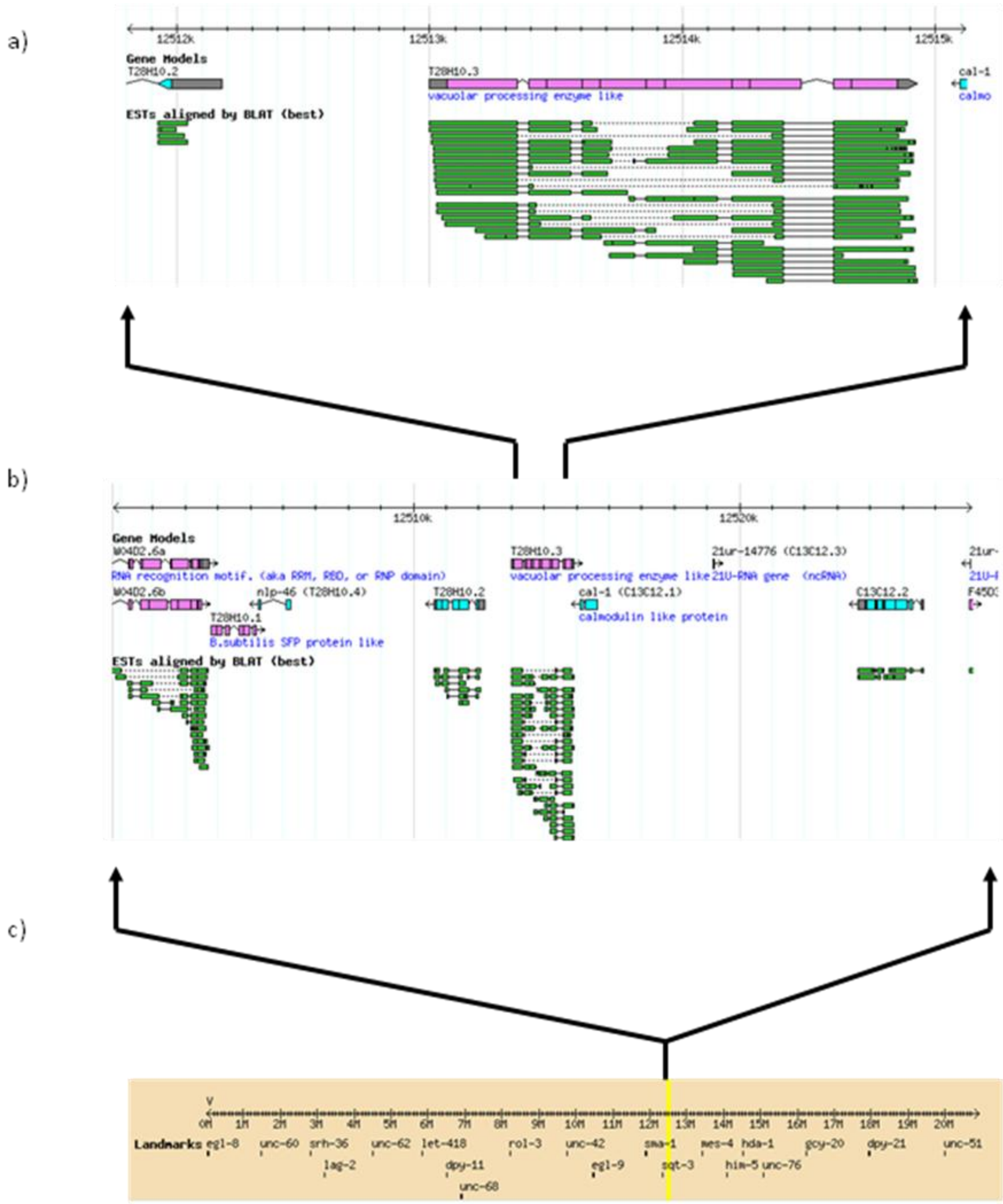



**Figure 3.5. Wormbase display of genomic region around *C. elegans* T28H10.3.** The gene's position along chromosome V, gene model (pink and blue rectangles) and know ESTs aligned by BLAT (green rectangles) are shown. Filled boxes represent the exons of genes in the gene model. The direction of transcription is indicated by arrows at the end of the gene models.

a) the display of region 1 kb upstream and 100 bp downstream of T28H10.3.
b) display of 25kb region around T28H10.3.
c) display of chromosome V. The position of T28H10.3 is indicated by the yellow line.

#### 3.3.2.2.2 The T28H10.3 construct from the Promoterome

The T28H10.3 promoter was present within the Promoterome as a Gateway entry clone with 868 bp of 5' upstream sequence inserted into a pDON_P4-P1r vector (Figure 3.6.a). This vector has a size of 3,515 bp and was tested with restriction enzymes EcoRV (single fragment) and SacI (double fragments of sizes 1,138 bp and 2,377 bp) (Figure 3.6.b). The promoter::GFP construct was made with LR Gateway recombination reaction into the GFP destination vector pDEST-DD04 (Figure 3.6.a). The construct contains an *unc-119* rescue gene which was used as a selective marker by the rescue of *unc-119* worms to wildtype (Figure 3.6.a). The T28H10.3 promoter::GFP was 11,347 bp and was digested with three restriction enzymes to confirm its size, which were with HindIII (single cutter), SacI (double cutter with fragment sizes 1,819 bp and 9,528 bp) and XbaI (triple cutter with fragments of 547 bp which appears as a faint band at the bottom of the gel, 5,081 bp and 5,719 bp in length) (Figure 3.6.c).

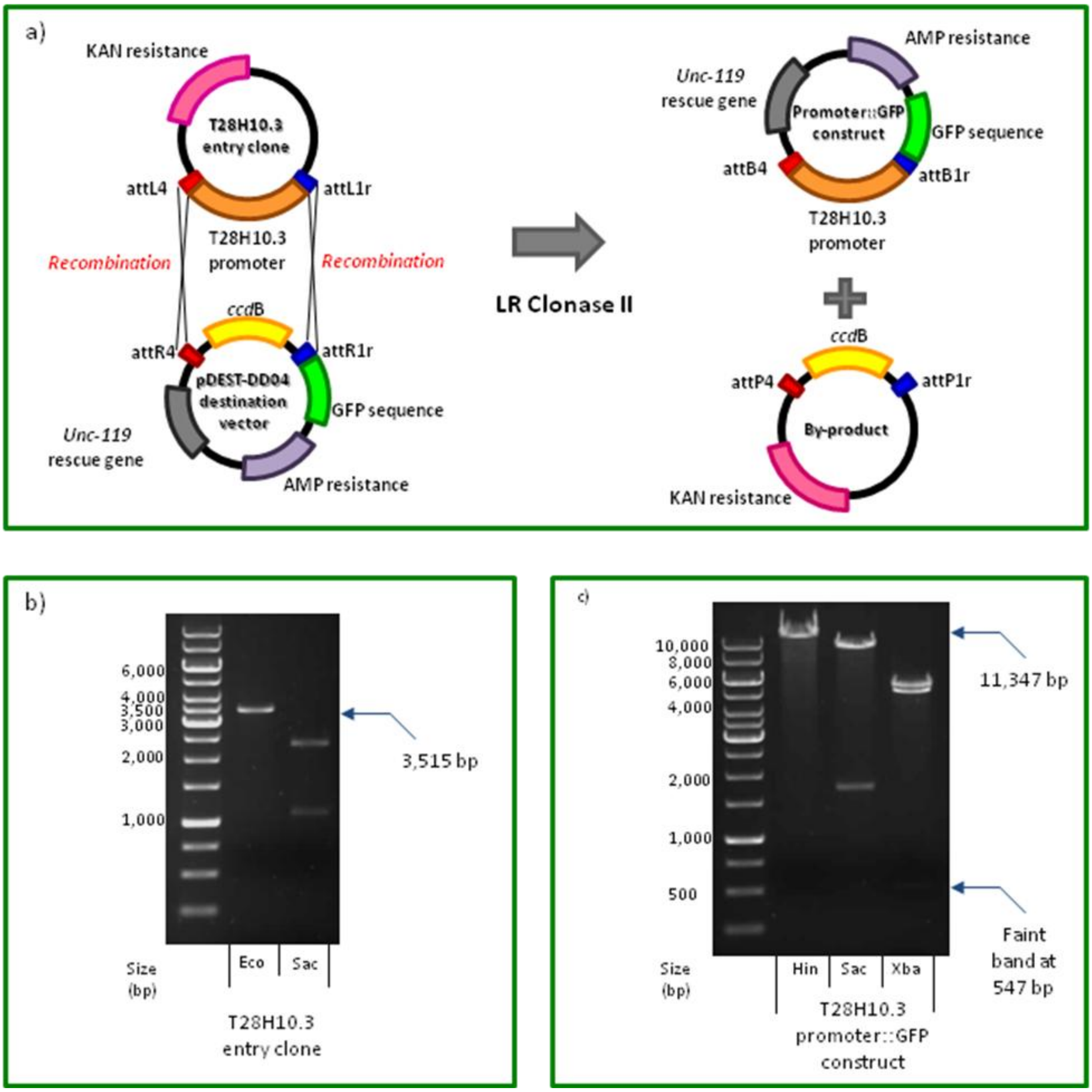



**Figure 3.6. Gateway LR reaction for the T28H10.3 Promoterome entry clone.**

a) The Gateway LR reaction between the T28H10.3 entry clone and the GFP containing pDEST-DD04 destination vector. More details of the LR reaction can be found in figure 3.12.

b) Restriction digests of T28H10.3 entry clone with EcoRV (Eco) and SalI (Sal). EcoRV linearises the plasmid to produce a single fragment of 3515 bp while SalI produces two fragments of 1138 bp and 2377 bp in length.

c) Restriction digests of T28H10.3 promoter::GFP construct. The total size of the construct is 11,347 bp. The restriction enzyme HindIII (Hin) linearises the plasmid, SacI (Sac) which produces two fragments of 1819 bp and 9528 bp, and XbaI (Xba) which produces three fragments of 547 bp, 5081 bp and 5719 bp. The band at 547 bp was present on the gel but was too faint to be photographed.

#### 3.3.2.2.3 Expression pattern of T28H10.3 promoter::GFP construct

The construct was inserted into *unc-119 C. elegans* worms via gold particle bombardment. Transformed worms were analyzed for GFP expression by fluorescence microscopy. *C. elegans* has a complex morphology and contains many tissue types for such a small organism (Figure 3.7.a). The T28H10.3 promoter::GFP construct was shown to be expressed in the intestinal cells of the worm (Figure 3.7.b). The expression started just after the worms reached the comma stage and shows a constantly strong level throughout its various developmental stages. The expression level was especially strong in cells at the ends of the intestinal tract and was ubiquitously strong within the adult intestine. The construct contained a nuclear localization signal as can be seen by the nuclear expression within the L3 worm (Figure 3.7.b).

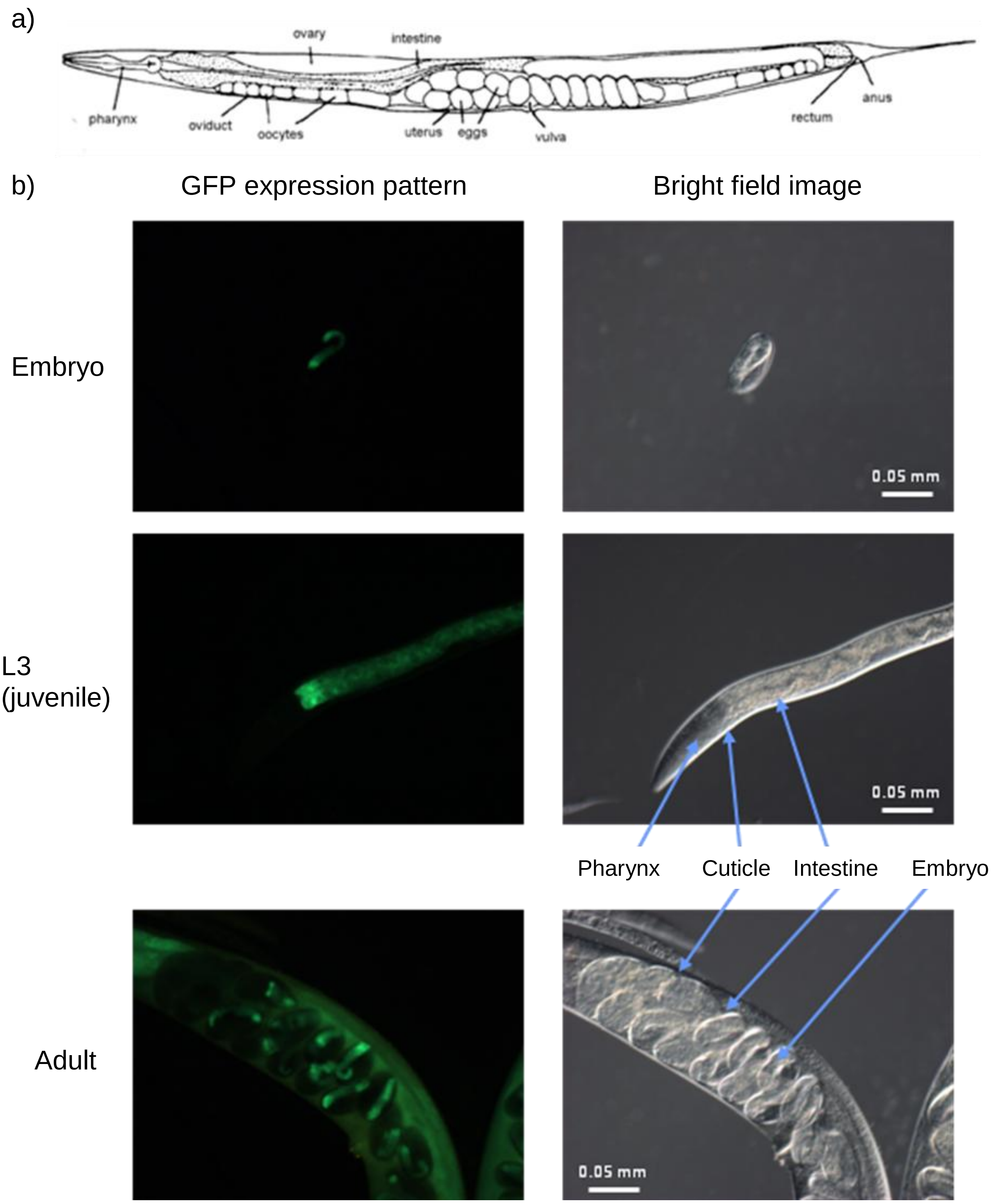



**Figure 3.7. Expression patterns generated with the T05E11.6 promoter::GFP construct.**

a) Diagram of C. *elegans* adult showing positions of major organs including the pharynx, ovary, intestines, and vulva. Adapted from http://avery.rutgers.edu/WSSP/StudentScholars/project/introduction/worms.html.

b) GFP expression pattern of transformed worms in the embryo, L3 and adult stages. The pictures on the left show the GFP expression and a bright field view of the same image are presented on the right. Expression was observed early during development in the intestine and continued throughout all life stages of the worm. Certain anatomical features are highlighted for the L3 and adult worms (blue arrows). Scale bar shows actual length in millimeters.

### 3.3.3 Analysis of *C. elegans PIG-A* homologue

#### 3.3.3.1 Sequence analysis

*PIG-A* is an important part of the enzyme complex involved in the first step of GPI biosynthesis. PIG-A catalyses the reaction between GlcNAc and PI to form GlcNAc-PI. Knockout of *PIG-A* orthologues has been shown to result in the loss of GPI anchoring in a variety of organisms (Alfieri *et al.*, 2003; Shichishima and Noji, 2002; Vossen *et al.*, 1997). *C. elegans* contains one homologue for *PIG-A* with the name D2085.6 (Table 3.1). Protein sequences for PIG-A and its homologues in yeast, *C. elegans* and *C. briggsae* display a large amount of sequence conservation with each other (Figure 3.8.a). The human sequence displays a 25 amino acid overhang at the N-terminus which is not present within the other sequences. *C. briggsae* also lacks a 43 amino acid domain (in between amino acid positions 129 and 172 in the PIG-A sequence) that is highly conserved in the other three sequences. Conservation between the amino acid positions of *C. elegans* and *C. briggsae* is higher than for the other two proteins (Figure 3.8.b) as can be expected from their relatively close evolutionary relationship.

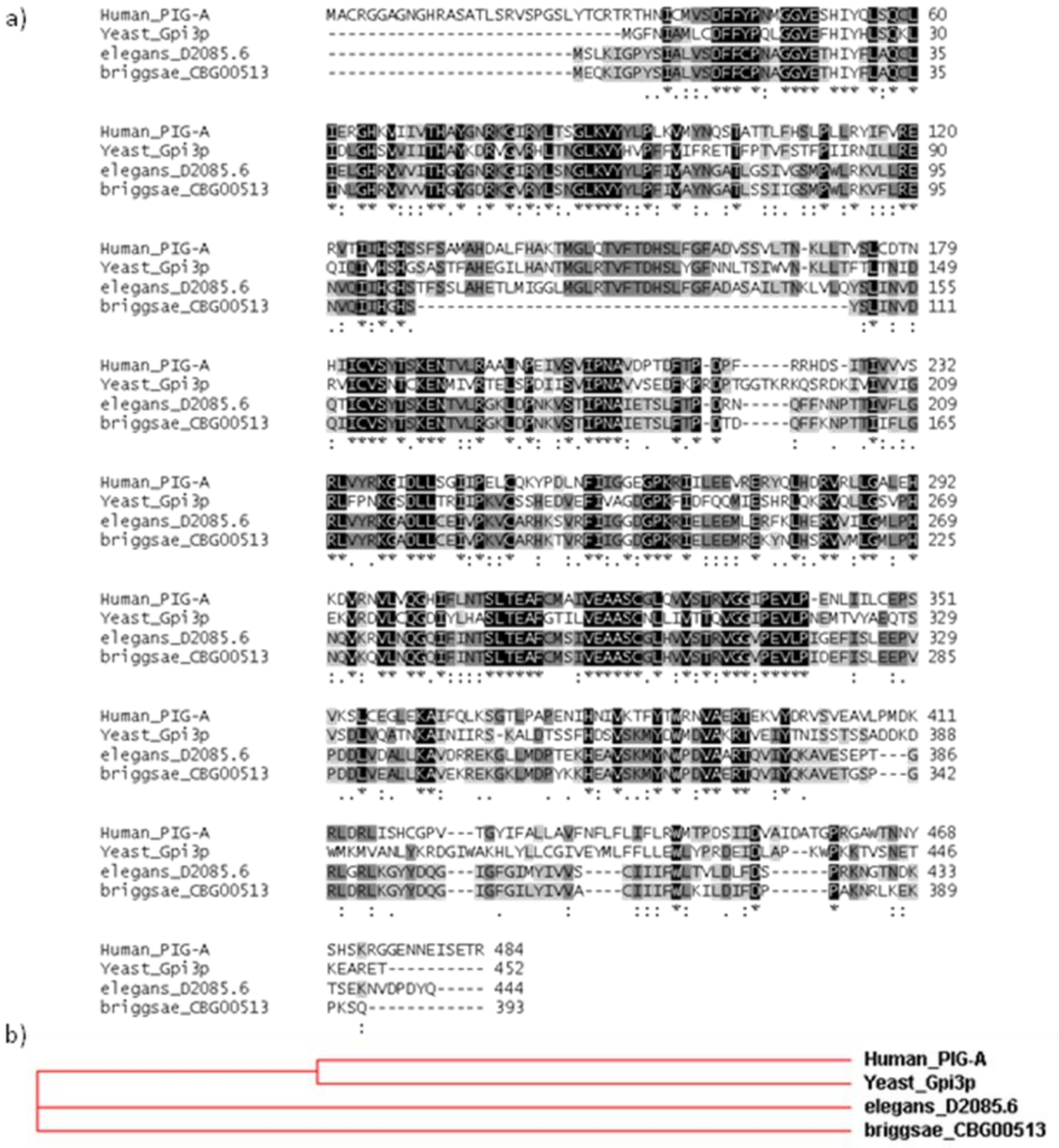



**Figure 3.8. Analysis of the protein sequences of PIG-A homologues.** Sequences for human PIG-A, yeast Gpi3p, and homologues in *C. elegans* and *C. briggsae* were analysed with CLUSTALX version 2.0.12.

a) Multiple sequence alignment of the four protein sequences. Light grey boxes indicate an alignment of two amino acids, darker grey boxes indicate three amino acids alignment, and black boxes indicate total conservation of amino acid sequence at the position. Symbols under the amino acids come from CLUSTALX output, with "." indicating semi-conservative substitution, ":" indicating conservative substitution according to amino acid type, and "*" indicating total conservation of the residue.

b) Cladogram of the four protein sequences.

### 3.3.3.2 Expression analysis of *C. elegans PIG-A* homologue D2085.6

#### 3.3.3.2.1 Selection of promoter region

The *C. elegans* gene D2085.6 was chosen for expression analysis with Gateway homologous recombination. D2085.6 is found near the centre of chromosome II between positions 8,661,644 and 8,659,714 (Figure 3.9.c). The sequence is found within a gene rich region, with five other genes inside a region of 25 kb that does not appear to include very much repetitive sequences (Figure 3.9.b). The sequence 5,152 bp upstream of the start codon was chosen for the production of the promoter::GFP reporter construct (Figure 3.9.a). The finished Gateway product is to be injected into the gonad of worms to induce transformation.

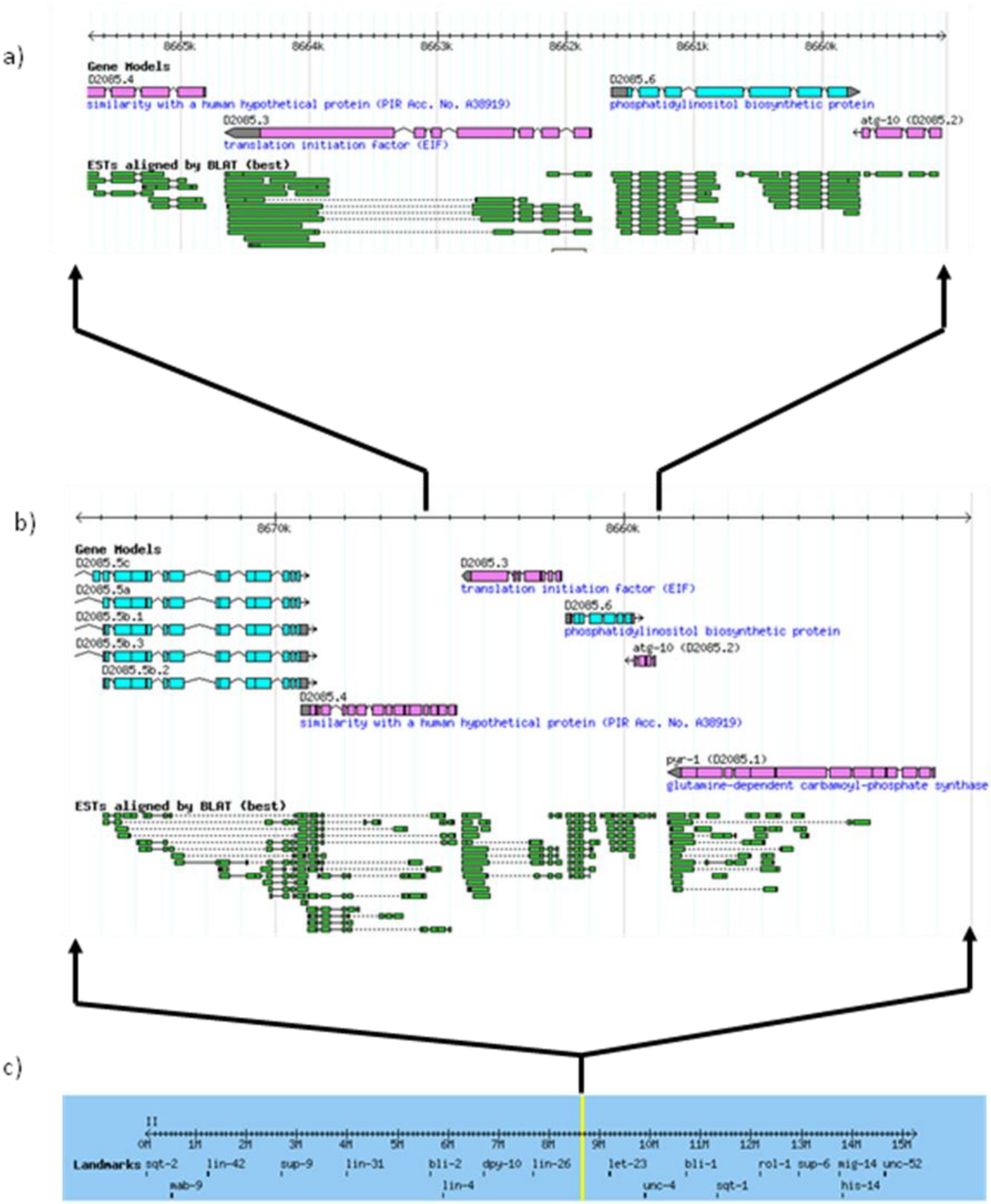


**Figure 3.9. Wormbase display of genomic region around *C. elegans* D2085.6.** Display consists of position along chromosome II, gene model (pink and blue rectangles), and know ESTs aligned by BLAT (green rectangles). Exons of genes are displayed as filled boxes in the gene model. Arrows at the end of gene models indicate direction of transcription.

a) The display of region 6 kb upstream and 1 kb downstream of D2085.6.

b) Display of 25kb region around D2085.6.

c) Display of chromosome II in its entirety. The yellow region indicates the position of D2085.6.

#### 3.3.3.2.2 PCR of attB flanked promoter

Oligonucleotide primers were designed for the promoter with the homologous recombination site attB4 added as an overhang onto the forward primer at the 5’ end of the promoter sequence and an attB1r site on the reverse primer at the promoter’s 3’ end (Figure 3.10.a). The start methionine codon was also inserted into the sequence on the reverse primer for compatibility with the subsequent GFP sequence. The D2085 cosmid (obtained from the Sanger Institute, Hinxton) was used to clone the promoter of D2085.6 via PCR and a 5,206 bp product was produced with attB4 and attB1r sites flanking at the 5’ and 3’ ends, respectively (Figure 3.10.b).

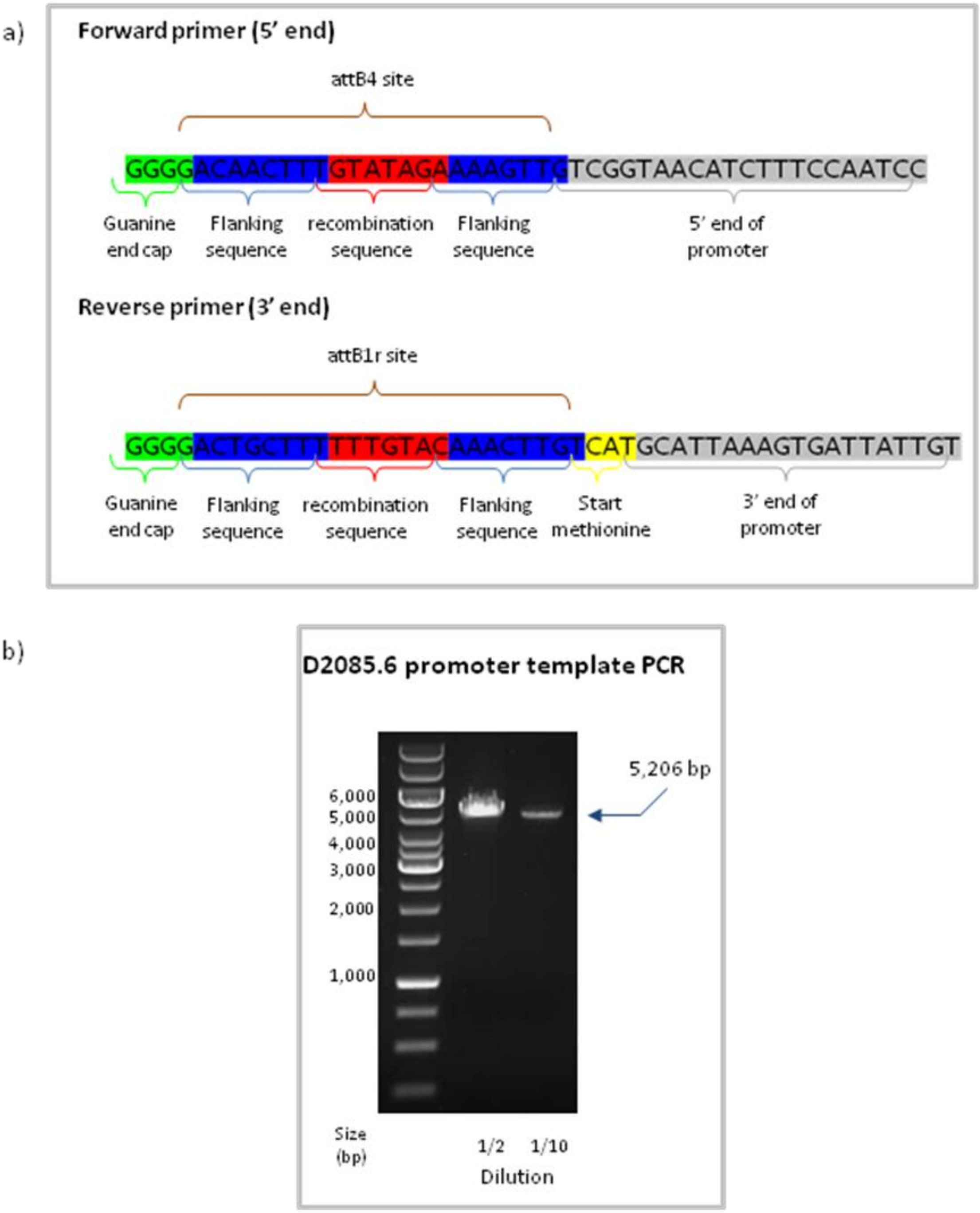


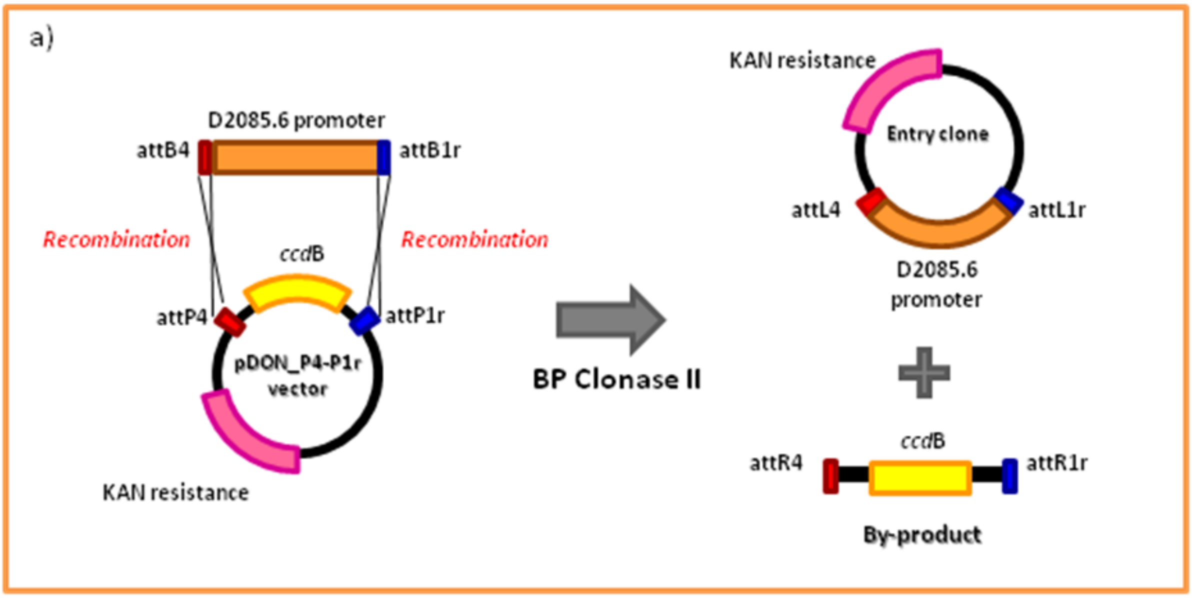

**Figure 3.10. Making of the D2085.6 promoter template for Gateway recombination.**

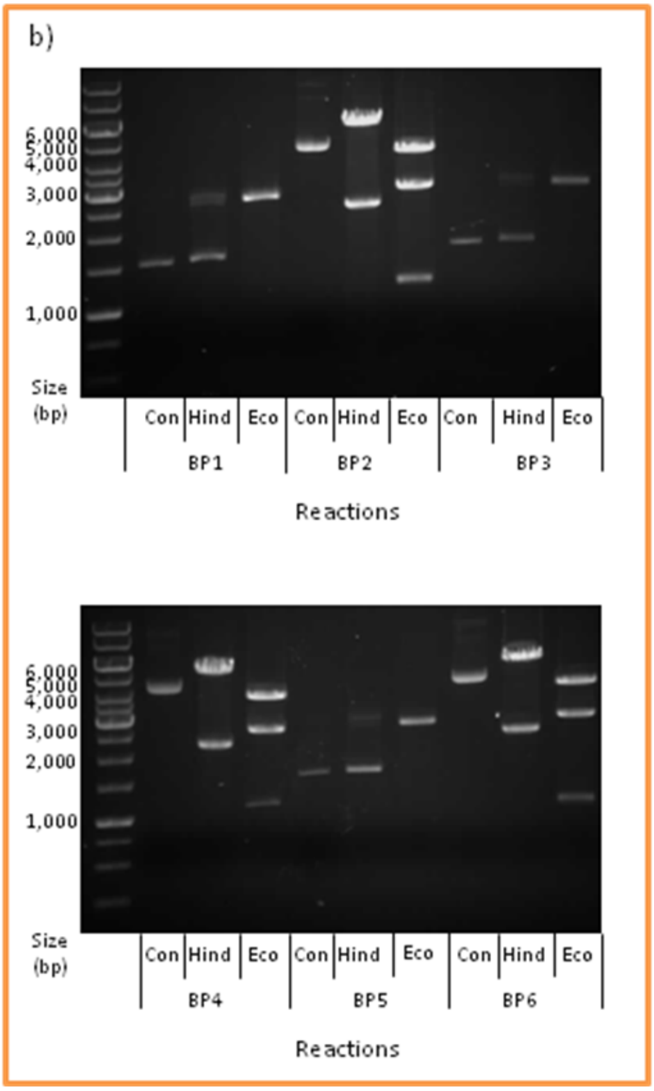
a) Forward and reverse primers of promoter region. 21 bp of sequences at the 5' and 3' ends of the desired promoter region were joined with attB4 and attB1r sites for subsequent BP reaction. The ends of the primers were capped with four guanine residues.

b) Gel purified results of promoter PCR, which shows the sequence at above 5 kb in length. Samples were diluted 1/2 and 1/10 fold before loading. The final concentration of the DNA was approx. 30ng/µl.

#### 3.3.3.2.3 Making of entry clone with BP reaction

The PCR product was then subjected to a BP reaction with the donor vector pDON_P4-P1r to produce an entry clone that contains a kanamycin resistance selection marker (Figure 3.11.a). After selection six colonies were chosen for miniprep (BP 1-6) and digested with restriction enzymes HindIII (cuts twice to give fragments of 2,350 bp and 5,448 bp) and EcoRV (cuts thrice to give fragments of 1,103 bp, 2,662 bp and 4,033 bp). Colonies BP2, BP4 and BP6 produced the expected fragments for each of the enzymes (Figure 3.11.b). Plasmids from BP2 were linearised with the restriction enzyme BstYI which produced a single band that corresponds to the expected length of the entry clone (7,798 bp, Figure 3.11.c). The unlinearised version of the BP2 plasmid was used for the subsequent LR reaction.



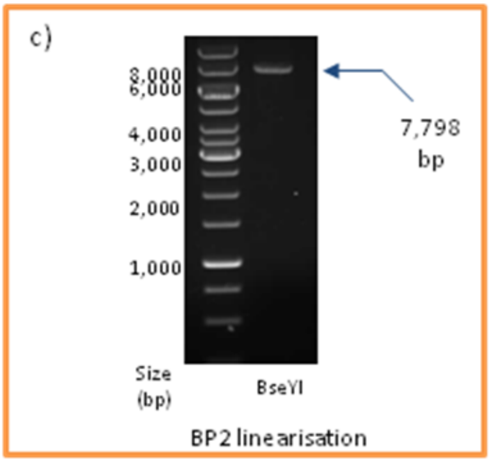




**Figure 3.11. Gateway BP reaction for the D2085.6 promoter.**

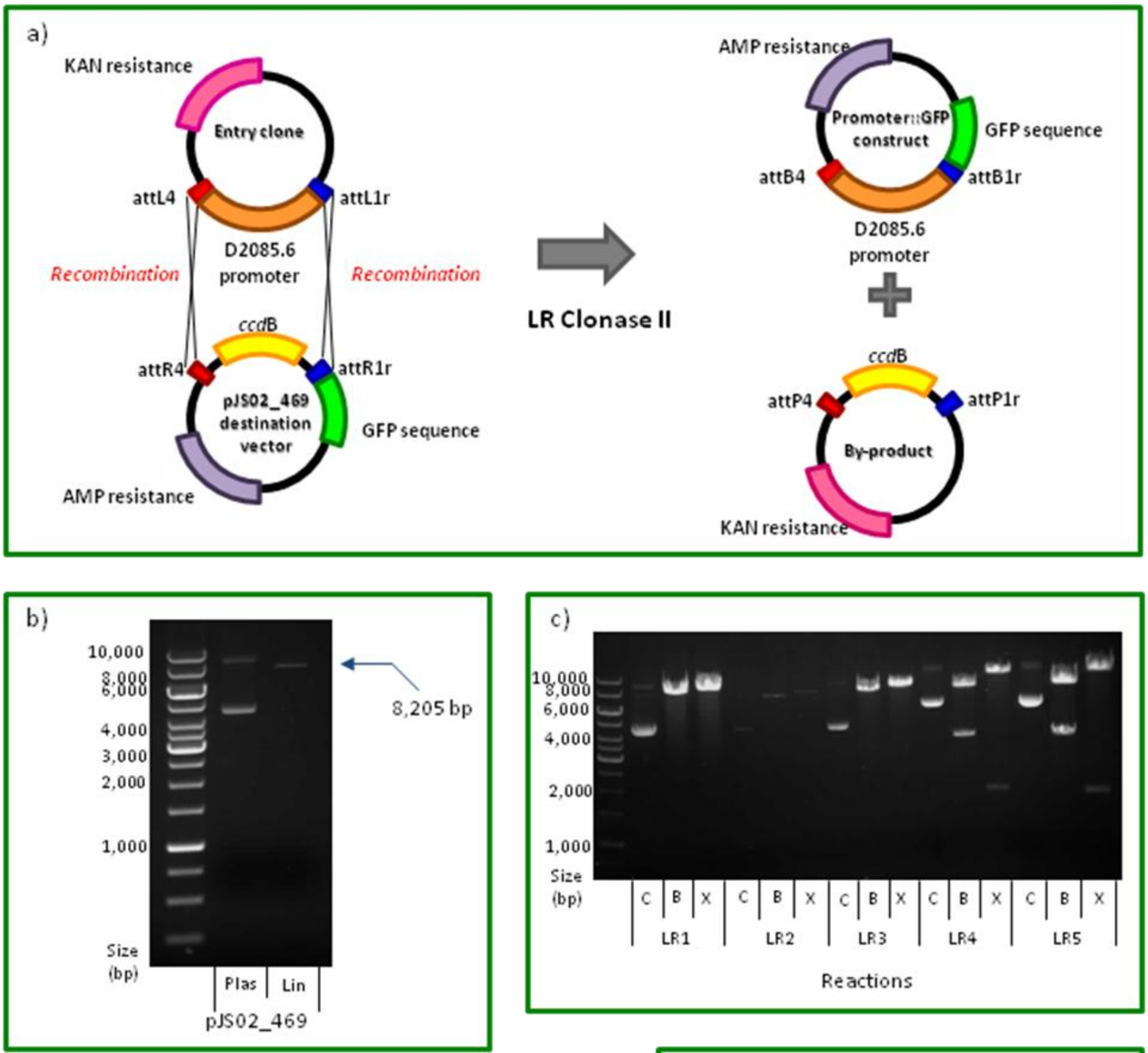
a) The Gateway BP reaction. AttB sites flanking the promoter react with attP sites on the vector and recombination of DNA occurs to produce attL and attR sites at the end of the reaction. The promoter sequence is inserted into the entry clone. The *ccd*B gene is a negative selection marker that causes lethality in *E. coli* and ensures bacteria transformed with the by-product do not survive.

b) Restriction digest of BP transformants named BP 1-6. Con stands for control (unlinearised plasmid), Hind stands for HindIII digestion, and Eco stands for digestion with EcoRV.

c) BP2 product linearised with BseYI. DNA concentration was approx. 30ng/µl.

#### 3.3.3.2.4 Production of Promoter::GFP construct with LR reaction

Entry clone BP2 was subjected to LR reaction with the destination vector pJS02_469 (Figure 3.12.a). pJS02_469 contains an ampicillin selection marker and a GFP sequence in frame with the attB1r site, which is joined in frame with the promoter sequence after the LR reaction to produce the promoter::GFP construct. The destination vector was first linearised with the SalI restriction enzyme to allow greater efficiency during the reaction (Figure 3.12.b). After selection colonies LR 1, 2, 3, 4, and 5 were checked with restriction enzymes BamHI (cuts twice, 3,629 bp and 7,597 bp fragments) and XbaI (cuts twice, 1,696 bp and 9,530 bp fragments) for correct insertion of the promoter (Figure 3.12.c). LR 4 and LR5 showed fragments of the correct sizes. Plasmids from both of the colonies were linearised with SacI which produced the expected size of the product (11,226 bp).



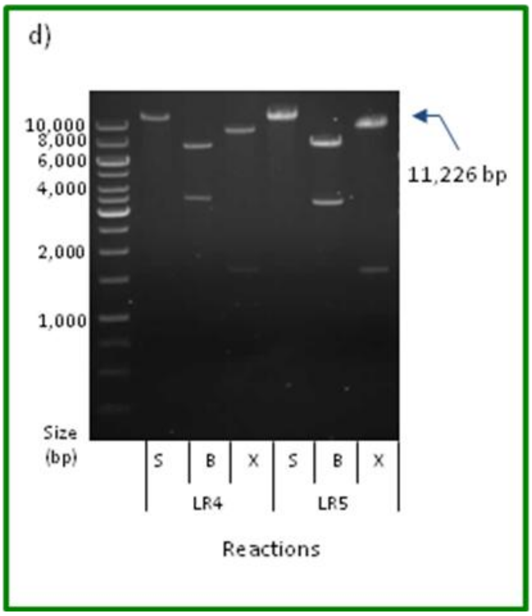




**Figure 3.12. Gateway LR reaction for the D2085.6 promoter.**

a) The Gateway LR reaction. In a reverse of the BP reaction attL and attR sites on the entry and destination vectors react to produce the promoter::GFP construct and a *ccd*B containing by-product. The destination vector contains an ampicillin resistance gene which is used for selection. The GFP sequence joins in frame to the promoter sequence at the end of the reaction.

b) Restriction digestion of pJS02_469 vector with SalI for linearisation. Plas indicate the non-linearised version and Lin indicate the linearised vector.

c) Restriction digests of transformed colonies LR 1-5. C stands for undigested control, B stands for digestion with BamHI and X stands for XbaI digestion.

d) Restriction digestion of LR4 and LR5. S stands for SacI which linearises the sequence. LR4 and LR5 are approximately 60ng/µl and 140ng/µl, respectively.

## 3.4 Discussion

Known human and yeast GPI biosynthesis genes were used in a bioinformatic search to find their homologues in *C. elegans* and *C. briggsae*. Of the 23 genes found in the human pathway 16 of them have *C. elegans* homologues, while *C. briggsae* contains an additional 2 more homologues in the pathway genes (Table 3.1). Other important components of GPI anchoring, such as Dol-P-Man synthesis (Table 3.2) and lipid remodelling (Table 3.3), also have homologues in both of the worms. An account of the nematode genes involved in the various steps of GPI biosynthesis is given below.

### 3.4.1 GPI biosynthesis genes

#### 3.4.1.1 Step 1

Seven human genes have been found so far for the first step of GPI biosynthesis, where they have been postulated to form a complex for their catalytic activity (Tiede *et al.*, 2000). Both *C. elegans* and *C. briggsae* contain homologues for four of the genes involved in this process. The *PIG-A* gene is the catalytic subunit of the complex and is one of the four genes that have a homologue in both nematodes. *PIG-C*, *PIG-Q* and *PIG-P* also have homologues within the nematodes and are important for the activity of the enzymatic complex in humans and yeast (Leidich *et al.*, 1995; Newman *et al.*, 2005; Tiede *et al.*, 2001). The PIG-H protein, which is postulated to form a complex with PIG-A, does not have homologues in the nematodes (Watanabe *et al.*, 1996). Homologues of PIG-Y were also absent from both nematodes; this relatively small protein interacts with PIG-A and the Ras pathway in yeast and may act as a regulator of GPI biosynthesis (Sobering *et al.*, 2004). PIG-Y has also been shown to be important for human PIG-A function but a complex can still be formed in its absence (Murakami *et al.*, 2005). PIG-Y appears to regulate the function of PIG-A, and this mode of regulation may be absent in both *C. elegans* and *C. briggsae*. Lastly *DMP2* is involved in GPI biosynthesis in humans and is absent in yeast, where it affects the rate of reaction of the first step of GPI anchor synthesis (Watanabe *et al.*, 2000b). *DMP2* is also involved in Dol-P-Man synthesis in humans and may act in a regulatory role to coordinate between the two biosynthetic processes. Both *C. elegans* and *C. briggsae* lack a homologue for *DMP2*, suggesting that the Dol-P-Man synthesis pathway is not involved in the regulation of GPI biosynthesis in both of the nematodes.

#### 3.4.1.2 Step 2

*PIG-L/GPI12* is the human/yeast gene responsible for deacetylation of the GLcNAc-PI in the second step of GPI biosynthesis. This step was shown to be crucial for GPI synthesis in a number of organisms including *Trypanosoma brucei*, yeast, and mammals (Urbaniak *et al.*, 2005; Watanabe *et al.*, 1999), but interestingly has no homologue in *C. elegans*. *C. briggsae* however was shown to contain a homologue to *PIG-L* called CBG07954, which also does not have a homologue in *C. elegans*, suggesting that the gene has been lost during the evolution of *C. elegans*. *PIG-L* is a zinc metalloenzyme (Urbaniak *et al.*, 2005), and its role in GPI anchor synthesis in *C. elegans* may have been taken up by an unrelated deacetylase.

#### 3.4.1.3 Step 3

*PIG-W* is the human gene responsible for the addition of an acyl group onto the inositol ring in the third step of GPI biosynthesis (Murakami *et al.*, 2003). Both the human and the yeast homologue have been shown to cause defective GPI anchoring and affect the maturation of GPI anchored proteins (Umemura *et al.*, 2003), and acylation is also a common feature in *Trypanosoma brucei* (Ferguson, 1999). Both *C. elegans* and *C. briggsae* have homologues for *PIG-W*, suggesting that inositol acylation might also be an important step in GPI anchor addition of both of these nematodes.

#### 3.4.1.4 Localisation to the luminal side of the ER and addition of mannoses to the GPI anchor: Steps 4, 5, 7, 8, and 9

Step 4 of GPI biosynthesis is carried out by a flippase which is still uncharacterised in human and yeast. Step 5 of GPI biosynthesis occurs within the lumen of the ER and involves the addition of the first mannose subunit, which is catalysed by *PIG-M* in humans (Maeda *et al.*, 2001b). Both *C. elegans* and *C. briggsae* contain one homologue for the gene. PIG-X/Pbn1p in human/yeast interacts with PIG-M/Gpi14p and acts to stabilise the protein in the ER via its chaperone-like activity (Ashida *et al.*, 2005a; Subramanian *et al.*, 2006). This gene however does not have a homologue in *C. elegans* or *C. briggsae*. It may be that the nematode *PIG-M* homologues do not require stabilisation for their function; alternatively an unrelated chaperone protein may stabilise the homologues within the ER of the nematodes.

Steps 7 and 8 in GPI biosynthesis involve the addition of the second and third mannoses to the GPI structure. The genes responsible for both of these steps in human/yeast are *PIG-V/GPI18* and *PIG-B/GPI10*, respectively. *C. elegans* and *C. briggsae* have homologues for both of these mannosylation genes. The three core mannose subunits are essential in GPI biosynthesis and is a common feature of all GPI anchors found so far (Ikezawa, 2002).

In human and yeast, a fourth mannose is sometimes added to the GPI structure via *PIG-Z/SMP3* in step 9 of GPI biosynthesis. This modification is not required in human cells but is essential for anchoring of proteins in yeast (Grimme *et al.*, 2001). This modification in humans appears to be tissue specific, and GPI anchors with three or four mannose subunits have been observed (Taron *et al.*, 2004b). Both of the nematode species analysed here do not contain a homologue for this process, suggesting that the addition of the fourth mannose does not occur within *C. elegans* and *C. briggsae* and that this may be a species specific modification.

#### 3.4.1.5 Addition of phosphoethanolamine to mannoses: steps 6, 10, and 11

In both humans and yeast, phosphoethanolamine is added to the three core mannose subunits via the genes *PIG-N/MCD4*, *PIG-G/GPI7* and *PIG-O/GPI13* in steps 6, 10 and 11, respectively (Benachour *et al.*, 1999; Hong *et al.*, 2000; Hong *et al.*, 1999a). Both *C. elegans* and *C. briggsae* have homologues for each of these genes, with the best result for *PIG-N* (Y54E10BR.1/CBG04200), *PIG-G* (F28C6.4/CBG00550) and *GPI-O* (C27A12.9/CBG20246) in *C. elegans*/*C. briggsae*, respectively. Interestingly the nematodes homologues for each individual gene are also homologues for the others, with Y54E10BR.1 found to be also homologous to *PIG-G* and *PIG-O*, F28C6.4 also homologous to *PIG-N* and *PIG-O*, and C27A12.9 also homologous to *PIG-G* (Table 3.1). A sequence comparison of the *C. elegans* genes with ClustalW shows conserved motifs within the three predicted proteins but otherwise poor conservation for the rest of their sequences (Figure 3.13); the conserved nematode motifs corresponds to similar motifs on the three human genes, which may represent sites of important biological function, such as ligand binding sites, for this class of enzymes. Further analysis will be needed to elucidate exactly which of the homologues in *C. elegans* and *C. briggsae* are responsible for each of the phosphoethanolamine addition reactions. The addition of the first phosphoethanolamine is important for GPI anchor synthesis in both human and yeast (Vainauskas and Menon, 2006; Zhu *et al.*, 2006) while the addition of the third phosphoethanolamine is essential as the protein is attached to the anchor via this moiety (Hong *et al.*, 2000). Addition of the second phosphoethanolamine however is only important in yeast (Fujita *et al.*, 2004), whereas in humans the modification is needed for just a subset of GPI anchors (Shishioh *et al.*, 2005). It will be interesting to see how important the presence of this moiety on each mannose subunit is in both *C. elegans* and *C. briggsae*, and elucidate their influence on different tissue types in development and other physiological processes.

PIG-F/Gpi11p in human and yeast interact with PIG-G/Gpi7p and PIG-O/Gpi13p in the addition of the second and third mannoses in GPI anchor biosynthesis. PIG-F is an essential interaction partner of PIG-O in humans (Hong *et al.*, 2000), however defects in Gpi11p in yeast was shown not to be a requirement for this step (Taron *et al.*, 2000). *C. elegans* does not contain a homologue for this gene, while *C. briggsae* has a homologue to *PIG-F* but the gene has a predicted size of more than double its human counterpart (Table 3.1). The difference between the two nematodes posses interesting questions from an evolutionary perspective. It may be that the gene is ancestral and has been lost in the *C. elegans* lineage and *not C. briggsae*. Alternatively the gene may have taken on different roles in the two nematodes, with the *C. briggsae* version still possibly retaining some of its original function in GPI anchor synthesis. It will also be interesting to investigate the properties of the *C. elegans* PIG-O homologue compared to the human protein to elucidate the mechanism with which PIG-F acts in PIG-O regulation.

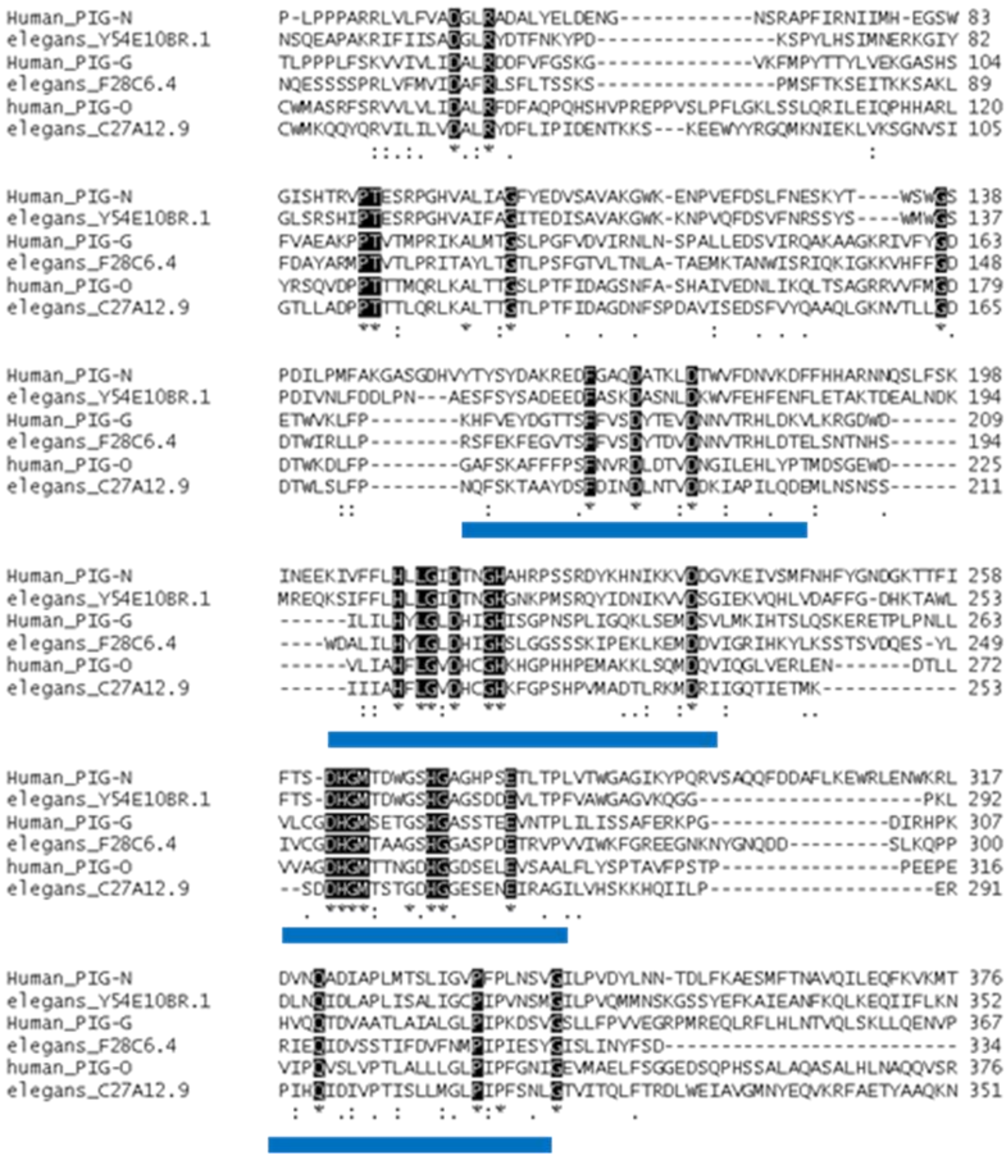


**Figure 3.13. ClustalW analysis of the three human phosphoethanoamine addition proteins and their *C. elegans* homologues.** The symbols in the graph are described in Figure 3.8. Blue bars indicate areas of high homology between all sequences, which may indicate areas of important functions. Only the partial sequences of the proteins with conservation between all of the genes are presented here.

### 3.4.1.6 Step 12

The last step in GPI biosynthesis involves the attachment of the protein to the anchor via the GPI transamidase (GPIT) complex. Each of the five subunits that make up the GPIT in human and yeast have homologues in both *C. elegans* and *C. briggsae*. PIG-K/Gpi8p, GPAA1/Gaa1p and PIG-T/Gpi16p are postulated to form the core structure of GPIT with PIG-K as the catalytic subunit, with *GPAA1* important for substrate recognition and *PIG-T* having a role in conferring specificity for the enzyme (Eisenhaber *et al.*, 2003b; Fraering *et al.*, 2001; Kang *et al.*, 2002; Vainauskas and Menon, 2004a). *PIG-S/GPI17* and *PIG-U/GAB1* are postulated to be responsible for structural and substrate recognition (Ohishi *et al.*, 2001). In yeast Gpi17p and Gab1p form a complex with each other and appear to associate transiently with the rest of the GPIT complex, suggesting that the whole complex functions as two different subunits (Grimme *et al.*, 2004; Zhu *et al.*, 2005). It will be interesting to see if the nematode homologues also form these complexes, and elucidate their mode of regulation with regards to different tissue types and developmental stages.

One of the most extensively characterised genes for the last step of GPI biosynthesis is *PIG-K*, the catalytic component of the GPIT complex. PIG-K is a cysteine protease and plays a crucial role in GPI biosynthesis (Spurway *et al.*, 2001). Both *C. elegans* and *C. briggsae* contain two homologues to this protein. The *C. elegans* homologues are T05E11.6 and T28H10.3; the T05E11.6 protein has a higher homology BLAST score (Table 3.1). Both of these proteins contain the two conserved residues, His157 and Cys199, within the PIG-K active site that are necessary for the enzymatic function of the protein (Figure 3.4.a) (Meyer *et al.*, 2000). PIG-K also contains a transmembrane domain at the C-terminus of the protein, which is believed to anchor the protein to the ER membrane. T28H10.3 contains a hydrophobic region at the C-terminus whereas T05E11.6 does not, however it has also been observed that the absence of the transmembrane domain in PIG-K does not impact on its activity *in vivo* (Ohishi *et al.*, 2000). PIG-K forms an intermolecular disulphide bridge with PIG-T in the GPIT complex at Cys92 which is important but not essential for full transamidase activity (Ohishi *et al.*, 2003); interestingly this residue is conserved in T05E11.6 but is absent in T28H10.3, where it is replaced by an asparagine; this also raises the possibility that the *C. elegans* PIG-K and PIG-T homologues form a part of the complex similar to the human proteins. Information from Wormbase reports only one partial EST assigned to T05E11.6 while T28H10.3 appears to be highly transcribed with 28 full length and partial ESTs attributed to it (Figure 3.14). Both of these genes have deletion mutants that generate sterile and lethal phenotypes, suggesting that they carry out essential processes within the worm. Both of the PIG-K homologues could potentially be a part of the GPI anchor synthesis pathway within *C. elegans*. An interesting possibility may be that the two genes are expressed in different temporal and spatial patterns, and that both proteins are needed for GPI anchoring during different stages of *C. elegans* development. An expression pattern has been generated for T28H10.3 from the Promoterome (Dupuy *et al.*, 2004) which will be discussed in detail below.

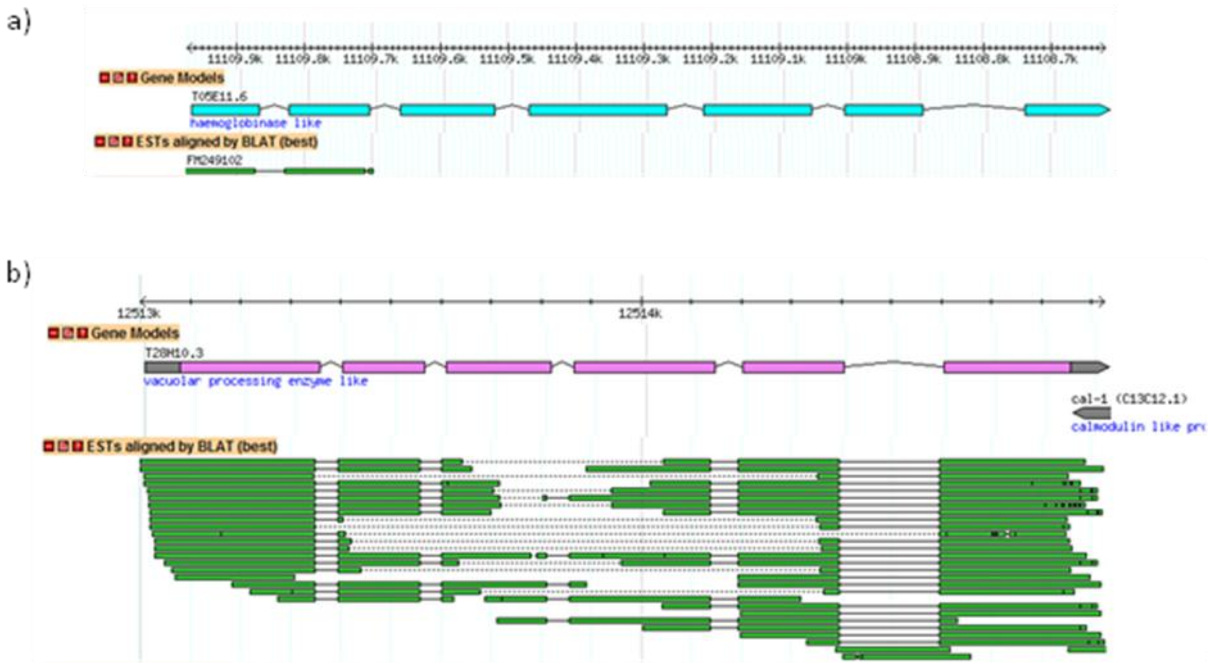



**Figure 3.14. Wormbase gene model for the *C. elegans* PIG-K homologues.** (a) shows the gene model for T05E11.6 and (b) shows the gene model for T28H10.3. The curated exons for each gene are shown as blue rectangles for T05E11.6 and pink rectangles for T28H10.3. ESTs attributed to the gene are displayed as green rectangles under the gene models. T05E11.6 contains one partial EST while T28H10.3 contains 28 full length and partial ESTs.

### 3.4.2 The Dol-P-Man synthesis genes

Dol-P-Man is an important mannose donor within the cell and is required for GPI anchor biosynthesis (Orlean, 1990b). Three genes are involved in Dol-P-Man production in humans with only one involved in yeast. The protein for the yeast *DPM1* gene contains a TM domain at its C-terminus that tethers the protein onto the ER membrane, while the human DPM1 lacks this domain and is instead stabilised by DPM3 to the ER, where DPM3 also prevents degradation of DPM1 (Ashida *et al.*, 2006). DPM2 in humans has two functions within the complex, the first for stabilisation of DPM1 and the second as a component of the first step of GPI biosynthesis, suggesting that there is a regulatory link between Dol-P-Man synthesis and the GPI anchor biosynthesis pathway in humans (Maeda *et al.*, 1998a; Watanabe *et al.*, 2000b). Both *C. elegans* and *C. briggsae* have homologues for *DPM1* and *DPM3* but lack homologues for *DPM2* (Table 3.2). Comparison of the human DPM1 and yeast Dpm1p sequences with *C. elegans* DPM-1 (Y66H1A.2) and *C. briggsae* CBR-DPM-1 (CBG13497) shows that *C. elegans* DPM-1 lacks the C-terminal TM domain similar to the human protein, while *C. briggsae* contains an extended C-terminal sequence that was predicted not to be a TM domain by the program TMHMM (Chen *et al.*, 2003b), and may in fact be a part of a different gene following other gene models (figure 3.15c). The human DPM1 sequence also has higher scores of homology to both of the nematode protein sequences than to Dmp1p in yeast. These together suggest that the synthesis of Dol-P-Man has greater similarity between nematodes and human than with yeast. The absence of *DPM2* homologues in both of the nematode species, however, suggests that there is no direct regulatory link between Dol-P-Man synthesis and the GPI anchor synthesis pathway, which is more similar to yeast. Taken together, it appears that the mechanisms of the Dol-P-Man synthesis pathway in nematode sits evolutionarily between that of the human and the yeast, with the human mechanism evolved to have a greater role within GPI anchor synthesis. Further evidence from genetic and expression analysis will be needed to test this hypothesis, as well as elucidate the role of Dol-P-Man synthesis pathway components in *C. elegans* and *C. briggsae*.

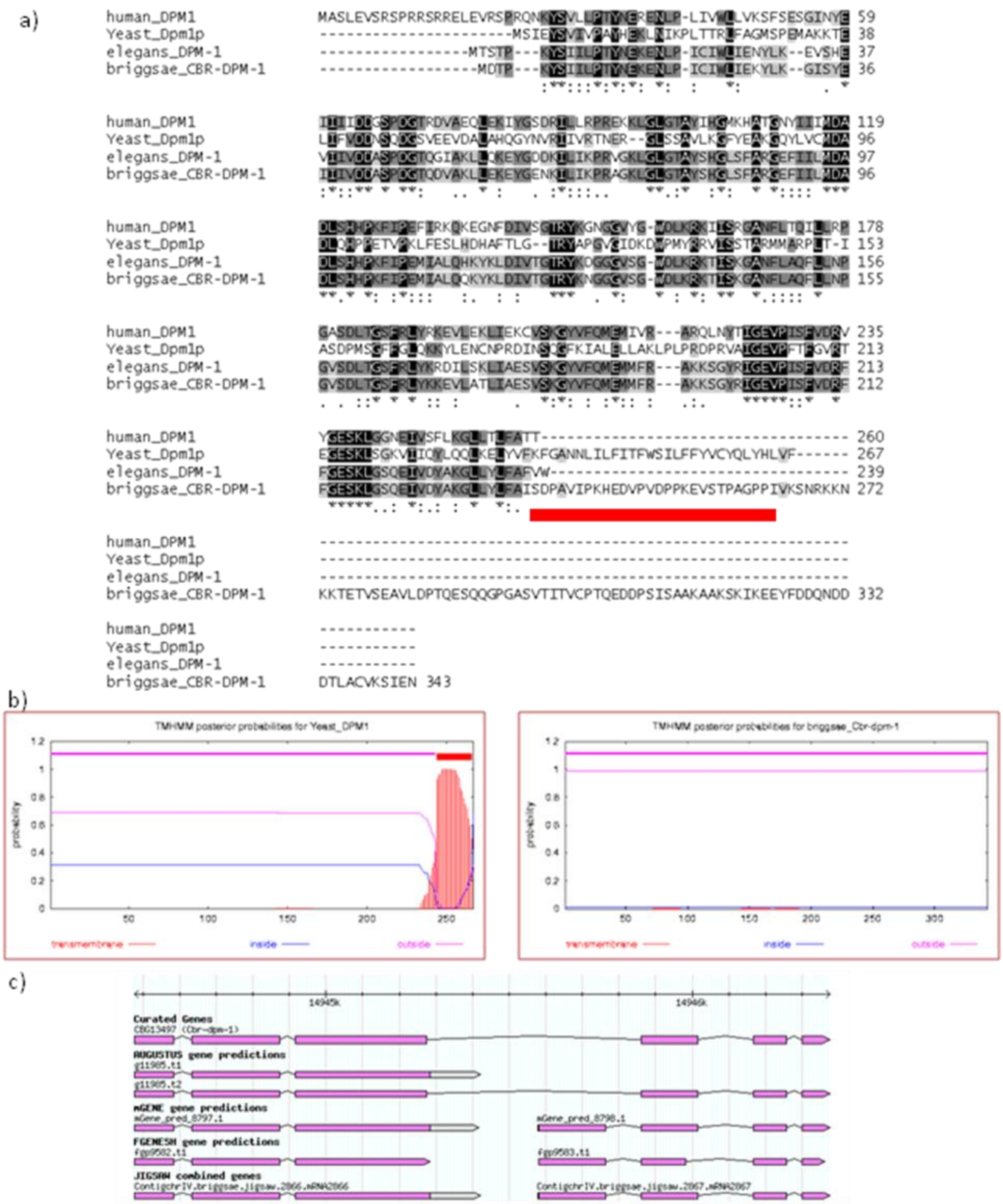



**Figure 3.15.** Analysis of the protein sequences of DPM1 homologues.

a) ClustalW alignment of human DPM1, yeast Dpm1p, *C. elegans* DPM-1 and *C. briggsae* CBR-DPM-1. The red bar indicates the position of the hydrophobic TM sequence in yeast Dpm1p.

b) TMHMM prediction results for yeast Dpm1p and *C. briggsae* CBR-DPM-1. The yeast sequence contains a prediction for TM domain at the C terminus while the *C. briggsae* sequence does not.

c) Wormbase gene view for *C. briggsae cbr-dpm-1*. At least four splice site prediction programs do not include the C-terminal region of the curated gene. Three of those programs place that region as a part of a different gene.

### 3.4.3 Lipid remodelling

Remodelling of the lipid portion of the GPI anchor occurs after the attachment of protein in human and yeast and is essential for its transport to the plasma membrane in both of these organisms (Maeda *et al.*, 2007). The anchor is first modified in the ER with removal of the acyl group on the inositol moiety with the deacetalyase *PGAP1/BST1* (Tanaka *et al.*, 2004). Both *C. elegans* and *C. briggsae* contain a homologue for this protein (Table 3.3), implying that the GPI anchored proteins expressed on the cell surface of these nematodes is also deacetylated. This has implications for the analysis of GPI anchored proteins within the worms with the commonly used enzyme phosphatidylinositol-specific phospholipase C (PIPLC), as this enzyme is only active against deacetylated versions of the GPI anchor (Roberts *et al.*, 1988).

GPI anchor fatty acid chains are modified in the Golgi before the protein is targeted to the surface of the cell. The remodelling process involves replacement of the relatively short and unsaturated lipid tail with a longer and saturated one, which is thought to allow greater packing of the GPI anchor with other saturated lipids in the plasma membrane that is essential for their incorporation into lipid rafts (Maeda *et al.*, 2007). The first step of remodelling involves the removal of the lipid tail at the sn-2 position and is carried out by *PGAP3/PER1* in humans and yeast (Fujita *et al.*, 2006a). Both of the nematodes species contain a homologue for this gene. Subsequent steps differ greatly between human and yeast. In humans a saturated C18:0 fatty acid is incorporated into the sn-2 position by the gene *PGAP2*, while in yeast the Gup1p protein adds a long saturated C26:0 chain to replace the lipid tail (Bosson *et al.*, 2006; Tashima *et al.*, 2006). *C. elegans* and *C. briggsae* both contain numerous homologues to *PGAP2*, which suggests that nematode GPI anchors might be modified in a similar manner to those in humans. The *C. elegans* homologue to *GUP1* has a low homology score in BLAST (*p*=0.026, Table 3.3) and is postulated to be a hedgehog acyltransferase (Burglin and Kuwabara, 2006). It is therefore likely that the *C. elegans* and *C. briggsae* lipid tail modifications are more closely related to human than yeast. Lipid modification is a relatively poorly understood process and several modifications are known to exist for the GPI anchor within the cell in a variety of organisms (Ernesto S Nakayasu *et al.*, 2009). The presence of multiple *PGAP2* homologues in both the nematodes raises the possibility that the GPI anchor can also be remodelled with a variety of lipid tails, and hints at interesting interactions of GPI anchored proteins within the two worms.

### 3.4.4 Expression patterns of homologues of *PIG-K* and *PIG-A*

Expression patterns for a particular gene can be generated in the worm which provides information on the temporal and spatial expression of the gene, giving us a better picture for its role in the various processes of development. The *PIG-K* and *PIG-A* homologues were chosen for expression pattern analysis due to the crucial role these proteins have in the synthesis of GPI anchors. The *C. elegans* PIG-K homologues are T05E11.6 and T28H10.3, with T05E11.6 having a higher homology score under BLAST alignment. Dupuy *et al.* have created a library of promoter::GFP DNA constructs for *C. elegans* genes called the Promoterome, which can be used for their expression analysis (Dupuy *et al.*, 2004). The *PIG-K* homologue T28H10.3 was available from the library (courtesy of Hope lab) as a plasmid with 868 bp of 5' upstream sequence in the promoter::GFP construct, and was used to elucidate the expression pattern of the gene *in vivo*. T28H10.3 is expressed early in the *C. elegans* embryo and had stable expression throughout the various life stages of the worm (Figure 3.7). The gene is strongly expressed in the intestine of the worm, especially at where the organ joins with the pharynx. GPI anchored gut enzymes may well be involved the digestion of ingested food, with other proteins having potential roles in cell adhesion, signalling and the prevention of pathogen entry (Harris and Siu, 2002; Sharom and Radeva, 2004; Sly and Hu, 1995; Yatsuda *et al.*, 2003). The GPI anchor is also an important apical sorting signal that allows proteins to be located to the correct surface of the cell within the gut, and may be the reason for the high level of T28H10.3 expression within the organ (Benting *et al.*, 1999). It would be interesting to also observe the expression pattern of T05E11.6 to see how much the *PIG-K* homologues overlap with each other within the worm. GPI anchored proteins have been shown to be important for certain neuronal functions (Karagogeos, 2003) and it may be that T05E11.6 is expressed within neurons and has its activity separated from T28H10.3 in a tissue specific manner. More research is needed to elucidate the exact mechanism with which the PIG-K homologues operate within *C. elegans*, which may shed light on the importance of GPI anchoring to the nematode in its growth and development.

The 5' promoter sequences used in the Promoterome constructs are typically 1 to 2 kb in length, which may not be the complete regulatory sequence of the gene (Dupuy *et al.*, 2004). It has been suggested that the use of a larger portion of the 5' promoter sequence may give a more accurate expression pattern for a given gene, which was attempted for the *C. elegans* PIG-A homologue D2085.6. 5 kb of the 5' upstream sequence of the gene was cloned into an appropriate vector with the Gateway expression system (Walhout *et al.*, 2000) to produce a promoter::GFP construct. The construct was tested with restriction digestion and produced bands of the expected sizes, which indicates that the Gateway recombination was carried out successfully (Figure 3.12.d). Trial injections were attempted with the *rol-6* marker only and produced transformed worms with the rolling phenotype (data not shown), however there was not enough time left in the project to attempt a transformation with the D2085.6 Promoter::GFP reporter construct. Hopefully this experiment can be attempted in the near future where it may shed light onto the expression pattern of the *C. elegans* PIG-A homologue, and infer on the importance of GPI anchoring in the nematode.

### 3.4.5 Conclusion

Most of the known steps of GPI synthesis are accounted for in both *C. elegans* and *C. briggsae*, suggesting that they possess the biosynthetic machinery needed for the production of GPI anchored proteins. GPI synthesis in the nematodes may be evolutionarily closer to the human pathway than to that of the yeast. This is suggested by the absence of a homologue for the *PIG-Z/SMP3* in nematodes for the addition of the fourth mannose, which is essential in yeast but non-essential in human. Both nematodes also contain homologues for human *DPM1* and *DPM3*, and sequence analysis of DPM1 homologues in nematodes and yeast suggests that an important C-terminal TM domain in yeast is absent in the nematode and human proteins. *DPM2* homologues however are not found in the nematode genomes, which is more similar to the situation in yeast. *C. elegans* and *C. briggsae* also contain homologues for the human lipid remodelling gene *PGAP2*, whereas only *C. elegans* has a weak homologue to the yeast lipid modification gene *GUP1*. Some differences also appear to exist between the GPI synthesis pathway of the nematodes when compared to human and yeast. The nematodes do not contain homologues for PIG-H and PIG-Y which bind to PIG-A in the first step of synthesis, suggesting that the worm *PIG-A* homologues may be less regulated than their human and yeast counterpart. The PIG-X protein which interacts with PIG-M in addition of the first mannose is also absent in nematodes. *PIG-F*, the interacting partner for the addition of the second and third ethanolamine is absent in *C. elegans* but has a homologue in *C. briggsae*, suggesting that there may be differences in GPI anchor synthesis between the two nematode species. Lastly, the absence of a homologue in *C. elegans* for the *PIG-L* gene raises a fundamental question about the GPI synthesis within the worm. PIG-L is a deacetylase responsible for the second step of GPI biosynthesis and was shown to be indispensible for GPI production in mammals and yeast (Nakamura *et al.*, 1997b; Watanabe *et al.*, 1999). The reaction for step 2 in *C. elegans* may be carried out by an as yet unknown deacetylase within the organism. Taken together a model for the production of GPI anchored proteins is given in figure 3.16, with the basic structure of a likely GPI anchor presented for both of the nematodes. GPI structures found in many organisms undergo extensive modifications depending on their environment (Ferguson, 1999). It will be interesting to see what modifications are present for GPI anchors within *C. elegans* and *C. briggsae*, where these modifications occur during development, and whether they contain tissue specific modifications that impact on the grown and behaviour of the worms.

*C. elegans* is a model organism that is very amenable to expression pattern analysis, which offers insight into the role of genes within a developmental context. A preliminary expression pattern was generated for one of the GPI synthesis pathway genes (*C. elegans* D2085.6, homologue of *PIG-K*) which showed that the gene is expressed in the intestine of the worm for most of its life cycle. Expression patterns for the other GPI biosynthesis genes can also be generated using the Gateway recombination process, which would allow the analysis of this important pathway in a developmental context that has hitherto only been examined in single cellular organisms and cell lines. *C. elegans* thus may provide a unique perspective on this important biological process. The presence of a homologue for the inositol deacetylase *PGAP1/BST1* also suggests that the *C. elegans* GPI anchor can be cleaved with PIPLC, which will allow the use of this enzyme in the analysis of GPI anchored proteins within the worm. *C. elegans* GPI anchoring is a poorly understood process but the model organism has shown great potential in the study of this important biological process, which may enrich the understanding of GPI anchored proteins in biology, especially within the context of development, growth, tissue specific processes and aging.

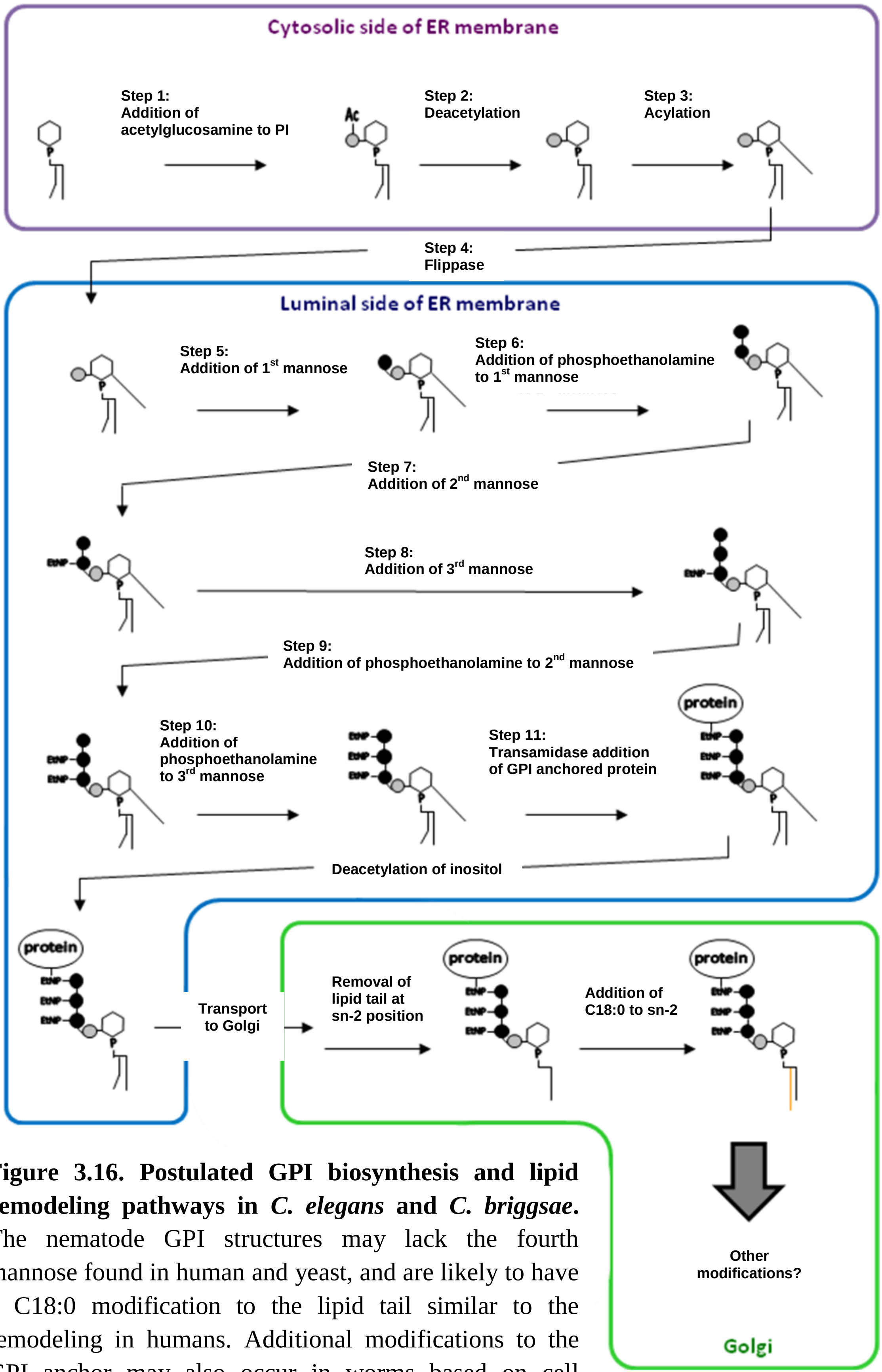



**Figure 3.16. Postulated GPI biosynthesis and lipid remodeling pathways in *C. elegans* and *C. briggsae*.** The nematode GPI structures may lack the fourth mannose found in human and yeast, and are likely to have a C18:0 modification to the lipid tail similar to the remodeling in humans. Additional modifications to the GPI anchor may also occur in worms based on cell specific processes and stage of development.

## *Chapter 4*

# *Caenorhabditis elegans* lipid raft and GPI anchored protein extraction

## 4.1 Introduction

### 4.1.1 The lipid raft membrane

The fluid mosaic model of membrane structure was proposed in 1972 and describes the membrane as an arrangement of globular proteins embedded within a bilayer of phospholipids, with freedom of movement for the proteins to carry out important cellular processes (Singer and Nicolson, 1972). This model, while broadly accurate, was later found to be inadequate to describe the multitude of interactions that proteins are able to form within the membrane environment. Proteins can be tethered into functional aggregates on the membrane by the action of the cytoskeleton, or by specific mechanisms such as clathrin coated pits (Kusumi and Sako, 1996; Ungewickell and Hinrichsen, 2007). One of the more controversial membrane protein-lipid interactions, considered by many to be functionally important, involves the formation of domains of glycolipids called lipid rafts. These domains contain collections of sphingolipids and cholesterol with a tight packing density that segregates them from the rest of the membrane phospholipids, creating distinct “rafts” of lipids that move as a unit within the lipid bilayer (Figure 4.1). Evidence for their existence and their functional significance has been hotly debated within the scientific literature ever since they were first discovered. In 1997 a model of the lipid raft was presented in the journal Nature, which was taken up as a semi-official definition of lipid rafts within the scientific community, and attracted comment from all sides of the debate (Simons and Ikonen, 1997). The paper defined rafts as a dynamic clustering of sphingolipids and cholesterol within the lipid bilayer that acts as a platform for protein-protein interaction, and protein attachment for transport within the cell. Lipid rafts have been postulated to be involved in a diverse number of important cellular processes, including transport, cell recognition, endocytosis and signalling (Anderson, 1993; Anderson *et al.*, 1992; Fiedler *et al.*, 1994; Solomon *et al.*, 2002).

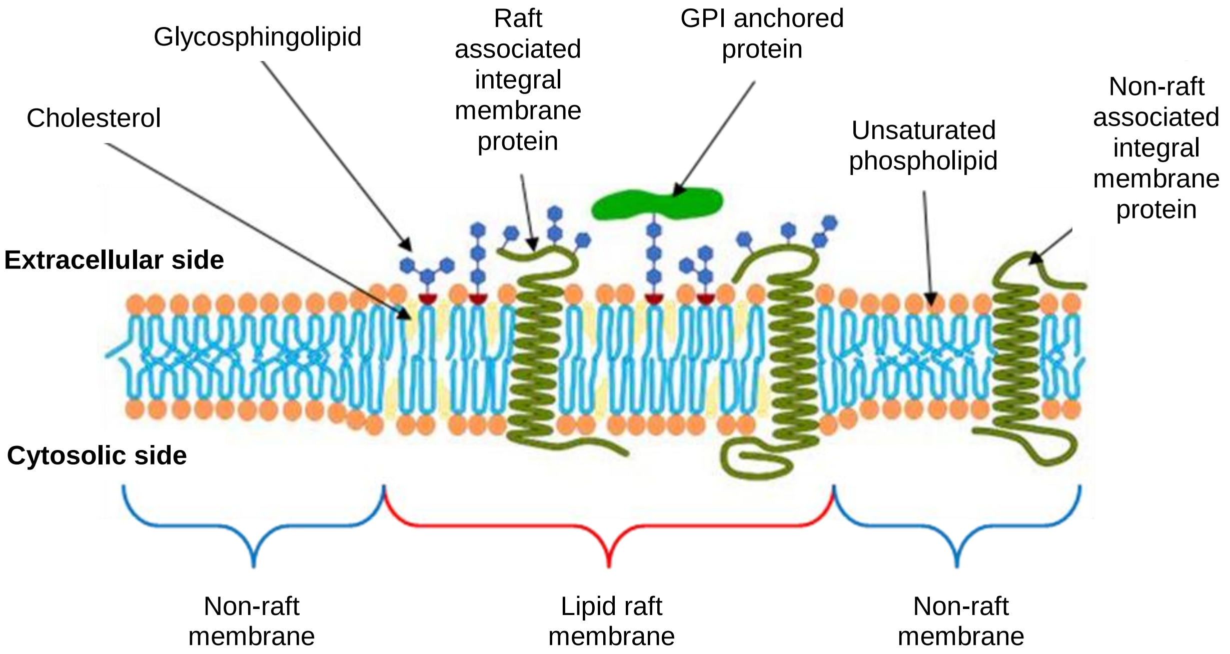



**Figure 4.1. Diagrammatic representation of lipid raft membranes.** Raft membranes contain an aggregation of sphingolipids with saturated fatty acid chains and cholesterol. GPI anchored proteins and other integral membrane proteins may associate with the raft domain, some of which may contain glycosylation. This diagram was adapted from http://cellbiology.med.unsw.edu.au/units/science/lecture0803.htm.

### 4.1.2 Extraction of lipid rafts from the cell

Lipid rafts are resistant to solubilisation when treated with cold non-ionic detergents. This property is thought to be due to the tight packing of the sphingolipids and cholesterol that are the major structural components of lipid rafts (Chamberlain, 2004), and forms the basis for the most popular methods for raft extraction. Lipid raft proteins are distinguished from non-raft integral membrane proteins in that they are not readily solubilised by detergents at low temperatures, and results in the extraction of a fraction commonly termed as the detergent resistant membrane (DRM). At higher concentrations of detergents or a higher temperature the protection gained from the tight packing is lost and lipid raft proteins become solubilised (Chamberlain and Gould, 2002). Raft proteins may also display varied levels of insolubility depending on the concentration of the detergent (Prior *et al.*, 2001). To date the most popular detergent used for the extraction of lipid rafts is Triton-X 100 (TX-100), though some researchers have opted for other detergents such as Brij 96, Brij 98, Lubrol WX, and others (Drevot *et al.*, 2002; Madore *et al.*, 1999; Roper *et al.*, 2000). The choice of detergent has been the subject of trial and error within the field, as each detergent has different solubilisation properties that allow them to dissolve different subsets of membrane proteins within the cell (Chamberlain, 2004). Detergent chemistry can be complex as each of them can have different properties regarding the size and propensity of micelle formation and phase separation, which directly influence lipid subdomain solubilisation. These properties can be hard to predict when more than one detergent is present, which explains why mixtures of detergents are rarely employed for lipid raft isolation (Linke, 2009). The raft isolation procedure involves a discontinuous sucrose gradient for the separation of detergent soluble and insoluble protein fractions (Brown and Rose, 1992). Rafts are found as a low density fraction that floats at the interface between the 5% and 30% sucrose layers of the density gradient (Hope and Pike, 1996).

#### 4.1.2.1 Detergents used for raft extraction

Early experiments with Brij 96 found this detergent could be used to extract lipid rafts from lymphoid cells (Draberova *et al.*, 1996), with Brij 98 chosen as a detergent by Drevot *et al.* for the extraction of T cell coupled receptors (TCR) from rafts (Drevot *et al.*, 2002). One of the advantages of Brij 96 and Brij 98 is that the detergent works at 37°C, which is thought to represent the extraction of a more physiologically relevant lipid raft fraction (Chamberlain, 2004). Brij 96 was shown to give better selectivity of raft domains than TX-100 when solubilising lipid rafts from neurons (Madore *et al.*, 1999). However, detergent-resistant fraction from myelin membranes extracted by Brij 96 was shown to float to a lower density compared to TX-100, which was postulated to represent a subpopulation of rafts within the membrane (Taylor *et al.*, 2002).

Lubrol WX was first used in the extraction of lipid rafts from epithelial cells and was shown to extract a distinct raft fraction from the microvilli of the cell (Roper *et al.*, 2000). Lubrol WX extracted rafts were also shown to give different solubilisation of raft proteins than TX-100 for proteins involved in apical trafficking (Slimane *et al.*, 2003), further reinforcing the idea of the presence of distinct “Lubrol rafts” within cell membranes.

TX-100 has been used extensively for the analysis of sphingolipid and cholesterol enriched domains from an early stage of lipid raft analysis (Brown and Rose, 1992; Hertz and Barenholz, 1977). The detergent has excellent properties when it comes to enrichment of the lipids found in rafts, with a 3-5 fold increase in cholesterol content, 15% increase in sphingolipids, and a marked decrease in non-raft lipids such as phosphatidylcholine, phosphatidylethanolamine and lipids of the inner membrane leaflet (Pike, 2003; Pike *et al.*, 2002; Prinetti *et al.*, 2000). Schuck *et al.* tested different detergents for their suitability for lipid raft extraction and showed that TX-100 was able to solubilise more non-raft proteins than Brij 96, Brij 98 and Lubrol WX. TX-100 was also able to concentrate raft lipids comprising cholesterol and sphingomyelin with greater selectively than the other detergents, and to produce a much ‘purer’ fraction of raft lipids from model membranes than Brij 96 and Lubrol WX (Garner *et al.*, 2008; Schuck *et al.*, 2003). The consensus seems to be that the different detergents used for lipid raft extraction are able to segregate rafts of different properties according to stringency. Weaker detergents such as Brij 96, Brij 98 and Lubrol WX are able to extract proteins which may only be transiently associated with rafts, while stronger detergents such as TX-100 extract a smaller subset of proteins that may represent the core lipid raft proteins found on the plasma membrane (Chamberlain, 2004; Schuck *et al.*, 2003).

#### 4.1.2.2 Non-detergent extraction methods

An important caveat with detergents comes from the finding that their use may encourage lipid domain formation in biological membranes (Heerklotz, 2002; Mayor and Maxfield, 1995), with the result that the lipid rafts extracted might be an artefact of the experimental procedure, and not be representative of physiological rafts that occur naturally within the cell. Some researchers have tried to alleviate this potential artefact by developing detergent-free methods of raft extraction. One of the first such methods was performed by Smart *et al.* and involved the separation of a caveolae-enriched raft fraction using their lighter density (Smart *et al.*, 1995). The unique features of caveolae have also been used to isolate rafts by the pulldown of caveolin containing membranes with antibody coated beads (Macdonald and Pike, 2005; Schnitzer *et al.*, 1995; Stan *et al.*, 1997). These protocols however require multiple sucrose gradient steps, and as a result produce low yields of proteins for further characterisation and analysis.

#### 4.1.2.3 Extraction methods used in proteomic projects

Studies of the protein constituents of lipid rafts with proteomic techniques have become increasingly frequent in the wake of the genomic era. One of the first proteomic analysis of lipid rafts was made in human T cells and identified over 70 proteins (von Haller *et al.*, 2001). Subsequent projects have looked at lipid rafts from a wide variety of cells and organisms, including *Candida albicanas*, rat liver cells, human HeLa cells, adipocytes, and others (Bae *et al.*, 2004; Foster *et al.*, 2003; Insenser *et al.*, 2006; Kim *et al.*, 2009). With a few exceptions (Bini *et al.*, 2003), the majority of lipid raft proteomic analysis used the now classical TX-100 detergent extraction method with flotation on sucrose gradients to extract their proteins, with some researchers using Opitprep$^{TM}$ medium to create the desired gradient (Blonder *et al.*, 2004; Li *et al.*, 2003; Li *et al.*, 2004a; Nebl *et al.*, 2002). TX-100 extraction has the advantage of a relatively easy set up, and an ability to be scaled up to purify the significant amount of proteins needed for proteomic studies. The higher stringency of TX-100 prepared rafts compared with Brij 96 or Lubrol WX is also an important factor for its widespread use in proteomic analysis, as the ubiquitous nature of proteomic studies means that contamination from other fractions can easily become misidentified as raft associated.

### 4.1.3 Extraction of GPI anchored proteins

It was found very early on that the GPI moiety of anchored proteins can become cleaved following digestion by the enzyme phosphatidylinositol-specific phospholipase C (PIPLC) (Ferguson *et al.*, 1985a; Ikezawa *et al.*, 1976). The enzyme was found to cleave the anchor at the P-O position of the phosphate group adjacent to the lipid backbone (Figure 4.2). PIPLC has been found in a number of organisms, including *Bacillus cereus, Bacillus thuringiensis, Trypanosoma brucei*, and others (Bulow and Overath, 1986; Ikezawa *et al.*, 1976; Taguchi *et al.*, 1980) . It was found that PIPLC cannot cleave GPI anchors with an acylation modification on the inositol ring; GPI anchors with this modification can however be cleaved by the enzyme phosphatidylinositol-specific phospholipase D (PIPLD), which was discovered in mammals and cleaves the GPI anchor on the phosphate group at the P-O position adjacent to the inositol ring (Figure 4.2) (Davitz *et al.*, 1987; Ikezawa, 2002). Cleaved GPI anchored proteins are no longer attached to the membrane and exhibit properties of soluble aqueous proteins upon release. This property and the specificity of the enzyme for GPI anchors has lead to the use of PIPLC as the *de-facto* route for the extraction of GPI anchored proteins from cells (Ikezawa, 2002).

One of the most popular methods for GPI anchored protein enrichment was created by Bordier and involves the use of Triton X-114 (TX-114) in their extraction (Bordier, 1981). The method utilises the property that TX-114 has a relatively low cloud point of 20°C that permits the separation of membrane proteins from their cytosolic counterparts into two phases, the detergent phase (detergent-rich) and the aqueous phase (detergent-poor). GPI anchored proteins usually partition into the detergent phase due to their amphipathic nature, however after treatment with PIPLC the proteins become hydrophilic and will partition instead to the aqueous phase (Hooper and Turner, 1988). This technique has been the basis of a number of proteomic studies into GPI anchored proteins, including studies of the GPI proteomes of *Arabidopsis thaliana, Plasmodium falciparum* and human HeLa cells (Borner *et al.*, 2003; Elortza *et al.*, 2003; Gilson *et al.*, 2006; Lalanne *et al.*, 2004; Sherrier *et al.*, 1999). These studies utilised what was described as a "shave and conquer" method (Elortza *et al.*, 2003), by enriching for membrane proteins, treating them with PIPLC , and finally extracting the released GPI anchored proteins with TX-114 phase separation. PIPLD was also used in one of these studies for *A. thaliana* and human HeLa cells (Elortza *et al.*, 2006)

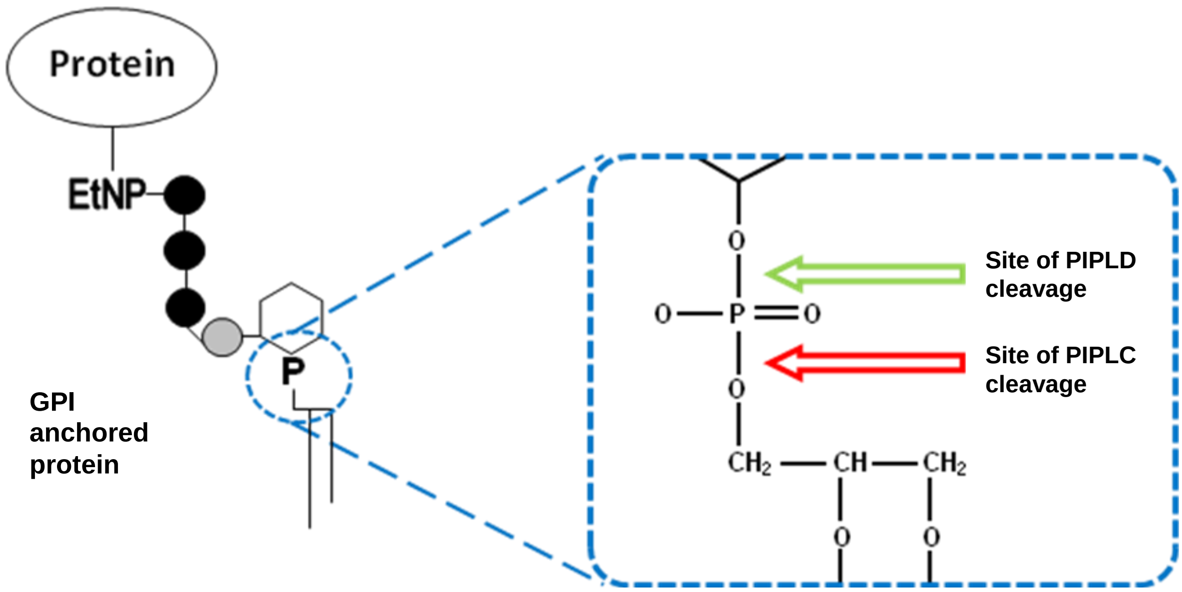



**Figure 4.2. Site of cleavage for PIPLC and PIPLD.** PIPLC cleaves the GPI anchor at the P-O bond next to the phospholipid backbone while PIPLD cleaves the anchor at the P-O bond adjacent to the inositol ring. Structure of the GPI anchored protein adapted from Chapter 3 figure 3.1.

### 4.1.4 *C. elegans* lipid raft and GPI anchor studies

*C. elegans* as a model organism has a relatively poor track record for membrane protein studies. Part of the reason comes from the worm's thick cuticle which makes protein extraction difficult. TX-100 and Lubrol PX were used in a solubilisation trial for the nicotinic acetylcholine receptor in *C. elegans* (Lewis and Berberich, 1992). Sedensky *et al.* were able to extract lipid raft from *C. elegans* by the use of TX-100 and show that the fraction contains two mammalian stomatin homologues UNC-1 and UNC-24, and a sodium channel subunit (UNC-8) which interacts with UNC-1 (Sedensky *et al.*, 2004). There is currently one GPI anchored protein identified in *C. elegans* called *phg-1* (alternative name *phas-1*), which is a homologue of the mammalian GPI anchored protein gas-1 involved in embryogenesis. PHG-1 was found to be released by PIPLC when expressed in a mammalian cell line (Agostoni *et al.*, 2002). *In silico* studies of *C. elegans* GPI anchored proteins have also been performed previously (Eisenhaber *et al.*, 2000; Fankhauser and Maser, 2005; Poisson *et al.*, 2007).

### 4.1.5 Outline for lipid raft and GPI anchored protein extraction in *C .elegans*

In this chapter I will detail the methods used for the extraction and enrichment of *C. elegans* lipid raft and GPI anchored proteins. Worms were grown in liquid culture and cleaned by flotation on a sucrose cushion (Hope, 1999). Protein extraction in *C. elegans* proteomic projects generally aim to break open the tough cuticle of the worm, which can be achieved by freeze-thawing of the worms, sonication, and glass homogenisation (Li *et al.*, 2009; Schrimpf *et al.*, 2001; Tabuse *et al.*, 2005). Membrane proteins can then be extracted via differential ultracentrifugation, with lipid rafts enriched from the crude membrane preparation by TX-100 solubilisation and sucrose gradient density centrifugation. Since GPI anchored proteins are already enriched in lipid rafts (Brown and Rose, 1992) we felt that there was no need for the TX-114 extraction procedure, as the previous proteomics studies of GPI anchored protein all used general membrane preparations as their starting material (Borner *et al.*, 2003; Elortza *et al.*, 2003). PIPLC was used on the lipid raft fraction and the released proteins were separated from the membrane fraction via ultracentrifugation. Presented in this chapter are the results from the extraction, which was applied later in chapter 5 for proteomic analysis.

## 4.2 Method

### 4.2.1 Worm strain

Wildtype N2 nematode strains were kept as described in Chapter 3.2.2.

### 4.2.2 Growth of bacteria

#### 4.2.2.1 OP50 strain

*E. coli* OP50 strain was acquired courtesy of Hope lab. OP50 stock was kept at 4°C on 140 mm diameter agar plates with Luria-Bertani (LB) agar formula (8.6 mM NaCl, 1% (w/v) peptone, 0.5% (w/v) yeast extract, 1.5% (w/v) bacteriological agar) and OP50 bacteria for NGM plates were grown in 100mL LB media (8.6 mM NaCl, 1% (w/v) peptone, 0.5% (w/v) yeast extract) at 37°C shaking overnight (o/n) and 5-6 drops were added to each NGM plate in a fume hood and left to dry for 24 hours.

#### 4.2.2.2 HB101 strain

HB101 *E. coli* strain was acquired courtesy of Hope lab. HB101 stock was kept on 140 mm diameter LB agar plates with streptomycin (50 μg/ml). Bacteria for worm liquid culture were grown in 1 l LB media at 37°C shaking o/n and spun at 3,000 g for 5 minutes. The supernatant was discarded and the bacterial pellet was resuspended in an equal volume of S. basal (0.1 M NaCl, 0.05 M Potassium phosphate pH 6, 5 μg/ml cholesterol) and stored at 4°C. Typically 12 ml of final bacterial suspension was made per 1 l of LB media.

### 4.2.3 Liquid culture of *C. elegans*

Worms from 2 fully populated (but not starved) NGM plates were washed into 100 ml of S Basal solution. 100 µl of Streptomycin (50 mg/ml), 100 µl of Nystatin (50 mg/ml) and 4.5 ml of HB101 bacterial suspension were added to the S Basal solution and the total mixture was incubated at 20°C shaking for 3 days, after which 1 ml of worms from the previous liquid culture was used to inoculate a new batch. The culture solution was checked daily and fresh bacteria were added as necessary.

### 4.2.4 Sucrose floatation extraction of *C. elegans*

Nematodes from four 100 ml liquid cultures were placed in 15 cm long test tubes and suspended on ice (4°C) for 10 minutes. The supernatant was discarded and the worm pellet was resuspended in 25 ml of 0.1 M NaCl in a 50ml falcon tube. An equal volume of 60% (w/v) sucrose was added to the worm suspension which was then centrifuged at 500 g for 2 minutes at 4°C. Worms floating on the surface were aspirated with a Pasteur pipette cut at the shoulder and diluted 10 times in cold 0.1 M NaCl. The suspension was centrifuged at 500g for 3 minutes and the supernatant discarded. Worms were then resuspended in 0.1 M NaCl and incubated for 1 hour at 20°C while shaking. Afterwards, the worms were placed on ice for 15 mins and centrifuged at 500g for 3 minutes and the pellet of worms was collected and the supernatant discarded. The worm pellet was then resuspended in an equal volume containing protease inhibitor solution (x2 concentration, Complete Protease Inhibitor Cocktail from Roche in 100 mM HEPES), flash frozen in liquid $N_2$, and stored at -70°C until required.

### 4.2.5 Extraction of membrane proteins

Washed *C. elegans* (18 ml) was taken from the freezer and left on ice to thaw. The worms were spun at approx. 10,000 g (13,000 rpm) for 1 minute on a Heraeus Biofuge Pico benchtop centrifuge and the supernatant was discarded. The worms were then flash frozen with liquid $N_2$ and ground with a pestle and mortar, subjected to sonication (ten bursts, 10 sec per burst, MSE Scientific Instruments), and further broken down in a glass homogenizer (10 plunges); all procedures were carried out at 4°C. The homogenate was spun at 500 g for 10 minutes and the supernatant was spun again at 3000 g for 15 minutes and the pellets discarded. The supernatant was then centrifuged at 50,000 g for 1 hour at 4°C, after which the remaining membrane pellet was taken and resuspended in 800 µl of protease inhibitor solution (x2 concentration, Complete Protease Inhibitor Cocktail from Roche in 100 mM HEPES). The remaining supernatant was spun again at 70,000 g for 1 hour at 4°C to produce a second membrane pellet, which was resuspended in 400 µl of protease inhibitor solution (x2 concentration). The membrane preparations from the first and second ultracentrifugation steps were flash frozen in liquid $N_2$ and stored at -20°C.

### 4.2.6 Lactose wash

Crude membrane proteins were thawed on ice and washed with cold 100 mM lactose made up to 1 ml (in protease inhibitor cocktail, x1 concentration) for 1 hour with occasional agitation. Membranes were collected by centrifugation (100,000g for 1 hour) and resuspended in the same volume of protease solution (x2 concentration) as before the wash.

### 4.2.7 Discontinuous sucrose gradient extraction of lipid rafts

All subsequent steps were performed at 4°C to maintain lipid raft integrity. Six batches of lactose washed membrane (6 x 200 µl aliquots) were each resuspended in 3.55 ml of MES (morpholineethanesulfonic acid) Buffered Saline (MBS, 25 mM MES, 150 mM NaCl, pH 6.5) with trials of 1%, 2% or 4% TX-100 (v/v). The solution was mixed with 3.75 ml of 80% sucrose solution (80% sucrose (w/v) in MBS) to make up a 40% sucrose solution containing the membrane samples. Sucrose gradients were set up in 6 centrifuge tubes (25x 89 mm, thin wall Ultra-clear, cat no. 344058, Beckman Coulter) by adding 15 ml of 5% sucrose solution (5% sucrose (w/v) in MBS) to the tubes and subsequent layering of 15 ml of 35% sucrose solution (35% sucrose (w/v) in MBS) under the first layer from the base of the tube using a blunted long syringe needle. The 40% sucrose membrane samples were loaded into the tubes from the base and the gradients were then centrifuged at 100,000 g for at least 18 hours in a SW25 swing out rotor (Beckman Coulter). Fractions were taken from the base of the centrifuge tubes at 3.75ml intervals. A total of 10 fractions were taken from each tube with the pellet resuspended in MBS. All fractions were subsequently diluted 10 fold with MBS and spun at 100,000 g for 1 hour 30 minutes with the pellet resuspended in 400 µl MBS.

### 4.2.8 Bicinchoninic acid (BCA) protein concentration assay

BCA assays were carried out as per the manufacturer's instructions (Pierce). BCA reagent A (1 mg sodium bicinchoninate, 2 mg sodium carbonate, 0.16 mg sodium tartrate, 0.4 mg NaOH, 0.95 mg sodium bicarbonate in 100 ml $dH_2O$, pH 11.25) and B (0.4 mg cupric sulfate (5 x hydrated) in 10 ml $dH_2O$) were mixed in the ratio 100:2 to make BSA working reagent. A dilution series of Bovine Serum Albumin (BSA, in 50 mM HEPES) standards were prepared from 0.1 mg/ml to 2.0 mg/ml. 50 µl of each standard, sample and one blank containing buffer were added to labelled tubes with 1.0 ml of BSA working reagent. The tubes were mixed via inversion and incubated at 60°C for 30 minutes. The absorbance of the final solution at 562nm wavelength was measured in a colorimeter for each tube. The protein concentrations of the samples were measured against the graph plot of the standards.

### 4.2.9 PIPLC digestion of lipid rafts

PIPLC was obtained from Molecular Probes (100 U/ml in 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 0.01% sodium azide and 50% glycerol). 5 µl of PIPLC (0.5U activity) and 5 µl of $dH_2O$ were added to two lots of 200 µl of lipid raft membrane (approx. 0.92 mg of protein each) and incubated at 4°C overnight with gentle shaking. The solution was centrifugated at 100,000g in a Beckman Optima benchtop ultracentrifuge at 4 °C for 1 hour and the supernatant was separated from the pellet. The pellets were resuspended in 200 µl of MBS.

### 4.2.10 1D-electrophoresis

1D-electrophoresis was carried out with the Protean III gel system from Bio-Rad. The glass plates were set up per manufacturers' instructions. 10% running gel (2.1 ml $dH_2O$, 1.67 ml polyacrylamide, 1.25ml of 1.5 M Tris HCl pH 8.8, 50 µl 10% SDS, 5 µl TEMED and 50 µl Ammonium persulfate (APS) per gel) was poured into the plates followed by 5% stacking gel (1.7 ml $dH_2O$, 0.42 ml polyacrylamide, 0.32 ml of 1.5 M Tris HCl pH 8.8, 25 µl 10% SDS, 2.5 µl TEMED and 25 µl APS per gel) on top with spacers and left on the bench to set. The gel was then placed in a gel tank with running buffer (3 g Tris base, 14.4 g glycine, 10 ml of 10% SDS in 1 l of $dH_2O$). Samples were prepared by making up each of the desired volume of samples up to 9 µl with $dH_2O$ and mixing them with 3 µl of 4x SDS sample buffer (4 ml glycerol, 0.8 g SDS, 2.5 ml of 1M Tris-HCl pH 6.8, 80 µl of 5 mg/ml bromophenol blue slurry, 0.2 ml β-mercaptoethanol (BME) and $dH_2O$ up to 10 ml), which were heated to 90°C for 10 minutes and then spun down briefly at 6,000 rpm on a benchtop centrifuge. 5 µl of size marker and all of the samples were added to the desired wells and the gels were ran at 100 V constant voltage for 10 minutes and subsequently 150 V until the blue front had reached the bottom of the gel.

### 4.2.11 Coomassie staining of gels

Gels were placed in 20 ml of Coomassie stain (40% methanol, 10% acetic acid, 50% water and 0.1 % (w/v) Coomassie Brilliant Blue R250) for 30 minutes while shaking. Gels were then washed with destaining solution (40% methanol, 10% acetic acid and

50% water) while shaking with regular replacement of the destaining solution at 15 minute intervals until the background stain has been mostly removed.

### 4.2.12 Silver staining

Gels from electrophoresis were placed in fixer solution (50% $H_2O$, 40% methanol and 10% acetic acid) for at least 1 hour. Each gel was then washed 3 times in 100ml of 30% ethanol for 20 minutes each with shaking. Gels were then each placed in 100ml of 0.02% sodium thiosulphate (in $dH_2O$) in  for 90 seconds with gentle shaking, washed 3 times in $dH_2O$ for 20 seconds each, and placed in 100 ml of silver stain solution (0.2 g silver nitrate and 20 μl formaldehyde in 100 ml $dH_2O$, made fresh) for 20 minutes while shaken. Gels were then washed 3 times with $dH_2O$ for 20 seconds each, placed in 100 ml of developer solution (3g sodium carbonate, 0.875 mg sodium thiosulphate and 100 μl formaldehyde in 100 ml $dH_2O$) and shaken for 3-5 minutes until the bands on the gel have developed to the desired intensity. The gels were washed twice again with $dH_2O$ for 30 seconds each and placed in 100 ml of stopper solution (0.5 g glycine in 100 ml $dH_2O$) for 10 minutes with shaking. The gels were kept at the end in 100ml of $dH_2O$.

### 4.2.13 Western blot

1D gels of desired protein samples were run as per instruction. Nitrocellulose membranes (0.45 μm pore size, Amersham Hybond ECL) were cut to the desired size and pre-soaked in transfer buffer (3 g Tris base, 14.4 g glycine, 20% methanol (v/v) in

1 l of $dH_2O$) for 10 minutes. Two thin sponges and four pieces of 3M paper cut to the size of the membrane were also pre-soaked in transfer buffer. The protein transfer cassette was then made in the following manner: place one sponge on the black side of the cassette followed by two pieces of 3M paper, then the gel was added with the membrane placed on top while making sure there were no air bubbles, and finally two more pieces of 3M paper were added with a sponge on top and the cassette was then closed. The cassette was placed in a western blot frame and added to a gel box with an ice pack, a magnetic flea and filled transfer buffer, and was run at 100 V at 4°C for 2 hours while being stirred. Transferred membranes were placed in 20 ml primary antibody solution (primary antibody at 1/1000 concentration and 4% powdered milk in phosphate buffered saline (PBST, 137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4, 0.05% Tween-20) and incubated at 4°C overnight. The primary antibody for *C. elegans* CAV-1 was raised in mouse against a peptide with the sequence CNFNIRKTGINQETTA, which covers a region at the C-terminus of the protein. The primary antibody for *C. elegans* ENT-1 was raised in rabbit (raised to a peptide with the sequence RAERQRNKNDEAVDSEGKV corrisponding to amino acid positions 245-263 in the ENT-1 protein, courtesy of Mrs. J. Ingram, Prof. Baldwin's lab). Membranes were then washed three times in PBST for 10 minutes each and incubated in secondary antibody solution (species specific secondary antibody conjugated to horse radish peroxidase (HRP) at 1/10,000 concentration and 4% powdered milk in PBST) for 2 hours at room temperature. Secondary antibodies for mouse were used for CAV-1, and rabbit used for ENT-1. The membranes were washed again three times in PBST for 10 minutes each and placed in 1.5 ml of HRP detection reagent (SuperSignal West Pico, Pierce) on top of Saran wrap and incubated at room temperature for 5 minutes. Excess detection reagent was dried off with 3M paper and

visualized with either developer film (Enhanced Chemiluminesence film, GE healthcare) or CCD detection camera.

## 4.3 Results

### 4.3.1 Extraction of membrane fraction

*C. elegans* membrane material was extracted via differential ultracentrifugation. *C. elegans* was grown in liquid culture in order to provide enough material for analysis by proteomics. Sucrose floatation was used to extract the worms as this method allows live nematodes to be separated from cell debris and dead worms, and to minimise bacterial contamination by allowing the worms to digest any remaining bacteria in their gut. The worms extracted were of a mixed stage in their life cycles when observed under a light microscope (data not shown). In order to break the tough cuticle, nematodes were subjected to three rounds of homogenisation by freeze-thawing, sonication and grinding with a glass homogeniser. Membrane extraction was performed at 50,000g in a Beckman ultracentrifuge. Another membrane fraction was made with 70,000g spin after it was found that some membrane material was still present in the supernatant after the first spin (Figure 4.3). The membrane was washed with lactose to remove excess galectin proteins before the isolation of lipid rafts (Figure 4.7).

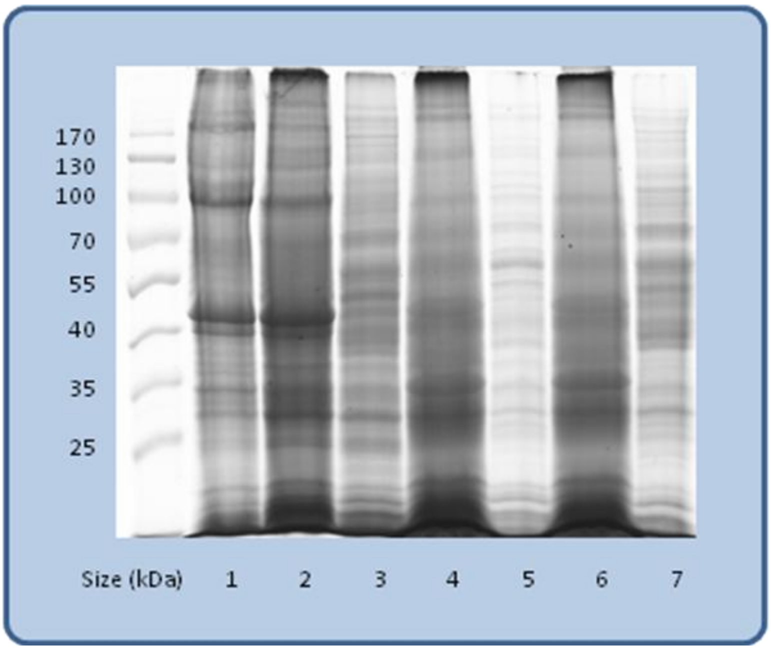



**Figure 4.3. Fractions isolated during membrane extraction.** Gel was stained with Coomassie Blue. Lane 1) insoluble pellet from 500g spin, 2) insoluble pellet from 3000g spin, 3) protein material before centrifugation, 4) membrane pellet from 50,000g, 5) supernatant from 50,000g, 6) membrane pellet from 70,000g and 7) supernatant from 70,000g.

### 4.3.2 Extraction of lipid rafts

*C. elegans* lipid raft was extracted using TX-100 in a discontinuous sucrose gradient from the 50,000g membrane preparation. Figure 4.4 shows the setup of the gradient with the membrane material at the bottom, which after ultracentrifugation separates the lipid raft components from the rest of the membrane. Trials of 1%, 2% and 4% TX-100 were performed to assess the concentration of detergent used for the extraction of lipid raft. 10 fractions were made from the sucrose gradient after ultracentrifugation. For the 1% TX-100 trial the majority of the proteins were confined to the TX-100 dissolved fractions 1-3, with a reduction of proteins in fraction 4, while protein concentration was increased again in fractions 5 and 6 and were virtually absent from fraction 7-10 (Figure 4.5). This is consistent with the observed presence of the light scattering band at the interface between the 5% and 30% sucrose concentrations caused by the floatation of TX-100 insoluble proteins and lipids, which corresponds to fractions 5 and 6 on in the protein extraction (figure 4.6). The light scattering band was not observed for TX-100 concentrations of 2% and 4% (data not shown) and the corresponding protein profiles for their respective fractions show the majority of the proteins to be present in fractions 1-3 with no enrichment in fractions 5 and 6 (figure 4.5b and 4.5c). Therefore a TX-100 concentration of 1% was used to for lipid raft extraction. Proteins from fractions 5 and 6 were pooled and referred to as the lipid raft fraction.

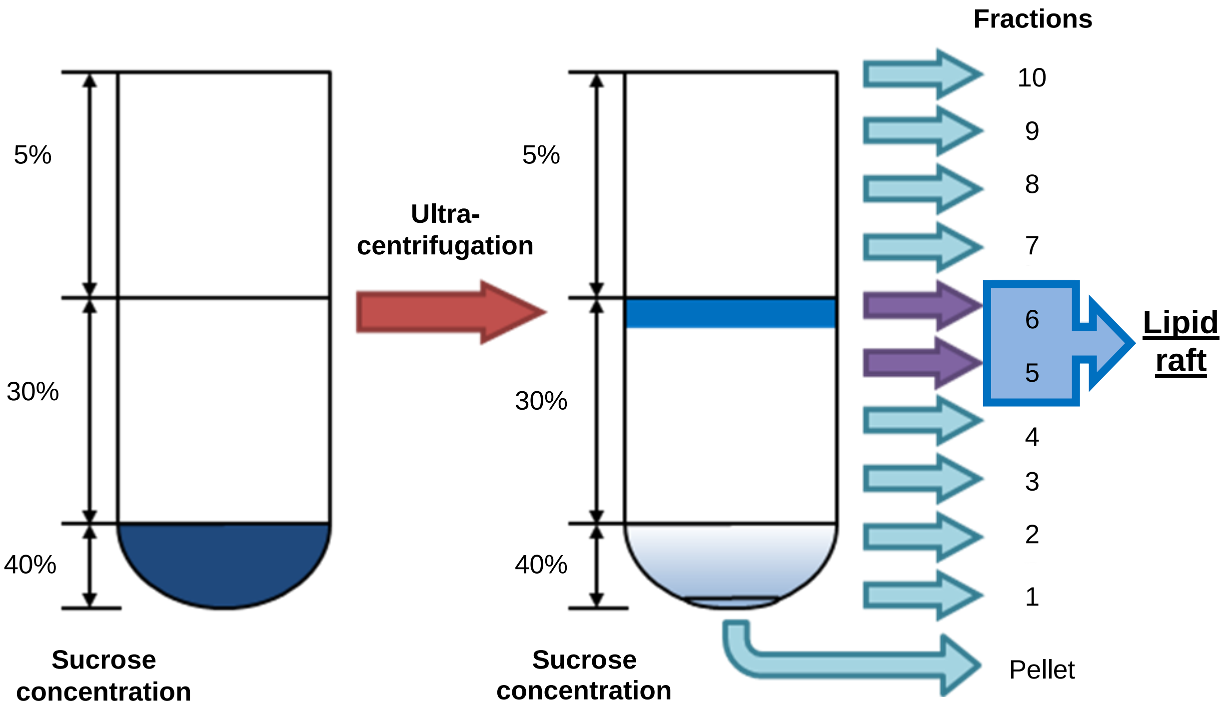



**Figure 4.4. Diagram of sucrose gradient density extraction of lipid rafts.** Membrane proteins are solubilised in a solution containing TX-100 and 40% sucrose and loaded into the bottom of the gradient. After ultracentrifugation at 100,000g o/n 10 sucrose fractions and 1 pellet fraction were taken from the gradient. Fractions 5 and 6 are at the boundary between 5% and 30% sucrose and contains purified lipid rafts.

Lipid raft extraction with 1% TX-100 was also performed for the 70,000g membranes. The proteins showed poor separation and a sizeable proportion ended up in fractions 7-10 of the sucrose gradient (figure 4.5d). There was also no light scattering band observed at the 5% to 30% sucrose interface. The membrane fraction from 70,000g was therefore not used and only membranes from 50,000g were used for lipid raft extraction.

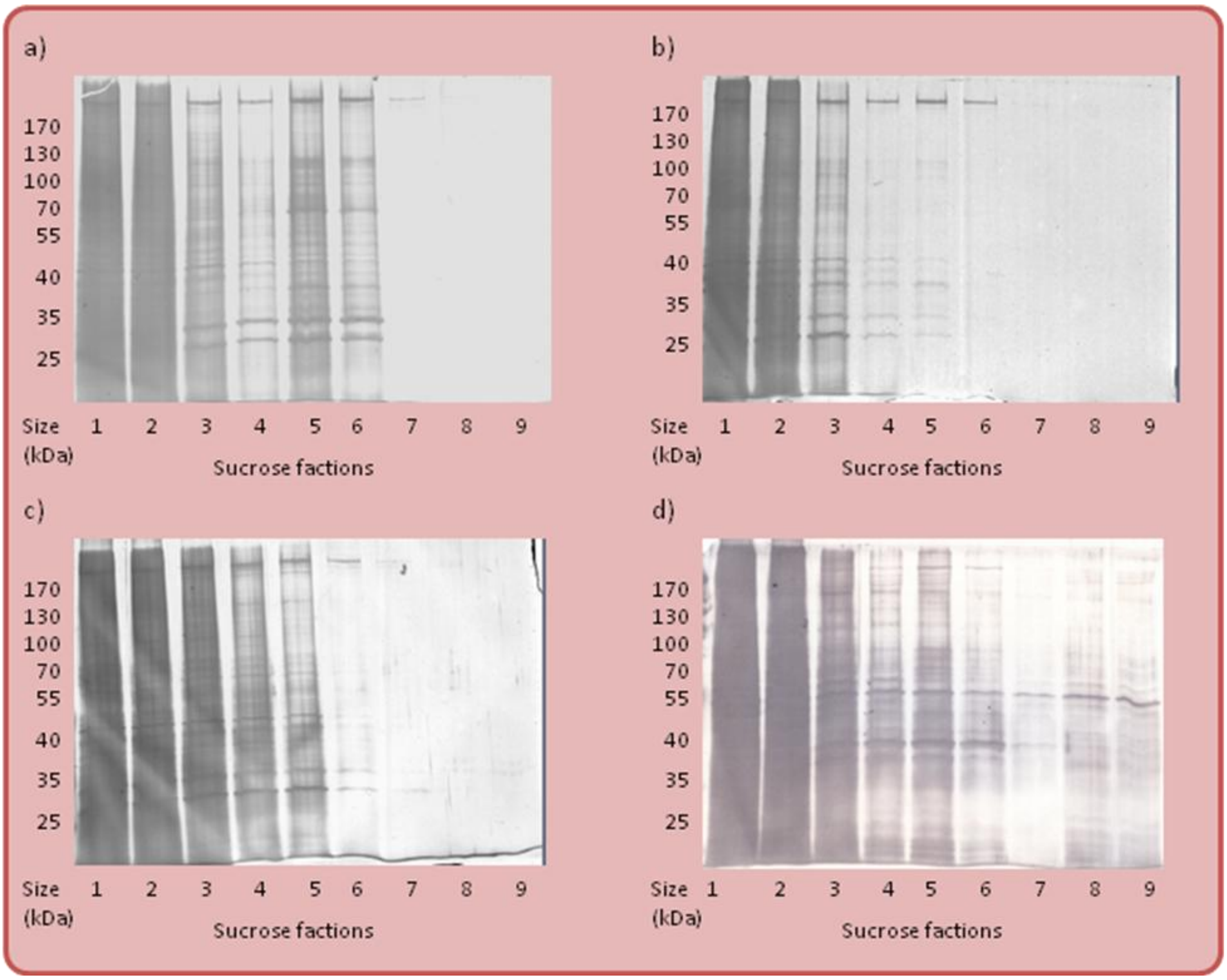



**Figure 4.5. Fractions 1-9 of sucrose density extractions from various experiments.** All gels were stained with silver nitrate. Gel a) 50,000g membrane protein extracted with 1% TX-100, b) 50,000g membrane protein extracted with 2% TX-100, c) 50,000g membrane protein extracted with 4% TX-100, d) 70,000g membrane protein extracted with 1% TX-100.

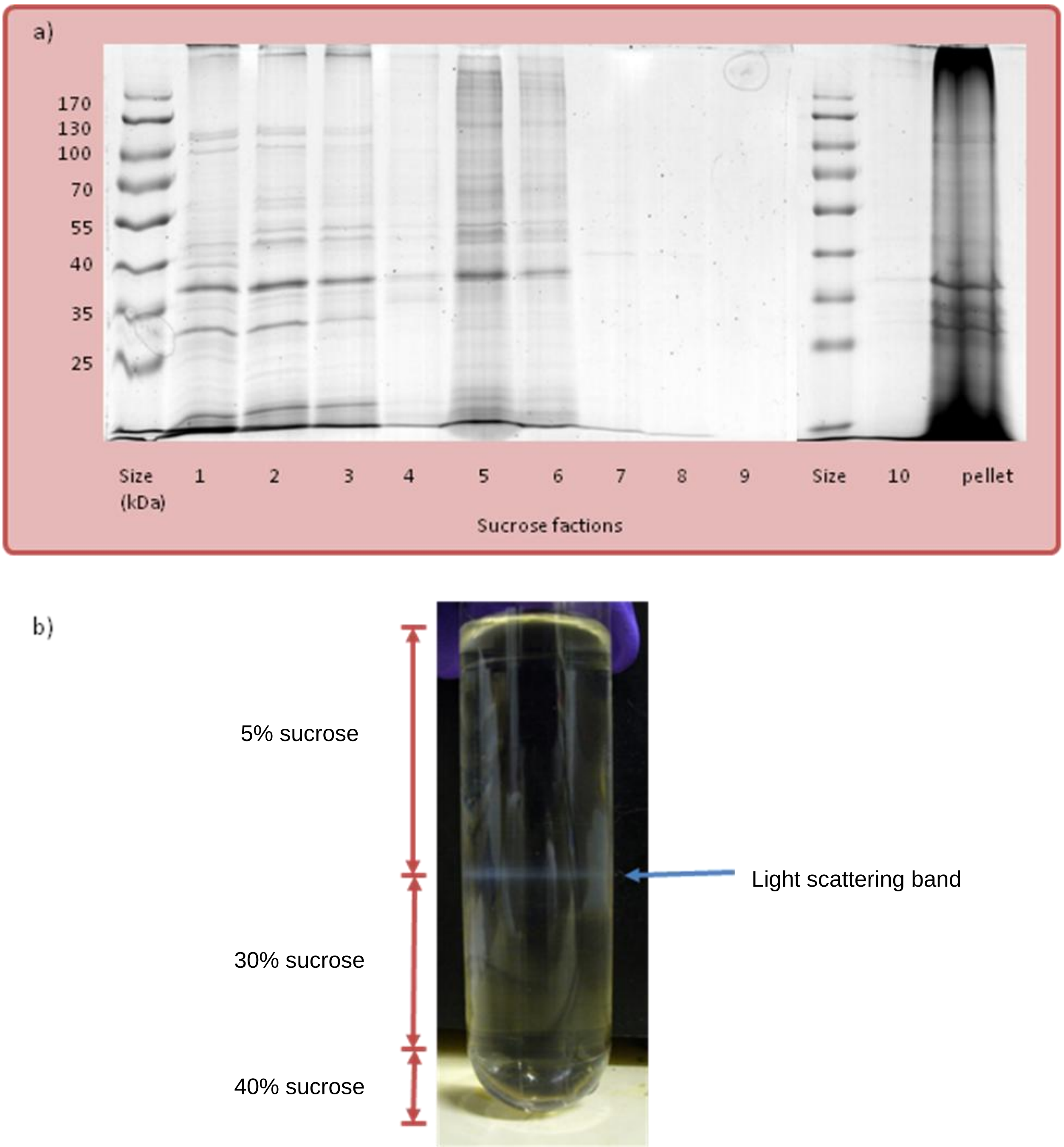



**Figure 4.6. Sucrose density extraction of lipid raft proteins from *C. elegans*.** a) fractions 1-10 of the sucrose density with the insoluble pellet run on two separate gels stained with Coomassie Blue, b) a photograph of the Beckman SW25 ultracentrifuge tube showing the presence of the light scattering band at the 5%/30% sucrose interface.

### 4.3.3 Washing of lipid raft fraction

In order to reduce the number of membrane associated proteins and keratin contamination from the samples the lipid raft fraction was sequentially washed with HPLC grade $H_2O$, 1M NaCl and then HPLC grade $H_2O$ again (figure 4.7). The remaining raft proteins were redissolved in HPLC grade $H_2O$. Keratin contamination was also minimised by carrying out all procedures in a fume hood. The concentration of washed lipid raft proteins was determined to be 4.6 mg/ml by BCA assay.

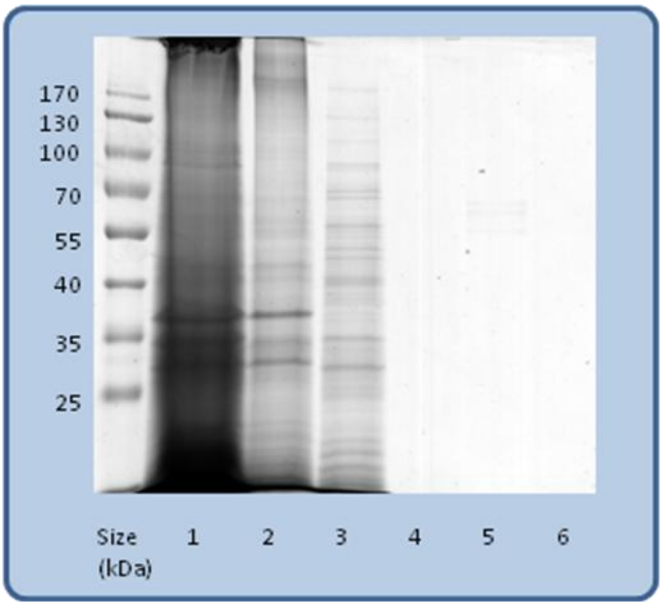



**Figure 4.7. Protein fractions from various wash stages.** Gel was stained with Coomassie Blue. Lane 1) total membrane from 50,000g, 2) total membrane after lactose wash, 3) supernatant from 100mM lactose wash 4) supernatant from first wash of lipid raft with HPLC water, 5) supernatant from lipid raft wash with 1M NaCl, 6) supernatant from second wash of lipid raft with HPLC water.

### 4.3.4 Verification of lipid raft fraction

Antibodies against a peptide sequence at the C-terminal section of *C. elegans* caveolin CAV-1 were raised in mice, and blotted against the fractions to verify the existence of lipid rafts. CAV-1 was observed to be enriched in lipid raft fraction compared to the total membrane (figure 4.8a and 4.8b). A blot of the sucrose gradient fractions shows the presence of CAV-1 in the TX-100 insoluble fractions 5-6 and TX-100 1-3, but not in fraction 4, suggesting that there are two distinct forms of caveolin within the protein sample examined, one of which is TX-100 soluble and the other TX-100 insoluble (Figure 4.8c). A control blot of CAV-1 was made in the presence of the peptide that was used to generate the antibody, which did not produce a band (data not shown). A blot of the sucrose gradient fractions against *C. elegans* ENT-1 shows the protein to be confined to the TX-100 soluble fractions 1-3 (figure 4.8d). ENT-1 is a nucleoside transporter and is not reported to be a lipid raft associated protein (Appleford *et al.*, 2004).

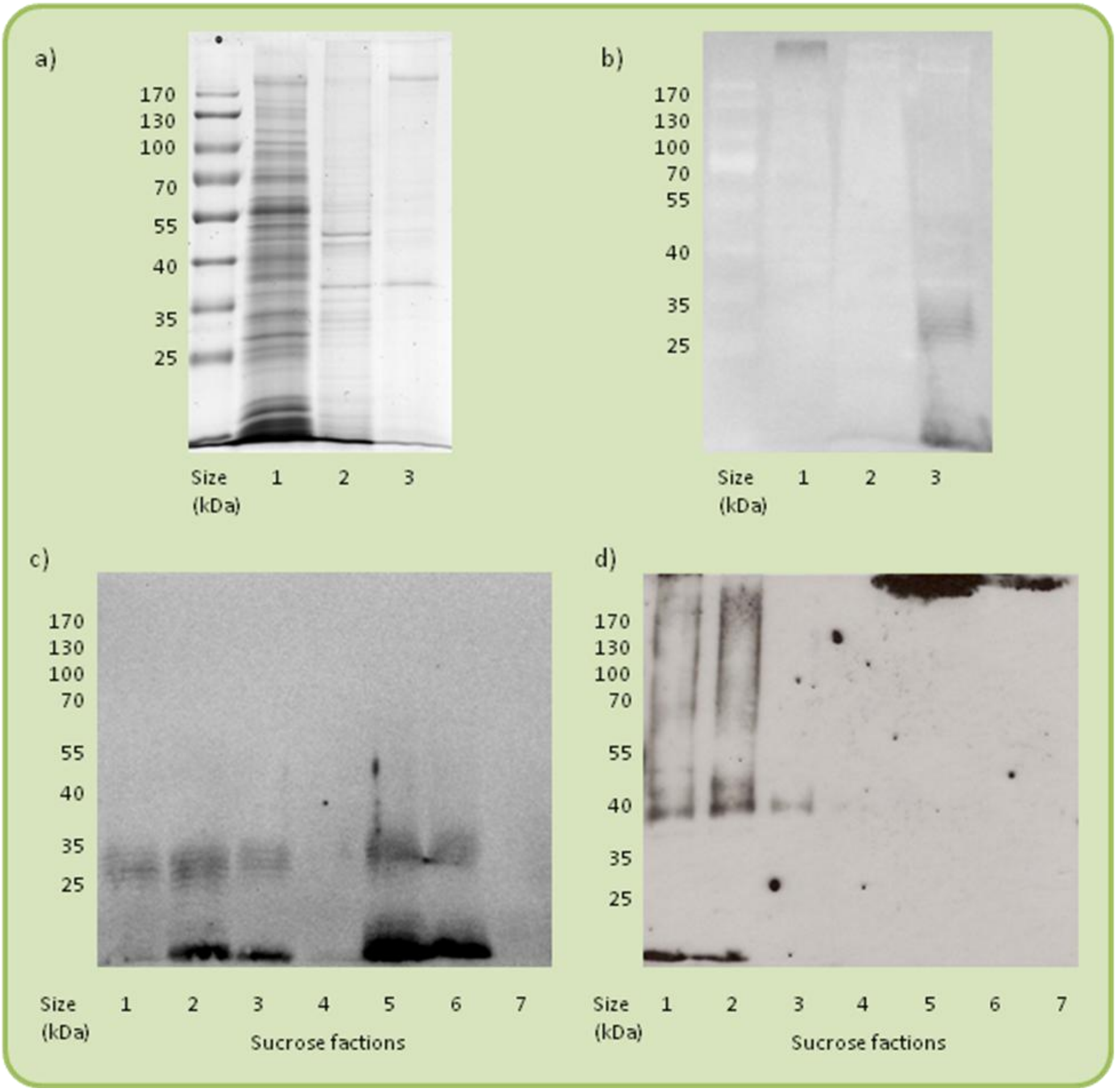



**Figure 4.8. Western blots of protein fractions.** a) Coomassie Blue staining of protein fractions. Lane 1- supernatant from membrane extraction at 50,000g, lane 2- membrane extracted at 50,000g (1/10 dilution), lane 3- lipid raft proteins(pooled fractions 5 and 6, 1/10 dilution), b) blot of gel (a) with CAV-1 antibody, c) blot of sucrose fractions 1-7 with CAV-1 antibody, d) blot of sucrose fractions 1-7 with ENT-1 antibody.

### 4.3.5 PIPLC digest of lipid raft fraction

PIPLC digestion was performed at 4°C overnight. Numerous proteins were released from the lipid raft fraction with only a small amount of high molecular weight proteins released from the control (figure 4.9.a). Fraction 5 from membrane proteins extracted with 4% TX-100 was also digested with PIPLC (figure 4.9b). This produced a relatively large release of proteins in the control digestion, indicating that fraction contains membrane associated protein contaminants and is unsuitable for GPI anchor protein analysis.

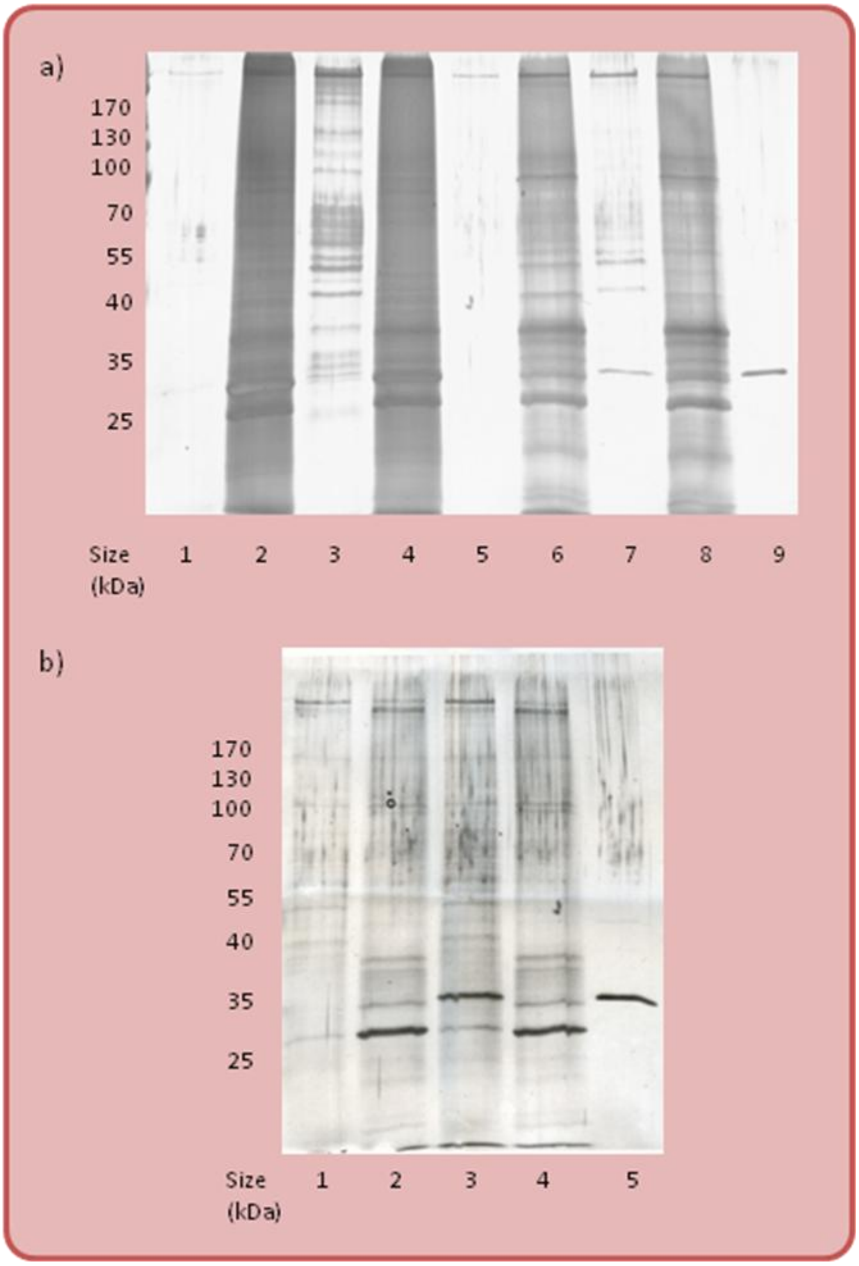



**Figure 4.9. Lipid raft fraction digested with PIPLC.**

a) Raft fraction extracted with 1% TX-100.

lane 1- supernatant of control of fraction 5, lane 2- membrane of control of fraction 5, lane 3- supernatant of digest of fraction 5, lane 4- membrane of digest of fraction 5, lane 5- supernatant of control of fraction 6, lane 6- membrane of control of fraction 6, lane 7- supernatant of digest of fraction 6, lane 8- membrane of digest of fraction 6, lane 9- PIPLC enzyme.

b) Raft fraction extracted with 4% TX-100.

lane 1- supernatant of control of fraction 5, lane 2- membrane of control of fraction 5, lane 3- supernatant of digest of fraction 5, lane 4- membrane of digest of fraction 5, lane 5- PIPLC enzyme.

### 4.4 Discussion

In this chapter the details of adapting a lipid raft isolation protocol to *C. elegans* are presented. *C. elegans* was grown in liquid culture rather than NGM plates to provide the large amount of proteins needed for the extraction of lipid rafts and associated GPI anchored proteins, and their downstream analysis with proteomics techniques. Liquid culture produces worms in a mixed stage of development with an increase in the number of small worms in the dauer stage. Sucrose density floatation was used to separate the dead worms from live ones and remove other contaminants such as cell debris and bacteria. The growth media contains the antifungal agent nystatin, which binds to sterols and is used as chemical for cholesterol depletion in lipid raft analysis (Stuart *et al.*, 2003). The growth media however also contain cholesterol as *C. elegans* cannot synthesize the compound *de-novo*, which may minimise the effect of nystatin on cholesterol depletion. Observation under a light microscope also confirms the presence of healthy worms at various stages.

#### 4.4.1 Membrane extraction

*C. elegans* is covered by a layer of tough cuticle that requires strong mechanical action to break apart. Previous protein extraction procedures for the nematode have used a combination of freeze-thaw, sonication and ground glass tissue grinders to homogenise the worms for proteomic studies (Kaji *et al.*, 2000; Li *et al.*, 2009; Tabuse *et al.*, 2005). All of these techniques were used here to ensure a thorough break up of worms, and the tissues were shown to be adequately homogenised when examined under a light microscope. Several stages of low level centrifugation were needed to remove the broken down cuticle material. Two different membrane preparations were extracted from the homogenate centrifuged at 50,000g and 70,000g respectively. The presence of the two different membrane fractions may have been due to the extraction of different sub cellular locales. The 70,000g membrane was subsequently shown to be unsuitable for lipid raft extraction via TX-100. Part of the reason may be that the fraction contains a relatively small amount of membranes that contain lipid rafts, such as the plasma membrane, and is therefore unsuitable for raft purification.

### 4.4.2 Lipid raft purification

Extraction of lipid rafts involves the solubilisation of the membrane proteins with 1% TX-100, based on their property of insolubility by weak non-ionic detergents at cold temperatures. Proteins extracted with a discontinuous sucrose gradient typically shows the presence of a light scattering band at the 5% to 30% sucrose concentration interface due to the lower buoyancy of lipid raft components (Sedensky *et al.*, 2004). Extraction of proteins with discontinuous sucrose density typically show a high concentration of proteins in the bottom fractions (1-3) that are solubilised by TX-100, a reduction of proteins in fraction 4, an increase in protein concentration corresponding to the light scattering band containing fractions 5 and 6, and little or no proteins in fractions 7-10 that corresponds to the 5% sucrose part of the gradient. This was shown to be the case for proteins extracted with 1% TX-100. The presence of *C. elegans* CAV-1 was confirmed with antibody blots in fractions 5 and 6 (Figure 4.8c). CAV-1 was also detected in fractions 1-3 of the blot, but was absent in fraction 4, indicating that there may be two distinct populations of caveolin within the membrane that are associated with raft and non-raft fractions, respectively. Figure 4.8b shows CAV-1 to be enriched in the pooled fractions 5 and 6 compared to the total membrane, which suggests that lipid rafts are enriched within these fractions. Caveolin is typically found as a marker of lipid raft fractions within the cell, and *C. elegans* CAV-1 was found to be localised selectively to the post-synaptic membrane of neurons in the worm, where it was shown to function in acetylcholine signalling (Parker *et al.*, 2007). Higher concentrations of TX-100 was suggested to improve the solubility of GPI anchored proteins specifically (Dr. Parkin and Prof. Hooper, personal communication) and concentrations of 2% and 4% TX-100 were used to extract rafts, which were found not to produce distinct lipid raft fractions (Figure 4.6b and 4.6c). This may have been due to the increased solubilisation of the raft fraction from the higher detergent content.

### 4.4.3 Washes and handling

Washes of the membrane material were performed to remove membrane associated proteins. It was found within preliminary proteomic analysis (Chapter 5) that there was an over abundance of galectins in the sample. Galectins are sugar binding proteins commonly associated with lipid rafts (Hansen *et al.*, 2005) and their presence is encouraging for the confirmation of the extraction of rafts; however it was found that the amount of the galectins present within the sample was having an adverse effect on the identification of other proteins. A lactose wash was carried out to remove most of the associated galectins before the sucrose gradient step in order to reduce their presence in the final raft preparation. Rafts were then washed with high salt concentrations (NaCl) to remove other proteins not directly associated with rafts. Washes with HPLC grade $dH_2O$ were intended to minimise keratin contamination, and the water used for the rest of the experimental procedures all came from MilliQ grade $dH_2O$, as keratin can become a major contaminant when concentrated from large volumes of water (Dr. Keen, personal communication). All procedures were carried out in flow-lamina fume hoods whenever possible to reduce airborne keratin contamination from dust and skin particles.

### 4.4.4 PIPLC release of proteins

Proteins were extracted from the lipid raft fraction by digestion with PIPLC, an enzyme which specifically cleaves the GPI anchor and allows the membrane bound proteins to be released into the aqueous phase. These released proteins were separated from the rest of the raft via ultracentrifugation. Results show that GPI anchored proteins were released from the 1% TX-100 extracted raft fraction, with the control digestion showing all but two contaminating bands with high molecular weights (Figure 4.9a). There was a greater number of contaminating bands found within the PIPLC digest for the 4% TX-100 extracted rafts (Figure 4.9b). The PIPLC enzyme from Molecular probes produced one band of the correct size for the protein indicating that the enzyme is of good quality (Figure 4.9a, lane 9). GPI anchored proteins released by PIPLC show a typical increase in apparent mass on SDS-PAGE gels due to the properties of the remaining sugar molecules attached to the protein after cleavage (Littlewood *et al.*, 1989). This effect was seen on several bands for the released proteins and indicates the presence of properly solubilised GPI anchored proteins (Figure 4.9a).

### 4.4.5 Future directions

One of the enduring controversies in lipid raft biology is the concerns the definition of the raft with respect to its method of extraction. It is observed that rafts extracted with different types and concentrations of detergents such as Brij96, Lubrol WX, and other non-detergent methods contain different subsets of proteins. Pike summarised three models for the makeup of lipid rafts that allows the existence of different raft domains (Pike, 2004) (Figure 4.10). In model I the lipid rafts are homogenous, but layered according to the selectivity of the various detergents such that the more selective amongst them (such as TX-100) extract the core raft components and the less selective (such as Brij 96) extracts a more general component. Model II proposes rafts to be entirely homogeneous and the different detergents extract sub-proteomes from the whole due to their specific properties on the membrane. Model III envisages the existence of wholly distinct heterogeneous sub-rafts with different properties that are susceptible to extraction by the various detergents used. The author proposed that current evidence points to model III being more likely to be valid as there are fundamental heterogeneities in the proteomes produced from the different detergents. An "Induced fit" hypothesis was offered where lipid rafts are grown from small "proto" rafts into larger stable structures consisting of a variety of different components. Lipid rafts can be very dynamic structures and changes in the concentrations of sphingolipids and cholesterol have been shown to drastically change their size in model membrane experiments (Prenner *et al.*, 2007). It will be interesting to extract lipid rafts from *C. elegans* using a variety of different detergent and non-detergent extraction methods to determine its raft constituents in detail, and to observe what kinds of rafts exist within this model organism.

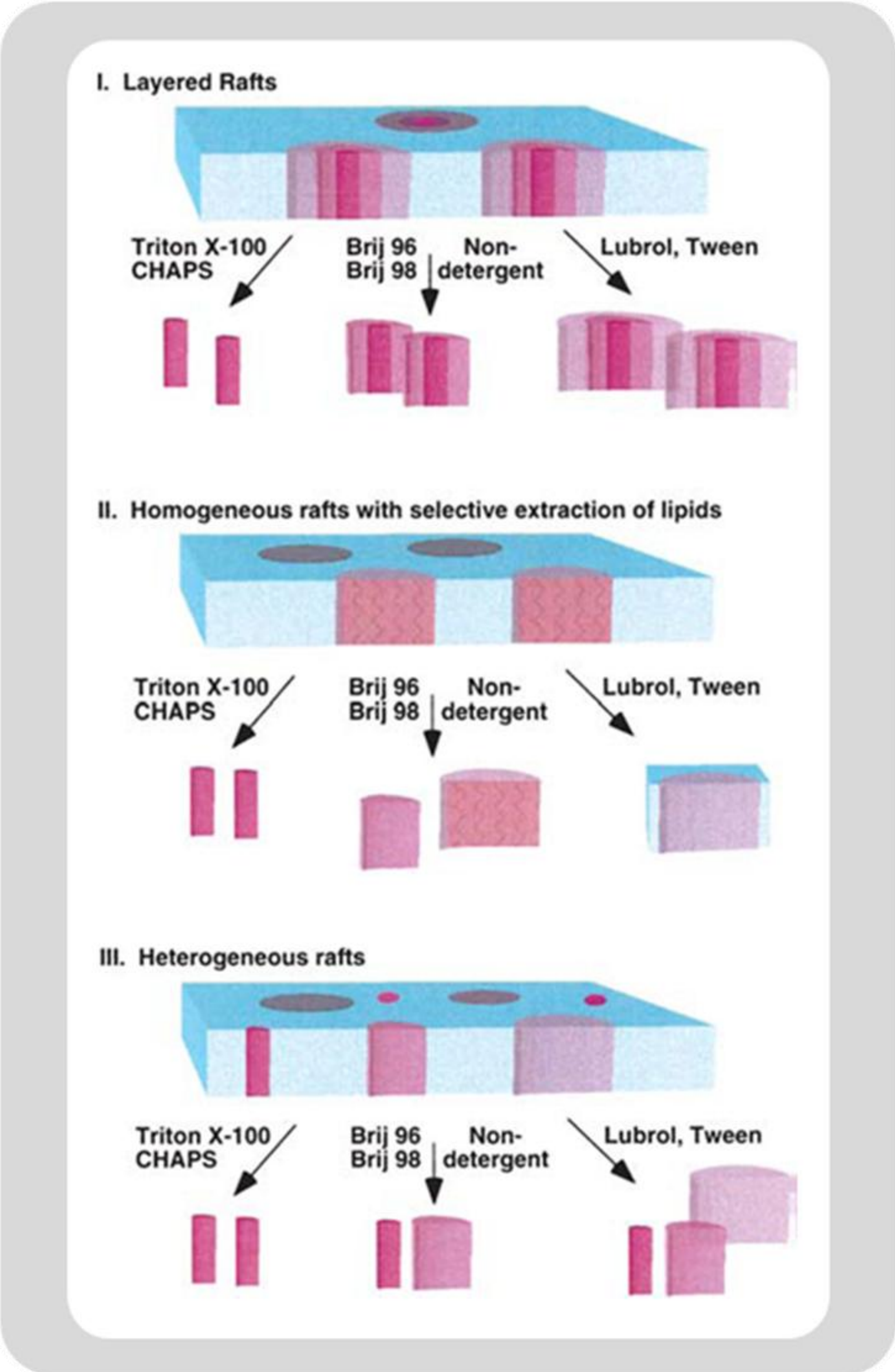



**Figure 4.10. Three postulated models for the existence of rafts in the membrane.** Model I) raft proteins and lipids form concentric layers around a core that can be extracted by detergents of different strength. Model II) rafts are homogenous and the detergents are selective in their extraction of components. Model III) different detergents extract distinct rafts with different properties. Diagram was adapted from Pike 2004.

The numerous roles of lipid raft have been studied in a variety of cell lines and single cellular organisms, but as yet have not been examined fully in a developmentally complex system. Raft domains have been shown to play a large role within important biological processes such as cell polarity and signal transduction that underpin animal development (Lajoie *et al.*, 2009). *C. elegans*, with its extensively annotated genome, well understood genetics and invariant cell lineage is well suited for looking at the role of rafts and GPI anchored proteins in important processes such as development, behaviour, locomotion and aging. *C. elegans* development is surprisingly complex for an organism of such a small size with four different molting stages during its life cycle. Raft proteins can be identified from the various stages of maturation in synchronised nematode populations to elucidate the roles they play within worm development. Disruption of lipid rafts may also be performed for *C. elegans* to assess their biological role within the organism; this may be achieved by growing the worms away from their cholesterol enriched media, and with the use of cholesterol depletion agents such as nystatin. The *C. elegans* genes Y57E12AL.1 and R11H6.2 both contain a serine incorporator (SERINC) domain that was shown to be involved in sphingolipid biosynthesis in mammalian and yeast cells (Inuzuka *et al.*, 2005). Y57E12AL.1 has already been shown to cause slow growth, abnormal egg laying and a patchy colouration in RNAi experiments (Kamath *et al.*, 2003) while a deletion mutant is available for R11H6.2 with no phenotype reported so far. These two genes may disrupt lipid rafts and allow the study of raft dynamics during the life cycle of the worm. *C. elegans* has the potential to become an invaluable tool for the study of lipid rafts, and may produce great insights into the biology of this important sub-cellular locale.

# ***Chapter 5***

# **Proteomic analysis of *Caenorhabditis elegans* lipid raft and GPI anchored proteins**

## 5.1 Introduction

Analyses of biological samples with proteomic techniques have evolved greatly since the 1980's, and today encompass a wide variety of protocols that are able to elucidate the proteomes of many different experimental systems. These techniques make use of multiple separation procedures to provide high fidelity and resolving power for the proteome of interest. Analytical methods today are based on two core technologies- 2 dimensional gel electrophoresis (Issaq and Veenstra, 2008) and multidimensional liquid chromatography (Motoyama and Yates, 2008)– from which efficient separation of proteins and peptides can be made for subsequent mass spectrometry analysis.

### 5.1.1 2D electrophoresis

2D electrophoresis is a protein separation technique with high resolving power and has been one of the workhorses for proteomic projects from an early age. The technique was first attempted in 1956 by the sequential application of two different electrophoretic processes at right angles to each other to produce a flat square shaped gel (Smithies and Poulik, 1956). The resulting gel not only improved the resolution of a complex blood serum sample but also was able to differentiate the various modifications of a protein present within the sample that would have otherwise been missed with a 1D gel. Many different combinations of techniques for the first and second dimensions were tried subsequently. In 1975 O'Farrell established what is now the standard configuration of 2D gel electrophoresis, by separating *E. coli* proteins according to their isoelectric point via isoelectric focusing (IEF) in the first dimension, followed by molecular weight via sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) in the second dimension (O'Farrell, 1975). In 1982 Bjellqvist *et al.* described the use of immobilised pH gradient (IPG) strips for the separation of proteins in the first dimension, which allowed the formation of a stable pH gradient for IEF and greatly improved the resolution and reproducibility of 2D gels (Bjellqvist *et al.*, 1982). One of the major advantages of 2D gel electrophoresis is the ability to compare quantitatively protein levels of spots between different protein samples. Reproducibility between gels however is poor due to the slightly different conditions that gels are subjected to during an experimental run, and a variety of computer programs have been made over the years to facilitate spot matching and quantitative comparisons between different samples (Righetti *et al.*, 2004). A method called differential in-gel electrophoresis (DIGE) was developed by Unlu *et al.* that allowed two or more samples of proteins to be run on the same gel. This technique used minimal labelling of Lys residues by different fluorophores for each sample, which were subsequently ran together and visualised separately using a fluorescence scanner (Unlu *et al.*, 1997); this resulted in a greater resolution of the proteins and an improved comparison between different samples. Solubilisation of membrane proteins is difficult with 2D electrophoresis because of the need to use weak non-ionic detergents compatible with IEF (Rabilloud, 2009); however membrane proteins can still be analysed with 2D gels when an optimal mixture of detergents for the first dimension is used (Churchward *et al.*, 2005).

### 5.1.2 Multidimensional liquid chromatography (MDLC)

Separation of proteins using combinations of two or more orthogonal high performance liquid chromatography (HPLC) techniques has gained steady momentum within the field and has become an increasingly popular method for the analysis of proteomes (Motoyama and Yates, 2008). The concept for MDLC existed in the 1980's, when Giddings outlined that the separation of proteins from two different HPLC systems would be orthogonal, with the result that the overall resolving power becomes the product of the resolution of each of the individual dimensions, which greatly increases the separation that can be achieved for the sample (Giddings, 1984). Progress within the field was overshadowed by improvements in 2D gel electrophoresis, until the invention of tandem mass spectrometry (MS/MS), which allowed direct peptide sequencing and accelerated the use of MDLC for proteomics (Yates *et al.*, 1995). MS/MS gave rise to a new branch of proteomic analysis called shotgun proteomics, which involves the separation of pre-digested peptides (rather than intact proteins) that are directly sequenced within the mass spectrometer, which are then matched to protein sequences *in silico*. This dramatically improved the number of proteins that can be identified for a given proteome, and has the added advantage that previously difficult proteomes such as membrane proteins can now be analysed with relative ease. The first large scale MDLC MS/MS project was called multidimensional protein identification technology (MudPIT) and has been a watershed in the application of this method for the study of proteomes (Washburn *et al.*, 2001). The standard setup for MDLC is for pre-digested peptides to enter the first dimension and separated into fractions, which are then applied to the second dimension and eluted directly into the mass spectrometer for sequencing. Reverse phase (RP) chromatography, which separates peptides based on hydrophobicity, is usually used for the second dimension as this system has very good resolving power, and can be fed directly to an electrospray ionisation (ESI) mass spectrometer for automation (Claessens and van Straten, 2004; Motoyama and Yates, 2008). The first dimension can be any method which gives good orthogonality with respect to the second dimension, with separation techniques such as size-exclusion chromatography (SEC), strong cation-exchange (SCX), IEF, SDS PAGE and others being used for a number of different projects (Chen *et al.*, 2002; Machtejevas *et al.*, 2004; Opiteck and Jorgenson, 1997; Peng *et al.*, 2008; Washburn *et al.*, 2001). One of the major disadvantages of MDLC over 2D electrophoresis is its reduced ability to effectively quantify protein levels and analyse post-translational modifications; improvements in these areas however have steadily been made, with new techniques such as isotope coded affinity tag (ICAT) and isobaric tags for relative and absolute quantification (iTRAQ) allowing better quantitative analysis within a given proteome (Gygi *et al.*, 1999; Ross *et al.*, 2004; van den Broek *et al.*, 2008).

### 5.1.3 MS protein identification by peptide mass fingerprint

The identification of proteins via mass spectrometry starts with limited cleavage of the protein via tryptic digestion. Trypsin cleaves the C-terminal peptide bond after Arg and Lys residues and has been shown to be extremely reliable in its peptidase action (Olsen *et al.*, 2004). The enzyme also digests proteins into fragments of a good range of masses that are compatible with mass spectrometry. For the analysis of a single protein (such as from a spot on a 2D gel) the mass to charge ratios (m/z) of all of the digested peptides are collected into a unique pattern for the protein, which is called a peptide mass fingerprint (PMF) (Pappin *et al.*, 1993). PMF allows rapid to protein identification by comparing the observed patterns with *in silico* digested fragments of predicted proteins, which are generated by specialised search algorithms such as MASCOT (Perkins *et al.,* 1999). Peptides from a mixture of proteins can confound the protein identification by giving conflicting PMFs for the search programs, which means that proteins need to be separated intact at high resolution before they can be subjected to tryptic digestion and MS analysis. This makes 2D electrophoresis the method of choice for protein identification by PMF. Proteins from 1D SDS-PAGE can also be analysed using this method, provided that the protein band in question is separated with sufficient resolution.

### 5.1.4 MS/MS sequencing

With the advent of MS/MS technology it became possible to directly sequence the peptides produced from a tryptic digest of proteins. Peptides analysed with the first MS are further fragmented by collision induced dissociation (CID) with inert gas to produce a set of partially broken peptide species, which are then analysed within a second MS instrument (Hunt *et al.*, 1986). Fragmentation of the peptide can occur at 3 positions for each amino acid on the peptide backbone, producing a neutral and a charged product that can be detected in the mass spectrometer. Depending on the position of the break and where the charge is assigned a total of six different types of peptide ions (a, b, c and x, y, z) can result for each amino acid position in the peptide (Figure. 5.1) (Hernandez *et al.*, 2006). The most common ions generated are from the b and y series, as they are formed after breakage of the amide bond. This allows a build up of the peptide for each amino acid lost in the collision, from the $b_1$, $b_2$,… and the $y_1$, $y_2$,… series of peptide peaks, allowing the production of a complete sequence of the peptide (Figure. 5.2) (Hunt *et al.*, 1986; Rioli *et al.*, 2003). Analysis of the sequence data with bioinformatic programs such as MASCOT and SEQUEST results in the identification of the protein (Perkins *et al.*, 1999; Wolters *et al.*, 2001). Protein identification from MS/MS results is much more precise than PMF due to the availability of sequence information for each of the peptides, and can be effectively used for protein mixtures in a shotgun proteomics experiment to generate a large number of identifications in a short amount of time.

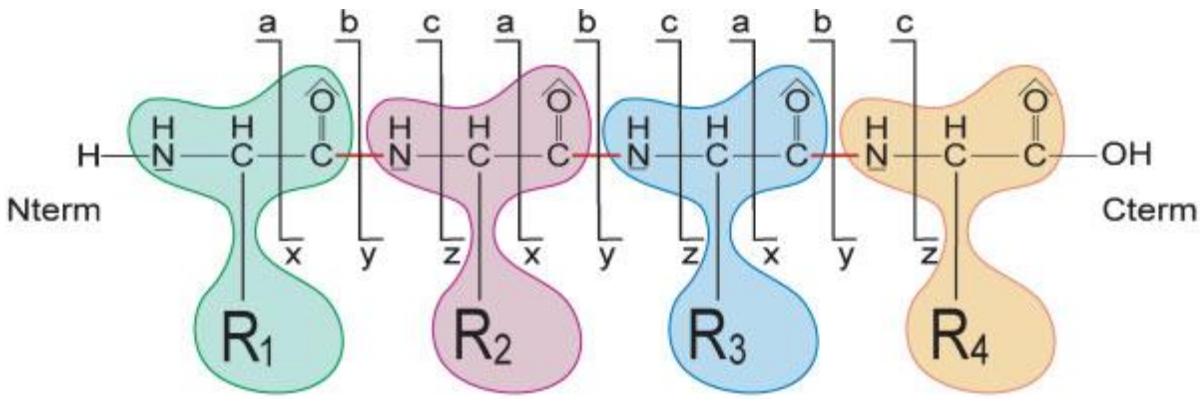



**Figure 5.1. Diagram showing MS/MS fragmentation of a peptide.** Six different kinds of fragments (a, b, c, x, y and z) can be produced for each amino acid in the peptide depending on which bond within the backbone is broken. The b and y represent the series of ions produced after breakage of the amide bond and are the ions most frequently seen in the mass analyser. Adapted from Hernandez *et al* (2006).

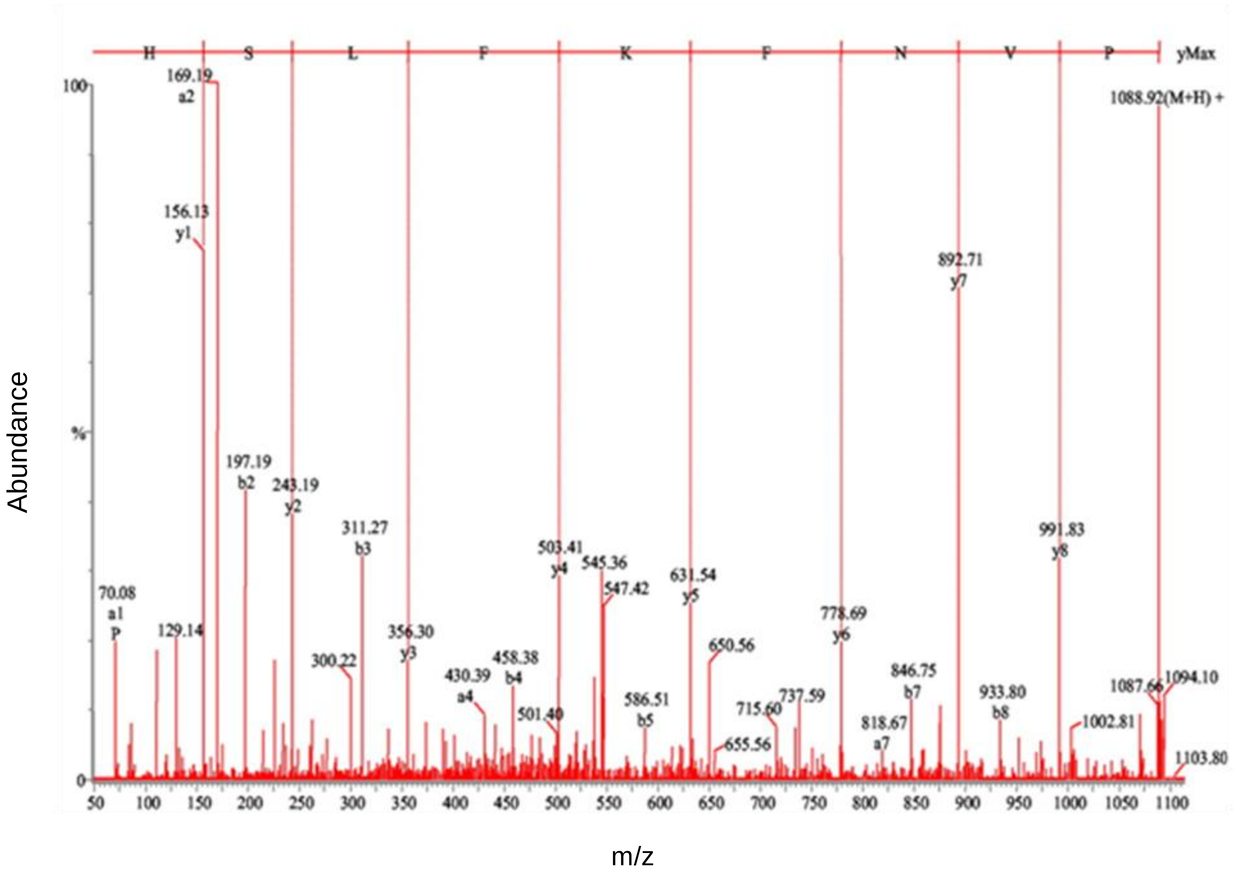



**Figure 5.2. Representative output of a typical MS/MS spectrum.** Output for a peptide with the sequence PVNFKFLSH is presented here. The peptide was identified with the y series of ions, with most of the b series and some of the a series also present. Adapted from Rioli *et al* (2003).

### 5.1.5 Previous work on lipid raft proteomics

There are a growing number of proteomics projects aimed at the identification of proteins within lipid rafts, which have uncovered a large number of genuine raft proteins within a number of model systems. Commonly identified proteins include cytoskeletal proteins such as F-actin, raft associated proteins such as hsp90 and lectins, V-ATPase, proteins involved in signal transduction such as Gα subunits, and GPI anchored proteins (Bini *et al.*, 2003; Foster *et al.*, 2003; Insenser *et al.*, 2006; Kim *et al.*, 2009; Li *et al.*, 2004a; von Haller *et al.*, 2001). An interesting result of these studies is the presence of mitochondria proteins in most of the studies, with some reporting as much as 24% of the total raft proteins identified as mitochondrial, prompting some researchers to suggest the existence of lipid rafts in this organelle (Bae *et al.*, 2004; Mellgren, 2008). Experiments with cholesterol depletion however do not support the notion that mitochondrial proteins are present within rafts (Foster, 2008; Zheng *et al.*, 2009). Nuclear membrane proteins have also been identified in these studies, which have been suggested to be a common contaminant of lipid raft preparations (Say and Hooper, 2007). In general, the ubiquitous nature of proteomic analysis means that the presence of some minor contaminating identification is expected in the final result.

### 5.1.6 GPI anchored protein proteomics

There have been a small number of proteomics studies of GPI anchored proteins, which were mainly carried out in humans, *Arabidopsis thaliana* and the malaria parasite *Plasmodium falciparum*. In *P. falciparum* 26 GPI anchored proteins were identified by PIPLC release (Gilson *et al.*, 2006). More than 40 proteins were found in *Arabidopsis* from a number of studies (Borner *et al.*, 2003; Elortza *et al.*, 2006; Elortza *et al.*, 2003). Elorza *et al.* analysed human HeLa cells in two studies with the release of GPI anchored proteins by PIPLC and PIPLD. PIPLC digestion yielded 6 protein identifications which included several known GPI anchored proteins, which are alkaline phosphatase, carboxypeptidase M, CD55 and CD59 (Elortza *et al.*, 2003). PIPLD treatment identified 5 more proteins, bringing the total of GPI anchored proteins identified in humans to 11 (Elortza *et al.*, 2006).

*In silico* prediction programs were also used complementarily as a part of the proteomic analyses of GPI anchored proteins. There was broad agreement between prediction programs and experimental results for human and *Arabidopsis*, with the use of multiple prediction programs found to be necessary to gain a comprehensive validation for the proteins (Elortza *et al.*, 2006). Protein prediction however matched poorly with results from *P. falciparum*, which may have been due to the phylogenetically distant protein training sets used for the prediction programs that made them less compatible with the *P. falciparum* genome (Gilson *et al.*, 2006).

### 5.1.7 Outline for this chapter

In this chapter the identification of lipid raft and GPI anchored proteins extracted from *C. elegans* is presented. Proteomic analysis of lipid raft proteins was carried out with the MDLC shotgun method at the Cambridge Centre for Proteomics (CCP), with SDS-PAGE separation of the proteins in the first dimension, followed by subsequent digestion of proteins with trypsin, before a second dimension separation with RPLC and final sequencing of the peptides with MS/MS. Protein identification was performed with MASCOT (www.matrixscience.com) by an in-house server at the CCP. Overall 41 proteins were identified from the preparation of lipid rafts from *C. elegans*. Three GPI anchored proteins were also identified with a combination of 1D and 2D electrophoresis followed by PMF. The identified proteins were also validated with GPI anchoring prediction programs. To date this is the largest analysis of lipid raft and GPI anchored proteins in the nematode *C. elegans*, and paves the way for further analysis of these two important classes of proteins within this model organism.

## 5.2 Method

### 5.2.1 Trichloroacetic acid (TCA) precipitation

Trichloroacetic acid (25 µl, 100% w/v) solution was added to 100 µl of a protein mixture and incubated at 4 °C for 10 min. The solution was centrifuged at approx. 10,000 g (13,000 rpm) on a table top microcentrifuge for 5 min and the supernatant was removed. The pellet was then washed with 200 µl of acetone at 4°C and spun at 13,000 rpm on a table top centrifuge for 5 min, the acetone discarded and the pellet washed again in the same manner. The pellet was finally dried at 95°C for 5 – 10 min.

### 5.2.2 1D-electrophoresis

The 1D-electrophoresis protocol was carried out as per instructions from Chapter 4.2.10.

### 5.2.3 2D-electrophoresis

Protein samples precipitated with TCA was solubilised in rehydration buffer (7 M urea, 2 M thiourea, 100 mM DTT, 0.5% (v/v) ampholytes, 4% (w/v) CHAPS, 1% (v/v) Triton X-100, trace of bromophenol blue) and equilibrated overnight on IPG strips (24 cm, pH 3-10, Bio-Rad). IEF was performed on a Protean IEF system (BioRad) at 8,000 V for 70,000 Vh. IEF strips were then incubated in equilibration buffer (50 mM Tris-HCl, pH 8.8, 6 M urea, 20% (v/v) glycerol, 2% (w/v) SDS) containing 0.5% (w/v) DTT for 15 min and again in equilibration buffer containing

4% (w/v) iodoacetamide for 15 min. IPG strips were placed onto precast Criterion 2D gels (8-16% resolving, Bio-Rad) with unstained molecular weight markers added adjacent to the anodic end of the strip and sealed with 1% (w/v) agarose. SDS-PAGE was performed in a Criterion electrophoresis tank (Bio-Rad) at 200 V for 1.5 h. Finished gels were silver stained as described in Chapter 4.2.12.

### 5.2.4 PMF of protein samples

PMF of protein samples was performed by Dr. J. N. Keen at the University of Leeds.

Polypeptide bands of interest from 1D gels were excised using a razor blade and chopped into pieces c. 1-2 $mm^2$. Spots from 2D gels were excised using a micropipette tip. Individual gel pieces were transferred to a microtitre plate for automated digestion using a MassPREP workstation (Waters).

The gel pieces were first subjected to automated destaining using 50 mM ammonium bicarbonate/50% (v/v) acetonitrile (for Coomassie blue stained gel pieces) or freshly-prepared 50 mM sodium thiosulphate/15 mM potassium ferricyanide (for silver-stained gel pieces). The proteins were reduced using 10 mM dithiothreitol (in 100 mM ammonium bicarbonate, 30 min) and alkylated using 55 mM iodoacetamide (in 100 mM ammonium bicarbonate, 20 min); then the gel pieces were washed with 100 mM ammonium bicarbonate and dehydrated using acetonitrile prior to the addition of 25 µl trypsin (Promega) solution (6 ng/µl in 50 mM ammonium bicarbonate). Digestion was allowed to proceed for 5 h at 37 ºC. Peptides were then extracted using 30 µl 1% (v/v) formic acid/2% (v/v) acetonitrile and an aliquot (1 µl) applied to a stainless steel MALDI plate together with 1 µl matrix solution (2 mg/ml α-cyano-4-hydroxycinnamic acid in 60% (v/v) acetonitrile/0.08% aqueous TFA). The dried plate was transferred to a mass spectrometer (M@LDI L/R, Waters) and each digest was analysed in reflectron mode using standard operating parameters. Briefly, the instrument used a $N^2$ laser at 337 nm, source voltage was set at 15000 V, microchannel plate detector voltage was set at 1950 V, pulse voltage was set at 2450 V, reflectron voltage was set at 2000 V, coarse laser energy was set to medium, with fine adjustment used for each sample to optimize signal. At least 100 laser shots were accumulated and combined to produce a raw spectrum. Spectra were processed (background subtraction, smoothing and peak centroiding) and calibrated externally using a tryptic digest of alcohol dehydrogenase and then internally using a trypsin autolysis product (m/z 2211.105 or 1045.564) as a "lockmass" point.

The set of monoisotopic peptide masses for each sample was used to search the SwissProt and/or NCBInr databases using the Mascot search engine (http://www.matrixscience.com) in order to identify the parent protein. Searches were typically performed using an unrestricted protein molecular mass range, variable modifications of carbamidomethyl-Cys, propionamido-Cys and oxidized-Met, searching tryptic peptides from all species, allowing one missed cleavage site and 100 ppm error tolerance in the peptide mass.

### 5.2.5 LC MS/MS

The LC MS/MS protocol was performed by Dr. Michael J Deery at the Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge.

Two aliquots of 50 µl of lipid raft proteins were ultracentrifuged at 50,000g and the pellet retained, which yielded approximately 75 µg of lipid raft proteins each. The pellets were dissolved in SDS sample buffer, run on a 1D gel in lanes 1 and 3 and visualised with Coomasie Blue staining. Ten gel bands were excised from lane 3 of the gel and transferred into a 96-well PCR plate with the labels 3a to 3j. Sample preparation was performed in a Mass Prep Station (Micromass, UK). The gel bands were destained, reduced with DTT, alkylated with iodoacetamide and digested with trypsin at 37°C overnight. Digested supernatant (10 µl) was loaded onto an autosampler for LC-MS/MS analysis using an Eksigent NanoLC-1D Plus (Eksigent Technologies, Dublin, CA) HPLC system and an LTQ Orbitrap mass spectrometer (ThermoFisher, Waltham, MA). Reverse-phase chromatography was used to separate the peptides at a flow rate of 300 nl/min in an LC-Packings (Dionex, Sunnyvale, CA) PepMap 100 column (C18, 75 µm i.d. x 150 mm, 3 µm particle size). Peptides were loaded onto a precolumn (Dionex Acclaim PepMap 100 C18, 5 µm particle size, 100 A, 300 µm i.d x 5mm) from the autosampler with 0.1% formic acid for 5 minutes at a flow rate of 10 µl/min. The ten port valve was then switched to allow peptide elution from the precolumn onto the analytical column. A mixture of solvent A (0.1% formic acid in HPLC grade $H_2O$) and solvent B (0.1% formic acid in acetonitrile) was used to elute the peptides with a gradient of 5-50% solution B in 40 minutes. The eluted peptides were sprayed into the mass spectrometer with a New Objective nanospray source. All the m/z values of eluted ions were measured at a resolution of 7500 in the Orbitrap mass analyzer. Peptide ions with charge states of 2+ and 3+ were isolated and fragmented in the LTQ linear ion trap by collision-induced dissociation and MS/MS spectra were taken from the peptides. Spectral data was analyzed using Bioworks Browser (version 3.3.1 SP1, ThermoFisher) by conversion to dta (text) files using the Sequest Batch Search tool (within Bioworks), which was then converted to a single mgf file using a SSH script in the SSH Secure Shell Client program (Version 3.2.9 Build 283, SSH Communications Corp.). Lastly the combined files were submitted to the Mascot search algorithm (Matrix Science, London UK) with a fixed modification of carbamidomethyl and a variable modification of oxidation (M) and searched against the Wormbase database for protein identification.

### 5.2.6 Western blot of DAF-21 protein

Western blot protocol was adapted from the method used in Chapter 4. 1D gel of protein samples were transferred to nitrocellulose membranes (0.45 μm pore size, Amersham Hybond ECL) at 100 V at 4°C for 2 hours while stirred. Membranes were then incubated in primary antibody solution (primary antibody raised in rabbit to a recombinant protein of the C-terminal 238 amino acid sequence of *B. pahangi* HSP90, known to cross react with *C. elegans* DAF-21 (Devaney *et al.*, 2005), courtesy of Prof. Devaney at the University of Glasgow) at 1/1000 concentration in PBST (137 mM NaCl, 12 mM phosphate, 2.7 mM KCl, pH 7.4, 0.05% Tween-20) with 4% powdered milk and at 4°C overnight. Membranes were washed for 10 minutes in PBST three times. The membranes were then incubated in rabbit secondary antibody conjugated to horse-radish peroxidase (HRP) at 1/10,000 concentration in PBST and 4% powdered milk for 2 hours at room temperature. The membranes were washed again in PBST for 10 minutes for three times. Washed membranes were incubated in 1.5 ml of HRP detection reagent (SuperSignal West Pico, Pierce) for 5 minutes at room temperature and visualised with CCD detection camera.

## 5.3 Results

### 5.3.1 1D gel electrophoresis and PMF identification of proteins

Protein identification with PMF involves the elucidation of an accurate tryptic digest for a single protein, which requires the protein to be separated at a high enough resolution to minimise cross contamination. Proteins from the PIPLC released fraction were subjected to1D gel electrophoresis, which after silver staining appeared to be of sufficiently low complexity to sequence with PMF (Figure. 5.3). Bands were cut from the 1D gel of PIPLC releasate and analysed with MALDI MS, two of which were identified as *C. elegans* ZK6.11 and DOD-19 (Figure. 5.4). ZK6.11 was predicted to be GPI anchored by all four prediction programs, and DOD-19 had GPI anchor predictions in both GPI-SOM and PredGPI (Table. 5.1). The 1D gel for the lipid raft membrane fraction was deemed to have an insufficient resolution, and was therefore not subjected to PMF analysis (Figure. 5.3).

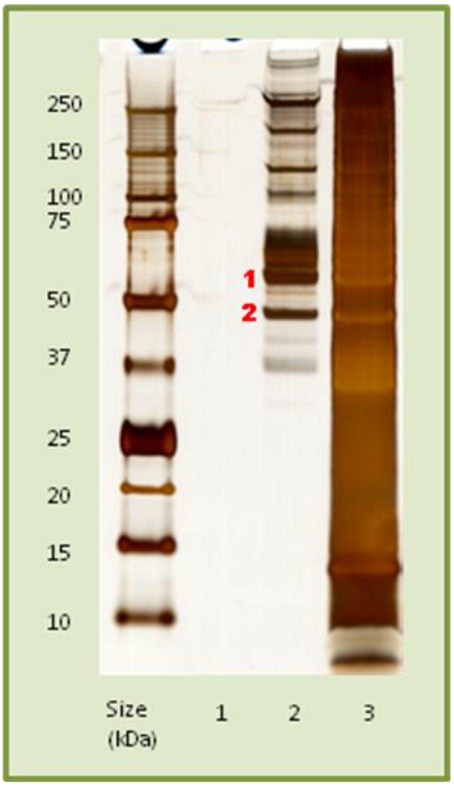
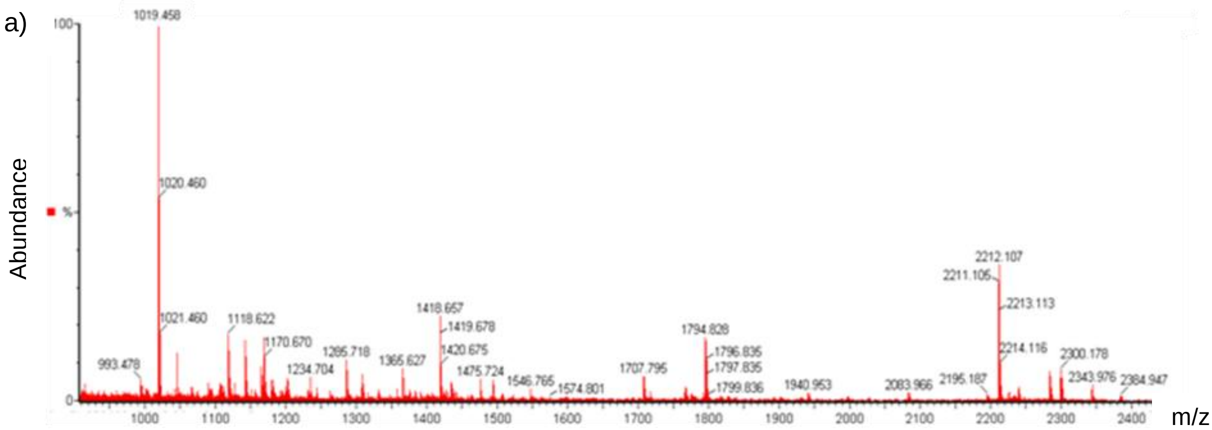



**Figure 5.3. 1D gel of lipid raft and PIPLC released proteins used for proteomic analysis.** Samples were visualised with silver stain. Lane 1: control releasate after incubation of raft fraction with $dH_2O$, lane 2: PIPLC released fraction, lane 3: lipid raft proteins. Two bands from the PIPLC released fraction (labelled 1 and 2 in red) were identified with PMF. Proteins from the lipid raft fraction were separated insufficiently for MS analysis.

| Band number | Protein identified | Score | Peptides matched | Sequence coverage (%) | Molecular weight (KDa) | p*I* | GPI anchor prediction | Big PI | GPI SOM | Frag Anchor | Pred GPI |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | dod-19 | 92 | 10 | 18 | 48.4 | 8.87 | y | ● | ● | ● | ● |
| 2 | ZK6.11a | 108 | 11 | 27 | 42.4 | 8.62 | y | ○ | ● | ○ | ● |

**Table 5.1. PMF protein identifications from 1D gel.** Proteins from bands 1 and 2 were identified following tryptic digest and MS analysis with a MALDI instrument. MASCOT search score, number of peptides used in the identification, percentage sequence coverage, molecular weight, p*I* values, prediction for GPI anchoring (predicted with two or more prediction programs), and the prediction result of the four programs (filled circle ● represents positive prediction while open circle ○ represents negative prediction) for each of the identified proteins is presented. Score is -10 log($p$) where $p$ is the probability of the match is a random event. Scores at > 50 indicate identification of the sequence at $p < 0.05$.

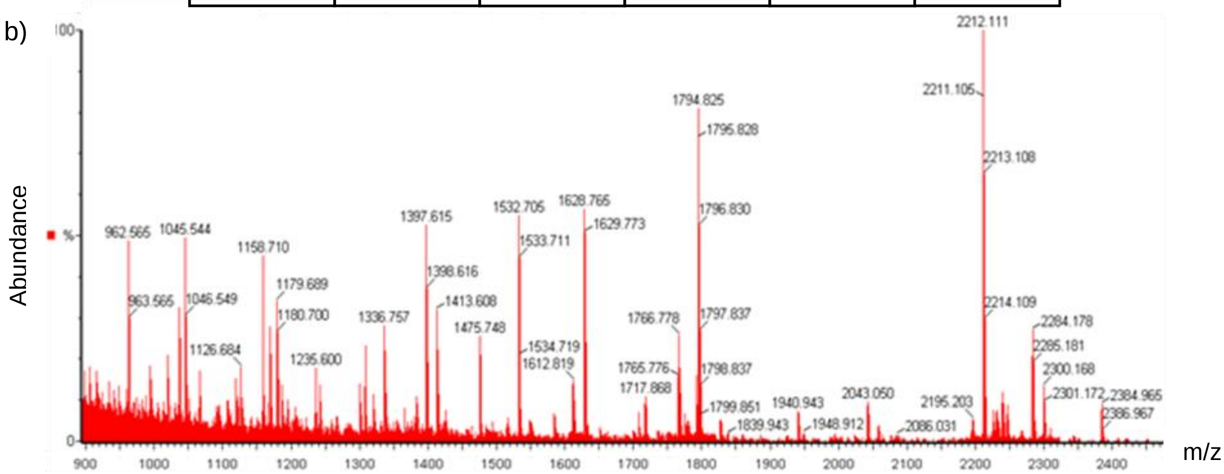
a)




Peptide masses (m/z) used in PMF for band 1 (DOD-19)

| | | | | | |
|---|---|---|---|---|---|
| 915.702 | 993.477 | 1001.484 | **1019.461** | **1041.471** | **1142.683** |
| **1169.667** | **1181.740** | 1202.646 | **1285.723** | **1418.656** | **1434.697** |
| **1546.762** | **1902.034** | 1929.965 | 1996.994 | 2342.982 | |

b)



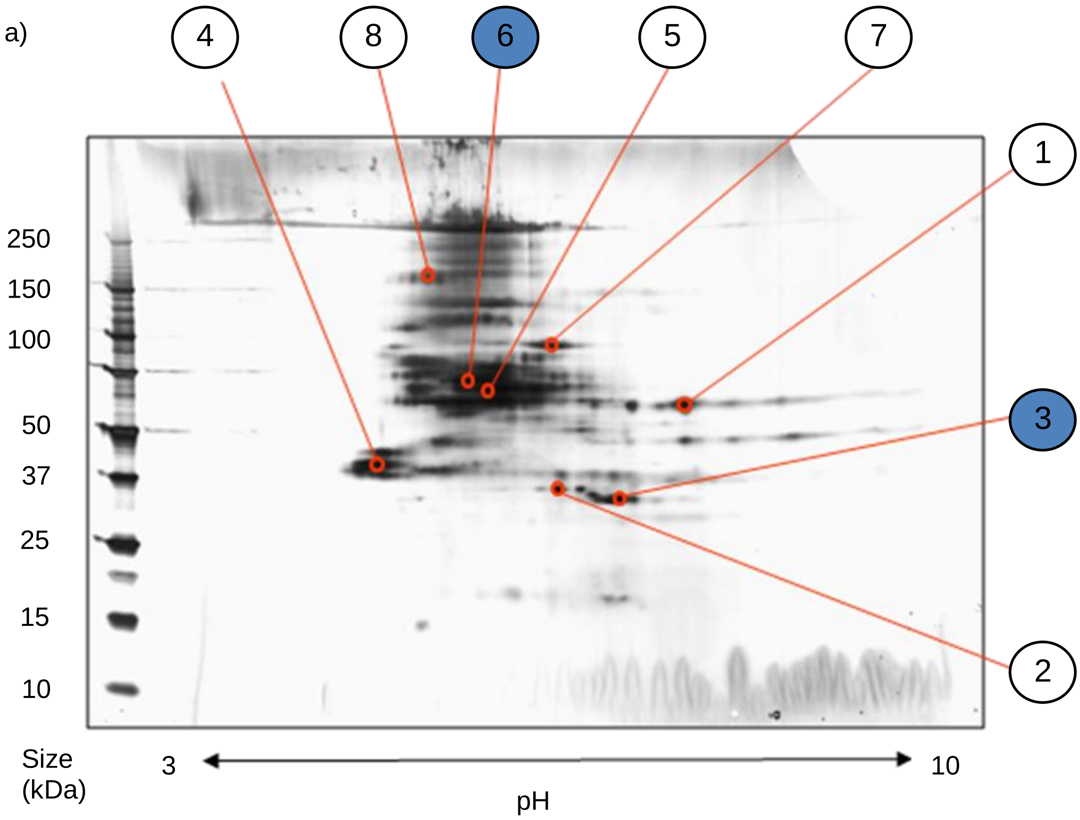

Peptide masses (m/z) used in PMF for band 2 (ZK6.11)

| | | | | | |
|---|---|---|---|---|---|
| **962.568** | **1019.465** | 1036.493 | 1158.711 | **1169.674** | **1185.727** |
| 1242.608 | **1336.757** | **1397.613** | **1413.600** | **1532.702** | 1611.767 |
| **1628.763** | **2042.039** | **2058.021** | | | |

**Figure 5.4. Output of MALDI MS data for PIPLC released protein bands from 1D gel.** Results for spot 1 are shown in (a) and Results for spot 2 are shown in (b). Mass peaks correspond to the m/z value of tryptically digested peptides. After automated and manual removal of common peaks, eg. from trypsin and contaminating keratin, the remaining sets of m/z values (shown in table for each band) were subjected to PMF analysis with MASCOT. Peptide masses in red were found to be part of the PMF that identified the protein. Ten peptides were used in the identification of DOD-19 from band 1 and eleven peptides were used in the identification of ZK6.11 from band 2.

### 5.3.2 2D gel electrophoresis and PMF identification of proteins

Both the PIPLC released fraction and the lipid raft fractions were subsequently analyzed with 2D gel electrophoresis to improve the resolution of the proteins for identification (Figure. 5.5). The resolution of the PIPLC releasate was better than the lipid raft samples, which may have been due to the relatively hydrophobic nature of the proteins of the lipid raft causing streaking within the IEF strip. Eight spots were taken from each gel and digested with trypsin for PMF analysis. Spots 1-8 were assigned to the PIPLC released proteins and spots 9-16 were assigned to the lipid raft fraction (Table. 5.2). Two *C. elegans* proteins were identified from the PIPLC released fractions, which were LEC-2 and F56F10.1 from spots 3 and 6, respectively. LEC-2 (galectin 2) is a cytosolic protein that is not predicted to be GPI anchored while F56F10.1 is a putative serine protease that contains GPI anchored predictions from four predictive programs. Four of the spots from the lipid raft samples contained *C. elegans* protein identifications, with spot 10, 11 and 12 identified as LEC-3, LEC-2 and LEC-4 respectively, and spot 16 identified a mixture of LEC-4, LEC-2 and LEC-1. Keratin contamination was present in spots 2 and 13, while all other spots produced insufficient data for identification.



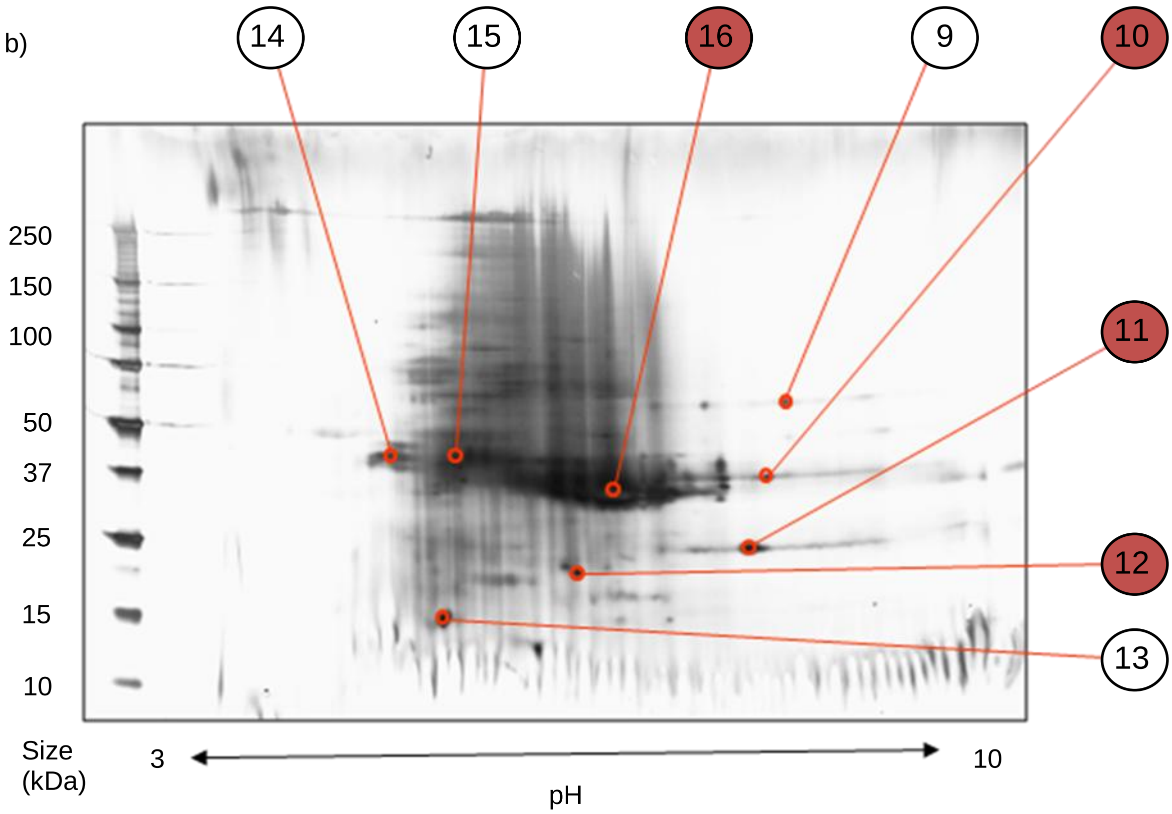




**Figure 5.5. 2D electrophoretic gels for proteomics analysis.** a) shows the gel for PIPLC released proteins and b) shows the gel for lipid raft proteins. Both of the gels were run with a 3 to 10 pH gradient in IEF. Eight protein spots were excised from each gel for tryptic digestion and P.M.F. analysis with spots 1-8 from the PIPLC release and spots 9- 16 from the lipid raft fraction. Circles indicate the position of excised spots and filled circles indicate spots with positive *C. elegans* protein identification.

a)

| Spot number | Protein identified | Score | Peptides matched | Sequence coverage (%) | Molecular weight (KDa) | p*I* | GPI anchor prediction |
|---|---|---|---|---|---|---|---|
| 1 | None | n/a | n/a | n/a | n/a | n/a | n/a |
| 2 | K2C1, Human Keratin | 80 | 22 | 38 | 66.0 | 8.16 | n/a |
| **3** | **LEC-2, *C. elegans*** | **116** | **10** | **50** | **31.3** | **6.19** | **n** |
| 4 | None | n/a | n/a | n/a | n/a | n/a | n/a |
| 5 | None | n/a | n/a | n/a | n/a | n/a | n/a |
| **6** | **F56F10.1, *C. elegans*** | **56** | **7** | **14** | **60.5** | **5.31** | **y** |
| 7 | None | n/a | n/a | n/a | n/a | n/a | n/a |
| 8 | None | n/a | n/a | n/a | n/a | n/a | n/a |

b)

| Spot number | Protein identified | Score | Peptides matched | Sequence coverage (%) | Molecular weight (KDa) | p*I* |
|---|---|---|---|---|---|---|
| 9 | None | n/a | n/a | n/a | n/a | n/a |
| **10** | **LEC-3, *C. elegans*** | **176** | **11** | **37** | **32.4** | **6.82** |
| **11** | **LEC-2, *C. elegans*** | **112** | **9** | **35** | **31.3** | **6.19** |
| **12** | **LEC-4, *C. elegans*** | **95** | **9** | **28** | **32.4** | **6.02** |
| 13 | K2C1, Human keratin | 70 | 17 | 33 | 66.0 | 8.16 |
| 14 | None | n/a | n/a | n/a | n/a | n/a |
| 15 | None | n/a | n/a | n/a | n/a | n/a |
| **16** | **LEC-4, *C. elegans*** | **100** | **15** | **52** | **32.4** | **6.02** |
| | **LEC-2, *C. elegans*** | **63** | **10** | **41** | **31.3** | **6.19** |
| | **LEC-1, *C. elegans*** | **63** | **10** | **37** | **31.8** | **6.12** |

**Table 5.2. PMF analysis results of protein spots from 2D gels.** Table a) shows results from the PIPLC released sample and table b) shows results from the lipid raft sample. MASCOT search score, number of peptides used in the identification, percentage sequence coverage, molecular weight, p*I* values and prediction for GPI anchoring (predicted with two or more prediction programs in Chapter2) for each of the identified proteins is presented. The relevant *C. elegans* proteins are highlighted for each table. Spot number 16 contained multiple identifications that may have been the result of incomplete separation during IEF. Score is -10 log(*p*) where *p* is the probability of the match is a random event. Scores at > 50 indicate identification of the protein at $p < 0.05$.

### 5.3.3 Liquid chromatography and MS/MS

Lipid raft proteins were sent to the Cambridge Centre for Proteomics (CCP) to be sequenced using LC MS/MS. Two protein samples were sent to the centre and separated on a 1D gel (Figure. 5.6). The SDS gel of sample 3 was cut into 10 strips (labelled bands 1 to 10) with each strip digested with trypsin and subjected to reverse phase chromatography, which eluted directly into an Orbitrap mass analyzer for MS/MS sequencing. A total of 287 proteins were identified from the raw proteomic analysis with significant hits from one or more peptides (Appendix 3). Many of the identified proteins contained non-significant and duplicated peptides in the analysis (Figure. 5.7), and this led to the imposition of a minimum of two or more unique peptides as a criterion for the positive identification of a protein. Forty five proteins from the initial list were found to satisfy this criterion, of which F52H3.7a and F52H3.7b were found to have the same set of identified peptides and encode for the same LEC-2 protein. F52H3.7b was chosen over F52H3.7a as it contained a larger number of uniquely identified peptides. The final list of validated proteins from LC MS/MS analysis was 44 (Table 5.3).

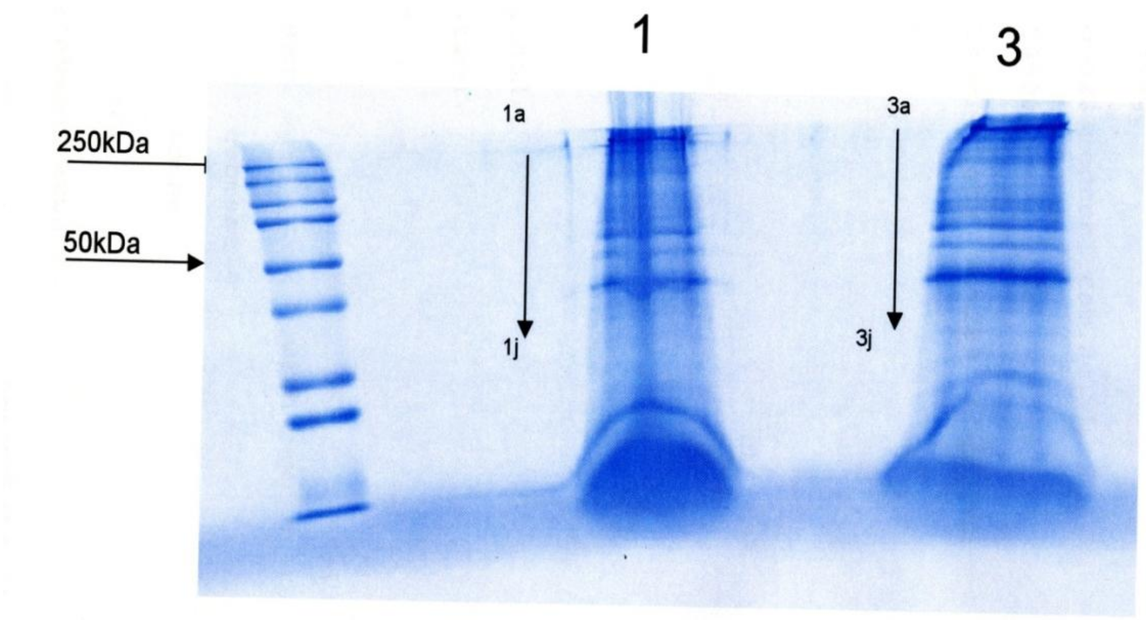



**Figure 5.6. 1D gel of lipid raft proteins used for LC MS/MS.** Lanes labelled 1 and 3 were identical samples sent to the CCP. Proteins from lane 3 were cut into ten gel bands (3a to 3j, labelled 1 to 10 respectively) which were subsequently digested with trypsin and separated with an RP column for MS/MS analysis.

a)

K10C2.1 CE37128 WBGene00019617 serine Mass: 258383 Score: 783 Queries matched: 18 emPAI: 0.18

K10C2.1 CE37128 WBGene00019617 serine carboxypeptidase status:Partially_confirmed UniProt:Q94269 protein_id:AAK39256.2

☐ Check to include this hit in error tolerant search or archive report

| | Query | Observed | Mr(expt) | Mr(calc) | Delta | Miss | Score | Expect | Rank | Peptide |
|---|---|---|---|---|---|---|---|---|---|---|
| ☑ | 54 | 423.2322 | 844.4499 | 844.4403 | 0.0096 | 0 | 39 | 0.022 | 1 | K.DNGLAVTR.Q |
| ☑ | 71 | 443.7459 | 885.4773 | 885.4668 | 0.0104 | 0 | (35) | 0.032 | 1 | K.VADLGQQR.F |
| ☑ | 72 | 443.7474 | 885.4802 | 885.4668 | 0.0134 | 0 | 61 | 9e-05 | 1 | K.VADLGQQR.F |
| ☑ | 123 | 508.2905 | 1014.5664 | 1014.5498 | 0.0165 | 0 | 46 | 0.0032 | 1 | R.SQFLAPPQK.T |
| ☑ | 151 | 548.8067 | 1095.5988 | 1095.5812 | 0.0177 | 0 | 70 | 8e-06 | 1 | R.TATDTYLALK.D |
| ☑ | 197 | 603.3633 | 1204.7120 | 1204.6928 | 0.0192 | 0 | 55 | 0.00035 | 1 | K.AAHILIIDSPR.G |
| ☑ | 266 | 700.8588 | 1399.7031 | 1399.6772 | 0.0259 | 0 | 72 | 5.8e-06 | 1 | K.TLFENVYSWNK.A |
| ☑ | 317 | 796.8903 | 1591.7660 | 1591.7447 | 0.0213 | 0 | 102 | 4.5e-09 | 1 | R.GMGIGNGMVSAVNDVR.T + Oxidation (M) |
| ☑ | 322 | 804.8899 | 1607.7652 | 1607.7396 | 0.0256 | 0 | (62) | 5.5e-05 | 1 | R.GMGIGNGMVSAVNDVR.T + 2 Oxidation (M) |
| ☑ | 327 | 811.4589 | 1620.9033 | 1620.8664 | 0.0369 | 0 | 66 | 1.8e-05 | 1 | R.VWNLPGITYGLNFK.Q |
| ☑ | 328 | 811.4605 | 1620.9065 | 1620.8664 | 0.0400 | 0 | (32) | 0.044 | 1 | R.VWNLPGITYGLNFK.Q |
| ☑ | 338 | 853.9973 | 1705.9800 | 1705.9515 | 0.0285 | 0 | (63) | 3.7e-05 | 1 | K.QLLPQYQPAPVTVPR.R |
| ☑ | 339 | 853.9982 | 1705.9818 | 1705.9515 | 0.0302 | 0 | 79 | 1e-06 | 1 | K.QLLPQYQPAPVTVPR.R |
| ☑ | 366 | 953.5225 | 1905.0304 | 1904.9884 | 0.0420 | 0 | (82) | 4.7e-07 | 1 | R.AADVSPFLPSTLFVDQAK.K |
| ☑ | 367 | 953.5243 | 1905.0340 | 1904.9884 | 0.0456 | 0 | 83 | 3e-07 | 1 | R.AADVSPFLPSTLFVDQAK.K |
| ☑ | 448 | 786.4055 | 2356.1946 | 2356.1376 | 0.0570 | 0 | 40 | 0.0058 | 1 | K.TALDTYTALEDFFVTYPPHR.N |
| ☑ | 492 | 1441.7382 | 2881.4618 | 2881.4215 | 0.0403 | 0 | 71 | 3.2e-06 | 1 | K.YYIQQYPDTTPVFQFLVDSGYPLK.V |
| ☑ | 493 | 1441.7449 | 2881.4752 | 2881.4215 | 0.0537 | 0 | (26) | 0.092 | 1 | K.YYIQQYPDTTPVFQFLVDSGYPLK.V |

b)

K11C4.5 CE41827 WBGene00006801 locus:unc-68 ryanodine Mass: 595342 Score: 47 Queries matched: 3

K11C4.5 CE41827 WBGene00006801 locus:unc-68 ryanodine receptor status:Partially_confirmed UniProt:Q94279 protein_id:AAB18318.3

☐ Check to include this hit in error tolerant search or archive report

| | Query | Observed | Mr(expt) | Mr(calc) | Delta | Miss | Score | Expect | Rank | Peptide |
|---|---|---|---|---|---|---|---|---|---|---|
| | 100 | 467.7774 | 933.5402 | 933.4590 | 0.0812 | 0 | (21) | 0.92 | 3 | K.VMNDLNTK.G |
| | 101 | 467.7775 | 933.5404 | 933.4590 | 0.0815 | 0 | 27 | 0.26 | 4 | K.VMNDLNTK.G |
| ☑ | 348 | 738.0497 | 2211.1272 | 2212.0405 | -0.9133 | 1 | 22 | 0.32 | 1 | K.DMVNAAEKMAEHSHLIWAK.K + 2 Oxidation (M) |

**Figure 5.7. (Previous page) Examples of MS/MS output from MASCOT.** The first line contains the name of the protein followed by its molecular weight, a non-probabilistic protein score derived from the ions scores and the number of peptide matches. An Exponentially Modified Protein Abundance Index (emPAI) value for an estimation of quantitation is provided if the number of queries is 100 or more. The table columns contain values for each individual peptide assigned to the protein, and starts off with the hyperlinked number of the peptide, followed by its experimental m/z value, molecular mass calculated from m/z, calculated relative molecular mass, difference (error) between experimental and calculated masses, number of missed enzyme cleavage sites, ions score (calculated as -10*Log(*p*), where individual ions scores > 33 indicate identity or extensive homology ($p<0.05$); duplicated matches with lower scoring are shown in brackets), expectation probability for the peptide match (significance $p<$ 0.05), rank of the ion (1 to 10,where 1 is the best match), and sequence of the peptide (residues adjacent to the peptide are shown either side of the periods. Modifications of any residues are underlined and listed after the sequence).

a) Results for the protein K10C2.1 from band 4. MASCOT attributes 18 peptides to the protein in which some are duplicates and others have an expected probability of >0.05. After inspection 12 unique peptides (K.DNGLAVTR.Q, K.VADLGQQR.F, R.SQFLAPPQK.T, R.TATDTYLALK.D, K.AAHILIIDSPR.G, K.TLFENVYSWNK.A, R.GMGIGNG<u>M</u>VSAVNDVR.T, R.VWNLPGITYGLNFK.Q, K.QLLPQYQPAPVTVPR.R, R.AADVSPFLPSTLFVDQAK.K, K.TALDTYTALEDFFVTYPPHR.N, and K.YYIQQYPDTTPVFQFLVDSGYPLK.V) were found that also had significant probability scores for each of them.
b) Results for the protein K11C4.5 from band 1, which is also known as unc-68 and encodes a ryanodine receptor in *C. elegans.* MASCOT assigned 3 peptides to the protein. Closer inspection, however, uncovers two identical peptide sequences within the analysis. Furthermore none of the peptides had a significant expected probability. This means that there are no unique significant peptides assigned to the protein and K11C4.5 is not counted towards the final total of identified lipid raft proteins.

| Public name | description | unique peptides | score | size | GPI prediction | Big PI | GPI SOM | Frag Anchor | Pred GPI |
|---|---|---|---|---|---|---|---|---|---|
| tag-10 | apical gut membrane protein | 3 | 245 | 473 | y | ○ | ● | ● | ● |
| act-4 | cytoskeleton | 3 | 169 | 332 | n | ○ | ○ | ○ | ○ |
| act-4 | cytoskeleton | 2 | 145 | 376 | n | ○ | ○ | ○ | ○ |
| daf-21 | molecular chaperone | 2 | 140 | 702 | n | ○ | ○ | ○ | ○ |
| pho-1 | phosphatase | 6 | 348 | 449 | y | ○ | ● | ● | ○ |
| F56F10.1 | carboxypeptidase | 4 | 358 | 540 | y | ● | ● | ● | ● |
| F32A5.3 | carboxypeptidase | 6 | 373 | 574 | y | ● | ● | ● | ● |
| K10C2.1 | carboxypeptidase | 12 | 783 | 2314 | y | ● | ● | ● | ● |
| Y16B4A.2 | carboxypeptidase | 8 | 597 | 2167 | y | ○ | ● | ○ | ● |
| Y40D12A.2 | carboxypeptidase | 2 | 107 | 512 | y | ○ | ● | ● | ● |
| pcp-2 | lysosomal carboxypeptidase | 5 | 449 | 1080 | y | ○ | ● | ● | ● |
| pcp-3 | lysosomal carboxypeptidase | 8 | 593 | 1080 | y | ● | ● | ● | ● |
| pcp-4 | lysosomal carboxypeptidase | 4 | 354 | 1042 | y | ○ | ● | ○ | ● |
| C26B9.5 | lysosomal carboxypeptidase | 4 | 252 | 516 | n | ○ | ○ | ○ | ○ |
| T25B6.2 | metalloprotease | 2 | 107 | 798 | n | ○ | ○ | ○ | ○ |
| F54F11.2 | metalloprotease | 11 | 734 | 1589 | n | ○ | ○ | ○ | ○ |
| dct-17 | insulin pathway daf-16 controlled proteins | 2 | 225 | 739 | y | ○ | ● | ● | ● |
| dod-19 | insulin pathway daf-16 controlled proteins | 4 | 333 | 406 | y | ○ | ● | ○ | ● |
| F57F4.4 | unc-68 ryanodine receptor associated proteins ($Ca^{2+}$ pathway) | 5 | 337 | 2090 | y | ● | ● | ● | ● |
| gfi-1 | unc-68 ryanodine receptor associated proteins ($Ca^{2+}$ pathway) | 3 | 181 | 2153 | y | ● | ● | ● | ● |
| lec-1 | galactoside binding lectin | 5 | 340 | 279 | n | ○ | ○ | ○ | ○ |
| lec-2 | galactoside binding lectin | 7 | 487 | 278 | n | ○ | ○ | ○ | ○ |
| lec-4 | galactoside binding lectin | 6 | 399 | 283 | n | ○ | ○ | ○ | ○ |
| lec-5 | galactoside binding lectin | 3 | 262 | 314 | n | ○ | ○ | ○ | ○ |
| tre-3 | sugar metabolism | 2 | 89 | 588 | y | ○ | ● | ○ | ● |
| stl-1 | stomatin like | 3 | 226 | 327 | n | ○ | ○ | ○ | ○ |
| vha-1 | vacuolar proton-translocating ATPase (V-ATPase) | 2 | 130 | 169 | n | ○ | ○ | ○ | ○ |
| vha-19 | vacuolar proton-translocating ATPase (V-ATPase) | 2 | 148 | 451 | n | ○ | ○ | ○ | ○ |
| vps-32.1 | vacuolar protein sorting | 3 | 255 | 221 | n | ○ | ○ | ○ | ○ |
| T19D12.4 | | 3 | 249 | 1028 | n | ○ | ○ | ○ | ○ |
| Y54G2A.18 | | 2 | 134 | 213 | n | ○ | ○ | ○ | ○ |
| C29F3.7 | | 7 | 594 | 491 | n | ○ | ● | ○ | ○ |
| ZK6.11 | | 6 | 395 | 386 | y | ● | ● | ● | ● |
| Y41D4B.16 | | 5 | 347 | 453 | y | ○ | ● | ● | ● |
| F54E2.1 | | 3 | 210 | 391 | y | ○ | ● | ● | ● |
| K08D8.6 | | 3 | 200 | 491 | y | ○ | ● | ● | ● |
| F35E12.10 | | 2 | 125 | 487 | y | ● | ● | ● | ● |
| F53C11.1 | | 2 | 144 | 494 | n | ○ | ● | ○ | ○ |
| B0024.4 | | 2 | 138 | 390 | y | ○ | ● | ● | ● |
| Y12A6A.1 | | 2 | 79 | 209 | y | ● | ● | ● | ● |
| R05G6.7 | channel Protein | 5 | 319 | 283 | n | ○ | ○ | ○ | ○ |
| npp-21 | nuclear Pore complex Protein | 2 | 85 | 1982 | n | ○ | ○ | ○ | ○ |
| eft-4 | translation elongation factor | 2 | 162 | 463 | n | ○ | ○ | ○ | ○ |
| F21D5.3 | copper oxidase | 5 | 338 | 743 | n | ○ | ○ | ○ | ○ |

**Table 5.3. (Previous page) Results of the LC MS/MS analysis of proteins from the lipid raft fraction.** All proteins were identified with two or more unique peptides with statistically significant scores. F52H3.7a was also found in the analysis but contained duplicated peptides with F52H3.7b and encodes the same protein lec-2, and as such was not included in this list. Protein scores are derived from ions scores as a ranking of protein hits on a non-probabilistic basis (Matrix Science). Public name, Wormbase ID, gene description, gene size and GO terms were taken from www.wormbase.org. GPI prediction was taken from Chapter 2 with confirmation when two or more prediction programs have validated the result (highlighted in orange). For each individual prediction program a ● denotes a positive prediction while a ○ indicates a negative prediction.

### 5.3.4 Western blot of lipid raft fraction

A literature search of the 44 lipid raft proteins identified from the LC MS/MS analysis revealed three proteins with available antibodies. These proteins are LEC-1, DAF-21 and VPS-32.1. The LEC-1 antibody was last used in a paper in 1996 and is unavailable from the authors (Arata *et al.*, 1996), while antibodies for both DAF-21 and VPS-32.1 were available from their respective authors (Devaney *et al.*, 2005; Michelet *et al.*, 2009). The DAF-21 antibody was raised against the HSP90 orthologue of the filarial nematode *Brugia pahangi* and was shown to have cross reactivity against the *C. elegans* protein. DAF-21 was shown to be enriched in the lipid raft fraction compared to total membrane (Figure. 5.8). No staining was observed for the VPS-32.1 antibody (data not shown).

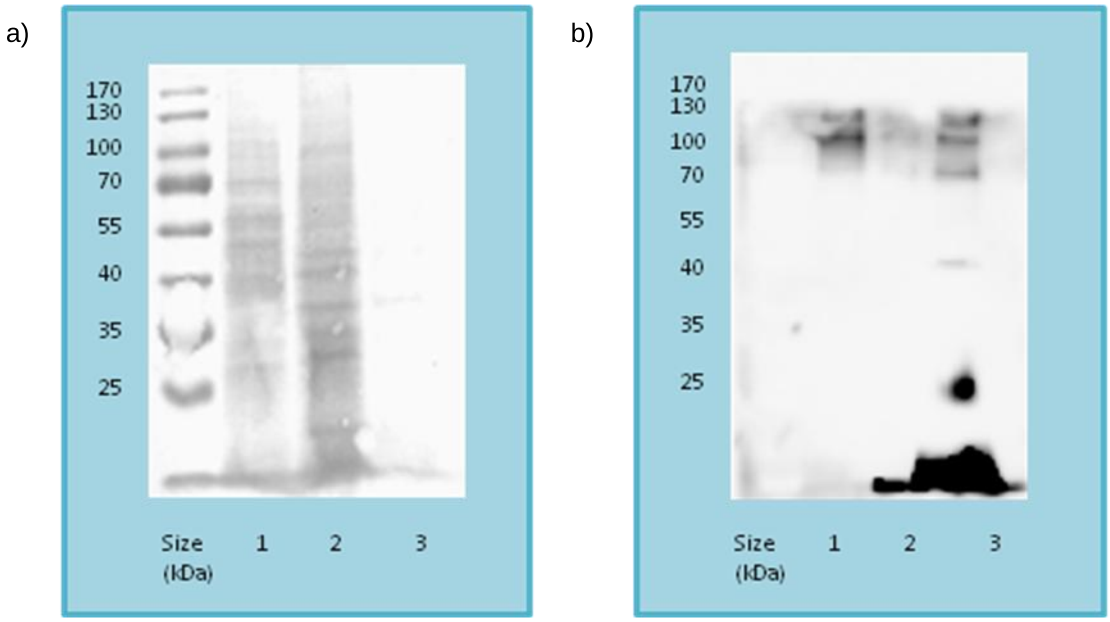



**Figure 5.8. Blot of protein fractions with DAF-21 specific antibody.** a) Ponceau staining of proteins before blot development and b) shows the results of the blot after probing with DAF-21 antibody at 1:1,000 concentration. Lane 1 contains the supernatant fraction after membrane extraction, lane 2 contains a 10 fold dilution of the membrane fraction and lane 3 contains the lipid raft fraction. All protein contents were diluted to their approximate cellular ratios.

## 5.4 Discussion

*C. elegans* lipid raft and GPI anchored proteins were identified and analyzed with proteomic techniques for the first time. Both gel based techniques (including 2D electrophoresis) and multidimensional LC were used to give adequate separation of proteins, which were then subjected to a combination of PMF and MS/MS peptide sequencing for identification.

### 5.4.1 Gel analysis and PMF of GPI anchored proteins and lipid rafts

Both lipid raft and GPI anchored proteins released by PIPLC were initially subjected to 1D and 2D gel electrophoresis for the identification of proteins. Identification with PMF requires high resolution of the protein of interest as the technique is very sensitive to the presence of peptide masses from other contaminating proteins. A preliminary analysis of GPI anchored proteins separated by 1D gels showed that certain bands were sufficiently separated for analysis with PMF, which produced identifications for two *C. elegans* proteins ZK6.11a and ZK6.10 (DOD-19). ZK6.11a is an uncharacterized protein with GPI anchoring prediction in four of the prediction programs used in Chapter 2 (Big PI, GPI SOM, FragAnchor and PredGPI), while *dod-19* (stands for down-stream of *daf-16*) is an unknown gene predicted by two programs (GPI SOM and PredGPI) and is regulated by *daf-16*, which acts within the insulin mediated pathway to affect development in dauer formation, life span and reproduction (Murphy *et al.*, 2003). Bands from the lipid raft fraction however were insufficiently separated due to its higher complexity and no protein identification was attempted from 1D gel analysis.

The GPI anchored proteins and lipid raft proteins were then separated with 2D gel electrophoresis in order to improve resolution and increase the number of proteins that can potentially be identified with PMF. A greater degree of separation was achieved for the released GPI anchored proteins, and individual spots were resolved which showed the presence of many spots with the same mass but different p*I*, which indicates the presence of possible post translational modifications (Figure. 5.5a). Lipid raft proteins were less well separated with 2D gel electrophoresis, with extensive smearing present on the gels (Figure. 5.5b). This may have been due to the first dimension of separation requiring mild non-ionic detergents so as to not interfere with native charge of the protein during isoelectric focusing, which are ill suited for solubilisation of the highly hydrophobic membrane proteins. The presence of raft lipids also compounds the problem. Lipids such as sphingolipids and sphingomyelin have saturated long chain fatty acids within their structure that allow tight packing and further reduce their solubility with mild detergents. Never-the-less the 2D analysis was able to offer greater resolving power that 1D electrophoresis for the lipid raft proteins, which were separated to an appropriate resolution for PMF identification. Overall two spots from the GPI anchored protein gel produced positive *C. elegans* identifications after analysis with mass spectrometry (table 5.2). F56F10.1 was identified from spot 6 and encodes a putative carboxypeptidase with predictions in all four GPI anchor prediction programs. Some mammalian carboxylpeptidases are found to be GPI anchored and are involved in signalling (Reverter *et al.*, 2004). Taken together the evidence suggests that F56F10.1 may be a genuine GPI anchored protein. LEC-2 identified from spot 3 is involved in the binding of sugar moieties on the cell surface of *C. elegans* (Nemoto-Sasaki *et al.*, 2008). LEC proteins are galectins with sugar binding domains and are a class of cytosolic proteins that are strongly associated with membrane sphingolipids. They are found to be a major component of lipid rafts in a number of eukaryotic species (Lajoie *et al.*, 2009). Galectins appear to be highly expressed in *C. elegans* (see below) and are likely to represent a common contamination within the PIPLC released sample due to the method of their extraction. Spot 2 from the GPI anchored protein gel identified human keratin which is a common contaminant of proteomic studies. Four proteins were identified from the lipid raft fraction. LEC-3, LEC-2 and LEC-4 were identified from spots 10, 11 and 12 respectively, while spot 16 contained identification from LEC-1, LEC-2 and LEC-4, with spot 13 identified as a human keratin. All of the LEC proteins were identified with good scores (above 50) and sequence coverage (28% or more) (Table. 5.2). LEC proteins are commonly found within lipid rafts and have been a well validated raft marker in a number of studies in a variety of mammalian cell lines (Hansen *et al.*, 2001; Hsu *et al.*, 2009; Li *et al.*, 2004a). However the level of abundance of the lectins in the sample appeared to have had an adverse effect on the identification of other lipid raft components.

### 5.4.2 2-dimensional LC MS/MS of lipid raft proteins

It was clear from the results of 2D electrophoresis that the technique was unsuitable for lipid raft proteins and a more sensitive method was need for their identification, and 2D LC MS/MS was chosen to separate and analyse the proteins for this fraction. LC based techniques have improved dramatically in the past few years and are able to separate proteomes with a high resolution and fidelity (Motoyama and Yates, 2008). MS/MS analysis is able to achieve direct sequencing of the peptide, which gives greater confidence in the assignment of peptide peaks to proteins and allows improved identification over PMF (Gage *et al.*, 2009). Excess galectin proteins from the lipid raft fraction were washed off with lactose, as the presence of extremely abundant proteins may affect the efficiency of identification by causing ionization suppression and detector saturation within the limited loading capacity of LC columns (Lasonder *et al.*, 2002). 1D SDS-PAGE was used in the first dimension as it allows better separation of the proteins before peptide separation with LC. RPLC was used in the second dimension after trypsin digestion as this technique offers good orthogonality with 1D SDS PAGE, and can be directly linked to the mass spectrometer for MS/MS sequencing with minimal loss of material during handling. A total of 287 proteins were identified with the results from the LC MS/MS, however not all of these proteins have appropriate predictions upon closer inspection. There are instances where several different proteins were predicted with the same peptide (data not shown), which may represent a conserved sequence, and many of the proteins were identified with only one peptide sequence- so called “one hit wonders” that lack specificity and do not present a confident prediction (Figure 5.7). Identification was considered valid when two or more unique peptides with significant sequence identity ($p$<0.05) have been attributed to the protein in question, which is an increasingly common criterion for the validation of MS/MS data in proteomics analyses (Gage *et al.*, 2009). This approach reduced the number of proteins identified with LC MS/MS to 44 for the *C. elegans* lipid raft fraction. Properties for each of the proteins were taken from Wormbase, and a comprehensive analysis of their function and their relationship to known lipid raft components from other systems is given below.

### 5.4.2.1 Apical gut membrane protein

The identified *C. elegans* protein comes from the *tag-10* gene and encodes a gut apical protein with homology to the GA1 gut apical protein of *Heamonchus contortus*, a blood parasite of ruminant animals. GA1 is a polyprotein processed into two isoforms, $p52^{GA1}$ which is GPI anchored and $p46^{GA1}$ which is associated to the membrane with a GPI anchored protein (Jasmer *et al.*, 1996). The protein is being actively developed as a target for vaccine production against the parasite (Jasmer *et al.*, 2007). Previous research has implicated the *tag-10* gene to have homology to the $p52^{GA1}$ form of GA1 (Rehman and Jasmer, 1998) and analysis from chapter 2 has shown that the protein is predicted to be GPI anchored with three prediction programs (GPI SOM, FragAnchor and PredGPI) indicating that TAG-10 may be a GPI anchored protein. GA1 is a part of a group of secreted proteins from *H. contortus*, and has also been identified in a proteomic search of such proteins in the nematode (Yatsuda *et al.*, 2003). Lipid rafts are involved extensively in apical sorting in epithelial cells (Hoekstra *et al.*, 2003) and the *C. elegans cav-2* homologue has also been shown to be localised on the apical surface of the intestine (Parker *et al.*, 2009), which point to the validity of this identification as a genuine raft protein.

### 5.4.2.2 Cytoskeletal protein

Both of the proteins identified are produced by alternate splicing of the *C. elegans* gene *act-4*, namely M03F4.2a and M03F4.2b. Analysis of the peptides that gave rise to these identification revealed unique hits for each of the isoforms, and justifies the inclusion of both on the list of predicted proteins. The actin cytoskeleton is involved in raft formation and maintenance, forming a lattice structure that associates with raft components, which creates greater stability in protein and lipid interactions (Chichili and Rodgers, 2009). Lipid rafts recruit the actin cytoskeleton in maintaining the T cell activation signal (Kabouridis and Jury, 2008), as well as establishing cell polarity in neuron axon growth and fission yeast mating (Kamiguchi, 2006; Wachtler and Balasubramanian, 2006); the actin cytoskeleton has also been shown as a regulator of endocytosis by caveolae (Lajoie and Nabi, 2007).

**5.4.2.3 Molecular chaperone**

*Daf-21* is the *C. elegans* homologue of the mammalian gene heat shock protein 90 (Hsp90). Hsp90 is a well studied cytosolic protein of 90 kDa and is up regulated in conditions of elevated temperature. The protein also has many functions in unstressed conditions, including protein folding, intracellular transport, protein degradation and signalling (Csermely *et al.*, 1998). *HSP90* has a major role in cancer biology, where it prevents apoptosis through stabilization of PI3K/AKT signalling (Mohsin *et al.*, 2005), promotes cancer cell proliferation (Calderwood *et al.*, 2006), induces angiogenesis via phosphorylation of eNOS (Fontana *et al.*, 2002), and has a role in many other key oncogenic processes.

Hsp90 interacts with many lipid raft proteins, particularly those within signalling pathways. The protein associates with the Dengue Virus Receptor within raft domains to facilitate its entry into cells (Reyes-Del Valle *et al.*, 2005). Hsp90 localises the heterotrimeric G protein $G\alpha_{12}$ to lipid rafts, where it functions to produce cytoskeletal rearrangements and induce oncogenic transformation (Waheed and Jones, 2002). Fever induction and maintenance is regulated by *HSP90* in humans, in association with caveolin and the JAK-STAT3 signalling pathway (Shah *et al.*, 2002). Recently it has been found that *HSP90* has a pro-apoptotic role by its interaction with c-Jun N-terminal kinase (JNK) in rafts (Nieto-Miguel *et al.*, 2008). *Daf-21* was also found to be relatively enriched within the raft fraction compared to total membrane via Western blotting (Figure. 5.8), which further validates the protein as a genuine raft component within *C. elegans*. The *C. elegans* caveolin homologues *cav-1* and *cav-2* have been found to be upregulated in heat shock conditions, which implicates a function for lipid rafts in this environmental response for the worm (Parker and Baylis, 2009). *Daf-21* has a number of functions within the worm such as chemosensation, cell cycle control, responses to heat shock and dauer formation (Ailion and Thomas, 2000; Inoue *et al.*, 2006; Vowels and Thomas, 1994; Wang and Kim, 2003). *Daf-21* has also been shown to be involved in a number of signalling pathways such as TGF-beta and heterotrimeric G protein pathways for the induction of the dauer stage and chemosensation (Bargmann, 2006; Bastiani and Mendel, 2006; Savage-Dunn, 2005). Interestingly the Hsp90 inhibitor geldanamicin does not bind to DAF-21 and has no observed effects on *C. elegans* phenotype, which was suggested to be the result of adaptive evolution within the worm (David *et al.*, 2003; Him *et al.*, 2009).

#### 5.4.2.4 Phosphatase

The *C. elegans* gene EGAP2.3 (also known as *pho-1*) encodes an intestinal acid phosphatase which may have a role in digestion. It is localized in the intestinal brush border in the worm (Beh *et al.*, 1991). Pho-1 expression in the intestines starts in late embryogenesis and is maintained at a high level throughout the worm's development (Maduro and Rothman, 2002). The intestinal brush border of other systems have well characterized lipid raft domains, and this may be due to its function as an absorptive surface for nutrients and a barrier for pathogen entry (Danielsen and Hansen, 2003, 2008). Prostatic acid phosphatase, a prostate cancer marker, has recently been shown to be raft associated (Quintero *et al.*, 2007). *Pho-1* also has GPI anchor predictions from two predictor programs (GPI SOM and FragAnchor).

### 5.4.2.5 Carboxypeptidase

Nine carboxypeptidases were found by the proteomic analysis, with five (F56F10.1, F32A5.3, K10C2.1, Y16B4A.2 and Y40D12A.2) involved in non-lysosomal compartments. Carboxypeptidases are enzymes that hydrolyse the C-terminal end of peptides. They were first studied in protein digestion, but were later found to have a large number of roles, including protein maturation and regulation of biological processes. Both Carboxypeptidase E and Prohormone convertase 2 are involved in prohormone targeting and is resident within rafts, where this feature is essential for their sorting into the regulated secretory pathway (RSP) (Assadi *et al.*, 2004; Dhanvantari and Loh, 2000). Carboxypeptidase M is also a regulator of hormones, where it can change the receptor specificity of kinins and the inflammatory response (Reverter *et al.*, 2004; Zhang *et al.*, 2008); it exists on the surface membrane via a GPI anchor linkage, and may also be released for its function (Li and Skidgel, 1999; Skidgel *et al.*, 1996). All of the five carboxylpeptidases have two or more predictions for GPI anchoring (Table. 4) with F56F10.1 also found in the PIPLC released fraction, suggesting that the proteins are likely to be true lipid raft residents within *C. elegans*.

### 5.4.2.6 Lysosomal carboxypeptidase

Four of the carboxypeptidases (*pcp-2*, *pcp-3*, *pcp-4* and C26B9.5) found in the lipid raft preparation are considered to come from the lysosomal compartment of *C. elegans* cells. The PCP-2, PCP-3 and PCP-4 proteins have been predicted by two or more programs for GPI anchoring, while C26B9.5 is not a predicted GPI anchored protein. Lysosomes are known to contain lipid rafts (Kobayashi *et al.*, 1998; Simons and Gruenberg, 2000), and rafts form a part of the endosome sorting pathway in conjunction with caveolae endocytosis (Helms and Zurzolo, 2004). Lysosomes also contain a number of carboxypeptidases (Skidgel and Erdos, 1998) that function in protein turnover and cell signalling, and these enzymes may also be associated with lipid rafts within the lysosome (Obermajer *et al.*, 2008; Roshy *et al.*, 2003).

#### 5.4.2.7 Metallopeptidase

Both of the metallopeptidases found within the study (T25B6.2 and F54F11.2a) are members of the *C. elegans* neprilysin family. Neprilysin (NEP) is a zinc dependent metalloprotease integral to the plasma membrane that functions by turning off certain peptide signalling at the cell surface and is involved in many nervous, cardiovascular, inflammatory and immune signalling pathways (Turner *et al.*, 2001). Interestingly both of the metallopeptidases found in *C. elegans* are also homologues of the *H. contortus* neprilysin protein MEP1 (Redmond *et al.*, 1997) after search with BLAST. One of the most intensively studied functions of NEP is its role in amyloid β peptide (Aβ) processing in Alzheimer's disease (Carson and Turner, 2002). NEP has been shown to have caveolae localization, with this feature possibly significant for its role in Aβ processing (Cordy *et al.*, 2006; Riemann *et al.*, 2001).

#### 5.4.2.8 Insulin pathway *daf-16* controlled proteins

F35E12.7a (*dct-17*) and ZK6.10 (*dod-19*) are found to act downstream of the insulin/insulin-like growth factor-1 pathway related transcription factor *daf-16* and are implicated to have functions within the development, innate immunity and aging of the worm (Murphy *et al.*, 2003; Pinkston-Gosse and Kenyon, 2007; Styer *et al.*, 2008). Both DCT-17 and DOD-19 proteins have predictions in two or more prediction programs for GPI anchoring and therefore may reside within the lipid raft component. While their functions are not known they may play a role in signal transduction pathways due to their implied functions within the growth and development of the worm.

#### 5.4.2.9 *unc-68* ryanodine receptor associated proteins ($Ca^{2+}$ pathway)

Both *C. elegans* genes F57F4.3 (*gfi-1*) and F57F4.4 have been shown to interact directly with UNC-68 in yeast two hybrid assays (www.wormbase.org, Sakube and Kagawa, 1999,). *Unc-68* encodes a ryanodine receptor membrane protein involved in $Ca^{2+}$ signalling (Maryon *et al.*, 1996). It is expressed in the sarcoplasmic reticulum (SR) of all muscle cells and is the major protein involved in the proliferation of Calcium induced Calcium release (CICR) in *C. elegans* (Maryon *et al.*, 1998). Lipid rafts have been shown to have a role in calcium signalling (Noble *et al.*, 2006), and ryanodine receptor was found in lipid raft fractions extracted with Triton X-100 along with other members of the signalling pathway in rat cells (Weerth *et al.*, 2007). Both F57F4.3 and F57F4.4 are predicted to be GPI anchored with all four prediction programs and this feature may help its interaction with UNC-68 in the SR; some proteins in the SR, such as carbonic anhydrase IV, have also been shown to be GPI anchored (Waheed *et al.*, 1992). Interestingly UNC-68 was also identified in the proteomic analysis but did not pass the threshold for significant hits (Figure 5.7b and Appendix 3).

#### 5.4.2.10 Sugar binding lectins

The *C. elegans* lectins LEC-1, 2, 4 and 5 were identified in the proteomic search, with LEC-3 not identified in any of the searches. The LEC proteins are galectins that bind β-galactosides, and are soluble proteins that exist within the cytoplasm. A study of 11 lectin genes in *C. elegans* showed that LEC-1,2 and 4 have β-galactoside binding activity as well as different affinities for other sugar molecules, while LEC-5 had a predicted ER targeting signal and was shown previously to be N-glycosylated, implicating it as a secreted protein (Fan *et al.*, 2005; Nemoto-Sasaki *et al.*, 2008).

LEC-1 is the most well studied of the *C. elegans* galectins. It was found to be a novel tandem repeat 32 kDa sugar binding protein, with two domains for binding that each had different affinities for the same target sugar molecule (Arata *et al.*, 2001; Arata *et al.*, 1997). LEC-1 was shown to be localized to the cuticle of the worm (Arata *et al.*, 1996). A proteomic study of gene expression with 2D DIGE on *C. elegans* development identified LEC-1 and LEC-2, where their expression increased sharply after hatching and was maintained at high levels throughout the life of the worm (Tabuse *et al.*, 2005).

Galectins exhibit a wide variety of functions in the cell, such as polarized sorting of proteins, axonal regeneration, apoptosis, signalling and immunity (Delacour *et al.*, 2008; Kohatsu *et al.*, 2006; Miura *et al.*, 2004; Paz *et al.*, 2001; Perillo *et al.*, 1995). Galectins are associated with lipid rafts (Hansen *et al.*, 2005) and some can bind to modified cholesterol (Ideo *et al.*, 2007). There is evidence that galectins form their own membrane microdomains by their interaction with glycoproteins (Ahmad *et al.*, 2004), which are called lectin-glycoprotein lattices (Lajoie *et al.*, 2009); this structure is postulated to have roles in signalling at the cell surface and may compete for signalling factors from lipid rafts (Lajoie *et al.*, 2007).

#### 5.4.2.11 Sugar metabolism

W05E10.4 was identified as a trehalase (*tre-3*) in *C. elegans* involved in sugar metabolism, where it was shown by RT-PCR to be expressed in all life stages of the worm (Pellerone *et al.*, 2003). Trehalase is a classical GPI anchored protein found in early studies of GPI anchoring in rabbits (Ruf *et al.*, 1990; Takesue *et al.*, 1986), and *C. elegans* TRE-3 is also predicted to possess a GPI anchor from two of the GPI prediction programs (GPI SOM and PredGPI).

#### 5.4.2.12 Stomatin-like protein

*Stl-1* in *C. elegans* encodes a stomatin-like protein and was shown to have an increased transcription level in the worm in response to the addition of ethanol (Kwon *et al.*, 2004). Stomatin is a 32kDa membrane bound protein with a role in the regulation of Na+/K- ion transport (Stewart, 1997). Its mutation causes the rare anaemic disease Overhydrated Hereditary Stomatocytosis (OHSt), and its mode of action involves regulation with cytoskeletal components (Stewart *et al.*, 1993). Stomatin is raft associated in erythrocytes, platelets and epithelial cells (Fricke *et al.*, 2003; Mairhofer *et al.*, 2002; Salzer and Prohaska, 2001), and is used as a marker for the presence of lipid rafts (Salzer *et al.*, 2008; Umlauf *et al.*, 2006). Recently it was shown that the stomatin-like protein STP-2 regulates T-cell activation, giving a role for such proteins in raft- associated signalling (Kirchhof *et al.*, 2008).

#### 5.4.2.13 Vacuolar proton-translocating ATPase (V-ATPase)

R10E11.8 (*vha-1*) and Y55H10A.1 (*vha-19*) were found to encode for components of the *C. elegans* V-ATPase complex. V-ATPase is related to the F-ATPase of mitochondria, and works as a membrane bound proton pump in a variety of organelles such as endosomes, lysosomes, the Golgi apparatus, and others (Anderson and Orci, 1988). It consists of two major complexes V1 (cytosolic) and V0 (membrane), which are both made up of multiple subunits (Saroussi and Nelson, 2009). The primary role of V-ATPase is to acidify the pH of various organelles, and to create a proton motive force to drive secondary transport processes within them (Beyenbach and Wieczorek, 2006). V-ATPase has been isolated from lipid raft preparations of endothelial cells, phagosomes and synaptic vesicles (Dermine *et al.*, 2001; Sprenger *et al.*, 2006; Yoshinaka *et al.*, 2004). Rafts were found to regulate the activity of V-ATPases by attenuating V1 and V0 subunit association (Lafourcade *et al.*, 2008). Both of the *C. elegans* proteins found in this analysis are related to subunits of the membrane bound V0 complex of V-ATPase, with *vha-1* homologous to subunit c and *vha-19* encoding a non-homologous replacement for the fungal subunit c' called Ac45, which is found only in multicellular organisms (Oka *et al.*, 1997; Schoonderwoert and Martens, 2002). V-ATPase is involved in cell fusion and apical sorting/secretion in *C. elegans*, and is required for ovulation and embryogenesis (Kontani *et al.*, 2005; Liegeois *et al.*, 2006; Oka and Futai, 2000).

#### 5.4.2.14 Vacuolar protein sorting

Another vacuolar protein found within the study was C56C10.3 (*vps-32.1*) which is related to the yeast vacuolar protein sorting (Vps) factor and is a part of the ESCRT-III complex within *C. elegans* (Michelet *et al.*, 2009). Lipid rafts are involved in protein sorting in vacuolar compartments and Vps is heavily implicated in protein sorting within endosomes (Kobayashi and Hirabayashi, 2000; Piper and Luzio, 2001). Interestingly the VPS-32.1 protein was found in *C. elegans* to occupy distinct domains within the endosome compared to other ESCRT-III proteins (Michelet *et al.*, 2009), which may be due to lipid raft partitioning. A small amount of antibody to VPS-32.1 was obtained courtesy of the Legouis lab, but the western blot experiment was unsuccessful in detecting any bands.

#### 5.4.2.15 Proteins without Wormbase descriptions

Eleven of the remaining proteins do not have clear descriptions of function from Wormbase. Of these T19D12.4a and Y54G2A.18 do not have prediction for GPI anchoring, while C29F3.7a and F53C11.1 possess GPI anchoring prediction in only one of the prediction programs (GPI SOM). The rest of the seven proteins (ZK6.11a, Y41D4B.16, F54E2.1, K08D8.6, F35E12.10, B0024.4 and Y12A6A.1) have GPI anchoring predicted in at least two prediction programs, with ZK6.11a also found within the PIPLC released fraction, making them likely raft components.

#### 5.4.2.16 Potential false positives

There are four proteins identified within the analysis that may be false positives in the light of their function. The first protein is R05G6.7, which functions as a channel protein with predicted localisation in mitochondria. Mitochondria have been shown to form lipid domains with non-raft properties (Grijalba *et al.*, 1999) and both raft associated and mitochondrial ion channels are involved in apoptosis, with sometimes a large amount of cross-talk (Garcia *et al.*, 2003; Szabo *et al.*, 2004). However there is no direct evidence that mitochondria contain rafts, with a recent study placing mitochondrial proteins as contaminants of raft extraction (Zheng *et al.*, 2009). The second potential contaminant is R07G3.3a (*npp-21*) which encodes a nuclear pore protein that is very unlikely to be resident in rafts, as the nuclear envelope is not likely to form raft-like domains. Previous reports have also indicated that certain raft associated proteins cause common cross-contamination during nuclear membrane extraction (Say and Hooper, 2007). R03G5.1a (*eft-4*) is the third identified protein that may have been falsely predicted to be raft associated. *eft-4* encodes a translation elongation factor in *C. elegans* and works in concert with ribosomes in the cytosol to ensure proper protein translation, so it is unlikely to be lipid raft associated (Proud, 1994). Lastly F21D5.3 encodes a laccase copper oxidase, which is secreted onto the cell wall of *Cryptococcus neoformans* and has a role in its virulence to humans (Zhu and Williamson, 2004). A recent study of lipid rafts in *Cryptococcus* has shown that laccase does not associate with either raft or non-raft membrane (Siafakas *et al.*, 2006). All of these proteins show compelling evidence in the literature for their non-raft association and are therefore excluded from the list of lipid raft proteins found in *C. elegans*.

### 5.4.3 Conclusion

Overall the 2-dimensional LC MS/MS analysis produced 40 likely candidates of lipid raft proteins in *C. elegans*, with 36 of them showing features of known lipid raft components and homology to lipid raft proteins from other organisms. This added with the LEC-3 protein identified with 2D electrophoresis makes the final number of lipid raft proteins found in *C. elegans* to be 41 (Table. 5.4). The proteins identified are involved in a number of processes including signalling, sugar binding, transport, proteolysis, molecular chaperone and the cytoskeleton. This study represents the largest number of raft proteins identified in the nematode to date, and sheds light on the importance of this sub-membrane proteome in the biology of *C. elegans*.

Of interest are the 21 proteins found within the lipid raft fraction that have GPI anchor prediction in at least two GPI anchored prediction programs, which represents just over 50% of the raft proteins identified (Table. 5.4). The caveat present is that prediction programs do not necessarily reflect anchoring of the protein *in vivo*; however a conservative estimate using only proteins with validation from four prediction programs still yields a high percentage of GPI anchoring in the 41 identified proteins (9 proteins, 22% of the total), while three of these proteins (F56F10.1, ZK6.11 and DOD-19) have had their GPI anchoring validated by PIPLC digestion. GPI anchored proteins have not been extensively studied in *C. elegans* and it is interesting that the organism has such a high proportion of GPI anchoring within its raft proteome. This may reflect the importance of this post translational modification on the biology of the worm. Most of the GPI anchor synthesis pathway is conserved in *C. elegans* (Chapter 3) with the two homologues of the catalytic subunit of the transamidase complex (*PIG-K* in humans) both containing the active site residues of the enzyme. GPI anchored proteins could play a major role within the biology of the worm and may be involved in a variety of processes such as development, various signalling processes, digestion, transport, sugar metabolism and organelle maintenance. It would be interesting to further study the role GPI anchored proteins and lipid rafts have within *C. elegans*, which coupled with its extensive genetic knowledge can offer a greater understanding of these important classes of proteins.

| Gene name | description | size (amino acids) | identified in LC MS/MS | identified with 2D electrophoresis | GPI prediction with two programs or more | released by PIPLC |
|---|---|---|---|---|---|---|
| tag-10 | apical gut membrane protein | 473 | y | n | y | n |
| act-4 | cytoskeleton | 332 | y | n | n | n |
| act-4 | cytoskeleton | 376 | y | n | n | n |
| daf-21 | molecular chaperone | 702 | y | n | n | n |
| pho-1 | phosphatase | 449 | y | n | y | n |
| F56F10.1 | carboxypeptidase | 540 | y | n | y | y |
| F32A5.3 | carboxypeptidase | 574 | y | n | y | n |
| K10C2.1 | carboxypeptidase | 2314 | y | n | y | n |
| Y16B4A.2 | carboxypeptidase | 2167 | y | n | y | n |
| Y40D12A.2 | carboxypeptidase | 512 | y | n | y | n |
| pcp-2 | lysosomal carboxypeptidase | 1080 | y | n | y | n |
| pcp-3 | lysosomal carboxypeptidase | 1080 | y | n | y | n |
| pcp-4 | lysosomal carboxypeptidase | 1042 | y | n | y | n |
| C26B9.5 | lysosomal carboxypeptidase | 516 | y | n | n | n |
| T25B6.2 | metalloprotease | 798 | y | n | n | n |
| F54F11.2 | metalloprotease | 1589 | y | n | n | n |
| dct-17 | insulin pathway daf-16 controlled proteins | 739 | y | n | y | n |
| dod-19 | insulin pathway daf-16 controlled proteins | 406 | y | n | y | y |
| F57F4.4 | unc-68 ryanodine receptor associated proteins ($Ca^{2+}$ pathway) | 2090 | y | n | y | n |
| gfi-1 | unc-68 ryanodine receptor associated proteins ($Ca^{2+}$ pathway) | 2153 | y | n | y | n |
| lec-1 | galactoside binding lectin | 279 | y | y | n | n |
| lec-2 | galactoside binding lectin | 278 | y | y | n | n |
| lec-3 | galactoside binding lectin | 297 | n | y | n | n |
| lec-4 | galactoside binding lectin | 283 | y | y | n | n |
| lec-5 | galactoside binding lectin | 314 | y | n | n | n |
| tre-3 | sugar metabolism | 588 | y | n | y | n |
| stl-1 | stomatin like | 327 | y | n | n | n |
| vha-1 | vacuolar proton-translocating ATPase (V-ATPase) | 169 | y | n | n | n |
| vha-19 | vacuolar proton-translocating ATPase (V-ATPase) | 451 | y | n | n | n |
| vps-32.1 | vacuolar protein sorting | 221 | y | n | n | n |
| T19D12.4 | | 1028 | y | n | n | n |
| Y54G2A.18 | | 213 | y | n | n | n |
| C29F3.7 | | 491 | y | n | n | n |
| ZK6.11 | | 386 | y | n | y | y |
| Y41D4B.16 | | 453 | y | n | y | n |
| F54E2.1 | | 391 | y | n | y | n |
| K08D8.6 | | 491 | y | n | y | n |
| F35E12.10 | | 487 | y | n | y | n |
| F53C11.1 | | 494 | y | n | n | n |
| B0024.4 | | 390 | y | n | y | n |
| Y12A6A.1 | | 209 | y | n | y | n |

**Table 5.4. Final list of identified lipid raft and GPI anchored proteins from *C. elegans*.** A total of 41 raft proteins were found in the *C. elegans* lipid raft fraction with LC MS/MS and 2D electrophoresis. Of these proteins, 21 were found to have GPI anchoring predicted by two or more programs (highlighted in light grey). Three GPI anchored proteins were experimentally verified with PIPLC digestion and also appear in the list of identified lipid raft proteins (highlighted in dark grey).

# *Chapter 6*

# General discussion

## 6.1 GPI anchored proteins

### 6.1.1 The function of GPI anchored proteins

The study of proteins is an area of immense interest within molecular biology. Almost all biological processes are carried out by proteins, and their biochemistry shapes our understanding of the various mechanisms and pathways that take place within the cell. Proteins have also been recently implicated in the passage of genetic information via the mechanism of epigenetics, which has challenged the idea that hereditary information is passed exclusively by DNA. The study of proteins has enriched our understanding of biology and evolution, and is likely to continue to have a large impact in the future.

Membrane proteins and protein modifications are important areas of study within the field of protein biochemistry. Membrane proteins are thought to make up approximately 30% of all proteins with a cell (Wallin and von Heijne, 1998). They are responsible for a large number of cellular processes and maintain the internal environment of the cell by allowing selective exchange of materials with the outside world. Membrane proteins are also critical for the transmission of information from outside of the cell, which allow the cell to respond to changes in the environment, adapt to various external stimuli, and communicate with other cells during development. Almost all proteins carry some level of post translational modification for their activity. Modifications such as phosphorylation may regulate the activity of a protein for a particular enzymatic reaction, and others such as palmitoylation and glycosylation may act as markers that allow the protein to be transported to the correct sub-cellular compartments for their function.

Modification with a GPI moiety allows an otherwise aqueous protein to become anchored to the membrane. Because of this GPI anchored proteins behave in a similar fashion to integrated membrane proteins and yet at the same time contain no transmembrane (TM) domains (Brown, 1992). The anchor itself acts as a signal that localises the protein to the apical part of polarised cells as well as lysosomal compartments during endocytosis (Fivaz *et al.*, 2002; Lisanti and Rodriguez-Boulan, 1991). GPI anchored proteins have a range of functions including catalysis, signal transduction, cell recognition, parasite invasion and others. They have been shown to be important in host invasion by *Trypanosoma brucei*, embryonic development in mice, and are responsible for onset of the haemophilic disease paroxysmal nocturnal hemoglobinuria (PNH) in humans (Alfieri *et al.*, 2003; Ferguson, 2000; Parker, 1996).

### 6.1.2 Roles within raft and endocytosis

GPI anchoring requires a complex biosynthetic machinery to be produced in the cell. The most well characterised pathways are found in humans and *Saccharomyces cerevisiae* (yeast). Both species require more than 20 genes for the production of a GPI anchored protein (Paulick and Bertozzi, 2008). The GPI anchor also undergo extensive fatty acid remodelling in the ER and the Golgi before the protein becomes located to its final destination (Fujita and Jigami, 2008). Why does a cell need such an energetically expensive method for associating a protein to the membrane when a less complex method, such as the inclusion of a hydrophobic TM domain at the C-terminus, will also achieve the same end? An important property of the GPI anchor comes from its association with the sphingolipid/cholesterol enriched membrane micro domains called lipid rafts (Brown, 1992). GPI anchored proteins such as the folate receptor aggregate in raft domains on the cell surface, in which replacement of the anchor with a TM domain abolishes this association and produces a random distribution of the protein on the plasma membrane (Varma and Mayor, 1998). It has been proposed that GPI anchored proteins participate in a novel pinocytotic pathway involving the GPI-anchored protein enriched endosomal compartment (GEEC), which is distinct from internalisation with clathrin coated pits or caveolae mediated endocytosis (Lakhan *et al.*, 2009). Endocytosis of GEECs is regulated by the GTPase Cdc42 as was seen in the uptake of folate via the folate receptor (Sabharanjak *et al.*, 2002). In the disease neurodegenerative spongiform encephalopathy GPI anchored prion proteins are converted from a soluble $PrP^{C}$ form to an insoluble infective $PrP^{Sc}$ form, which causes amyloid plaques to form in the neurones of patients (Prusiner, 1996). While the GPI anchor has been shown not to be necessary for the conversion of prion proteins to their infective form, their unique endocytotic mechanisms have been implicated in the maintenance of infectivity within this disease (Priola and McNally, 2009). Raft association is also implicated in the role of signalling for GPI anchored proteins. The GPI anchored protein uPAR (uPA Receptor), which binds to uPA (urokinase type Plasminogen Activator) and facilitates cell migration via phosphorylation of focal adhesion kinase and epidermal growth factor receptor (EGFR) in cancer cells (Tang and Wei, 2008), was shown to be disrupted by the action of elevated lipid raft gangliosides GT1b and GM3 that may have acted to sequester the protein from its targets (Wang *et al.*, 2005). GPI anchored proteins may also be released from the cell surface by phospholipases via the cleavage of the anchor, and this mechanism is used by *CR-1* (*Cripto-1*) for signalling in development and tumour progression (Watanabe *et al.*, 2007). The biology of GPI anchored proteins is intimately associated with their presence within lipid rafts, and they are thus able to take on roles within the cell that would not be possible if the protein was bound to the membrane via a TM domain.

### 6.1.3 Lipid raft and GPI anchored proteins in *C. elegans*

GPI anchored proteins have important roles in development and signalling, however most of the research carried out for this class of proteins have been made in mammalian cell lines such as human HeLa cells (Metz *et al.*, 1994), Madin-Darby canine kidney (MDCK) cells (Urquhart *et al.*, 2005) and Chinese hamster ovarian (CHO) cells (Priola and McNally, 2009), single cellular organisms such as yeast (Pittet and Conzelmann, 2007), and protozoan internal parasites such as *Trypanosoma brucei* and *Trypanosoma cruzi* (Ferguson, 1999; Tarleton, 2007). *C. elegans* is a model organism with an extensive history of study within development, in which all of its cell fates have been determined using microscopy. There has been limited research in lipid rafts and GPI anchored proteins for *C. elegans*. Sedensky e*t al.* had found the stomatin homologue UNC-1, the stomatin-like protein UNC-24 and the sodium ion channel UNC-8 in a Triton X-100 (TX-100) extracted nematode raft preparation (Sedensky *et al.*, 2004). Agostoni *et al.* were able to express the *C. elegans* protein PHG-1 (PHAS-1) in a mammalian cell system and showed that it was GPI anchored via cleavage with phosphatidylinositol-specific phospholipase C (PIPLC) (Agostoni *et al.*, 2002). In this thesis I have explored the use of *C. elegans* as model for the study of lipid raft and GPI anchored proteins. *C. elegans* homologues of all known genes involved in the GPI synthesis pathway were elucidated and analysed, with a possible pathway and final GPI anchor structure postulated for the nematode. The *C. elegans* genome was put through four GPI anchoring prediction programs with different algorithms to produce a comprehensive list of hypothetical GPI anchored proteins for the worm. Finally a lipid raft fraction was extracted from *C. elegans* membrane preparations using TX-100 sucrose density floatation, which was then treated with PIPLC to release GPI anchored proteins; these two samples were then subjected to separation with 2D gel electrophoresis and multi-dimensional liquid chromatography (MDLC), with the separated proteins identified using mass spectrometry. To date this is the largest number of lipid raft and GPI anchored proteins identified within *C. elegans*. A discussion of the results obtained, what they mean to lipid raft and GPI anchored protein research, as well as their implications for research within the nematode model system is given below.

## 6.2 GPI anchored synthesis pathway and lipid modifications in *C. elegans*

### 6.2.1 GPI anchored synthesis and lipid modification in *T. brucei*

The GPI synthesis pathway is a well studied system and the majority of the discoveries of its components were found in human and yeast, with several genes also elucidated in *T. Brucei* (Ferguson, 1999). The core structure of the GPI anchor is conserved within all eukaryotic species found so far, with prokaryotic organisms lacking the modification completely (Ikezawa, 2002). Several archaebacterial species were also proposed to contain GPI anchored proteins via a bioinformatics search (Eisenhaber *et al.*, 2001). Both humans and yeast contain 12 steps for GPI anchored synthesis (outlined in Chapter 3), with the majority of the genes within each of the steps conserved between the two species. The biosynthesis pathway however is markedly different for *T. Brucei*, which is a protozoan parasite that causes African sleeping sickness in humans. *T. brucei* has two distinct proliferative stages, the blood stream form which is resident within the host mammal, and the procyclic from which resides inside its vector the tsetse fly. GPI anchored proteins (known as variant surface glycoproteins (VSG) in the blood stage and procyclins in the vector stage of the parasite) on the surface of *T. brucei* are thought to be important for both of its life cycle stages and have been shown to be essential for its infectivity in humans (Hong and Kinoshita, 2009). The parasite is densely coated with VSGs on the surface, which creates antigenic variation on the organism that is thought to allow the blood stream form to evade host immune responses (Pays and Nolan, 1998). The GPI structures are different within each of the life stages, with the procyclic form containing an acyl group on the inositol ring that makes it resistant to PIPLC (Field *et al.*, 1991). There are a total of seven genes found so far in *T. brucei* GPI biosynthesis, with three found also in fatty acid remodelling. The *T. brucei* gene for the second step of biosynthesis (TbGPI12) contains different substrate and inhibitor specificity with respect to the human version of the gene (*PIG-L*) (Sharma *et al.*, 1999; Smith *et al.*, 2001). TbGPI10 was found to be the *T. brucei* gene responsible for the addition of the third mannose onto the GPI structure and is able to substitute for its orthologues in human and yeast (Nagamune *et al.*, 2000). Five subunits for the transamidase complex of step 12 were found in *T. brucei*, in which three of the components, TbGAA1, TbGPI8 and TbGPI16, were found to be homologues to the human genes *GAA1*, *PIG-K* and *PIG-T* respectively, but the other two genes (TTA1 and TTA2) were found to have homologues only in other protozoan species (Nagamune *et al.*, 2003). GAA1, PIG-K and PIG-T are proposed to form a small subunit which interacts with another small subunit composed of PIG-S and PIG-U to form the transamidase complex in humans (Zhu *et al.*, 2005), which indicates that there may have been an evolutionary spilt between the protozoans and higher eukaryotes for their TTA1/TTA2 and PIG-S/PIG-U part of the transamidase complex. *T. brucei* contains two GPI inositol deacylases TbGPIdeAc and TbGPIdeAc2, in which TbGPIdeAc was found to be non-essential for GPI anchor production (Guther *et al.*, 2001), while TbGPIdeAc2 was found to be essential (Hong *et al.*, 2006). A homologue for the yeast sn-2 acyltransferase *GUP1* is also present within *T. brucei* (TbGup1), with the protozoan enzyme demonstrated to prefer the addition of a myristate (C14:0) moiety onto the anchor instead of the C26:0 moiety that is added by yeast (Hong *et al.*, 2006). Lipid remodelling of the GPI anchor occurs on both the sn-1 and sn-2 positions of *T. brucei*, in contrast to mammalian systems and yeast where the anchors are usually only modified in the sn-2 position (Hong and Kinoshita, 2009). *T. brucei* GPI anchors also may contain side chain modifications such as galactose and sialic acid that are not present within mammalian or yeast GPI structures (Ferguson *et al.*, 1993; Ikezawa, 2002). The GPI anchored synthesis machinery in *T. brucei* appear to have essential differences to the ones in human and yeast, which may be due mainly to its specialised role as a parasite, dual stage life cycle characterised by a procyclic vector and an invasive blood cycle stage, and a difference in evolutionary complexity between protozoan and higher eukaryotes. Since *C. elegans* is a metazoan with a relatively complex developmental process it would be more likely that its GPI biosynthesis pathway would be closer to the ones present in human and yeast than that of *T. brucei*. A bioinformatic search for homologues of *T. brucei* TTA1 and TTA2 in *C. elegans* returned no results (data not shown), while the worm does contain homologues for the human *PIG-S* and *PIG-U* genes, which further underscores the similarity of the nematode GPI biosynthesis pathway with that of other higher eukaryotes.

### 6.2.2 The *C. elegans* GPI synthesis pathway

The GPI biosynthesis pathway in *C. elegans* contains 16 of the 23 genes found in the human pathway. Most of the human synthesis steps have homologues in *C. elegans*, with the exception of the GlcNAc-PI deacytalase of step 2 and the fourth mannosyltransferase of step 9. The gene for step 2 has been shown to be essential in human, yeast and *T. brucei* (Sharma *et al.*, 1999; Watanabe *et al.*, 1999) and creates a bottleneck for the production of GPI anchors. The closely related nematode species *C. briggsae* does contain a homologue for *PIG-L*, the human gene for this step, but the *C. briggsae* gene also unusually does not have a homologue in *C. elegans*. It could be that the *C. elegans* version of the gene was lost in evolution and another unrelated

GlcNAc deacytalase has since taken up the role for the second step of GPI synthesis. A Wormbase search with the GO term GlcNAc deacytalase found the *C. elegans* gene F59B2.3 with this biological process, which may be a potential candidate for the second step of GPI biosynthesis. For the mannosyltransferase in step 9 both *C. elegans* and C. *briggsae* lack a homologue for the enzyme involved in this reaction. The fourth mannose is an essential addition for GPI anchors in yeast (Grimme *et al.*, 2001) but appears to be tissue specific in humans (Taron *et al.*, 2004b), which implies that the modification may not be essential in metazoans and may have been lost during the evolution of the nematodes.

### 6.2.3 The GPI transamidase complex

There is a remarkable amount of conservation in *C. elegans* for the 12$^{th}$ and last step of GPI biosynthesis, which involves the transamidase reaction that attaches the protein to the GPI anchor. Five components of the complex responsible for the transamidase reaction have been found so far in both human and yeast (*PIG-K/GPI8*, *GPAA1/GAA1*, *PIG-T/GPI16*, *PIG-S/GPI17* and *PIG-U/GAB1* for human/yeast, respectively) and all of the genes have homologues within *C. elegans*. *PIG-K* in humans is the catalytic subunit within the transamidase complex and has two homologues in *C. elegans*, T05E11.6 and T28H10.3. The proteins for these two genes both have high blast scores for the PIG-K protein and they also possess the two conserved residues of its active site (Ohishi *et al.*, 2000), which indicates that both of the homologues may be able to attach proteins to GPI anchors within *C. elegans*. T28H10.3 was shown in this study to be expressed strongly in the intestine of the worm in all life cycle stages, from the early embryo to the adult. RNAi studies for T05E11.6 have yielded no phenotypes, while RNAi on T28H10.3 has resulted in embryonic lethality within the worm (Maeda *et al.*, 2001a). Recently a deletion mutant became available for each of the genes and they both have shown an embryonically lethal phenotype, indicating that both of the genes may be important for worm viability. It would be interesting to see if this effect on worm survival is due to the lack of GPI anchoring of proteins disrupting processes such as signalling within the worm, which may have a profound effect on its development. Recently a wealth of research has been made that implicates the GPI transamidase components as oncogenes in a variety of human cancers. The *PIG-U* gene was first found to be unregulated in human bladder and is associated with an overexpression of the GPI anchored protein uPAR, which caused an increase in STAT-3 signalling and is thought to mediate the oncogenic properties of *PIG-U* (Guo *et al.*, 2004). This upregulation was later confirmed to exist for both the mRNA and protein of *PIG-U* in bladder urothelial cell carcinoma (Shen *et al.*, 2008). *GPAA1*, *PIG-T* and *PIG-U* were found to be involved in breast cancer, with *GPAA1* and *PIG-T* implicated in tumorigenesis and invasiveness of the cancer, possibly through interactions with paxillin (Wu *et al.*, 2006). *GPAA1* expression was also found to be upregulated in head and neck squamous carcinoma, with an increase in copy number in these tumours (Jiang *et al.*, 2007). A large study of all five GPI transamidase subunits in 19 different cancers showed that all of the components have roles in a variety of cancers (Nagpal *et al.*, 2008). *PIG-U* was found to be overexpressed in colon and ovarian cancer, while *PIG-T* was upregulated in uterine, thyroid, melanoma, and breast cancers. *GPAA1* showed increased expression in uterine cancer and *PIG-S* expression was shown to be increased in lung, thyroid, ovarian and liver cancers. The catalytic unit *PIG-K* showed overexpression in ovarian and breast cancers but was significantly downregulated in bladder, liver and colon carcinoma cases. The study also found a significant increase in *PIG-U* and *PIG-K* expression in lymphoma, where in normal lymph node tissues the GPI transamidase subunits showed universally low levels of expression. *PIG-K* and *PIG-S* increased proliferation of SKBR3 breast cancer cells after transfection. The study also observed a large amount of variability in expression for all the tissue types tested, and GPI transamidase components were also found in the cytoplasm of cancer cells; the GPI transamidase complex normally acts inside the ER lumen for the attachment of GPI anchors to proteins, and their presence within the cytosol of cancer cells may point to additional roles for these proteins in the cell. *C. elegans* is well positioned for the study of these transamidase components within the role of development, and knock-ins of overexpressed versions of these genes will also be possible for the study of their effect in growth and tissue formation, which will hopefully aid in the understanding of the role they play within human cancers.

### 6.2.4 Lipid remodelling, Dolichol phosphate mannose (Dol-P-Man) synthesis and similarities with the human GPI anchor synthesis pathway

Both human and yeast GPI anchors are modified after the attachment of the protein via the GPI transamidase, while for *T. brucei* these modifications comes before this step (Fujita and Jigami, 2008; Hong and Kinoshita, 2009). The first step of remodelling for both human and yeast takes place within the ER and involves the deacetylation of the inositol ring by *PGAP1/BST1* (Tanaka *et al.*, 2004). This reaction has been shown to be important for the translocation of the protein in to the Golgi apparatus (Vashist *et al.*, 2001), downstream remodelling of other fatty acid chains (Maeda *et al.*, 2007), and for quality control of misfolded GPI anchored proteins in yeast (Fujita *et al.*, 2006b). After the deacetylation reaction the human and yeast pathways take a divergence in their modes of action. In humans the protein is transported to the Golgi, and the acyl chain at the sn-2 position of the anchor is removed by the GPI-phospholipase A2 enzyme *PGAP3* (Fujita and Jigami, 2008), while in yeast the same reaction occurs in the ER via the homologue *PER1* (Fujita *et al.*, 2006a). After this reaction in yeast a C26:0 fatty acid is added to the sn-2 position in the ER via *GUP1* (Bosson *et al.*, 2006) while a C18:0 species is added to the sn-2 position in human cells in the Golgi by the unrelated *GPAP2* (Tashima *et al.*, 2006). Yeast also contains a homologue to *GPAP2* called *CWH43* that was shown to be involved in the addition of ceramides to the anchor (Ghugtyal *et al.*, 2007). *C. elegans* contains a large number of homologues for human *GPAP2* but only one weak homologue for the yeast specific *GUP1*, suggesting that the nematode lipid remodelling pathway is more similar to the one in mammals than the one in yeast. Lipid remodelling of GPI anchored proteins is essential for their association with lipid rafts (Maeda *et al.*, 2007), and the disruption of their *C. elegans* homologues may be a method for the study of raft association of GPI anchored proteins in the worm.

Dol-P-Man is the mannose donor molecule for steps 6, 7, 8 and 9 of GPI biosynthesis and is synthesised on the luminal side of the ER membrane. Dol-P-Man synthesis in humans require three genes *DPM1*, *DPM2* and *DPM3,* while in yeast only *DPM1* is required (Maeda and Kinoshita, 2008). Yeast Dpm1p protein differs from human DPM1 by the presence of a C-terminal TM domain that tethers the protein onto the ER membrane, and represents two classes for the structure of the enzyme (Colussi *et* *al.*, 1997; Tomita *et al.*, 1998). Human DPM1 requires interaction with the membrane bound DPM3 protein in order to become stably associated with the ER membrane, and the lack of this association leads to the degradation of the DPM1 protein via the proteosome (Ashida *et al.*, 2006). The *C. elegans* DPM-1 appears to possess the sequence features more similar to the structure of human DPM1, and this is reinforced by the presence of a DPM3 homologue in the worm, which further point to the increased similarity of the nematode's GPI synthesis machinery to the one present in humans.

Overall the various processes involved in the production of a GPI anchored protein in *C. elegans* is presented here. Many of these genes have immense interest within biology and medicine, especially for the GPI transamidase subunits that have been implicated as potential oncogenes in various cancers. The *C. elegans* GPI synthesis and modification components show a great degree of similarity to the human pathways based on bioinformatics analysis, which may improve the relevance of discoveries within this organism to human diseases. The study of expression patterns, behaviour traits and knockout models in the worm will hopefully give use a greater understanding of the roles these genes play within growth and development.

## 6.3 Predictions of GPI anchored proteins from the *C. elegans* genome

One of the major advantages of working with *C. elegans* is the availability of one of the most comprehensively annotated genomes for bioinformatics studies. GPI anchored proteins contain two signal sequences, one at the N-terminus for ER targeting, and another at the C-terminal end for recognition and cleavage by the GPI transamidase complex (Udenfriend and Kodukula, 1995a). The C-terminal sequence has become the subject of special interest within recent years. This GPI anchored protein specific signal does not have a consensus sequence but contains specific motifs of amino acids centred on the ω site, which is the amino acid residue of anchor attachment. The work of Eisenhaber *et al*. established the requirements for amino acid and sequence properties within the C-terminal signal peptide (Eisenhaber *et al.*, 1998), which was followed up with their use in the first GPI anchored protein prediction program, BIG PI (Eisenhaber *et al.*, 1999). Subsequently a number of programs were also developed based on machine learning algorithms, such as GPI SOM (Fankhauser and Maser, 2005), DGPI (http://129.194.185.165/dgpi/, unavailable at time of writing), FragAnchor (Poisson *et al.*, 2007), and PredGPI (Pierleoni *et al.*, 2008). Due to the need for the presence of the N-terminal ER sequence in a GPI anchored protein, predictions from genomes usually follow a two stage stringency method, with positive predictions for both the N-terminal secretion signal and C-terminal GPI anchoring signal needed before the protein can be considered to be potentially GPI anchored. The N-terminal prediction is usually carried out with SignalP 3.0 (Bendtsen *et al.*, 2004), which was shown to have a high degree of accuracy in previous studies (Emanuelsson *et al.*, 2007).

### 6.3.1 A method of GPI anchor prediction using four programs

GPI anchored protein prediction was first used on an early version of the *C. elegans* genome as a test for the BIG-PI predictor, which found 86 proteins with potential GPI anchoring C-terminal sequences (Eisenhaber *et al.*, 2000). A recent proteomic study of GPI anchored proteins in human and *Arabidopsis thaliana* showed that an integrated approach of the usage of several GPI prediction programs gave the most stringent results, which matched experimentally identified GPI anchored proteins (Elortza *et al.*, 2006). A novel approach of using SignalP 3.0 and the four available GPI prediction programs (BIG-PI, GPI-SOM, FragAnchor and PredGPI) was developed for this project in order to assess, with a high degree of accuracy, the total number of potential GPI anchored proteins within the *C. elegans* genome. Prediction results from each of the programs were correlated against each other, such that proteins were grouped into categories of increasing stringency based on the number of prediction programs that validated them. After analysis it was found that the stringency of each individual prediction program differed considerably, with BIG-PI having the most strict criteria returning the lowest number of predictions (125 genes), and GPI-SOM containing the most relaxed criteria with the highest number of predictions (657 genes). The prediction results however correlated well between programs, and it was decided that a cut off point of simultaneous prediction by at least two different prediction programs would be used for a protein to be counted as a candidate GPI anchored protein. Overall 327 proteins from *C. elegans* were found to fit this criterion and represent the final list of potential GPI anchored proteins in the worm. This accounts for 1.39% of the total number of genes within the genome. In an effort to validate these results further available orthologues of these genes were taken from *C. briggsae* and subjected to the same analysis to see if they correlated with the *C. elegans* data. Of these 201 genes were found to fit the criterion for their *C. briggsae* orthologues. GO term analysis for these genes did not differ greatly from that of the 327 predicted genes, indicating that this approach of correlation between related species may not be strictly necessary for accurate GPI anchoring prediction, but does add extra stringency to the results. GPI anchored proteins predicted with the method developed in this thesis also have different levels of confidence, with proteins predicted with all four programs have a higher likelihood of GPI anchoring than proteins with three predictions, which in turn are more likely to be GPI anchored than proteins with only two predictions. This is the first time that such a method has been used for the genome wide prediction of GPI anchored proteins in a model organism. It will be interesting to test the validity of such an approach with further experimental data for the verification of the predictions.

### 6.3.2 The predicted GPI anchored proteins in *C. elegans*

Among the prediction results were proteins that have well documented GPI anchored homologues in other systems, such as acetylcholine esterase (*C. elegans ace-2, 3* and *4*) (Nalivaeva and Turner, 2001), trehalase (*tre-3* and *tre-5*) (Netzer and Gstraunthaler, 1993), apical membrane protein of gut epithelial cells (*tag-10*) (Jasmer *et al.*, 1996), Ly-6 superfamily of GPI-linked signalling proteins (*odr-2*, *hot-3, 4,* and *7*) (Chou *et al.*, 2001), and a large number of carboxypeptidases (Skidgel *et al.*, 1996). Interestingly the well known GPI anchored protein alkaline phosphatase was not represented in this list. BLAST search with both human and yeast alkaline phosphatases produced no homologues in the *C. elegans* genome; however an assay for the enzyme in the worm was able to produce a positive result (data not shown). It may be that *C. elegans* contains an unrelated phosphatase that is able to carry out the same reaction. GO terms of biological processes were available for 93 of the predicted proteins. Of these genes, 15% were involved in development, 8% in regulation, and 6% were classified as signalling proteins, which indicates that a substantial proportion of the GPI anchored proteins in *C. elegans* may be involved in signal transduction pathways. 48% of the genes were grouped as having metabolic activity, some of which such as *tag-10* may be involved in the digestion of nutrients on the apical surface of the intestine. This hypothesis is also suggested by the result that T28H10.3, the GPI biosynthesis gene *PIG-K* homologue, had shown strong expression within the intestine of the worm. 19% of the genes were found to have roles in cellular transport, which may correlate with an involvement with the GEEC endocytic pathway. 2% of the proteins were classified as defence while 1% was grouped with a role in cell adhesion, which was observed in some GPI anchored proteins in neuronal cells (Karagogeos, 2003). Overall *C. elegans* GPI anchored proteins show a diverse range of functions and may be involved in many different processes within the worm.

### 6.4 Proteomic analysis of GPI anchored and lipid raft proteins in *C. elegans*

The field of proteomics has been progressing at a rapid pace within the last 10 years. Technological improvements in protein separation, mass spectrometry (MS), and bioinformatics have greatly improved the fidelity of protein identifications, with the rising use of multi-dimensional liquid chromatography (MDLC) and tandem MS/MS fragmentation allowing more data to be extracted from proteomic samples than ever before (Motoyama and Yates, 2008). Older methods such as 2D electrophoresis have also been updated to keep up with the speed of innovations within the field (Issaq and Veenstra, 2008). These techniques are used for the elucidation of increasingly complex proteomes such as organelles, subcellular compartments, signalling cascades and protein modifications (Dunkley *et al.*, 2004; Rogers and Foster, 2009; Voshol *et al.*, 2009), offering a global view of their protein interactions and a greater insight into the roles they play within the organism.

Both lipid raft and GPI anchored proteins have been the subject of proteomic analysis in a diverse range of organisms. Lipid rafts are patches of lipids on the membrane composed of sphingolipids and cholesterol that are proposed to form distinct domains from the rest of the membrane lipids. Research into rafts have been fraught with controversy as many different definitions exist based on the method of their extraction from the cell (Pike, 2004). Rafts have been observed to form spontaneously in model membranes with physiological levels of the various lipids present within the plasma membrane (Prenner *et al.*, 2007), however domains observed *in vivo* are generally of much smaller sizes and are formed much more transiently compared to their *in vitro* models (Lagerholm *et al.*, 2005). Over the course of lipid raft research numerous definitions of the subdomain have been proposed, with the most recent consensus describing rafts as heterogeneous membrane domains of 10-200 nm in diameter, which are dynamic structures composed of sterol- and sphingolipids that compartmentalise cellular processes (Pike, 2006). Rafts may also coalesce to form larger platforms for cell signalling via protein-protein and protein-lipid interactions, such as in T-cell activation where rafts are proposed to recruit signalling partners and cytoskeletal components for the maturation of the immunological synapse (Meiri, 2005).

Extractions of lipid rafts were first attempted with non-ionic detergents such as Triton X-100 (TX-100) under cold conditions (Brown and Rose, 1992). The extracted fraction was insoluble in TX-100 at 4°C and floated to a characteristic density in a sucrose density gradient. The fraction was called detergent resistant membrane (DRM) and showed an enrichment of raft components such as GPI anchored proteins and sphingolipids. The method was also sensitive to cholesterol depletion and has been used for the analysis lipid rafts in a variety of systems. As the field of lipid raft research matured it became apparent that detergent extraction may have several shortcomings as the *de-facto* method of raft extraction. Criticisms come from the procedure of extracting rafts at 4°C, which may not represent actual raft structures at the physiological temperature of 37°C. Detergents have also been shown to induce the formation of domains in cell membranes that may not reflect actual structures within the cell (Shogomori and Brown, 2003). However, despite the artefactual nature of detergent extraction for lipid raft analyses it is still one of the workhorse techniques within the field, and is often the first port of call for the isolation of raft proteins in a novel system. This is especially apparent in the relatively new field of lipid raft proteomics, in which the majority of projects use TX-100 insolubility as the method for raft extraction (Insenser *et al.*, 2006; Kim *et al.*, 2009; Nebl *et al.*, 2002). Proteomic projects tend to require relatively large amounts of proteins for analysis, which detergent extraction methods are able to provide.

The most common method for the extraction of GPI anchored proteins involves cleavage of the GPI anchor with PIPLC from crude membrane fractions. This procedure is relatively straightforward with commercial sources of the enzyme available purified from bacteria (such as *Bacillus thuringiensis*). Proteomic studies however require greater stringency as the sensitivity of mass spectrometry instruments are likely to pick up even trace amounts of contaminating proteins from the digestion, which leads to falsely identified proteins. Most proteomic projects on GPI anchored proteins therefore perform an additional sucrose density purification step on the crude membrane before PIPLC digestion to improve specificity and reduce false positive results (Borner *et al.*, 2003; Elortza *et al.*, 2003; Gilson *et al.*, 2006).

### 6.4.1 Lipid raft proteomics in *C. elegans*

The *C. elegans* lipid raft proteome was analysed in this project with a combination of 2D electrophoresis and MDLC, with both MS peptide mass fingerprinting (PMF) and tandem MS/MS methods used for the identification of the proteins. Overall 45 proteins were identified with these techniques from TX-100 extracted nematode DRM. Four of these proteins were found to belong to subcellular fractions that are unlikely to contain lipid rafts, such as mitochondria (Zheng *et al.*, 2009), nuclear membrane (Say and Hooper, 2007), ribosomes (Proud, 1994) and a secreted protein (Siafakas *et al.*, 2006), and were therefore removed from the final list. In the end 41 potential lipid raft proteins were identified in *C. elegans,* which makes this the largest study of raft associated proteins in the nematode to date. Five *C. elegans* galectins (LEC-1, 2, 3, 4 and 5) were found within the study, which has been found in other systems to be a group of proteins commonly associated with rafts. Galectins may also form distinct lattices with glycoproteins on the plasma membrane that act in concert with lipid rafts for their function (Lajoie *et al.*, 2009). Genes that may be involved in raft mediated signalling were also present in the analysis, such as two ryanodine receptor associated proteins of the $Ca^{2+}$ pathway and two proteins that act downstream of the insulin/insulin like growth factor pathway. Other proteins such as carboxypeptidases, stomatin-like proteins, apical gut protein, the *HSP90* homologue *daf-21*, trehalase, actin, components of the V-ATPase complex and vacuolar protein sorting proteins were also found within the study, which corresponds well with the results of lipid raft proteomic studies in other systems (Foster *et al.*, 2003; Insenser *et al.*, 2006; Kim *et al.*, 2009; von Haller *et al.*, 2001). One of these vacuolar genes was found to be the *C. elegans* vacuolar protein sorting factor *vps-32.1*, which had been shown to be localised in distinct domains to other proteins within endosomes (Michelet *et al.*, 2009). Of special interest is the finding that 21 of the identified proteins are in the list of predicted GPI anchored proteins generated for *C. elegans*. These accounts for over 50% of the raft proteins identified and may point to a significant role for GPI anchoring within the biology of the nematode. One of these proteins, TAG-10 is a homologue of the GA1 apical gut protein of the ruminant parasite *Haemonchus contortus*. GA1 was shown to have a GPI anchored form (Jasmer *et al.*, 1996) and has been demonstrated as a valid target for vaccination against the parasite (Yatsuda *et al.*, 2003). It will be interesting to see what the role of TAG-10 is in *C. elegans* and what function the protein has within the worm intestine, which may also lead to a greater understanding of the biology of GA1 in *H. contortus*.

### 6.4.2 GPI anchored proteomics in *C. elegans*

*C. elegans* GPI anchored proteins were also analysed specifically with the PIPLC digestion of extracted raft preparations. Due to the low yield of proteins we were unable to analyse them with MDLC, and instead identified them from 1D and 2D gel electrophoresis. Three proteins were identified from gel bands and spots with PMF. These were F56F10.1, a carboxypeptidase, ZK6.10 (DOD-19), a protein that acts downstream of the insulin pathway gene *daf-16*, and ZK6.11a. All three of these proteins were also present within the list of predicted nematode GPI anchored proteins, which indicate the validity of using a combinatorial *in silico* and *in vitro* approach for the identification of GPI anchored proteins. The number of GPI anchored proteins identified in proteomic projects have been generally low, with 11 identified in human HeLa cells (Elortza *et al.*, 2006) and 11 proteins in the malarial parasite *Plasmodium falciparum* (Gilson *et al.*, 2006). The number of identification of GPI anchored proteins in *A. thaliana* have been relatively high, with some projects reporting up to 44 GPI anchored proteins identified in their proteomic analysis (Elortza *et al.*, 2003). The results here present a tentative first look at the GPI anchored proteome of *C. elegans*, and offer a technique for further refinement, which may potentially yield a higher number of identified proteins in the future.

### 6.5 Future directions and conclusion

Studies of GPI anchored proteins and lipid rafts have been steadily gathering pace in recent years. *C. elegans* makes a compelling model organism for their study. The ease of making GFP expression patterns within the worm allows the study of the GPI biosynthesis genes within the context of development, which has hitherto not been possible with the common model organisms used to study this process. Expression profiles of the different transamidase genes could be made in the worm, as they appear to have important roles for the regulation of growth in many human cancers and are very well conserved within the *C. elegans* genome. The presence of transamidase components in the cytosol of many cancers also suggests additional roles for these genes within the cell beyond the attachment of GPI anchors (Nagpal *et al.*, 2008), which may also be investigated within the worm with RNAi knockout and deletion mutants. Currently the *C. elegans PIG-K* homologues T05E11.6 and T28H10.3, and one of the *PIG-U* homologues B0491.1, have deletion mutants according to Wormbase, and they all show an embryonically lethal phenotype, suggesting that the genes play important roles within the biology of the worm. Genetic analysis of the other *C. elegans* GPI synthesis and lipid modification genes may also be performed to give us a more robust understanding of the role of GPI anchoring within the nematode. Knockouts of the lipid modification genes with RNAi may also disrupt the association of GPI anchored proteins to lipid rafts, which would allow the analysis of the importance of lipid rafts on this class of proteins for nematode growth and development.

Lipid rafts may also be disrupted within the worm to find exactly how this subdomain functions within development. *C. elegans* does not have *de-novo* cholesterol synthesis and requires extracellular sources of the sterol for their normal development (Brenner, 1974). Analysis of the sterol requirements of *C. elegans* found that the worm does not need a large amount of cholesterol to survive, and the level of cholesterol intake was apparently not large enough for it to have a role in lipid raft formation (Entchev and Kurzchalia, 2005). If sterols are not present in large amounts in *C. elegans* membranes, then does the worm contain physiologically relevant rafts? Distribution studies of cholesterol with the fluorescent cholesterol analog dehydroergsterol (DHE) and the cholesterol stain filipin have shown the accumulation of the sterol in specific cells of the nematode, such as pharynx, nerve ring, excretory gland cells, gut apical surface cells, oocytes and spermatozoa (Matyash *et al.*, 2001; Merris *et al.*, 2003). This raises the possibility that rafts are not uniformly distributed in all cell types within the worm and that important properties of rafts, such as signal complex formation and apical sorting, may be used by the nematode in a tissue specific manner. This is also supported by previous work with *C. elegans cav-1* and *cav-2,* which showed that the genes were expressed in a cell specific manner after the embryonic stage of development (Parker *et al.*, 2007; Parker *et al.*, 2009). Alternatively *C. elegans* may be able to produce heterogeneity within its membranes using a cholesterol-independent method, such as the LEC-4 mediated microdomains that exist in the brush border membrane of enterocytes (Hansen *et al.*, 2001). *C. elegans* contains two homologues (R11H6.2 and Y57E12AL.1) for the gene *SERINC*, which incorporates serines into lipids and is a highly conserved gene for the production of sphingolipids (Inuzuka *et al.*, 2005). Knockouts of these genes could potentially disrupt lipid rafts within the worm, giving us a unique insight into the way these lipid domains act within a developmentally complex organism.

There is also scope for the expansion of proteomic studies for lipid raft and GPI anchored proteins in *C. elegans*. Proteomics projects of nematodes have become increasingly popular within recent years with many subcellular fractions such as glycoproteins and mitochondria been the subject of research (Audhya and Desai, 2008; Kaji *et al.*, 2007; Li *et al.*, 2009). The analysis of the lipid raft proteome presented here is unlikely to be complete as common components such as caveolin were not present within the final list of identified proteins, even though antibody staining had shown the presence of CAV-1 within the raft fraction. Previous work with the *C. elegans* CAV-1 showed that the protein is differentially localised on the post-synaptic membrane of neurons (Parker *et al.*, 2007), while *cav-2* was found to be involved in apical lipid trafficking in worm intestinal cells (Parker *et al.*, 2009). Raft components have been found to be responsible for polarised membrane formation in neurons (Kamiguchi, 2006) and apical sorting in epithelial cells (Schuck and Simons, 2004), which further suggests that CAV-1 and CAV-2 are a part of lipid rafts within the worm. Other techniques for the separation of peptides such as Strong Cation Exchange (SCX) or size exclusion chromatography (SEC) can be used in the first dimension to better separate the peptides (Motoyama and Yates, 2008), and more sensitive mass spectrometry instruments such as Orbitrap may also be used on the *C.elegans* lipid raft proteome for an improved quality of peptide sequencing (Han *et al.*, 2008), which may lead to a higher number of proteins identified. Nematodes can be grown in a synchronised manner, and raft proteins can be conceivably extracted from defined stages of their life cycle for proteomics analysis, which will give us insight into the global changes of the lipid raft proteome during the development and molting of the worms. Quantitative analysis of proteins can also be achieved with sample labelling such as isotope-coded affinity tags (ICAT) and isobaric tag for relative and absolute quantitation (iTRAQ) (Gygi *et al.*, 1999; Ross *et al.*, 2004); alternatively worms metabolically labelled with $^{15}N$ have been described in the literature which may be used for quantitative proteomics (Krijgsveld *et al.*, 2003). A larger sample size of *C. elegans* GPI anchored proteins could be obtained to allow MDLC analysis, which may produce a larger list of identified proteins. *C. elegans* GPI anchored proteins can also be cleaved from the membrane fraction with the enzyme phosphatidylinositol-specific phospholipase D (PIPLD), which cleaves the GPI anchor at a different point to PIPLC and allows the release of proteins from GPI anchors that have retained the acyl moiety on their inositol ring (Davitz *et al.*, 1987). Previous analysis with PIPLD have shown a different subset of proteins released in both human and *A. thaliana* cells (Elortza *et al.*, 2006; Elortza *et al.*, 2003), and it would be interesting to see if a different set of GPI anchored proteins would be released by this enzyme in the worm. Studies can also be performed for other important proteomes within the nematode, such as phosphorylated proteins and organelles, which would open up new doors for protein biochemistry within *C. elegans*.

In this study an analysis of GPI anchor biosynthesis, the GPI anchored proteome and the lipid raft proteome of *C. elegans* was performed. A comprehensive list of *C. elegans* homologues involved in all know aspects of GPI biosynthesis was presented here. An analysis of all potential GPI anchored proteins was also performed with four specialised prediction programs on the *C. elegans* genome, which yielded a list of 327 proteins that may be of value for further GPI anchored protein research. 41 lipid raft and 3 PIPLC released GPI anchored proteins were found from enriched fractions of the *C. elegans* membrane, which represents the largest number of identifications for these classes of proteins in the nematode to date. *C. elegans* can offer a unique perspective on the functions of GPI anchored proteins and lipid rafts in the context of tissue types, growth, aging, and development, and there is great potential for the nematode to become an important model organism in the study of these proteins and subcellular domains.

# References

Agostoni, E., Gobessi, S., Petrini, E., Monte, M., and Schneider, C. (2002). Cloning and characterization of the C. elegans gas1 homolog: phas-1. Biochim Biophys Acta *1574*, 1-9.

Ahmad, N., Gabius, H.J., Andre, S., Kaltner, H., Sabesan, S., Roy, R., Liu, B., Macaluso, F., and Brewer, C.F. (2004). Galectin-3 precipitates as a pentamer with synthetic multivalent carbohydrates and forms heterogeneous cross-linked complexes. J Biol Chem *279*, 10841-10847.

Ailion, M., and Thomas, J.H. (2000). Dauer formation induced by high temperatures in Caenorhabditis elegans. Genetics *156*, 1047-1067.

Alban, A., David, S.O., Bjorkesten, L., Andersson, C., Sloge, E., Lewis, S., and Currie, I. (2003). A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. Proteomics *3*, 36-44.

Alberts, B., Wilson, J., and Hunt, T. (2008). Molecular biology of the cell, 5th edn (New York, Garland Science).

Alfalah, M., Jacob, R., Preuss, U., Zimmer, K.P., Naim, H., and Naim, H.Y. (1999). O-linked glycans mediate apical sorting of human intestinal sucrase-isomaltase through association with lipid rafts. Curr Biol *9*, 593-596.

Alfieri, J.A., Martin, A.D., Takeda, J., Kondoh, G., Myles, D.G., and Primakoff, P. (2003). Infertility in female mice with an oocyte-specific knockout of GPI-anchored proteins. J Cell Sci *116*, 2149-2155.

Almeida, I.C., Camargo, M.M., Procopio, D.O., Silva, L.S., Mehlert, A., Travassos, L.R., Gazzinelli, R.T., and Ferguson, M.A. (2000). Highly purified glycosylphosphatidylinositols from Trypanosoma cruzi are potent proinflammatory agents. EMBO J *19*, 1476-1485.

America, A.H., and Cordewener, J.H. (2008). Comparative LC-MS: a landscape of peaks and valleys. Proteomics *8*, 731-749.

Anderson, H.A., Chen, Y., and Norkin, L.C. (1996). Bound simian virus 40 translocates to caveolin-enriched membrane domains, and its entry is inhibited by drugs that selectively disrupt caveolae. Mol Biol Cell *7*, 1825-1834.

Anderson, R.G. (1993). Caveolae: where incoming and outgoing messengers meet. Proc Natl Acad Sci U S A *90*, 10909-10913.

Anderson, R.G. (1998). The caveolae membrane system. Annu Rev Biochem *67*, 199-225.

Anderson, R.G., Kamen, B.A., Rothberg, K.G., and Lacey, S.W. (1992). Potocytosis: sequestration and transport of small molecules by caveolae. Science *255*, 410-411.

Anderson, R.G., and Orci, L. (1988). A view of acidic intracellular compartments. J Cell Biol *106*, 539-543.

Appleford, P.J., Griffiths, M., Yao, S.Y., Ng, A.M., Chomey, E.G., Isaac, R.E., Coates, D., Hope, I.A., Cass, C.E., Young, J.D., and Baldwin, S.A. (2004). Functional redundancy of two nucleoside transporters of the ENT family (CeENT1, CeENT2) required for development of Caenorhabditis elegans. Mol Membr Biol *21*, 247-259.

Arata, A., Sekiguchi, M., Hirabayashi, J., and Kasai, K. (2001). Effects of substitution of conserved amino acid residues on the sugar-binding property of the tandem-repeat 32-kDa galectin of the nematode Caenorhabditis elegans. Biol Pharm Bull *24*, 14-18.

Arata, Y., Akimoto, Y., Hirabayashi, J., Kasai, K., and Hirano, H. (1996). An immunohistochemical study of the 32-kDa galectin (beta-galactoside-binding lectin) in the nematode Caenorhabditis elegans. Histochem J *28*, 201-207.

Arata, Y., Hirabayashi, J., and Kasai, K. (1997). Structure of the 32-kDa galectin gene of the nematode Caenorhabditis elegans. J Biol Chem *272*, 26669-26677.
Ashida, H., Hong, Y., Murakami, Y., Shishioh, N., Sugimoto, N., Kim, Y.U., Maeda, Y., and Kinoshita, T. (2005a). Mammalian PIG-X and yeast Pbn1p are the essential components of glycosylphosphatidylinositol-mannosyltransferase I. Mol Biol Cell *16*, 1439-1448.
Ashida, H., Hong, Y., Murakami, Y., Shishioh, N., Sugimoto, N., Kim, Y.U., Maeda, Y., and Kinoshita, T. (2005b). Mammalian PIG-X and Yeast Pbn1p Are the Essential Components of Glycosylphosphatidylinositol-Mannosyltransferase I. Molecular Biology of the Cell *16*, 1439-1448.
Ashida, H., Maeda, Y., and kinoshita, T. (2005c). DPM1, the catalytic subunit of dolichol-phosphate mannose synthase, is tethered to and stabilized on the endoplasmic reticulum membrane by DPM3. The Journal of Biological Chemistry *281*, 896-904.
Ashida, H., Maeda, Y., and Kinoshita, T. (2006). DPM1, the catalytic subunit of dolichol-phosphate mannose synthase, is tethered to and stabilized on the endoplasmic reticulum membrane by DPM3. J Biol Chem *281*, 896-904.
Assadi, M., Sharpe, J.C., Snell, C., and Loh, Y.P. (2004). The C-terminus of prohormone convertase 2 is sufficient and necessary for Raft association and sorting to the regulated secretory pathway. Biochemistry *43*, 7798-7807.
Audhya, A., and Desai, A. (2008). Proteomics in Caenorhabditis elegans. Brief Funct Genomic Proteomic *7*, 205-210.
Bae, T.J., Kim, M.S., Kim, J.W., Kim, B.W., Choo, H.J., Lee, J.W., Kim, K.B., Lee, C.S., Kim, J.H., Chang, S.Y., Kang, C.Y., Lee, S.W., and Ko, Y.G. (2004). Lipid raft proteome reveals ATP synthase complex in the cell surface. Proteomics *4*, 3536-3548.
Baggerman, G., Boonen, K., Verleyen, P., De Loof, A., and Schoofs, L. (2005). Peptidomic analysis of the larval Drosophila melanogaster central nervous system by two-dimensional capillary liquid chromatography quadrupole time-of-flight mass spectrometry. J Mass Spectrom *40*, 250-260.
Balen, B., Krsnik-Rasol, M., Zamfir, A.D., Milosevic, J., Vakhrushev, S.Y., and Peter-Katalinic, J. (2006). Glycoproteomic survey of Mammillaria gracillis tissues grown in vitro. J Proteome Res *5*, 1658-1666.
Bangs, J.D., Hereld, D., Krakow, J.L., Hart, G.W., and Englund, P.T. (1985). Rapid processing of the carboxyl terminus of a trypanosome variant surface glycoprotein. Proc Natl Acad Sci U S A *82*, 3207-3211.
Barberan-Soler, S., and Zahler, A.M. (2008). Alternative splicing regulation during C. elegans development: splicing factors as regulated targets. PLoS Genet *4*, e1000001.
Bargmann, C.I. (2006). Chemosensation in C. elegans. WormBook, 1-29.
Bastiani, C., and Mendel, J. (2006). Heterotrimeric G proteins in C. elegans. WormBook, 1-25.
Beh, C.T., Ferrari, D.C., Chung, M.A., and McGhee, J.D. (1991). An acid phosphatase as a biochemical marker for intestinal development in the nematode Caenorhabditis elegans. Dev Biol *147*, 133-143.
Benachour, A., Sipos, G., Flury, I., Reggiori, F., Canivenc-Gansel, E., Vionnet, C., Conzelmann, A., and Benghezal, M. (1999). Deletion of GPI7, a yeast gene required for addition of a side chain to the glycosylphosphatidylinositol (GPI) core structure, affects GPI protein transport, remodeling, and cell wall integrity. J Biol Chem *274*, 15251-15261.

Bendtsen, J.D., Nielsen, H., von Heijne, G., and Brunak, S. (2004). Improved prediction of signal peptides: SignalP 3.0. J Mol Biol *340*, 783-795.
Benting, J.H., Rietveld, A.G., and Simons, K. (1999). N-Glycans mediate the apical sorting of a GPI-anchored, raft-associated protein in Madin-Darby canine kidney cells. J Cell Biol *146*, 313-320.
Beyenbach, K.W., and Wieczorek, H. (2006). The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. J Exp Biol *209*, 577-589.
Bickel, P.E. (2002). Lipid rafts and insulin signaling. Am J Physiol Endocrinol Metab *282*, E1-E10.
Bini, L., Pacini, S., Liberatori, S., Valensin, S., Pellegrini, M., Raggiaschi, R., Pallini, V., and Baldari, C.T. (2003). Extensive temporally regulated reorganization of the lipid raft proteome following T-cell antigen receptor triggering. Biochem J *369*, 301-309.
Bjellqvist, B., Ek, K., Righetti, P.G., Gianazza, E., Gorg, A., Westermeier, R., and Postel, W. (1982). Isoelectric focusing in immobilized pH gradients: principle, methodology and some applications. J Biochem Biophys Methods *6*, 317-339.
Blonder, J., Hale, M.L., Lucas, D.A., Schaefer, C.F., Yu, L.R., Conrads, T.P., Issaq, H.J., Stiles, B.G., and Veenstra, T.D. (2004). Proteomic analysis of detergent-resistant membrane rafts. Electrophoresis *25*, 1307-1318.
Blumenthal, T., Evans, D., Link, C.D., Guffanti, A., Lawson, D., Thierry-Mieg, J., Thierry-Mieg, D., Chiu, W.L., Duke, K., Kiraly, M., and Kim, S.K. (2002). A global analysis of Caenorhabditis elegans operons. Nature *417*, 851-854.
Boggs, J.M. (1987). Lipid intermolecular hydrogen bonding: influence on structural organization and membrane function. Biochim Biophys Acta *906*, 353-404.
Boothroyd, J.C. (1985). Antigenic variation in African trypanosomes. Annu Rev Microbiol *39*, 475-502.
Bordier, C. (1981). Phase separation of integral membrane proteins in Triton X-114 solution. J Biol Chem *256*, 1604-1607.
BORMAN, S., RUSSELL, H., and SIUZDAK, G. (2003). A mass spectrometry timeline. Today's Chemist at Work *13*, 47-49.
Borner, G.H., Lilley, K.S., Stevens, T.J., and Dupree, P. (2003). Identification of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A proteomic and genomic analysis. Plant Physiol *132*, 568-577.
Borner, G.H., Sherrier, D.J., Stevens, T.J., Arkin, I.T., and Dupree, P. (2002). Prediction of glycosylphosphatidylinositol-anchored proteins in Arabidopsis. A genomic analysis. Plant Physiol *129*, 486-499.
Borner, G.H., Sherrier, D.J., Weimar, T., Michaelson, L.V., Hawkins, N.D., Macaskill, A., Napier, J.A., Beale, M.H., Lilley, K.S., and Dupree, P. (2005). Analysis of detergent-resistant membranes in Arabidopsis. Evidence for plasma membrane lipid rafts. Plant Physiol *137*, 104-116.
Bosson, R., Jaquenoud, M., and Conzelmann, A. (2006). GUP1 of Saccharomyces cerevisiae encodes an O-acyltransferase involved in remodeling of the GPI anchor. Mol Biol Cell *17*, 2636-2645.
Brenner, S. (1974). The genetics of Caenorhabditis elegans. Genetics *77*, 71-94.
Brewis, I.A., Ferguson, M.A., Mehlert, A., Turner, A.J., and Hooper, N.M. (1995). Structures of the glycosyl-phosphatidylinositol anchors of porcine and human renal membrane dipeptidase. Comprehensive structural studies on the porcine anchor and interspecies comparison of the glycan core structures. J Biol Chem *270*, 22946-22956.

Brizard, J.P., Ramos, J., Robert, A., Lafitte, D., Bigi, N., Sarda, P., Laoudj-Chenivesse, D., Navarro, F., Blanc, P., Assenat, E., Maurel, P., Pascussi, J.M., and Vilarem, M.J. (2009). Identification of proteomic changes during human liver development by 2D-DIGE and mass spectrometry. J Hepatol *51*, 114-126.
Brodsky, R.A., and Hu, R. (2006). PIG-A mutations in paroxysmal nocturnal hemoglobinuria and in normal hematopoiesis. Leuk Lymphoma *47*, 1215-1221.
Brown, D.A. (1992). Interactions between GPI-anchored proteins and membrane lipids. Trends Cell Biol *2*, 338-343.
Brown, D.A., and Rose, J.K. (1992). Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. Cell *68*, 533-544.
Bulow, R., and Overath, P. (1986). Purification and characterization of the membrane-form variant surface glycoprotein hydrolase of Trypanosoma brucei. J Biol Chem *261*, 11918-11923.
Burglin, T.R., and Kuwabara, P.E. (2006). Homologs of the Hh signalling network in C. elegans. WormBook, 1-14.
Calderwood, S.K., Khaleque, M.A., Sawyer, D.B., and Ciocca, D.R. (2006). Heat shock proteins in cancer: chaperones of tumorigenesis. Trends Biochem Sci *31*, 164-172.
Callera, G.E., Montezano, A.C., Yogi, A., Tostes, R.C., and Touyz, R.M. (2007). Vascular signaling through cholesterol-rich domains: implications in hypertension. Curr Opin Nephrol Hypertens *16*, 90-104.
Canas, B., Lopez-Ferrer, D., Ramos-Fernandez, A., Camafeita, E., and Calvo, E. (2006). Mass spectrometry technologies for proteomics. Brief Funct Genomic Proteomic *4*, 295-320.
Caras, I.W., Weddell, G.N., and Williams, S.R. (1989). Analysis of the signal for attachment of a glycophospholipid membrane anchor. J Cell Biol *108*, 1387-1396.
Carcy, B., Precigout, E., Schetters, T., and Gorenflot, A. (2006). Genetic basis for GPI-anchor merozoite surface antigen polymorphism of Babesia and resulting antigenic diversity. Vet Parasitol *138*, 33-49.
Cargile, B.J., Sevinsky, J.R., Essader, A.S., Stephenson, J.L., Jr., and Bundy, J.L. (2005). Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. J Biomol Tech *16*, 181-189.
Carson, J.A., and Turner, A.J. (2002). Beta-amyloid catabolism: roles for neprilysin (NEP) and other metallopeptidases? J Neurochem *81*, 1-8.
Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W., and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. Science *263*, 802-805.
Chamberlain, L.H. (2004). Detergents as tools for the purification and classification of lipid rafts. FEBS Lett *559*, 1-5.
Chamberlain, L.H., and Gould, G.W. (2002). The vesicle- and target-SNARE proteins that mediate Glut4 vesicle fusion are localized in detergent-insoluble lipid rafts present on distinct intracellular membranes. J Biol Chem *277*, 49750-49754.
Chen, J., Lee, C.S., Shen, Y., Smith, R.D., and Baehrecke, E.H. (2002). Integration of capillary isoelectric focusing with capillary reversed-phase liquid chromatography for two-dimensional proteomics separation. Electrophoresis *23*, 3143-3148.
Chen, R., Anderson, V., Hiroi, Y., and Medof, M.E. (2003a). Proprotein interaction with the GPI transamidase. J Cell Biochem *88*, 1025-1037.
Chen, Y., Yu, P., Luo, J., and Jiang, Y. (2003b). Secreted protein prediction system combining CJ-SPHMM, TMHMM, and PSORT. Mamm Genome *14*, 859-865.

Chichili, G.R., and Rodgers, W. (2009). Cytoskeleton-membrane interactions in membrane raft structure. Cell Mol Life Sci *66*, 2319-2328.
Choi, B.K., Chitwood, D.J., and Paik, Y.K. (2003). Proteomic changes during disturbance of cholesterol metabolism by azacoprostane treatment in Caenorhabditis elegans. Mol Cell Proteomics *2*, 1086-1095.
Choi, Y.W., Tan, Y.J., Lim, S.G., Hong, W., and Goh, P.Y. (2004). Proteomic approach identifies HSP27 as an interacting partner of the hepatitis C virus NS5A protein. Biochem Biophys Res Commun *318*, 514-519.
Chou, J.H., Bargmann, C.I., and Sengupta, P. (2001). The Caenorhabditis elegans odr-2 gene encodes a novel Ly-6-related protein required for olfaction. Genetics *157*, 211-224.
Chou, K.C., and Shen, H.B. (2007). MemType-2L: a web server for predicting membrane proteins and their types by incorporating evolution information through Pse-PSSM. Biochem Biophys Res Commun *360*, 339-345.
Churchward, M.A., Butt, R.H., Lang, J.C., Hsu, K.K., and Coorssen, J.R. (2005). Enhanced detergent extraction for analysis of membrane proteomes by two-dimensional gel electrophoresis. *3*, 5.
Claessens, H.A., and van Straten, M.A. (2004). Review on the chemical and thermal stability of stationary phases for reversed-phase liquid chromatography. J Chromatogr A *1060*, 23-41.
Colussi, P.A., Taron, C.H., Mack, J.C., and Orlean, P. (1997). Human and Saccharomyces cerevisiae dolichol phosphate mannose synthases represent two classes of the enzyme, but both function in Schizosaccharomyces pombe. Proc Natl Acad Sci U S A *94*, 7873-7878.
Conrad, D.H., Goyette, J., and Thomas, P.S. (2008). Proteomics as a method for early detection of cancer: a review of proteomics, exhaled breath condensate, and lung cancer screening. J Gen Intern Med *23 Suppl 1*, 78-84.
Consortium, T.*C.e.*S. (1998). Genome sequence of the nematode C. elegans: a platform for investigating biology. Science *282*, 2012-2018.
Conzelmann, A., Spiazzi, A., and Bron, C. (1987). Glycolipid anchors are attached to Thy-1 glycoprotein rapidly after translation. Biochem J *246*, 605-610.
Cordy, J.M., Hooper, N.M., and Turner, A.J. (2006). The involvement of lipid rafts in Alzheimer's disease. Mol Membr Biol *23*, 111-122.
Coyne, K.E., Crisci, A., and Lublin, D.M. (1993). Construction of synthetic signals for glycosyl-phosphatidylinositol anchor attachment. Analysis of amino acid sequence requirements for anchoring. J Biol Chem *268*, 6689-6693.
Cram, E.J., Shang, H., and Schwarzbauer, J.E. (2006). A systematic RNA interference screen reveals a cell migration gene network in C. elegans. J Cell Sci *119*, 4811-4818.
Csermely, P., Schnaider, T., Soti, C., Prohaszka, Z., and Nardai, G. (1998). The 90-kDa molecular chaperone family: structure, function, and clinical applications. A comprehensive review. Pharmacol Ther *79*, 129-168.
Danielsen, E.M., and Hansen, G.H. (2003). Lipid rafts in epithelial brush borders: atypical membrane microdomains with specialized functions. Biochim Biophys Acta *1617*, 1-9.
Danielsen, E.M., and Hansen, G.H. (2008). Lipid raft organization and function in the small intestinal brush border. J Physiol Biochem *64*, 377-382.
David, C.L., Smith, H.E., Raynes, D.A., Pulcini, E.J., and Whitesell, L. (2003). Expression of a unique drug-resistant Hsp90 ortholog by the nematode Caenorhabditis elegans. Cell Stress Chaperones *8*, 93-104.

Davitz, M.A., Hereld, D., Shak, S., Krakow, J., Englund, P.T., and Nussenzweig, V. (1987). A glycan-phosphatidylinositol-specific phospholipase D in human serum. Science *238*, 81-84.

Davydenko, S.G., Feng, D., Jantti, J., and Keranen, S. (2005). Characterization of GPI14/YJR013w mutation that induces the cell wall integrity signalling pathway and results in increased protein production in Saccharomyces cerevisiae. Yeast *22*, 993-1009.

Deeg, M.A., and Davitz, M.A. (1995). Glycosylphosphatidylinositol-phospholipase D: a tool for glycosylphosphatidylinositol structural analysis. Methods Enzymol *250*, 630-640.

Deeg, M.A., Humphrey, D.R., Yang, S.H., Ferguson, T.R., Reinhold, V.N., and Rosenberry, T.L. (1992). Glycan components in the glycoinositol phospholipid anchor of human erythrocyte acetylcholinesterase. Novel fragments produced by trifluoroacetic acid. J Biol Chem *267*, 18573-18580.

Delacour, D., Koch, A., Ackermann, W., Eude-Le Parco, I., Elsasser, H.P., Poirier, F., and Jacob, R. (2008). Loss of galectin-3 impairs membrane polarisation of mouse enterocytes in vivo. J Cell Sci *121*, 458-465.

Delahunty, C., and Yates III, J.R. (2005). Protein identification using 2D-LC-MS/MS. Methods *35*, 248–255.

Dermine, J.F., Duclos, S., Garin, J., St-Louis, F., Rea, S., Parton, R.G., and Desjardins, M. (2001). Flotillin-1-enriched lipid raft domains accumulate on maturing phagosomes. J Biol Chem *276*, 18507-18512.

Devaney, E., O'Neill, K., Harnett, W., Whitesell, L., and Kinnaird, J.H. (2005). Hsp90 is essential in the filarial nematode Brugia pahangi. Int J Parasitol *35*, 627-636.

Dhanvantari, S., and Loh, Y.P. (2000). Lipid raft association of carboxypeptidase E is necessary for its function as a regulated secretory pathway sorting receptor. J Biol Chem *275*, 29887-29893.

Draberova, L., Amoui, M., and Draber, P. (1996). Thy-1-mediated activation of rat mast cells: the role of Thy-1 membrane microdomains. Immunology *87*, 141-148.

Drevot, P., Langlet, C., Guo, X.J., Bernard, A.M., Colard, O., Chauvin, J.P., Lasserre, R., and He, H.T. (2002). TCR signal initiation machinery is pre-assembled and activated in a subset of membrane rafts. EMBO J *21*, 1899-1908.

Dunkley, T.P., Watson, R., Griffin, J.L., Dupree, P., and Lilley, K.S. (2004). Localization of organelle proteins by isotope tagging (LOPIT). Mol Cell Proteomics *3*, 1128-1134.

Dupuy, D., Li, Q.R., Deplancke, B., Boxem, M., Hao, T., Lamesch, P., Sequerra, R., Bosak, S., Doucette-Stamm, L., Hope, I.A., Hill, D.E., Walhout, A.J., and Vidal, M. (2004). A first version of the Caenorhabditis elegans Promoterome. Genome Res *14*, 2169-2175.

Eisenhaber, B., Bork, P., and Eisenhaber, F. (1998). Sequence properties of GPI-anchored proteins near the omega-site: constraints for the polypeptide binding site of the putative transamidase. Protein Eng *11*, 1155-1161.

Eisenhaber, B., Bork, P., and Eisenhaber, F. (1999). Prediction of potential GPI-modification sites in proprotein sequences. J Mol Biol *292*, 741-758.

Eisenhaber, B., Bork, P., and Eisenhaber, F. (2001). Post-translational GPI lipid anchor modification of proteins in kingdoms of life: analysis of protein sequence data from complete genomes. Protein Eng *14*, 17-25.

Eisenhaber, B., Bork, P., Yuan, Y., Loffler, G., and Eisenhaber, F. (2000). Automated annotation of GPI anchor sites: case study C. elegans. Trends Biochem Sci *25*, 340-341.
Eisenhaber, B., Maurer-Stroh, S., Novatchkova, M., Schneider, G., and eisenhaber, F. (2003a). Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer to proteins. BioEssays *25*, 367-385.
Eisenhaber, B., Maurer-Stroh, S., Novatchkova, M., Schneider, G., and Eisenhaber, F. (2003b). Enzymes and auxiliary factors for GPI lipid anchor biosynthesis and post-translational transfer to proteins. Bioessays *25*, 367-385.
Eisenhaber, B., Wildpaner, M., Schultz, C.J., Borner, G.H., Dupree, P., and Eisenhaber, F. (2003c). Glycosylphosphatidylinositol lipid anchoring of plant proteins. Sensitive prediction from sequence- and genome-wide studies for Arabidopsis and rice. Plant Physiol *133*, 1691-1701.
Eisenhaber, F., Eisenhaber, B., Kubina, W., Maurer-Stroh, S., Neuberger, G., Schneider, G., and Wildpaner, M. (2003d). Prediction of lipid posttranslational modifications and localization signals from protein sequences: big-Pi, NMT and PTS1. Nucleic Acids Res *31*, 3631-3634.
Elortza, F., Mohammed, S., Bunkenborg, J., Foster, L.J., Nuhse, T.S., Brodbeck, U., Peck, S.C., and Jensen, O.N. (2006). Modification-specific proteomics of plasma membrane proteins: identification and characterization of glycosylphosphatidylinositol-anchored proteins released upon phospholipase D treatment. J Proteome Res *5*, 935-943.
Elortza, F., Nuhse, T.S., Foster, L.J., Stensballe, A., Peck, S.C., and Jensen, O.N. (2003). Proteomic analysis of glycosylphosphatidylinositol-anchored membrane proteins. Mol Cell Proteomics *2*, 1261-1270.
Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP and related tools. Nat Protoc *2*, 953-971.
Engstrom, Y., Loseva, O., and Theopold, U. (2004). Proteomics of the Drosophila immune response. Trends Biotechnol *22*, 600-605.
Entchev, E.V., and Kurzchalia, T.V. (2005). Requirement of sterols in the life cycle of the nematode Caenorhabditis elegans. Semin Cell Dev Biol *16*, 175-182.
Ernesto S Nakayasu, E., Dmitry V Yashunsky, D.V., Lilian L Nohara, L.L., Ana Claudia T Torrecilhas, A.C.T., Andrei V Nikolaev, A.V., and Igor C Almeida, I.C. (2009). GPIomics global analysis of glycosylphosphatidylinositol-anchored molecules of Trypanosoma cruzi. Molecular Systems Biology, 261.
Evans, C.A., Tonge, R., Blinco, D., Pierce, A., Shaw, J., Lu, Y., Hamzah, H.G., Gray, A., Downes, C.P., Gaskell, S.J., Spooncer, E., and Whetton, A.D. (2004). Comparative proteomics of primitive hematopoietic cell populations reveals differences in expression of proteins regulating motility. Blood *103*, 3751-3759.
Fabre, A.L., Orlean, P., and Taron, C.H. (2005). Saccharomyces cerevisiae Ybr004c and its human homologue are required for addition of the second mannose during glycosylphosphatidylinositol precursor assembly. FEBS J *272*, 1160-1168.
Fan, X., She, Y.M., Bagshaw, R.D., Callahan, J.W., Schachter, H., and Mahuran, D.J. (2005). Identification of the hydrophobic glycoproteins of Caenorhabditis elegans. Glycobiology *15*, 952-964.
Fankhauser, N., and Maser, P. (2005). Identification of GPI anchor attachment signals by a Kohonen self-organizing map. Bioinformatics *21*, 1846-1852.

Fenn, J.B., Mann, M., Meng, C.K., Wong, S.F., and Whitehouse, C.M. (1989). Electrospray ionization for mass spectrometry of large biomolecules. Science *246*, 64-71.
Ferguson, M.A. (1999). The structure, biosynthesis and functions of glycosylphosphatidylinositol anchors, and the contributions of trypanosome research. J Cell Sci *112 ( Pt 17)*, 2799-2809.
Ferguson, M.A. (2000). Glycosylphosphatidylinositol biosynthesis validated as a drug target for African sleeping sickness. Proc Natl Acad Sci U S A *97*, 10673-10675.
Ferguson, M.A., Brimacombe, J.S., Brown, J.R., Crossman, A., Dix, A., Field, R.A., Guther, M.L., Milne, K.G., Sharma, D.K., and Smith, T.K. (1999). The GPI biosynthetic pathway as a therapeutic target for African sleeping sickness. Biochim Biophys Acta *1455*, 327-340.
Ferguson, M.A., Haldar, K., and Cross, G.A. (1985a). Trypanosoma brucei variant surface glycoprotein has a sn-1,2-dimyristyl glycerol membrane anchor at its COOH terminus. J Biol Chem *260*, 4963-4968.
Ferguson, M.A., Homans, S.W., Dwek, R.A., and Rademacher, T.W. (1988). Glycosyl-phosphatidylinositol moiety that anchors Trypanosoma brucei variant surface glycoprotein to the membrane. Science *239*, 753-759.
Ferguson, M.A., Low, M.G., and Cross, G.A. (1985b). Glycosyl-sn-1,2-dimyristylphosphatidylinositol is covalently linked to Trypanosoma brucei variant surface glycoprotein. J Biol Chem *260*, 14547-14555.
Ferguson, M.A., Murray, P., Rutherford, H., and McConville, M.J. (1993). A simple purification of procyclic acidic repetitive protein and demonstration of a sialylated glycosyl-phosphatidylinositol membrane anchor. Biochem J *291 ( Pt 1)*, 51-55.
Ferguson, M.A., and Williams, A.F. (1988). Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu Rev Biochem *57*, 285-320.
Feuk-Lagerstedt, E., Movitz, C., Pellme, S., Dahlgren, C., and Karlsson, A. (2007). Lipid raft proteome of the human neutrophil azurophil granule. Proteomics *7*, 194-205.
Fiedler, K., Parton, R.G., Kellner, R., Etzold, T., and Simons, K. (1994). VIP36, a novel component of glycolipid rafts and exocytic carrier vesicles in epithelial cells. EMBO J *13*, 1729-1740.
Field, M.C., Menon, A.K., and Cross, G.A. (1991). A glycosylphosphatidylinositol protein anchor from procyclic stage Trypanosoma brucei: lipid structure and biosynthesis. EMBO J *10*, 2731-2739.
Fire, A. (1986). Integrative transformation of Caenorhabditis elegans. EMBO J *5*, 2673-2680.
Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811.
Fivaz, M., Vilbois, F., Thurnheer, S., Pasquali, C., Abrami, L., Bickel, P.E., Parton, R.G., and van der Goot, F.G. (2002). Differential sorting and fate of endocytosed GPI-anchored proteins. EMBO J *21*, 3989-4000.
Fontaine, T., Magnin, T., Melhert, A., Lamont, D., Latge, J.P., and Ferguson, M.A. (2003). Structures of the glycosylphosphatidylinositol membrane anchors from Aspergillus fumigatus membrane proteins. Glycobiology *13*, 169-177.
Fontana, J., Fulton, D., Chen, Y., Fairchild, T.A., McCabe, T.J., Fujita, N., Tsuruo, T., and Sessa, W.C. (2002). Domain mapping studies reveal that the M domain of hsp90 serves as a molecular scaffold to regulate Akt-dependent phosphorylation of endothelial nitric oxide synthase and NO release. Circ Res *90*, 866-873.

Foster, L.J. (2008). Lessons learned from lipid raft proteomics. Expert Rev Proteomics *5*, 541-543.
Foster, L.J., De Hoog, C.L., and Mann, M. (2003). Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. Proc Natl Acad Sci U S A *100*, 5813-5818.
Fraering, P., Imhof, I., Meyer, U., Strub, J.M., van Dorsselaer, A., Vionnet, C., and Conzelmann, A. (2001). The GPI transamidase complex of Saccharomyces cerevisiae contains Gaa1p, Gpi8p, and Gpi16p. Mol Biol Cell *12*, 3295-3306.
Fricke, B., Stewart, G.W., Treharne, K.J., Mehta, A., Knopfle, G., Friedrichs, N., Muller, K.M., and von During, M. (2003). Stomatin immunoreactivity in ciliated cells of the human airway epithelium. Anat Embryol (Berl) *207*, 1-7.
Fujita, M., and Jigami, Y. (2008). Lipid remodeling of GPI-anchored proteins and its function. Biochim Biophys Acta *1780*, 410-420.
Fujita, M., Umemura, M., Yoko-o, T., and Jigami, Y. (2006a). PER1 is required for GPI-phospholipase A2 activity and involved in lipid remodeling of GPI-anchored proteins. Mol Biol Cell *17*, 5253-5264.
Fujita, M., Yoko-o, T., Okamoto, M., and Jigami, Y. (2004). GPI7 involved in glycosylphosphatidylinositol biosynthesis is essential for yeast cell separation. J Biol Chem *279*, 51869-51879.
Fujita, M., Yoko, O.T., and Jigami, Y. (2006b). Inositol deacylation by Bst1p is required for the quality control of glycosylphosphatidylinositol-anchored proteins. Mol Biol Cell *17*, 834-850.
Furukawa, Y., Tsukamoto, K., and Ikezawa, H. (1997). Mutational analysis of the C-terminal signal peptide of bovine liver 5'-nucleotidase for GPI anchoring: a study on the significance of the hydrophilic spacer region. Biochim Biophys Acta *1328*, 185-196.
Gage, M.C., Keen, J.N., Buxton, A.T., Bedi, M.K., and Findlay, J.B. (2009). Proteomic analysis of IgE-mediated secretion by LAD2 mast cells. J Proteome Res *8*, 4116-4125.
Gallegos, A.M., Storey, S.M., Kier, A.B., Schroeder, F., and Ball, J.M. (2006). Structure and cholesterol dynamics of caveolae/raft and nonraft plasma membrane domains. Biochemistry *45*, 12100-12116.
Garcia, A., Cayla, X., Fleischer, A., Guergnon, J., Alvarez-Franco Canas, F., Rebollo, M.P., Roncal, F., and Rebollo, A. (2003). Rafts: a simple way to control apoptosis by subcellular redistribution. Biochimie *85*, 727-731.
Garner, A.E., Smith, D.A., and Hooper, N.M. (2008). Visualization of detergent solubilization of membranes: implications for the isolation of rafts. Biophys J *94*, 1326-1340.
Gaynor, E.C., Mondesert, G., Grimme, S.J., Reed, S.I., Orlean, P., and Emr, S.D. (1999). MCD4 encodes a conserved endoplasmic reticulum membrane protein essential for glycosylphosphatidylinositol anchor synthesis in yeast. Mol Biol Cell *10*, 627-648.
Gerber, L.D., Kodukula, K., and Udenfriend, S. (1992). Phosphatidylinositol glycan (PI-G) anchored membrane proteins. Amino acid requirements adjacent to the site of cleavage and PI-G attachment in the COOH-terminal signal peptide. J Biol Chem *267*, 12168-12173.
Ghugtyal, V., Vionnet, C., Roubaty, C., and Conzelmann, A. (2007). CWH43 is required for the introduction of ceramides into GPI anchors in Saccharomyces cerevisiae. Molecular Microbiology *65*, 1493-1502.

Giddings, J.C. (1984). Two-dimensional separations: concept and promise. Anal Chem *56*, 1258A-1260A, 1262A, 1264A passim.
Gilson, P.R., Nebl, T., Vukcevic, D., Moritz, R.L., Sargeant, T., Speed, T.P., Schofield, L., and Crabb, B.S. (2006). Identification and stoichiometry of glycosylphosphatidylinositol-anchored membrane proteins of the human malaria parasite Plasmodium falciparum. Mol Cell Proteomics *5*, 1286-1299.
Gottschalk, A., Almedom, R.B., Schedletzky, T., Anderson, S.D., Yates, J.R., 3rd, and Schafer, W.R. (2005). Identification and characterization of novel nicotinic receptor-associated proteins in Caenorhabditis elegans. EMBO J *24*, 2566-2578.
Greetham, D., Morgan, C., Campbell, A.M., van Rossum, A.J., Barrett, J., and Brophy, P.M. (2004). Evidence of glutathione transferase complexing and signaling in the model nematode Caenorhabditis elegans using a pull-down proteomic assay. Proteomics *4*, 1989-1995.
Grijalba, M.T., Vercesi, A.E., and Schreier, S. (1999). Ca2+-induced increased lipid packing and domain formation in submitochondrial particles. A possible early step in the mechanism of Ca2+-stimulated generation of reactive oxygen species by the respiratory chain. Biochemistry *38*, 13279-13287.
Grimme, S.J., Gao, X.D., Martin, P.S., Tu, K., Tcheperegine, S.E., Corrado, K., Farewell, A.E., Orlean, P., and Bi, E. (2004). Deficiencies in the endoplasmic reticulum (ER)-membrane protein Gab1p perturb transfer of glycosylphosphatidylinositol to proteins and cause perinuclear ER-associated actin bar formation. Mol Biol Cell *15*, 2758-2770.
Grimme, S.J., Westfall, B.A., Wiedman, J.M., Taron, C.H., and Orlean, P. (2001). The essential Smp3 protein is required for addition of the side-branching fourth mannose during assembly of yeast glycosylphosphatidylinositols. J Biol Chem *276*, 27731-27739.
Grunfelder, C.G., Engstler, M., Weise, F., Schwarz, H., Stierhof, Y.D., Boshart, M., and Overath, P. (2002). Accumulation of a GPI-anchored protein at the cell surface requires sorting at multiple intracellular levels. Traffic *3*, 547-559.
GuhaThakurta, D., Palomar, L., Stormo, G.D., Tedesco, P., Johnson, T.E., Walker, D.W., Lithgow, G., Kim, S., and Link, C.D. (2002). Identification of a novel cis-regulatory element involved in the heat shock response in Caenorhabditis elegans using microarray gene expression and computational methods. Genome Res *12*, 701-712.
Guo, Z., Linn, J.F., Wu, G., Anzick, S.L., Eisenberger, C.F., Halachmi, S., Cohen, Y., Fomenkov, A., Hoque, M.O., Okami, K., Steiner, G., Engles, J.M., Osada, M., Moon, C., Ratovitski, E., Trent, J.M., Meltzer, P.S., Westra, W.H., Kiemeney, L.A., Schoenberg, M.P., Sidransky, D., and Trink, B. (2004). CDC91L1 (PIG-U) is a newly discovered oncogene in human bladder cancer. Nat Med *10*, 374-381.
Guther, M.L., Leal, S., Morrice, N.A., Cross, G.A., and Ferguson, M.A. (2001). Purification, cloning and characterization of a GPI inositol deacylase from Trypanosoma brucei. EMBO J *20*, 4923-4934.
Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., and Aebersold, R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol *17*, 994-999.
Halaschek-Wiener, J., Khattra, J.S., McKay, S., Pouzyrev, A., Stott, J.M., Yang, G.S., Holt, R.A., Jones, S.J., Marra, M.A., Brooks-Wilson, A.R., and Riddle, D.L. (2005). Analysis of long-lived C. elegans daf-2 mutants using serial analysis of gene expression. Genome Res *15*, 603-615.

Han, M.J., and Lee, S.Y. (2006). The Escherichia coli proteome: past, present, and future prospects. Microbiol Mol Biol Rev *70*, 362-439.
Han, X., Aslanian, A., and Yates, J.R., 3rd (2008). Mass spectrometry for proteomics. Curr Opin Chem Biol *12*, 483-490.
Hansen, G.H., Immerdal, L., Thorsen, E., Niels-Christiansen, L.L., Nystrom, B.T., Demant, E.J., and Danielsen, E.M. (2001). Lipid rafts exist as stable cholesterol-independent microdomains in the brush border membrane of enterocytes. J Biol Chem *276*, 32338-32344.
Hansen, G.H., Pedersen, E.D., Immerdal, L., Niels-Christiansen, L.L., and Danielsen, E.M. (2005). Anti-glycosyl antibodies in lipid rafts of the enterocyte brush border: a possible host defense against pathogens. Am J Physiol Gastrointest Liver Physiol *289*, G1100-1107.
Harder, T., and Simons, K. (1997). Caveolae, DIGs, and the dynamics of sphingolipid-cholesterol microdomains. Curr Opin Cell Biol *9*, 534-542.
Harris, T.J., and Siu, C.H. (2002). Reciprocal raft-receptor interactions and the assembly of adhesion complexes. Bioessays *24*, 996-1003.
Hartley, J.L., Temple, G.F., and Brasch, M.A. (2000). DNA cloning using in vitro site-specific recombination. Genome Res *10*, 1788-1795.
He, H., Cai, L., Skogerbo, G., Deng, W., Liu, T., Zhu, X., Wang, Y., Jia, D., Zhang, Z., Tao, Y., Zeng, H., Aftab, M.N., Cui, Y., Liu, G., and Chen, R. (2006). Profiling Caenorhabditis elegans non-coding RNA expression with a combined microarray. Nucleic Acids Res *34*, 2976-2983.
Heerklotz, H. (2002). Triton promotes domain formation in lipid raft mixtures. Biophys J *83*, 2693-2701.
Helms, J.B., and Zurzolo, C. (2004). Lipids as targeting signals: lipid rafts and intracellular trafficking. Traffic *5*, 247-254.
Hernandez, P., Muller, M., and Appel, R.D. (2006). Automated protein identification by tandem mass spectrometry: issues and strategies. Mass Spectrom Rev *25*, 235-254.
Hertz, R., and Barenholz, Y. (1977). Relations between Composition of Liposomes and Their Interaction with Triton X-100. J Colloid Interf Sci *60*, 188-200.
Hillenkamp, F., Karas, M., Beavis, R.C., and Chait, B.T. (1991). Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers. Anal Chem *63*, 1193A-1203A.
Him, N.A., Gillan, V., Emes, R.D., Maitland, K., and Devaney, E. (2009). Hsp-90 and the biology of nematodes. BMC Evol Biol *9*, 254.
Hirose, S., Prince, G.M., Sevlever, D., Ravi, L., Rosenberry, T.L., Ueda, E., and Medof, M.E. (1992). Characterization of putative glycoinositol phospholipid anchor precursors in mammalian cells. Localization of phosphoethanolamine. J Biol Chem *267*, 16968-16974.
Hoekstra, D., Maier, O., van der Wouden, J.M., Slimane, T.A., and van, I.S.C. (2003). Membrane dynamics and cell polarity: the role of sphingolipids. J Lipid Res *44*, 869-877.
Holder, A.A. (1983). Carbohydrate is linked through ethanolamine to the C-terminal amino acid of Trypanosoma brucei variant surface glycoprotein. Biochem J *209*, 261-262.
Homans, S.W., Edge, C.J., Ferguson, M.A., Dwek, R.A., and Rademacher, T.W. (1989). Solution structure of the glycosylphosphatidylinositol membrane anchor glycan of Trypanosoma brucei variant surface glycoprotein. Biochemistry *28*, 2881-2887.

Homans, S.W., Ferguson, M.A., Dwek, R.A., Rademacher, T.W., Anand, R., and Williams, A.F. (1988). Complete structure of the glycosyl phosphatidylinositol membrane anchor of rat brain Thy-1 glycoprotein. Nature *333*, 269-272.
Hong, Y., and Kinoshita, T. (2009). Trypanosome glycosylphosphatidylinositol biosynthesis. Korean J Parasitol *47*, 197-204.
Hong, Y., Maeda, Y., Watanabe, R., Inoue, N., Ohishi, K., and Kinoshita, T. (2000). Requirement of PIG-F and PIG-O for transferring phosphoethanolamine to the third mannose in glycosylphosphatidylinositol. J Biol Chem *275*, 20911-20919.
Hong, Y., Maeda, Y., Watanabe, R., Ohishi, K., Mishkind, M., Riezman, H., and Kinoshita, T. (1999a). Pig-n, a mammalian homologue of yeast Mcd4p, is involved in transferring phosphoethanolamine to the first mannose of the glycosylphosphatidylinositol. J Biol Chem *274*, 35099-35106.
Hong, Y., Nagamune, K., Morita, Y.S., Nakatani, F., Ashida, H., Maeda, Y., and Kinoshita, T. (2006). Removal or maintenance of inositol-linked acyl chain in glycosylphosphatidylinositol is critical in trypanosome life cycle. J Biol Chem *281*, 11595-11602.
Hong, Y., Ohishi, K., Kang, J.Y., Tanaka, S., Inoue, N., Nishimura, J., Maeda, Y., and kinoshita, T. (2003). Human PIG-U and Yeast Cdc91p Are the Fifth Subunit of GPI Transamidase That Attaches GPI-Anchors to Proteins. Molecular Biology of the Cell *14*, 1780-1789.
Hong, Y., Ohishi, K., Watanabe, R., Endo, Y., Maeda, Y., and Kinoshita, T. (1999b). GPI1 stabilizes an enzyme essential in the first step of glycosylphosphatidylinositol biosynthesis. journal of Biological Chemistry *274*, 18582-18588.
Hooper, N.M., and Turner, A.J. (1988). Ectoenzymes of the kidney microvillar membrane. Differential solubilization by detergents can predict a glycosyl-phosphatidylinositol membrane anchor. Biochem J *250*, 865-869.
Hooper, N.M., and Turner, A.J. (1992). Lipid modification of proteins : a practical approach (Oxford, IRL Press at Oxford University Press).
Hope, H.R., and Pike, L.J. (1996). Phosphoinositides and phosphoinositide-utilizing enzymes in detergent-insoluble lipid domains. Mol Biol Cell *7*, 843-851.
Hope, I.A. (1991). 'Promoter trapping' in Caenorhabditis elegans. Development *113*, 399-408.
Hope, I.A. (1999). C. elegans : a practical approach (Oxford, Oxford University Press).
Hortsch, M., and Goodman, C.S. (1990). Drosophila fasciclin I, a neural cell adhesion molecule, has a phosphatidylinositol lipid membrane anchor that is developmentally regulated. J Biol Chem *265*, 15104-15109.
Howell, S., Lanctot, C., Boileau, G., and Crine, P. (1994). A cleavable N-terminal signal peptide is not a prerequisite for the biosynthesis of glycosylphosphatidylinositol-anchored proteins. J Biol Chem *269*, 16993-16996.
Hsu, D.K., Chernyavsky, A.I., Chen, H.Y., Yu, L., Grando, S.A., and Liu, F.T. (2009). Endogenous galectin-3 is localized in membrane lipid rafts and regulates migration of dendritic cells. J Invest Dermatol *129*, 573-583.
Hunt, D.F., Yates, J.R., 3rd, Shabanowitz, J., Winston, S., and Hauer, C.R. (1986). Protein sequencing by tandem mass spectrometry. Proc Natl Acad Sci U S A *83*, 6233-6237.
Husson, S.J., Janssen, T., Baggerman, G., Bogert, B., Kahn-Kirby, A.H., Ashrafi, K., and Schoofs, L. (2007). Impaired processing of FLP and NLP peptides in

carboxypeptidase E (EGL-21)-deficient Caenorhabditis elegans as analyzed by mass spectrometry. J Neurochem *102*, 246-260.
Husson, S.J., Landuyt, B., Nys, T., Baggerman, G., Boonen, K., Clynen, E., Lindemans, M., Janssen, T., and Schoofs, L. (2009). Comparative peptidomics of Caenorhabditis elegans versus C. briggsae by LC-MALDI-TOF MS. Peptides *30*, 449-457.
Ideo, H., Seko, A., and Yamashita, K. (2007). Recognition mechanism of galectin-4 for cholesterol 3-sulfate. J Biol Chem *282*, 21081-21089.
Ikezawa, H. (2002). Glycosylphosphatidylinositol (GPI)-anchored proteins. Biol Pharm Bull *25*, 409-417.
Ikezawa, H., Yamanegi, M., Taguchi, R., Miyashita, T., and Ohyabu, T. (1976). Studies on phosphatidylinositol phosphodiesterase (phospholipase C type) of Bacillus cereus. I. purification, properties and phosphatase-releasing activity. Biochim Biophys Acta *450*, 154-164.
Ikonomou, G., Samiotaki, M., and Panayotou, G. (2009). Proteomic methodologies and their application in colorectal cancer research. Crit Rev Clin Lab Sci *46*, 319-342.
Inoue, N., Kinoshita, T., Orii, T., and Takeda, J. (1993). Cloning of a human gene, PIG-F, a component of glycosylphosphatidylinositol anchor biosynthesis, by a novel expression cloning strategy. J Biol Chem *268*, 6882-6885.
Inoue, N., Watanabe, R., Takeda, J., and Kinoshita, T. (1996). PIG-C, one of the three human genes involved in the first step of glycosylphosphatidylinositol biosynthesis is a homologue of Saccharomyces cerevisiae GPI2. Biochem Biophys Res Commun *226*, 193-199.
Inoue, T., Hirata, K., Kuwana, Y., Fujita, M., Miwa, J., Roy, R., and Yamaguchi, Y. (2006). Cell cycle control by daf-21/Hsp90 at the first meiotic prophase/metaphase boundary during oogenesis in Caenorhabditis elegans. Dev Growth Differ *48*, 25-32.
Insel, P.A., and Patel, H.H. (2009). Membrane rafts and caveolae in cardiovascular signaling. Curr Opin Nephrol Hypertens *18*, 50-56.
Insenser, M., Nombela, C., Molero, G., and Gil, C. (2006). Proteomic analysis of detergent-resistant membranes from Candida albicans. Proteomics *6 Suppl 1*, S74-81.
Inuzuka, M., Hayakawa, M., and Ingi, T. (2005). Serinc, an activity-regulated protein family, incorporates serine into membrane lipid synthesis. J Biol Chem *280*, 35776-35783.
Ishitsuka, R., Sato, S.B., and Kobayashi, T. (2005). Imaging lipid rafts. J Biochem *137*, 249-254.
Issaq, H., and Veenstra, T. (2008). Two-dimensional polyacrylamide gel electrophoresis (2D-PAGE): advances and perspectives. Biotechniques *44*, 697-698, 700.
Ivakhno, S., and Kornelyuk, A. (2006). Quantitative proteomics and its applications for systems biology. Biochemistry (Mosc) *71*, 1060-1072.
Iwabe, N., Kuma, K., Hasegawa, M., Osawa, S., and Miyata, T. (1989). Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc Natl Acad Sci U S A *86*, 9355-9359.
Jasmer, D.P., Lahmers, K.K., and Brown, W.C. (2007). Haemonchus contortus intestine: a prominent source of mucosal antigens. Parasite Immunol *29*, 139-151.
Jasmer, D.P., Perryman, L.E., and McGuire, T.C. (1996). Haemonchus contortus GA1 antigens: related, phospholipase C-sensitive, apical gut membrane proteins encoded as a polyprotein and released from the nematode during infection. Proc Natl Acad Sci U S A *93*, 8642-8647.

Jiang, W.W., Zahurak, M., Zhou, Z.T., Park, H.L., Guo, Z.M., Wu, G.J., Sidransky, D., Trink, B., and Califano, J.A. (2007). Alterations of GPI transamidase subunits in head and neck squamous carcinoma. Mol Cancer *6*, 74.

Jones, S.J., Riddle, D.L., Pouzyrev, A.T., Velculescu, V.E., Hillier, L., Eddy, S.R., Stricklin, S.L., Baillie, D.L., Waterston, R., and Marra, M.A. (2001). Changes in gene expression associated with developmental arrest and longevity in Caenorhabditis elegans. Genome Res *11*, 1346-1352.

Kabouridis, P.S., and Jury, E.C. (2008). Lipid rafts and T-lymphocyte function: implications for autoimmunity. FEBS Lett *582*, 3711-3718.

Kaji, H., Kamiie, J., Kawakami, H., Kido, K., Yamauchi, Y., Shinkawa, T., Taoka, M., Takahashi, N., and Isobe, T. (2007). Proteomics reveals N-linked glycoprotein diversity in Caenorhabditis elegans and suggests an atypical translocation mechanism for integral membrane proteins. Mol Cell Proteomics *6*, 2100-2109.

Kaji, H., Tsuji, T., Mawuenyega, K.G., Wakamiya, A., Taoka, M., and Isobe, T. (2000). Profiling of Caenorhabditis elegans proteins using two-dimensional gel electrophoresis and matrix assisted laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis *21*, 1755-1765.

Kamath, R.S., and Ahringer, J. (2003). Genome-wide RNAi screening in Caenorhabditis elegans. Methods *30*, 313-321.

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., and Ahringer, J. (2003). Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature *421*, 231-237.

Kamiguchi, H. (2006). The region-specific activities of lipid rafts during axon growth and guidance. J Neurochem *98*, 330-335.

Kang, J.Y., Hong, Y., Ashida, H., Shishioh, N., Murakami, Y., Morita, Y.S., Maeda, Y., and Kinoshita, T. (2005). PIG-V involved in transferring the second mannose in glycosylphosphatidylinositol. The Journal of Biological Chemistry *280*, 9489-9497.

Kang, X., Szallies, A., Rawer, M., Echner, H., and Duszenko, M. (2002). GPI anchor transamidase of Trypanosoma brucei: in vitro assay of the recombinant protein and VSG anchor exchange. J Cell Sci *115*, 2529-2539.

Karagogeos, D. (2003). Neural GPI-anchored cell adhesion molecules. Front Biosci *8*, s1304-1320.

Karas, M., and Hillenkamp, F. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem *60*, 2299-2301.

Kelley, S. (2000). Getting started with Acedb. Brief Bioinform *1*, 131-137.

Kerwin, J.L., Tuininga, A.R., and Ericsson, L.H. (1994). Identification of molecular species of glycerophospholipids and sphingomyelin using electrospray mass spectrometry. J Lipid Res *35*, 1102-1114.

Kim, K.B., Kim, B.W., Choo, H.J., Kwon, Y.C., Ahn, B.Y., Choi, J.S., Lee, J.S., and Ko, Y.G. (2009). Proteome analysis of adipocyte lipid rafts reveals that gC1qR plays essential roles in adipogenesis and insulin signal transduction. Proteomics *9*, 2373-2382.

Kimble, J., Hodgkin, J., Smith, T., and Smith, J. (1982). Suppression of an amber mutation by microinjection of suppressor tRNA in C. elegans. Nature *299*, 456-458.

Kirchhof, M.G., Chau, L.A., Lemke, C.D., Vardhana, S., Darlington, P.J., Marquez, M.E., Taylor, R., Rizkalla, K., Blanca, I., Dustin, M.L., and Madrenas, J. (2008). Modulation of T cell activation by stomatin-like protein 2. J Immunol *181*, 1927-1936.

Kobayashi, T., and Hirabayashi, Y. (2000). Lipid membrane domains in cell surface and vacuolar systems. Glycoconj J *17*, 163-171.

Kobayashi, T., Nishizaki, R., and Ikezawa, H. (1997). The presence of GPI-linked protein(s) in an archaeobacterium, Sulfolobus acidocaldarius, closely related to eukaryotes. Biochim Biophys Acta *1334*, 1-4.

Kobayashi, T., Stang, E., Fang, K.S., de Moerloose, P., Parton, R.G., and Gruenberg, J. (1998). A lipid associated with the antiphospholipid syndrome regulates endosome structure and function. Nature *392*, 193-197.

Kohara, Y. (1996). [Large scale analysis of C. elegans cDNA]. Tanpakushitsu Kakusan Koso *41*, 715-720.

Kohatsu, L., Hsu, D.K., Jegalian, A.G., Liu, F.T., and Baum, L.G. (2006). Galectin-3 induces death of Candida species expressing specific beta-1,2-linked mannans. J Immunol *177*, 4718-4726.

Komatsu, S., Konishi, H., and Hashimoto, M. (2007). The proteomics of plant cell membranes. Journal of Experimental Botany *58*, 103-112.

Kontani, K., Moskowitz, I.P., and Rothman, J.H. (2005). Repression of cell-cell fusion by components of the C. elegans vacuolar ATPase complex. Dev Cell *8*, 787-794.

Kreeger, P.K., and Lauffenburger, D.A. (2010). Cancer systems biology: a network modeling perspective. Carcinogenesis *31*, 2-8.

Krijgsveld, J., Ketting, R.F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C.P., Plasterk, R.H., and Heck, A.J. (2003). Metabolic labeling of C. elegans and D. melanogaster for quantitative proteomics. Nat Biotechnol *21*, 927-931.

Kurzchalia, T.V., Dupree, P., and Monier, S. (1994). VIP21-Caveolin, a protein of the trans-Golgi network and caveolae. FEBS Lett *346*, 88-91.

Kusumi, A., and Sako, Y. (1996). Cell surface organization by the membrane skeleton. Curr Opin Cell Biol *8*, 566-574.

Kwon, J.Y., Hong, M., Choi, M.S., Kang, S., Duke, K., Kim, S., Lee, S., and Lee, J. (2004). Ethanol-response genes and their regulation analyzed by a microarray and comparative genomic approach in the nematode Caenorhabditis elegans. Genomics *83*, 600-614.

Lafourcade, C., Sobo, K., Kieffer-Jaquinod, S., Garin, J., and van der Goot, F.G. (2008). Regulation of the V-ATPase along the endocytic pathway occurs through reversible subunit association and membrane localization. PLoS One *3*, e2758.

Lagerholm, B.C., Weinreb, G.E., Jacobson, K., and Thompson, N.L. (2005). Detecting microdomains in intact cell membranes. Annu Rev Phys Chem *56*, 309-336.

Lajoie, P., Goetz, J.G., Dennis, J.W., and Nabi, I.R. (2009). Lattices, rafts, and scaffolds: domain regulation of receptor signaling at the plasma membrane. J Cell Biol *185*, 381-385.

Lajoie, P., and Nabi, I.R. (2007). Regulation of raft-dependent endocytosis. J Cell Mol Med *11*, 644-653.

Lajoie, P., Partridge, E.A., Guay, G., Goetz, J.G., Pawling, J., Lagana, A., Joshi, B., Dennis, J.W., and Nabi, I.R. (2007). Plasma membrane domain organization regulates EGFR signaling in tumor cells. J Cell Biol *179*, 341-356.

Lake, J.A., Henderson, E., Oakes, M., and Clark, M.W. (1984). Eocytes: a new ribosome structure indicates a kingdom with a close relationship to eukaryotes. Proc Natl Acad Sci U S A *81*, 3786-3790.

Lakhan, S.E., Sabharanjak, S., and De, A. (2009). Endocytosis of glycosylphosphatidylinositol-anchored proteins. J Biomed Sci *16*, 93.

Lalanne, E., Honys, D., Johnson, A., Borner, G.H., Lilley, K.S., Dupree, P., Grossniklaus, U., and Twell, D. (2004). SETH1 and SETH2, two components of the glycosylphosphatidylinositol anchor biosynthetic pathway, are required for pollen germination and tube growth in Arabidopsis. Plant Cell *16*, 229-240.
Landry, Y., Niederhoffer, N., Sick, E., and Gies, J.P. (2006). Heptahelical and other G-protein-coupled receptors (GPCRs) signaling. Curr Med Chem *13*, 51-63.
Lang, T. (2007). SNARE proteins and 'membrane rafts'. J Physiol *585*, 693-698.
Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.
Lasonder, E., Ishihama, Y., Andersen, J.S., Vermunt, A.M., Pain, A., Sauerwein, R.W., Eling, W.M., Hall, N., Waters, A.P., Stunnenberg, H.G., and Mann, M. (2002). Analysis of the Plasmodium falciparum proteome by high-accuracy mass spectrometry. Nature *419*, 537-542.
Leidich, S.D., Drapp, D.A., and Orlean, P. (1994). A conditionally lethal yeast mutant blocked at the first step in glycosyl phosphatidylinositol anchor synthesis. J Biol Chem *269*, 10193-10196.
Leidich, S.D., Kostova, Z., Latek, R.R., Costello, L.C., Drapp, D.A., Gray, W., Fassler, J.S., and Orlean, P. (1995). Temperature-sensitive yeast GPI anchoring mutants gpi2 and gpi3 are defective in the synthesis of N-acetylglucosaminyl phosphatidylinositol. Cloning of the GPI2 gene. J Biol Chem *270*, 13029-13035.
Lewis, J.A., and Berberich, S. (1992). A detergent-solubilized nicotinic acetylcholine receptor of Caenorhabditis elegans. Brain Res Bull *29*, 667-674.
Li, J., Cai, T., Wu, P., Cui, Z., Chen, X., Hou, J., Xie, Z., Xue, P., Shi, L., Liu, P., Yates, J.R., 3rd, and Yang, F. (2009). Proteomic analysis of mitochondria from Caenorhabditis elegans. Proteomics.
Li, N., Mak, A., Richards, D.P., Naber, C., Keller, B.O., Li, L., and Shaw, A.R. (2003). Monocyte lipid rafts contain proteins implicated in vesicular trafficking and phagosome formation. Proteomics *3*, 536-548.
Li, N., Shaw, A.R., Zhang, N., Mak, A., and Li, L. (2004a). Lipid raft proteomics: analysis of in-solution digest of sodium dodecyl sulfate-solubilized lipid raft proteins by liquid chromatography-matrix-assisted laser desorption/ionization tandem mass spectrometry. Proteomics *4*, 3156-3166.
Li, S., Armstrong, C.M., Bertin, N., Ge, H., Milstein, S., Boxem, M., Vidalain, P.O., Han, J.D., Chesneau, A., Hao, T., Goldberg, D.S., Li, N., Martinez, M., Rual, J.F., Lamesch, P., Xu, L., Tewari, M., Wong, S.L., Zhang, L.V., Berriz, G.F., Jacotot, L., Vaglio, P., Reboul, J., Hirozane-Kishikawa, T., Li, Q., Gabel, H.W., Elewa, A., Baumgartner, B., Rose, D.J., Yu, H., Bosak, S., Sequerra, R., Fraser, A., Mango, S.E., Saxton, W.M., Strome, S., Van Den Heuvel, S., Piano, F., Vandenhaute, J., Sardet, C., Gerstein, M., Doucette-Stamm, L., Gunsalus, K.C., Harper, J.W., Cusick, M.E., Roth, F.P., Hill, D.E., and Vidal, M. (2004b). A map of the interactome network of the metazoan C. elegans. Science *303*, 540-543.
Li, X.Y., and Skidgel, R.A. (1999). Release of glycosylphosphatidylinositol-anchored carboxypeptidase M by phosphatidylinositol-specific phospholipase C upregulates enzyme synthesis. Biochem Biophys Res Commun *258*, 204-210.
Liegeois, S., Benedetto, A., Garnier, J.M., Schwab, Y., and Labouesse, M. (2006). The V0-ATPase mediates apical secretion of exosomes containing Hedgehog-related proteins in Caenorhabditis elegans. J Cell Biol *173*, 949-961.

Lillico, S., Field, M.C., Blundell, P., Coombs, G.H., and Mottram, J.C. (2002). Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. Molecular Biology of the Cell *14*, 1182-1194.
Lillico, S., Field, M.C., Blundell, P., Coombs, G.H., and Mottram, J.C. (2003). Essential roles for GPI-anchored proteins in African trypanosomes revealed using mutants deficient in GPI8. Mol Biol Cell *14*, 1182-1194.
Link, A.J., Eng, J., Schieltz, D.M., Carmack, E., Mize, G.J., Morris, D.R., Garvik, B.M., and Yates, J.R., 3rd (1999). Direct analysis of protein complexes using mass spectrometry. Nat Biotechnol *17*, 676-682.
Linke, D. (2009). Detergents: an overview. Methods Enzymol *463*, 603-617.
Lisanti, M.P., and Rodriguez-Boulan, E. (1991). Polarized sorting of GPI-linked proteins in epithelia and membrane microdomains. Cell Biol Int Rep *15*, 1023-1049.
Littlewood, G.M., Hooper, N.M., and Turner, A.J. (1989). Ectoenzymes of the kidney microvillar membrane. Affinity purification, characterization and localization of the phospholipase C-solubilized form of renal dipeptidase. Biochem J *257*, 361-367.
Lopez, J.L. (2007). Two-dimensional electrophoresis in proteome expression analysis. J Chromatogr B Analyt Technol Biomed Life Sci *849*, 190-202.
Macdonald, J.L., and Pike, L.J. (2005). A simplified method for the preparation of detergent-free lipid rafts. J Lipid Res *46*, 1061-1067.
Macfarlane, R.D., and Torgerson, D.F. (1976). Californium-252 plasma desorption mass spectroscopy. Science *191*, 920-925.
Machtejevas, E., John, H., Wagner, K., Standker, L., Marko-Varga, G., Forssmann, W.G., Bischoff, R., and Unger, K.K. (2004). Automated multi-dimensional liquid chromatography: sample preparation and identification of peptides from human blood filtrate. J Chromatogr B Analyt Technol Biomed Life Sci *803*, 121-130.
MacLellan, D.L., Steen, H., Adam, R.M., Garlick, M., Zurakowski, D., Gygi, S.P., Freeman, M.R., and Solomon, K.R. (2005). A quantitative proteomic analysis of growth factor-induced compositional changes in lipid rafts of human smooth muscle cells. Proteomics *5*, 4733-4742.
Madi, A., Mikkat, S., Ringel, B., Ulbrich, M., Thiesen, H.J., and Glocker, M.O. (2003). Mass spectrometric proteome analysis for profiling temperature-dependent changes of protein expression in wild-type Caenorhabditis elegans. Proteomics *3*, 1526-1534.
Madore, N., Smith, K.L., Graham, C.H., Jen, A., Brady, K., Hall, S., and Morris, R. (1999). Functionally different GPI proteins are organized in different domains on the neuronal surface. EMBO J *18*, 6917-6926.
Maduro, M., and Pilgrim, D. (1995). Identification and cloning of unc-119, a gene expressed in the Caenorhabditis elegans nervous system. Genetics *141*, 977-988.
Maduro, M.F., and Rothman, J.H. (2002). Making worm guts: the gene regulatory network of the Caenorhabditis elegans endoderm. Dev Biol *246*, 68-85.
Maeda, I., Kohara, Y., Yamamoto, M., and Sugimoto, A. (2001a). Large-scale analysis of gene function in Caenorhabditis elegans by high-throughput RNAi. Curr Biol *11*, 171-176.
Maeda, Y., and Kinoshita, T. (2008). Dolichol-phosphate mannose synthase: structure, function and regulation. Biochim Biophys Acta *1780*, 861-868.
Maeda, Y., Takeda, J., Hino, J., Kangawa, K., and kinoshita, T. (2000). Human dolichol-phosphate-mannose synthase consists of three subunits, DPM1, DPM2 and DPM3. EMBO Journal *19*, 2475-2482.

Maeda, Y., Tashima, Y., Houjou, T., Fujita, M., Yoko-o, T., Jigami, Y., Taguchi, R., and Kinoshita, T. (2007). Fatty acid remodeling of GPI-anchored proteins is required for their raft association. Mol Biol Cell *18*, 1497-1506.

Maeda, Y., Tomita, S., Watanabe, R., Ohishi, K., and Kinoshita, T. (1998a). DPM2 regulates biosynthesis of dolichol phosphate-mannose in mammalian cells: correct subcellular localization and stabilization of DPM1, and binding of dolichol phosphate. EMBO J *17*, 4920-4929.

Maeda, Y., Tomita, S., Watanabe, R., Ohishi, K., and Kinoshita, T. (1998b). DPM2 regulates correct subcellular localization and stabilization of DPM1, and binding of dolichol phosphate. EMBO Journal *17*, 4920-4929.

Maeda, Y., Watanabe, R., Harris, C.L., Hong, Y., Ohishi, K., Kinoshita, K., and Kinoshita, T. (2001b). PIG-M transfers the first mannose to glycosylphosphatidylinositol on the lumenal side of the ER. EMBO J *20*, 250-261.

Maeda, Y., Watanabe, R., Harris, C.L., Hong, Y., Ohishi, K., Kinoshita, K., and Kinoshita, T. (2001c). PIG-M transfers the first mannose to glycosylphosphatidylinositol on the lumenal side of the ER. EMBO Journal *20*, 250-261.

Mairhofer, M., Steiner, M., Mosgoeller, W., Prohaska, R., and Salzer, U. (2002). Stomatin is a major lipid-raft component of platelet alpha granules. Blood *100*, 897-904.

Marengo, E., Robotti, E., Antonucci, F., Cecconi, D., Campostrini, N., and Righetti, P.G. (2005). Numerical approaches for quantitative analysis of two-dimensional maps: a review of commercial software and home-made systems. Proteomics *5*, 654-666.

Marouga, R., David, S., and Hawkins, E. (2005). The development of the DIGE system: 2D fluorescence difference gel analysis technology. Anal Bioanal Chem *382*, 669-678.

Maryon, E.B., Coronado, R., and Anderson, P. (1996). unc-68 encodes a ryanodine receptor involved in regulating C. elegans body-wall muscle contraction. J Cell Biol *134*, 885-893.

Maryon, E.B., Saari, B., and Anderson, P. (1998). Muscle-specific functions of ryanodine receptor channels in Caenorhabditis elegans. J Cell Sci *111 ( Pt 19)*, 2885-2895.

Mathy, G., and Sluse, F.E. (2008). Mitochondrial comparative proteomics: strengths and pitfalls. Biochim Biophys Acta *1777*, 1072-1077.

Matyash, V., Geier, C., Henske, A., Mukherjee, S., Hirsh, D., Thiele, C., Grant, B., Maxfield, F.R., and Kurzchalia, T.V. (2001). Distribution and transport of cholesterol in Caenorhabditis elegans. Mol Biol Cell *12*, 1725-1736.

Maurya, M.R., Benner, C., Pradervand, S., Glass, C., and Subramaniam, S. (2007). Systems biology of macrophages. Adv Exp Med Biol *598*, 62-79.

Mayor, S., and Maxfield, F.R. (1995). Insolubility and redistribution of GPI-anchored proteins at the cell surface after detergent treatment. Mol Biol Cell *6*, 929-944.

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

McConville, M.J., and Ferguson, M.A. (1993). The structure, biosynthesis and function of glycosylated phosphatidylinositols in the parasitic protozoa and higher eukaryotes. Biochem J *294 ( Pt 2)*, 305-324.

Meiri, K.F. (2005). Lipid rafts and regulation of the cytoskeleton during T cell activation. Philos Trans R Soc Lond B Biol Sci *360*, 1663-1672.
Mellgren, R.L. (2008). Detergent-resistant membrane subfractions containing proteins of plasma membrane, mitochondrial, and internal membrane origins. J Biochem Biophys Methods *70*, 1029-1036.
Mello, C.C., Kramer, J.M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. EMBO J *10*, 3959-3970.
Merris, M., Wadsworth, W.G., Khamrai, U., Bittman, R., Chitwood, D.J., and Lenard, J. (2003). Sterol effects and sites of sterol accumulation in Caenorhabditis elegans: developmental requirement for 4alpha-methyl sterols. J Lipid Res *44*, 172-181.
Metz, C.N., Brunner, G., Choi-Muira, N.H., Nguyen, H., Gabrilove, J., Caras, I.W., Altszuler, N., Rifkin, D.B., Wilson, E.L., and Davitz, M.A. (1994). Release of GPI-anchored membrane proteins by a cell-associated GPI-specific phospholipase D. EMBO J *13*, 1741-1751.
Metzner, C., Salmons, B., Gunzburg, W.H., and Dangerfield, J.A. (2008). Rafts, anchors and viruses--a role for glycosylphosphatidylinositol anchored proteins in the modification of enveloped viruses and viral vectors. Virology *382*, 125-131.
Meyer, U., Benghezal, M., Imhof, I., and Conzelmann, A. (2000). Active site determination of Gpi8p, a caspase-related enzyme required for glycosylphosphatidylinositol anchor addition to proteins. Biochemistry *39*, 3461-3471.
Michelet, X., Alberti, A., Benkemoun, L., Roudier, N., Lefebvre, C., and Legouis, R. (2009). The ESCRT-III protein CeVPS-32 is enriched in domains distinct from CeVPS-27 and CeVPS-23 at the endosomal membrane of epithelial cells. Biol Cell *101*, 599-615.
Miura, T., Takahashi, M., Horie, H., Kurushima, H., Tsuchimoto, D., Sakumi, K., and Nakabeppu, Y. (2004). Galectin-1beta, a natural monomeric form of galectin-1 lacking its six amino-terminal residues promotes axonal regeneration but not cell death. Cell Death Differ *11*, 1076-1083.
Mohsin, S.K., Weiss, H.L., Gutierrez, M.C., Chamness, G.C., Schiff, R., Digiovanna, M.P., Wang, C.X., Hilsenbeck, S.G., Osborne, C.K., Allred, D.C., Elledge, R., and Chang, J.C. (2005). Neoadjuvant trastuzumab induces apoptosis in primary breast cancers. J Clin Oncol *23*, 2460-2468.
Moran, P., and Caras, I.W. (1991). Fusion of sequence elements from non-anchored proteins to generate a fully functional signal for glycophosphatidylinositol membrane anchor attachment. J Cell Biol *115*, 1595-1600.
Moran, P., Raab, H., Kohr, W.J., and Caras, I.W. (1991). Glycophospholipid membrane anchor attachment. Molecular analysis of the cleavage/attachment site. J Biol Chem *266*, 1250-1257.
Morita, N., Nakazato, H., Okuyama, H., Kim, Y., and Thompson, G.A., Jr. (1996). Evidence for a glycosylinositolphospholipid-anchored alkaline phosphatase in the aquatic plant Spirodela oligorrhiza. Biochim Biophys Acta *1290*, 53-62.
Morita, Y.S., Paul, K.S., and Englund, P.T. (2000). Specialized fatty acid synthesis in African trypanosomes: myristate for GPI anchors. Science *288*, 140-143.
Morris, H.R., Panico, M., Barber, M., Bordoli, R.S., Sedgwick, R.D., and Tyler, A. (1981). Fast atom bombardment: a new mass spectrometric method for peptide sequence analysis. Biochem Biophys Res Commun *101*, 623-631.

Motoyama, A., Xu, T., Ruse, C.I., Wohlschlegel, J.A., and Yates, J.R., 3rd (2007). Anion and cation mixed-bed ion exchange for enhanced multidimensional separations of peptides and phosphopeptides. Anal Chem *79*, 3623-3634.

Motoyama, A., and Yates, J.R., 3rd (2008). Multidimensional LC separations in shotgun proteomics. Anal Chem *80*, 7187-7193.

Mukherjee, S., Tessema, M., and Wandinger-Ness, A. (2006). Vesicular trafficking of tyrosine kinase receptors and associated proteins in the regulation of signaling and vascular function. Circ Res *98*, 743-756.

Murakami, Y., Siripanyaphinyo, U., Hong, Y., Tashima, Y., Maeda, Y., and Kinoshita, T. (2005). The initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-Y, a seventh component. Mol Biol Cell *16*, 5236-5246.

Murakami, Y., Siripanyapinyo, U., Hong, Y., Kang, J.Y., Ishihara, S., Nakakuma, H., Maeda, Y., and Kinoshita, T. (2003). PIG-W is critical for inositol acylation but not for flipping of glycosylphosphatidylinositol-anchor. Mol Biol Cell *14*, 4285-4295.

Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., and Kenyon, C. (2003). Genes that act downstream of DAF-16 to influence the lifespan of Caenorhabditis elegans. Nature *424*, 277-283.

Nagamune, K., Nozaki, T., Maeda, Y., Ohishi, K., Fukuma, T., Hara, T., Schwarz, R.T., Sutterlin, C., Brun, R., Riezman, H., and Kinoshita, T. (2000). Critical roles of glycosylphosphatidylinositol for Trypanosoma brucei. Proc Natl Acad Sci U S A *97*, 10336-10341.

Nagamune, K., Ohishi, K., Ashida, H., Hong, Y.C., Hino, J., Kangawa, K., Inoue, N., Maeda, Y., and Kinoshita, T. (2003). GPI transamidase of Trypanosoma brucei has two previously uncharacterized (trypanosomatid transamidase 1 and 2) and three common subunits. P Natl Acad Sci USA *100*, 10682-10687.

Nägele, E., Vollmer, M., Hörth, P., and Vad, C. (2004). 2D-LC/MS techniques for the identification of proteins in highly complex mixtures. Expert Review of Proteomics *1*, 37-46

Nagpal, J.K., Dasgupta, S., Jadallah, S., Chae, Y.K., Ratovitski, E.A., Toubaji, A., Netto, G.J., Eagle, T., Nissan, A., Sidransky, D., and Trink, B. (2008). Profiling the expression pattern of GPI transamidase complex subunits in human cancer. Mod Pathol *21*, 979-991.

Naik, R.S., Krishnegowda, G., and Gowda, D.C. (2003). Glucosamine inhibits inositol acylation of the glycosylphosphatidylinositol anchors in intraerythrocytic Plasmodium falciparum. J Biol Chem *278*, 2036-2042.

Nakamura, N., Inoue, N., Watanabe, R., Takahashi, M., Takeda, J., Stevens, V.L., and Kinoshita, T. (1997a). Expression cloning of PIG-L, a candidate N-acetylglucosaminyl-phosphatidylinositol deacetylase. journal of Biological Chemistry *272*, 15834-15840.

Nakamura, N., Inoue, N., Watanabe, R., Takahashi, M., Takeda, J., Stevens, V.L., and Kinoshita, T. (1997b). Expression cloning of PIG-L, a candidate N-acetylglucosaminyl-phosphatidylinositol deacetylase. J Biol Chem *272*, 15834-15840.

Nakano, Y., Noda, K., Endo, T., Kobata, A., and Tomita, M. (1994). Structural study on the glycosyl-phosphatidylinositol anchor and the asparagine-linked sugar chain of a soluble form of CD59 in human urine. Arch Biochem Biophys *311*, 117-126.

Nalivaeva, N.N., and Turner, A.J. (2001). Post-translational modifications of proteins: acetylcholinesterase as a model system. Proteomics *1*, 735-747.

Nebl, T., Pestonjamasp, K.N., Leszyk, J.D., Crowley, J.L., Oh, S.W., and Luna, E.J. (2002). Proteomic analysis of a detergent-resistant membrane skeleton from neutrophil plasma membranes. J Biol Chem *277*, 43399-43409.
Nemoto-Sasaki, Y., Hayama, K., Ohya, H., Arata, Y., Kaneko, M.K., Saitou, N., Hirabayashi, J., and Kasai, K. (2008). Caenorhabditis elegans galectins LEC-1-LEC-11: structural features and sugar-binding properties. Biochim Biophys Acta *1780*, 1131-1142.
Netzer, A., and Gstraunthaler, G. (1993). Selective release of apical membrane enzymes from cultured renal epithelia by phosphatidylinositol-specific phospholipase C. Ren Physiol Biochem *16*, 299-310.
Newman, H.A., Romeo, M.J., Lewis, S.E., Yan, B.C., Orlean, P., and Levin, D.E. (2005). Gpi19, the Saccharomyces cerevisiae homologue of mammalian PIG-P, is a subunit of the initial enzyme for glycosylphosphatidylinositol anchor biosynthesis. Eukaryot Cell *4*, 1801-1807.
Nichols, B. (2003). Caveosomes and endocytosis of lipid rafts. J Cell Sci *116*, 4707-4714.
Nichols, B. (2005). Cell biology: without a raft. Nature *436*, 638-639.
Nielsen, H., Engelbrecht, J., Brunak, S., and von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng *10*, 1-6.
Nielsen, H., and Krogh, A. (1998). Prediction of signal peptides and signal anchors by a hidden Markov model. Proc Int Conf Intell Syst Mol Biol *6*, 122-130.
Nieto-Miguel, T., Gajate, C., Gonzalez-Camacho, F., and Mollinedo, F. (2008). Proapoptotic role of Hsp90 by its interaction with c-Jun N-terminal kinase in lipid rafts in edelfosine-mediated antileukemic therapy. Oncogene *27*, 1779-1787.
Noble, K., Zhang, J., and Wray, S. (2006). Lipid rafts, the sarcoplasmic reticulum and uterine calcium signalling: an integrated approach. J Physiol *570*, 29-35.
O'Farrell, P.H. (1975). High resolution two-dimensional electrophoresis of proteins. J Biol Chem *250*, 4007-4021.
Obermajer, N., Repnik, U., Jevnikar, Z., Turk, B., Kreft, M., and Kos, J. (2008). Cysteine protease cathepsin X modulates immune response via activation of beta2 integrins. Immunology *124*, 76-88.
Ohishi, K., Inoue, N., and Kinoshita, T. (2001). PIG-S and PIG-T, essential for GPI anchor attachment to proteins, form a complex with GAA1 and GPI8. EMBO J *20*, 4088-4098.
Ohishi, K., Inoue, N., Maeda, Y., Takeda, J., Riezman, H., and Kinoshita, T. (2000). Gaa1p and gpi8p are components of a glycosylphosphatidylinositol (GPI) transamidase that mediates attachment of GPI to proteins. Mol Biol Cell *11*, 1523-1533.
Ohishi, K., Nagamune, K., Maeda, Y., and Kinoshita, T. (2003). Two subunits of glycosylphosphatidylinositol transamidase, GPI8 and PIG-T, form a functionally important intermolecular disulfide bridge. J Biol Chem *278*, 13959-13967.
Oka, T., and Futai, M. (2000). Requirement of V-ATPase for ovulation and embryogenesis in Caenorhabditis elegans. J Biol Chem *275*, 29556-29561.
Oka, T., Yamamoto, R., and Futai, M. (1997). Three vha genes encode proteolipids of Caenorhabditis elegans vacuolar-type ATPase. Gene structures and preferential expression in an H-shaped excretory cell and rectal cells. J Biol Chem *272*, 24387-24392.

Olsen, J.V., Ong, S.E., and Mann, M. (2004). Trypsin cleaves exclusively C-terminal to arginine and lysine residues. Mol Cell Proteomics *3*, 608-614.
Opiteck, G.J., and Jorgenson, J.W. (1997). Two-dimensional SEC/RPLC coupled to mass spectrometry for the analysis of peptides. Anal Chem *69*, 2283-2291.
Oriol, R., Martinez-Duncker, I., Chantret, I., Mollicone, R., and Codogno, P. (2002). Common origin and evolution of glycosyltransferases using Dol-P-monosaccharides as donor substrate. Mol Biol Evol *19*, 1451-1463.
Orlean, P. (1990a). Dolichol Phosphate Mannose synthase is required in vivo for glycosyl phosphatidylinositol membrane anchoring, O mannosylation, and N glycosylation of protein in Saccharomyces cerevisiae. Molecular and Cellular Biology *10*, 5796-5805.
Orlean, P. (1990b). Dolichol phosphate mannose synthase is required in vivo for glycosyl phosphatidylinositol membrane anchoring, O mannosylation, and N glycosylation of protein in Saccharomyces cerevisiae. Mol Cell Biol *10*, 5796-5805.
Orlean, P., and Menon, A.K. (2007). Thematic review series: lipid posttranslational modifications. GPI anchoring of protein in yeast and mammalian cells, or: how we learned to stop worrying and love glycophospholipids. J Lipid Res *48*, 993-1011.
Pappin, D.J., Hojrup, P., and Bleasby, A.J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. Curr Biol *3*, 327-332.
Parker, C.J. (1996). Molecular basis of paroxysmal nocturnal hemoglobinuria. Stem Cells *14*, 396-411.
Parker, S., and Baylis, H.A. (2009). Overexpression of caveolins in Caenorhabditis elegans induces changes in egg-laying and fecundity. Commun Integr Biol *2*, 382-384.
Parker, S., Peterkin, H.S., and Baylis, H.A. (2007). Muscular dystrophy associated mutations in caveolin-1 induce neurotransmission and locomotion defects in Caenorhabditis elegans. Invert Neurosci *7*, 157-164.
Parker, S., Walker, D.S., Ly, S., and Baylis, H.A. (2009). Caveolin-2 is required for apical lipid trafficking and suppresses basolateral recycling defects in the intestine of Caenorhabditis elegans. Mol Biol Cell *20*, 1763-1771.
Parton, R.G. (1994). Ultrastructural localization of gangliosides; GM1 is concentrated in caveolae. J Histochem Cytochem *42*, 155-166.
Patel, H.H., Murray, F., and Insel, P.A. (2008). G-protein-coupled receptor-signaling components in membrane raft and caveolae microdomains. Handb Exp Pharmacol, 167-184.
Patra, S.K. (2008). Dissecting lipid raft facilitated cell signaling pathways in cancer. Biochim Biophys Acta *1785*, 182-206.
Paulick, M.G., and Bertozzi, C.R. (2008). The glycosylphosphatidylinositol anchor: a complex membrane-anchoring structure for proteins. Biochemistry *47*, 6991-7000.
Pays, E., and Nolan, D.P. (1998). Expression and function of surface proteins in Trypanosoma brucei. Mol Biochem Parasitol *91*, 3-36.
Paz, A., Haklai, R., Elad-Sfadia, G., Ballan, E., and Kloog, Y. (2001). Galectin-1 binds oncogenic H-Ras to mediate Ras membrane anchorage and cell transformation. Oncogene *20*, 7486-7493.
Pellerone, F.I., Archer, S.K., Behm, C.A., Grant, W.N., Lacey, M.J., and Somerville, A.C. (2003). Trehalose metabolism genes in Caenorhabditis elegans and filarial nematodes. Int J Parasitol *33*, 1195-1206.
Peng, L., Rawson, P., McLauchlan, D., Lehnert, K., Snell, R., and Jordan, T.W. (2008). Proteomic analysis of microsomes from lactating bovine mammary gland. J Proteome Res *7*, 1427-1432.

Perillo, N.L., Pace, K.E., Seilhamer, J.J., and Baum, L.G. (1995). Apoptosis of T cells mediated by galectin-1. Nature *378*, 736-739.
Perkins, D.N., Pappin, D.J., Creasy, D.M., and Cottrell, J.S. (1999). Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis *20*, 3551-3567.
Peuravuori, J., Bursakova, P., and Pihlaja, K. (2007). ESI-MS analyses of lake dissolved organic matter in light of supramolecular assembly. Anal Bioanal Chem *389*, 1559-1568.
Pierleoni, A., Martelli, P.L., and Casadio, R. (2008). PredGPI: a GPI-anchor predictor. BMC Bioinformatics *9*, 392.
Pike, L.J. (2003). Lipid rafts: bringing order to chaos. J Lipid Res *44*, 655-667.
Pike, L.J. (2004). Lipid rafts: heterogeneity on the high seas. Biochem J *378*, 281-292.
Pike, L.J. (2006). Rafts defined: a report on the Keystone Symposium on Lipid Rafts and Cell Function. J Lipid Res *47*, 1597-1598.
Pike, L.J., Han, X., Chung, K.N., and Gross, R.W. (2002). Lipid rafts are enriched in arachidonic acid and plasmenylethanolamine and their composition is independent of caveolin-1 expression: a quantitative electrospray ionization/mass spectrometric analysis. Biochemistry *41*, 2075-2088.
Pinkston-Gosse, J., and Kenyon, C. (2007). DAF-16/FOXO targets genes that regulate tumor growth in Caenorhabditis elegans. Nature Genetics *39*, 1403-1409.
Piper, R.C., and Luzio, J.P. (2001). Late endosomes: sorting and partitioning in multivesicular bodies. Traffic *2*, 612-621.
Pittet, M., and Conzelmann, A. (2007). Biosynthesis and function of GPI proteins in the yeast Saccharomyces cerevisiae. Biochim Biophys Acta *1771*, 405-420.
Poisson, G., Chauve, C., Chen, X., and Bergeron, A. (2007). FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. Genomics Proteomics Bioinformatics *5*, 121-130.
Pomorski, T., and Menon, A.K. (2006). Lipid flippases and their biological functions. Cell Mol Life Sci *63*, 2908-2921.
Pottekat, A., and Menon, A.K. (2004). Subcellular localization and targeting of N-acetylglucosaminyl phosphatidylinositol de-N-acetylase, the second enzyme in the glycosylphosphatidylinositol biosynthetic pathway. J Biol Chem *279*, 15743-15751.
Praitis, V., Casey, E., Collar, D., and Austin, J. (2001). Creation of low-copy integrated transgenic lines in Caenorhabditis elegans. Genetics *157*, 1217-1226.
Prenner, E., Honsek, G., Honig, D., Mobius, D., and Lohner, K. (2007). Imaging of the domain organization in sphingomyelin and phosphatidylcholine monolayers. Chem Phys Lipids *145*, 106-118.
Prinetti, A., Chigorno, V., Tettamanti, G., and Sonnino, S. (2000). Sphingolipid-enriched membrane domains from rat cerebellar granule cells differentiated in culture. A compositional study. J Biol Chem *275*, 11658-11665.
Priola, S.A., and McNally, K.L. (2009). The role of the prion protein membrane anchor in prion infection. Prion *3*, 134-138.
Prior, I.A., Harding, A., Yan, J., Sluimer, J., Parton, R.G., and Hancock, J.F. (2001). GTP-dependent segregation of H-ras from lipid rafts is required for biological activity. Nat Cell Biol *3*, 368-375.
Proud, C.G. (1994). Peptide-chain elongation in eukaryotes. Mol Biol Rep *19*, 161-170.
Prusiner, S.B. (1996). Molecular biology and pathogenesis of prion diseases. Trends Biochem Sci *21*, 482-487.

Puoti, A., and Conzelmann, A. (1993). Characterization of abnormal free glycophosphatidylinositols accumulating in mutant lymphoma cells of classes B, E, F, and H. J Biol Chem *268*, 7215-7224.

Quintero, I.B., Araujo, C.L., Pulkka, A.E., Wirkkala, R.S., Herrala, A.M., Eskelinen, E.L., Jokitalo, E., Hellstrom, P.A., Tuominen, H.J., Hirvikoski, P.P., and Vihko, P.T. (2007). Prostatic acid phosphatase is not a prostate specific target. Cancer Res *67*, 6549-6554.

Rabilloud, T. (2009). Membrane proteins and proteomics: love is possible, but so difficult. Electrophoresis *30 Suppl 1*, S174-180.

Reboul, J., Vaglio, P., Rual, J.F., Lamesch, P., Martinez, M., Armstrong, C.M., Li, S., Jacotot, L., Bertin, N., Janky, R., Moore, T., Hudson, J.R., Jr., Hartley, J.L., Brasch, M.A., Vandenhaute, J., Boulton, S., Endress, G.A., Jenna, S., Chevet, E., Papasotiropoulos, V., Tolias, P.P., Ptacek, J., Snyder, M., Huang, R., Chance, M.R., Lee, H., Doucette-Stamm, L., Hill, D.E., and Vidal, M. (2003). C. elegans ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. Nat Genet *34*, 35-41.

Redmond, D.L., Knox, D.P., Newlands, G., and Smith, W.D. (1997). Molecular cloning and characterisation of a developmentally regulated putative metallopeptidase present in a host protective extract of Haemonchus contortus. Mol Biochem Parasitol *85*, 77-87.

Reece-Hoyes, J.S., Shingles, J., Dupuy, D., Grove, C.A., Walhout, A.J., Vidal, M., and Hope, I.A. (2007). Insight into transcription factor gene duplication from Caenorhabditis elegans Promoterome-driven expression patterns. BMC Genomics *8*, 27.

Reggiori, F., Canivenc-Gansel, E., and Conzelmann, A. (1997). Lipid remodeling leads to the introduction and exchange of defined ceramides on GPI proteins in the ER and Golgi of Saccharomyces cerevisiae. EMBO J *16*, 3506-3518.

Rehman, A., and Jasmer, D.P. (1998). A tissue specific approach for analysis of membrane and secreted protein antigens from Haemonchus contortus gut and its application to diverse nematode species. Mol Biochem Parasitol *97*, 55-68.

Reid, P.C., Urano, Y., Kodama, T., and Hamakubo, T. (2007). Alzheimer's disease: cholesterol, membrane rafts, isoprenoids and statins. J Cell Mol Med *11*, 383-392.

Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S., and Kim, S.K. (2000). A global profile of germline gene expression in C. elegans. Mol Cell *6*, 605-616.

Resh, M.D. (2004). Membrane targeting of lipid modified signal transduction proteins. Subcell Biochem *37*, 217-232.

Resh, M.D. (2006). Palmitoylation of ligands, receptors, and intracellular signaling molecules. Sci STKE *2006*, re14.

Reverter, D., Maskos, K., Tan, F., Skidgel, R.A., and Bode, W. (2004). Crystal structure of human carboxypeptidase M, a membrane-bound enzyme that regulates peptide hormone activity. J Mol Biol *338*, 257-269.

Reyes-Del Valle, J., Chavez-Salinas, S., Medina, F., and Del Angel, R.M. (2005). Heat shock protein 90 and heat shock protein 70 are components of dengue virus receptor complex in human cells. J Virol *79*, 4557-4567.

Richard, M., De Groot, P., Courtin, O., Poulain, D., Klis, F., and Gaillardin, C. (2002). GPI7 affects cell-wall protein anchorage in Saccharomyces cerevisiae and Candida albicans. Microbiology *148*, 2125-2133.

Riemann, D., Hansen, G.H., Niels-Christiansen, L., Thorsen, E., Immerdal, L., Santos, A.N., Kehlen, A., Langner, J., and Danielsen, E.M. (2001). Caveolae/lipid rafts in fibroblast-like synoviocytes: ectopeptidase-rich membrane microdomains. Biochem J *354*, 47-55.

Righetti, P.G., Castagna, A., Antonucci, F., Piubelli, C., Cecconi, D., Campostrini, N., Antonioli, P., Astner, H., and Hamdan, M. (2004). Critical survey of quantitative proteomics in two-dimensional electrophoretic approaches. Journal of Chromatography A *1051*, 3-17.

Rioli, V., Gozzo, F.C., Heimann, A.S., Linardi, A., Krieger, J.E., Shida, C.S., Almeida, P.C., Hyslop, S., Eberlin, M.N., and Ferro, E.S. (2003). Novel natural peptide substrates for endopeptidase 24.15, neurolysin, and angiotensin-converting enzyme. J Biol Chem *278*, 8547-8555.

Roberts, W.L., Myher, J.J., Kuksis, A., Low, M.G., and Rosenberry, T.L. (1988). Lipid analysis of the glycoinositol phospholipid membrane anchor of human erythrocyte acetylcholinesterase. Palmitoylation of inositol results in resistance to phosphatidylinositol-specific phospholipase C. J Biol Chem *263*, 18766-18775.

Rogers, A., Antoshechkin, I., Bieri, T., Blasiar, D., Bastiani, C., Canaran, P., Chan, J., Chen, W.J., Davis, P., Fernandes, J., Fiedler, T.J., Han, M., Harris, T.W., Kishore, R., Lee, R., McKay, S., Muller, H.M., Nakamura, C., Ozersky, P., Petcherski, A., Schindelman, G., Schwarz, E.M., Spooner, W., Tuli, M.A., Van Auken, K., Wang, D., Wang, X., Williams, G., Yook, K., Durbin, R., Stein, L.D., Spieth, J., and Sternberg, P.W. (2008). WormBase 2007. Nucleic Acids Res *36*, D612-617.

Rogers, L.D., and Foster, L.J. (2009). Phosphoproteomics--finally fulfilling the promise? Mol Biosyst *5*, 1122-1129.

Roper, K., Corbeil, D., and Huttner, W.B. (2000). Retention of prominin in microvilli reveals distinct cholesterol-based lipid micro-domains in the apical plasma membrane. Nat Cell Biol *2*, 582-592.

Roshy, S., Sloane, B.F., and Moin, K. (2003). Pericellular cathepsin B and malignant progression. Cancer Metastasis Rev *22*, 271-286.

Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hattan, S., Khainovski, N., Pillai, S., Dey, S., Daniels, S., Purkayastha, S., Juhasz, P., Martin, S., Bartlet-Jones, M., He, F., Jacobson, A., and Pappin, D.J. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics *3*, 1154-1169.

Rothberg, K.G., Heuser, J.E., Donzell, W.C., Ying, Y.S., Glenney, J.R., and Anderson, R.G. (1992). Caveolin, a protein component of caveolae membrane coats. Cell *68*, 673-682.

Rual, J.F., Ceron, J., Koreth, J., Hao, T., Nicot, A.S., Hirozane-Kishikawa, T., Vandenhaute, J., Orkin, S.H., Hill, D.E., van den Heuvel, S., and Vidal, M. (2004). Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. Genome Res *14*, 2162-2168.

Ruf, J., Wacker, H., James, P., Maffia, M., Seiler, P., Galand, G., von Kieckebusch, A., Semenza, G., and Matei, N. (1990). Rabbit small intestinal trehalase. Purification, cDNA cloning, expression, and verification of glycosylphosphatidylinositol anchoring. J Biol Chem *265*, 15034-15039.

Sabharanjak, S., Sharma, P., Parton, R.G., and Mayor, S. (2002). GPI-anchored proteins are delivered to recycling endosomes via a distinct cdc42-regulated, clathrin-independent pinocytic pathway. Dev Cell *2*, 411-423.

Salzer, U., and Prohaska, R. (2001). Stomatin, flotillin-1, and flotillin-2 are major integral proteins of erythrocyte lipid rafts. Blood *97*, 1141-1143.
Salzer, U., Zhu, R., Luten, M., Isobe, H., Pastushenko, V., Perkmann, T., Hinterdorfer, P., and Bosman, G.J. (2008). Vesicles generated during storage of red cells are rich in the lipid raft marker stomatin. Transfusion *48*, 451-462.
Sankaram, M.B., and Thompson, T.E. (1990). Interaction of cholesterol with various glycerophospholipids and sphingomyelin. Biochemistry *29*, 10670-10675.
Santoni, V., Kieffer, S., Desclaux, D., Masson, F., and Rabilloud, T. (2000). Membrane proteomics: use of additive main effects with multiplicative interaction model to classify plasma membrane proteins according to their solubility and electrophoretic properties. Electrophoresis *21*, 3329-3344.
Saroussi, S., and Nelson, N. (2009). The little we know on the structure and machinery of V-ATPase. J Exp Biol *212*, 1604-1610.
Savage-Dunn, C. (2005). TGF-beta signaling. WormBook, 1-12.
Say, Y.H., and Hooper, N.M. (2007). Contamination of nuclear fractions with plasma membrane lipid rafts. Proteomics *7*, 1059-1064.
Schaaij-Visser, T.B., Brakenhoff, R.H., Jansen, J.W., O'Flaherty, M.C., Smeets, S.J., Heck, A.J., and Slijper, M. (2009). Comparative proteome analysis to explore p53 pathway disruption in head and neck carcinogenesis. J Proteomics *72*, 803-814.
Scheel, J., Srinivasan, J., Honnert, U., Henske, A., and Kurzchalia, T.V. (1999). Involvement of caveolin-1 in meiotic cell-cycle progression in Caenorhabditis elegans. Nat Cell Biol *1*, 127-129.
Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.
Schlegel, A., Volonte, D., Engelman, J.A., Galbiati, F., Mehta, P., Zhang, X.L., Scherer, P.E., and Lisanti, M.P. (1998). Crowded little caves: structure and function of caveolae. Cell Signal *10*, 457-463.
Schmid, S.L. (1997). Clathrin-coated vesicle formation and protein sorting: an integrated process. Annu Rev Biochem *66*, 511-548.
Schnitzer, J.E., McIntosh, D.P., Dvorak, A.M., Liu, J., and Oh, P. (1995). Separation of caveolae from associated microdomains of GPI-anchored proteins. Science *269*, 1435-1439.
Schoonderwoert, V.T., and Martens, G.J. (2002). Structural gene organization and evolutionary aspects of the V-ATPase accessory subunit Ac45. Biochim Biophys Acta *1574*, 245-254.
Schrimpf, S.P., Langen, H., Gomes, A.V., and Wahlestedt, C. (2001). A two-dimensional protein map of Caenorhabditis elegans. Electrophoresis *22*, 1224-1232.
Schuck, S., Honsho, M., Ekroos, K., Shevchenko, A., and Simons, K. (2003). Resistance of cell membranes to different detergents. Proc Natl Acad Sci U S A *100*, 5795-5800.
Schuck, S., and Simons, K. (2004). Polarized sorting in epithelial cells: raft clustering and the biogenesis of the apical membrane. J Cell Sci *117*, 5955-5964.
Sedensky, M.M., Siefker, J.M., Koh, J.Y., Miller, D.M., 3rd, and Morgan, P.G. (2004). A stomatin and a degenerin interact in lipid rafts of the nervous system of Caenorhabditis elegans. Am J Physiol Cell Physiol *287*, C468-474.
Shah, M., Patel, K., Fried, V.A., and Sehgal, P.B. (2002). Interactions of STAT3 with caveolin-1 and heat shock protein 90 in plasma membrane raft and cytosolic

complexes. Preservation of cytokine signaling during fever. J Biol Chem *277*, 45662-45669.
Sharma, D.K., Smith, T.K., Weller, C.T., Crossman, A., Brimacombe, J.S., and Ferguson, M.A.J. (1999). Differences between the trypanosomal and human GlcNAc-PI de-N-acetylases of glycosylphosphatidylinositol membrane anchor biosynthesis. Glycobiology *9*, 415-422.
Sharom, F.J., and Radeva, G. (2004). GPI-anchored protein cleavage in the regulation of transmembrane signals. Subcell Biochem *37*, 285-315.
Shaw, A.S. (2006). Lipid rafts: now you see them, now you don't. Nat Immunol *7*, 1139-1142.
Shen, Y.J., Ye, D.W., Yao, X.D., Trink, B., Zhou, X.Y., Zhang, S.L., Dai, B., Zhang, H.L., Zhu, Y., Guo, Z., Wu, G., and Nagpal, J. (2008). Overexpression of CDC91L1 (PIG-U) in bladder urothelial cell carcinoma: correlation with clinical variables and prognostic significance. BJU Int *101*, 113-119.
Sherrier, D.J., Prime, T.A., and Dupree, P. (1999). Glycosylphosphatidylinositol-anchored cell-surface proteins from Arabidopsis. Electrophoresis *20*, 2027-2035.
Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996). Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. Anal Chem *68*, 850-858.
Shichishima, T., and Noji, H. (2002). A new aspect of the molecular pathogenesis of paroxysmal nocturnal hemoglobinuria. Hematology *7*, 211-227.
Shishioh, N., Hong, Y., Ohishi, K., Ashida, H., Maeda, Y., and Kinoshita, T. (2005). GPI7 is the second partner of PIG-F and involved in modification of glycosylphosphatidylinositol. J Biol Chem *280*, 9728-9734.
Shogomori, H., and Brown, D.A. (2003). Use of detergents to study membrane rafts: the good, the bad, and the ugly. Biol Chem *384*, 1259-1263.
Siafakas, A.R., Wright, L.C., Sorrell, T.C., and Djordjevic, J.T. (2006). Lipid rafts in Cryptococcus neoformans concentrate the virulence determinants phospholipase B1 and Cu/Zn superoxide dismutase. Eukaryot Cell *5*, 488-498.
Sickmann, A., Mreyen, M., and Meyer, H.E. (2002). Identification of modified proteins by mass spectrometry. IUBMB Life *54*, 51-57.
Simmer, F., Moorman, C., van der Linden, A.M., Kuijk, E., van den Berghe, P.V., Kamath, R.S., Fraser, A.G., Ahringer, J., and Plasterk, R.H. (2003). Genome-wide RNAi of C. elegans using the hypersensitive rrf-3 strain reveals novel gene functions. PLoS Biol *1*, E12.
Simons, K., and Gruenberg, J. (2000). Jamming the endosomal system: lipid rafts and lysosomal storage diseases. Trends Cell Biol *10*, 459-462.
Simons, K., and Ikonen, E. (1997). Functional rafts in cell membranes. Nature *387*, 569-572.
Singer, S.J., and Nicolson, G.L. (1972). The fluid mosaic model of the structure of cell membranes. Science *175*, 720-731.
Sipos, G., Reggiori, F., Vionnet, C., and Conzelmann, A. (1997). Alternative lipid remodelling pathways for glycosylphosphatidylinositol membrane anchors in Saccharomyces cerevisiae. EMBO J *16*, 3494-3505.
Skidgel, R.A., and Erdos, E.G. (1998). Cellular carboxypeptidases. Immunol Rev *161*, 129-141.
Skidgel, R.A., McGwire, G.B., and Li, X.Y. (1996). Membrane anchoring and release of carboxypeptidase M: implications for extracellular hydrolysis of peptide hormones. Immunopharmacology *32*, 48-52.

Slimane, T.A., Trugnan, G., Van, I.S.C., and Hoekstra, D. (2003). Raft-mediated trafficking of apical resident proteins occurs in both direct and transcytotic pathways in polarized hepatic cells: role of distinct lipid microdomains. Mol Biol Cell *14*, 611-624.

Sly, W.S., and Hu, P.Y. (1995). Human carbonic anhydrases and carbonic anhydrase deficiencies. Annu Rev Biochem *64*, 375-401.

Smaby, J.M., Momsen, M., Kulkarni, V.S., and Brown, R.E. (1996). Cholesterol-induced interfacial area condensations of galactosylceramides and sphingomyelins with identical acyl chains. Biochemistry *35*, 5696-5704.

Smart, E.J., Ying, Y.S., Mineo, C., and Anderson, R.G. (1995). A detergent-free method for purifying caveolae membrane from tissue culture cells. Proc Natl Acad Sci U S A *92*, 10104-10108.

Smith, T.K., Crossman, A., Borissow, C.N., Paterson, M.J., Dix, A., Brimacombe, J.S., and Ferguson, M.A. (2001). Specificity of GlcNAc-PI de-N-acetylase of GPI biosynthesis and synthesis of parasite-specific suicide substrate inhibitors. EMBO J *20*, 3322-3332.

Smithies, O., and Poulik, M.D. (1956). Two-dimensional electrophoresis of serum proteins. Nature *177*, 1033.

Sobering, A.K., Watanabe, R., Romeo, M.J., Yan, B.C., Specht, C.A., Orlean, P., Riezman, H., and Levin, D.E. (2004). Yeast Ras regulates the complex that catalyzes the first step in GPI-anchor biosynthesis at the ER. Cell *117*, 637-648.

Solomon, S., Masilamani, M., Rajendran, L., Bastmeyer, M., Stuermer, C.A., and Illges, H. (2002). The lipid raft microdomain-associated protein reggie-1/flotillin-2 is expressed in human B cells and localized at the plasma membrane and centrosome in PBMCs. Immunobiology *205*, 108-119.

Sonnichsen, B., Koski, L.B., Walsh, A., Marschall, P., Neumann, B., Brehm, M., Alleaume, A.M., Artelt, J., Bettencourt, P., Cassin, E., Hewitson, M., Holz, C., Khan, M., Lazik, S., Martin, C., Nitzsche, B., Ruer, M., Stamford, J., Winzi, M., Heinkel, R., Roder, M., Finell, J., Hantsch, H., Jones, S.J., Jones, M., Piano, F., Gunsalus, K.C., Oegema, K., Gonczy, P., Coulson, A., Hyman, A.A., and Echeverri, C.J. (2005). Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. Nature *434*, 462-469.

Sprenger, R.R., Fontijn, R.D., van Marle, J., Pannekoek, H., and Horrevoets, A.J. (2006). Spatial segregation of transport and signalling functions between human endothelial caveolae and lipid raft proteomes. Biochem J *400*, 401-410.

Spurway, T.D., Dalley, J.A., High, S., and Bulleid, N.J. (2001). Early events in glycosylphosphatidylinositol anchor addition. substrate proteins associate with the transamidase subunit gpi8p. J Biol Chem *276*, 15975-15982.

Stahl, N., Baldwin, M.A., Hecker, R., Pan, K.M., Burlingame, A.L., and Prusiner, S.B. (1992). Glycosylinositol phospholipid anchors of the scrapie and cellular prion proteins contain sialic acid. Biochemistry *31*, 5043-5053.

Stan, R.V., Roberts, W.G., Predescu, D., Ihida, K., Saucan, L., Ghitescu, L., and Palade, G.E. (1997). Immunoisolation and partial characterization of endothelial plasmalemmal vesicles (caveolae). Mol Biol Cell *8*, 595-605.

Stein, L., Sternberg, P., Durbin, R., Thierry-Mieg, J., and Spieth, J. (2001). WormBase: network access to the genome and biology of Caenorhabditis elegans. Nucleic Acids Res *29*, 82-86.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Mardis, E.R., Marra, M.A., Miner, T.L., Minx, P., Mullikin, J.C., Plumb, R.W., Rogers, J., Schein, J.E., Sohrmann, M., Spieth, J., Stajich, J.E., Wei, C., Willey, D., Wilson, R.K., Durbin, R., and Waterston, R.H. (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol *1*, E45.

Stewart, G.W. (1997). Stomatin. Int J Biochem Cell Biol *29*, 271-274.

Stewart, G.W., Argent, A.C., and Dash, B.C. (1993). Stomatin: a putative cation transport regulator in the red cell membrane. Biochim Biophys Acta *1225*, 15-25.

Stinchcomb, D.T., Shaw, J.E., Carr, S.H., and Hirsh, D. (1985). Extrachromosomal DNA transformation of Caenorhabditis elegans. Mol Cell Biol *5*, 3484-3496.

Stuart, E.S., Webley, W.C., and Norkin, L.C. (2003). Lipid rafts, caveolae, caveolin-1, and entry by Chlamydiae into host cells. Exp Cell Res *287*, 67-78.

Styer, K.L., Singh, V., Macosko, E., Steele, S.E., Bargmann, C.I., and Aballay, A. (2008). Innate immunity in Caenorhabditis elegans is regulated by neurons expressing NPR-1/GPCR. Science *322*, 460-464.

Subramanian, S., Woolford, C.A., Drill, E., Lu, M., and Jones, E.W. (2006). Pbn1p: an essential endoplasmic reticulum membrane protein required for protein processing in the endoplasmic reticulum of budding yeast. Proc Natl Acad Sci U S A *103*, 939-944.

Sugiyama, E., DeGasperi, R., Urakaze, M., Chang, H.M., Thomas, L.J., Hyman, R., Warren, C.D., and Yeh, E.T. (1991). Identification of defects in glycosylphosphatidylinositol anchor biosynthesis in the Thy-1 expression mutants. J Biol Chem *266*, 12119-12122.

Sutterlin, C., Escribano, M.V., Gerold, P., Maeda, Y., Mazon, M.J., Kinoshita, T., Schwarz, R.T., and Riezman, H. (1998). Saccharomyces cerevisiae GPI10, the functional homologue of human PIG-B, is required for glycosylphosphatidylinositol-anchor synthesis. Biochemical Journal *332*, 153-159.

Szabo, I., Adams, C., and Gulbins, E. (2004). Ion channels and membrane rafts in apoptosis. Pflugers Arch *448*, 304-312.

Tabuse, Y., Nabetani, T., and Tsugita, A. (2005). Proteomic analysis of protein expression profiles during Caenorhabditis elegans development using two-dimensional difference gel electrophoresis. Proteomics *5*, 2876-2891.

Taguchi, R., Asahi, Y., and Ikezawa, H. (1980). Purification and properties of phosphatidylinositol-specific phospholipase C of Bacillus thuringiensis. Biochim Biophys Acta *619*, 48-57.

Takahashi, M., Inoue, N., Ohishi, K., Maeda, Y., Nakamura, N., Endo, Y., Fujita, T., Takeda, J., and Kinoshita, T. (1996). PIG-B, a membrane protein of the endoplasmic reticulum with a large lumenal domain, is involved in transferring the third mannose of the GPI anchor. EMBO Journal *15*, 4254-4261.

Takesue, Y., Yokota, K., Nishi, Y., Taguchi, R., and Ikezawa, H. (1986). Solubilization of trehalase from rabbit renal and intestinal brush-border membranes by a phosphatidylinositol-specific phospholipase C. FEBS Lett *201*, 5-8.

Tanaka, K., Waki, H., Ido, Y., Akita, S., Yoshida, Y., and Yoshida, T. (1988). Protein and Polymer Analyses up to m/z 100 000 by Laser Ionization Time-of flight Mass Spectrometry. Rapid Commun Mass Sp *2*, 151-153.

Tanaka, S., Maeda, Y., Tashima, Y., and Kinoshita, T. (2004). Inositol deacylation of glycosylphosphatidylinositol-anchored proteins is mediated by mammalian PGAP1 and yeast Bst1p. J Biol Chem *279*, 14256-14263.

Tang, C.H., and Wei, Y. (2008). The urokinase receptor and integrins in cancer progression. Cell Mol Life Sci *65*, 1916-1932.
Tarleton, R.L. (2007). Immune system recognition of Trypanosoma cruzi. Curr Opin Immunol *19*, 430-434.
Taron, B.W., Colussi, P.A., Wiedman, J.M., Orlean, P., and Taron, C.H. (2004a). Human Smp3p adds a fourth mannose to yeast and human glycosylphosphatidylinositol precursors in vivo. The Journal of Biological Chemistry *279*, 36083-36092.
Taron, B.W., Colussi, P.A., Wiedman, J.M., Orlean, P., and Taron, C.H. (2004b). Human Smp3p adds a fourth mannose to yeast and human glycosylphosphatidylinositol precursors in vivo. J Biol Chem *279*, 36083-36092.
Taron, C.H., Wiedman, J.M., Grimme, S.J., and Orlean, P. (2000). Glycosylphosphatidylinositol biosynthesis defects in Gpi11p- and Gpi13p-deficient yeast suggest a branched pathway and implicate gpi13p in phosphoethanolamine transfer to the third mannose. Mol Biol Cell *11*, 1611-1630.
Tashima, Y., Taguchi, R., Murata, C., Ashida, H., Kinoshita, T., and Maeda, Y. (2006). PGAP2 is essential for correct processing and stable expression of GPI-anchored proteins. Mol Biol Cell *17*, 1410-1420.
Taylor, C.M., Coetzee, T., and Pfeiffer, S.E. (2002). Detergent-insoluble glycosphingolipid/cholesterol microdomains of the myelin membrane. J Neurochem *81*, 993-1004.
Taylor, D.R., and Hooper, N.M. (2006). The prion protein and lipid rafts. Mol Membr Biol *23*, 89-99.
Taylor, R.C., and Coorssen, J.R. (2006). Proteome Resolution by Two-Dimensional Gel Electrophoresis Varies with the Commercial Source of IPG Strips. Journal of Proteome Research *5*, 2919-2927.
Tiede, A., Daniels, R.J., Higgs, D.R., Mehrein, Y., Schmidt, R.E., and Schubert, J. (2001). The human GPI1 gene is required for efficient glycosylphosphatidylinositol biosynthesis. Gene *271*, 247-254.
Tiede, A., Nischan, C., Schubert, J., and Schmidt, R.E. (2000). Characterisation of the enzymatic complex for the first step in glycosylphosphatidylinositol biosynthesis. Int J Biochem Cell Biol *32*, 339-350.
Timmons, L., and Fire, A. (1998). Specific interference by ingested dsRNA. Nature *395*, 854.
Tomita, S., Inoue, N., Maeda, Y., Ohishi, K., Takeda, J., and Kinoshita, T. (1998). A homologue of Saccharomyces cerevisiae Dpm1p is not sufficient for synthesis of dolichol-phosphate-mannose in mammalian cells. J Biol Chem *273*, 9249-9254.
Trelle, M.B., Salcedo-Amaya, A.M., Cohen, A.M., Stunnenberg, H.G., and Jensen, O.N. (2009). Global Histone Analysis by Mass Spectrometry Reveals a High Content of Acetylated Lysine Residues in the Malaria Parasite Plasmodium falciparum. J Proteome Res *8*, 3439-3450.
Treumann, A., Lifely, M.R., Schneider, P., and Ferguson, M.A. (1995). Primary structure of CD52. J Biol Chem *270*, 6088-6099.
Turner, A.J., Isaac, R.E., and Coates, D. (2001). The neprilysin (NEP) family of zinc metalloendopeptidases: genomics and function. Bioessays *23*, 261-269.
Udenfriend, S., and Kodukula, K. (1995a). How glycosylphosphatidylinositol-anchored membrane proteins are made. Annu Rev Biochem *64*, 563-591.
Udenfriend, S., and Kodukula, K. (1995b). Prediction of omega site in nascent precursor of glycosylphosphatidylinositol protein. Methods Enzymol *250*, 571-582.

Umemura, M., Okamoto, M., Nakayama, K., Sagane, K., Tsukahara, K., Hata, K., and Jigami, Y. (2003). GWT1 gene is required for inositol acylation of glycosylphosphatidylinositol anchors in yeast. J Biol Chem *278*, 23639-23647.
Umlauf, E., Mairhofer, M., and Prohaska, R. (2006). Characterization of the stomatin domain involved in homo-oligomerization and lipid raft association. J Biol Chem *281*, 23349-23356.
Ungewickell, E.J., and Hinrichsen, L. (2007). Endocytosis: clathrin-mediated membrane budding. Curr Opin Cell Biol *19*, 417-425.
Unlu, M., Morgan, M.E., and Minden, J.S. (1997). Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. Electrophoresis *18*, 2071-2077.
Urbaniak, M.D., Crossman, A., Chang, T., Smith, T.K., van Aalten, D.M., and Ferguson, M.A. (2005). The N-acetyl-D-glucosaminylphosphatidylinositol De-N-acetylase of glycosylphosphatidylinositol biosynthesis is a zinc metalloenzyme. J Biol Chem *280*, 22831-22838.
Urquhart, P., Pang, S., and Hooper, N.M. (2005). N-glycans as apical targeting signals in polarized epithelial cells. Biochem Soc Symp, 39-45.
Vainauskas, S., Maeda, Y., Kurniawan, H., Kinoshita, T., and Menon, A.K. (2002). Structural requirements for the recruitment of Gaa1 into a functional glycosylphosphatidylinositol transamidase complex. J Biol Chem *277*, 30535-30542.
Vainauskas, S., and Menon, A.K. (2004a). A conserved proline in the last transmembrane segment of Gaa1 is required for glycosylphosphatidylinositol (GPI) recognition by GPI transamidase. J Biol Chem *279*, 6540-6545.
Vainauskas, S., and Menon, A.K. (2004b). A conserved proline in the last transmembrane segment of Gaa1 is required for glycosylphosphatidylinositol (GPI) recognition by GPI transamidase. The Journal of Biological Chemistry *279*, 6540-6545.
Vainauskas, S., and Menon, A.K. (2006). Ethanolamine phosphate linked to the first mannose residue of glycosylphosphatidylinositol (GPI) lipids is a major feature of the GPI structure that is recognized by human GPI transamidase. J Biol Chem *281*, 38358-38364.
van den Broek, I., Sparidans, R.W., Schellens, J.H., and Beijnen, J.H. (2008). Quantitative bioanalysis of peptides by liquid chromatography coupled to (tandem) mass spectrometry. J Chromatogr B Analyt Technol Biomed Life Sci *872*, 1-22.
Varma, R., and Mayor, S. (1998). GPI-anchored proteins are organized in submicron domains at the cell surface. Nature *394*, 798-801.
Vashist, S., Kim, W., Belden, W.J., Spear, E.D., Barlowe, C., and Ng, D.T. (2001). Distinct retrieval and retention mechanisms are required for the quality control of endoplasmic reticulum protein folding. J Cell Biol *155*, 355-368.
Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995). Serial analysis of gene expression. Science *270*, 484-487.
Vidugiriene, J., and Menon, A.K. (1993). Early lipid intermediates in glycosyl-phosphatidylinositol anchor assembly are synthesized in the ER and located in the cytoplasmic leaflet of the ER membrane bilayer. J Cell Biol *121*, 987-996.
Vishwakarma, R.A., and Menon, A.K. (2005). Flip-flop of glycosylphosphatidylinositols (GPI's) across the ER. Chem Commun (Camb), 453-455.

von Haller, P.D., Donohoe, S., Goodlett, D.R., Aebersold, R., and Watts, J.D. (2001). Mass spectrometric characterization of proteins extracted from Jurkat T cell detergent-resistant membrane domains. Proteomics *1*, 1010-1021.
von Heijne, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Res *14*, 4683-4690.
von Heijne, G. (1990). The signal peptide. J Membr Biol *115*, 195-201.
Voshol, H., Ehrat, M., Traenkle, J., Bertrand, E., and van Oostrum, J. (2009). Antibody-based proteomics: analysis of signaling networks using reverse protein arrays. FEBS J *276*, 6871-6879.
Vossen, J.H., Muller, W.H., Lipke, P.N., and Klis, F.M. (1997). Restrictive glycosylphosphatidylinositol anchor synthesis in cwh6/gpi3 yeast cells causes aberrant biogenesis of cell wall proteins. J Bacteriol *179*, 2202-2209.
Vowels, J.J., and Thomas, J.H. (1994). Multiple Chemosensory Defects in Daf-11 and Daf-21 Mutants of Caenorhabditis-Elegans. Genetics *138*, 303-316.
Wachtler, V., and Balasubramanian, M.K. (2006). Yeast lipid rafts?--an emerging view. Trends Cell Biol *16*, 1-4.
Waheed, A., Zhu, X.L., Sly, W.S., Wetzel, P., and Gros, G. (1992). Rat skeletal muscle membrane associated carbonic anhydrase is 39-kDa, glycosylated, GPI-anchored CA IV. Arch Biochem Biophys *294*, 550-556.
Waheed, A.A., and Jones, T.L. (2002). Hsp90 interactions and acylation target the G protein Galpha 12 but not Galpha 13 to lipid rafts. J Biol Chem *277*, 32409-32412.
Walhout, A.J., Temple, G.F., Brasch, M.A., Hartley, J.L., Lorson, M.A., van den Heuvel, S., and Vidal, M. (2000). GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. Methods Enzymol *328*, 575-592.
Wallin, E., and von Heijne, G. (1998). Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. Protein Sci *7*, 1029-1038.
Walter, P., and Johnson, A.E. (1994). Signal sequence recognition and protein targeting to the endoplasmic reticulum membrane. Annu Rev Cell Biol *10*, 87-119.
Wang, J., and Kim, S.K. (2003). Global analysis of dauer gene expression in Caenorhabditis elegans. Development *130*, 1621-1634.
Wang, X.Q., Sun, P., and Paller, A.S. (2005). Gangliosides inhibit urokinase-type plasminogen activator (uPA)-dependent squamous carcinoma cell migration by preventing uPA receptor/alphabeta integrin/epidermal growth factor receptor interactions. J Invest Dermatol *124*, 839-848.
Washburn, M.P., Wolters, D., and Yates III, J.R. (2001). Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nature Biotechnology *19*, 242-247.
Watanabe, K., Bianco, C., Strizzi, L., Hamada, S., Mancino, M., Bailly, V., Mo, W., Wen, D., Miatkowski, K., Gonzales, M., Sanicola, M., Seno, M., and Salomon, D.S. (2007). Growth factor induction of Cripto-1 shedding by glycosylphosphatidylinositol-phospholipase D and enhancement of endothelial cell migration. J Biol Chem *282*, 31643-31655.
Watanabe, R., Kinoshita, T., Masaki, R., Yamamoto, A., Takeda, J., and Inoue, N. (1996). PIG-A and PIG-H, which participate in glycosylphosphatidylinositol anchor biosynthesis, form a protein complex in the endoplasmic reticulum. J Biol Chem *271*, 26868-26875.

Watanabe, R., Murakami, Y., Marmor, M.D., Inoue, N., Maeda, Y., Hino, J., Kangawa, K., Julius, M., and Kinoshita, T. (2000a). Initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-P and is regulated by DPM2. EMBO Journal *19*, 4402-4411.
Watanabe, R., Murakami, Y., Marmor, M.D., Inoue, N., Maeda, Y., Hino, J., Kangawa, K., Julius, M., and Kinoshita, T. (2000b). Initial enzyme for glycosylphosphatidylinositol biosynthesis requires PIG-P and is regulated by DPM2. EMBO J *19*, 4402-4411.
Watanabe, R., Ohishi, K., Maeda, Y., Nakamura, N., and Kinoshita, T. (1999). Mammalian PIG-L and its yeast homologue Gpi12p are N-acetylglucosaminylphosphatidylinositol de-N-acetylases essential in glycosylphosphatidylinositol biosynthesis. Biochem J *339 ( Pt 1)*, 185-192.
Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., and et al. (1992). A survey of expressed genes in Caenorhabditis elegans. Nat Genet *1*, 114-123.
Weerth, S.H., Holtzclaw, L.A., and Russell, J.T. (2007). Signaling proteins in raft-like microdomains are essential for Ca2+ wave propagation in glial cells. Cell Calcium *41*, 155-167.
Whelan, S.A., Lu, M., He, J., Yan, W., Saxton, R., Faull, K.F., Whitelegge, J.P., and Chang, H. (2009). Mass spectrometry (LC-MS/MS) site-mapping of N-glycosylated membrane proteins for breast cancer biomarkers. J Proteome Res.
Whitfield, J.F. (2006). Can statins put the brakes on Alzheimer's disease? Expert Opin Investig Drugs *15*, 1479-1485.
Wiedman, J.M., Fabre, A.L., Taron, B.W., Taron, C.H., and Orlean, P. (2007). In vivo characterization of the GPI assembly defect in yeast mcd4-174 mutants and bypass of the Mcd4p-dependent step in mcd4Delta cells. FEMS Yeast Res *7*, 78-83.
Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., and et al. (1994). 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans. Nature *368*, 32-38.
Woese, C.R., Kandler, O., and Wheelis, M.L. (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A *87*, 4576-4579.
Wollscheid, B., von Haller, P.D., Yi, E., Donohoe, S., Vaughn, K., Keller, A., Nesvizhskii, A.I., Eng, J., Li, X.J., Goodlett, D.R., Aebersold, R., and Watts, J.D. (2004). Lipid raft proteins and their identification in T lymphocytes. Subcell Biochem *37*, 121-152.
Wolters, D.A., Washburn, M.P., and Yates, J.R., 3rd (2001). An automated multidimensional protein identification technology for shotgun proteomics. Anal Chem *73*, 5683-5690.
Wong, S.C., Chan, C.M., Ma, B.B., Lam, M.Y., Choi, G.C., Au, T.C., Chan, A.S., and Chan, A.T. (2009). Advanced proteomic technologies for cancer biomarker discovery. Expert Rev Proteomics *6*, 123-134.
Wu, C.C., and Yates, J.R., 3rd (2003). The application of mass spectrometry to membrane proteomics. Nat Biotechnol *21*, 262-267.
Wu, G., Guo, Z., Chatterjee, A., Huang, X., Rubin, E., Wu, F., Mambo, E., Chang, X., Osada, M., Sook Kim, M., Moon, C., Califano, J.A., Ratovitski, E.A., Gollin, S.M., Sukumar, S., Sidransky, D., and Trink, B. (2006). Overexpression of glycosylphosphatidylinositol (GPI) transamidase subunits phosphatidylinositol glycan

class T and/or GPI anchor attachment 1 induces tumorigenesis and contributes to invasion in human breast cancer. Cancer Res *66*, 9829-9836.

Yamada, E. (1955). The fine structure of the renal glomerulus of the mouse. J Biophys Biochem Cytol *1*, 551-566.

Yan, W., Shen, F., Dillon, B., and Ratnam, M. (1998). The hydrophobic domains in the carboxyl-terminal signal for GPI modification and in the amino-terminal leader peptide have similar structural requirements. J Mol Biol *275*, 25-33.

Yates, J.R., 3rd, Eng, J.K., McCormack, A.L., and Schieltz, D. (1995). Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem *67*, 1426-1436.

Yates, J.R., Ruse, C.I., and Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng *11*, 49-79.

Yatsuda, A.P., Krijgsveld, J., Cornelissen, A.W., Heck, A.J., and de Vries, E. (2003). Comprehensive analysis of the secreted proteins of the parasite Haemonchus contortus reveals extensive sequence variation and differential immune recognition. J Biol Chem *278*, 16941-16951.

Yoon, H.J., Park, S.W., Lee, H.B., Im, S.Y., Hooper, N.M., and Park, H.S. (2007). Release of renal dipeptidase from glycosylphosphatidylinositol anchor by insulin-triggered phospholipase C/intracellular Ca2+. Arch Pharm Res *30*, 608-615.

Yoshinaka, K., Kumanogoh, H., Nakamura, S., and Maekawa, S. (2004). Identification of V-ATPase as a major component in the raft fraction prepared from the synaptic plasma membrane and the synaptic vesicle of rat brain. Neurosci Lett *363*, 168-172.

Zhang, X., Tan, F., Zhang, Y., and Skidgel, R.A. (2008). Carboxypeptidase M and kinin B1 receptors interact to facilitate efficient b1 signaling from B2 agonists. J Biol Chem *283*, 7994-8004.

Zhao, Y., Lee, W.N., and Xiao, G.G. (2009). Quantitative proteomics and biomarker discovery in human cancer. Expert Rev Proteomics *6*, 115-118.

Zheng, Y.Z., Berg, K.B., and Foster, L.J. (2009). Mitochondria do not contain lipid rafts, and lipid rafts do not contain mitochondrial proteins. J Lipid Res *50*, 988-998.

Zhu, X., and Williamson, P.R. (2004). Role of laccase in the biology and virulence of Cryptococcus neoformans. FEMS Yeast Res *5*, 1-10.

Zhu, Y., Fraering, P., Vionnet, C., and Conzelmann, A. (2005). Gpi17p does not stably interact with other subunits of glycosylphosphatidylinositol transamidase in Saccharomyces cerevisiae. Biochim Biophys Acta *1735*, 79-88.

Zhu, Y., Vionnet, C., and Conzelmann, A. (2006). Ethanolaminephosphate side chain added to glycosylphosphatidylinositol (GPI) anchor by mcd4p is required for ceramide remodeling and forward transport of GPI proteins from endoplasmic reticulum to Golgi. J Biol Chem *281*, 19830-19839.

# Appendix

## Appendix 1.

All *C. elegans* proteins predicted by prediction programs containing 778 predicted proteins, 327 of which have predictions with two or more programs. Columns 1 to 3 contain the Wormbase gene ID, gene name and a brief description of the protein. Columns 4 to 6 contain the GO terms for the proteins where available. Columns 7 to 10 contain the programs with which the protein was predicted, with ● indicating a positive prediction and ○ a negative. Column 11 denotes the number of prediction programs that gave the protein a positive result.

| Wormbase gene ID | gene name | brief description | Molecular function | Biological process | Cellular component | Big PI | GPI SOM | Frag Anchor | Pred GPI | No. of hits |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009700** | F44F4.1 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00017969** | F32A5.3 | Serine carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00016627** | C44B7.5 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018043** | F35D11.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00020248** | T05C1.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00015803** | C15H9.9 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00004370** | rig-3 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00008509** | F01G10.6 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00017978** | F32B5.4 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ● | ● | ● | ● | 4 |
| **WBGene00001581** | gfi-1 | n/a | n/a | regulation | n/a | ● | ● | ● | ● | 4 |
| **WBGene00019017** | F57F4.4 | n/a | n/a | regulation | n/a | ● | ● | ● | ● | 4 |
| **WBGene00019660** | K11H12.4 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00019663** | K11H12.7 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00008870** | F15H9.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00013969** | ZK337.1 | Alpha-2-macroglobulin family (3 domains) | Catalytic | metabolism | extracellular | ● | ● | ● | ● | 4 |
| **WBGene00014194** | ZK1037.6 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00017416** | F13B6.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00017494** | F15E11.5 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00020995** | W03F8.6 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00012439** | Y12A6A.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018507** | F46F5.16 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00007299** | C04F12.5 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018787** | cutl-20 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00021452** | Y39F10A.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00008868** | F15G9.5 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009679** | F44D12.2 | n/a | binding | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00021880** | Y54G2A.1 5 | n/a | n/a | development | n/a | ● | ● | ● | ● | 4 |
| **WBGene00003956** | pcp-1 | n/a | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00022246** | acp-7 | n/a | Catalytic | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00000038** | ace-4 | Acetylcholine-esterase | Catalytic | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00006869** | vab-2 | n/a | binding | Signalling | anchored | ● | ● | ● | ● | 4 |
| **WBGene00007911** | C34B7.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00017296** | F09E10.6 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00020195** | T03G6.3 | plasma cell membrane protein and phosphor-diesterase I (weak) | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00000037** | ace-3 | n/a | Catalytic | Signalling | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018576** | F47G3.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00001988** | hot-3 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00001989** | hot-4 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00016979** | C56G2.4 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00006987** | zmp-1 | matrix metalloproteinase | Catalytic | metabolism | cell surface | ● | ● | ● | ● | 4 |
| **WBGene00019320** | K02E10.6 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00007652** | C17G1.5 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00011879** | pho-7 | histidine acid phosphatase | Catalytic | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00013959** | ZK265.7 | n/a | n/a | development | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018115** | F36H9.7 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00022645** | ZK6.11 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00011498** | T05G5.1 | Caldesmon-like repeats | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00010150** | F56D5.6 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00018984** | F56F10.1 | peptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00017594** | F19C7.4 | lysosomal carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00007722** | C25D7.15 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00006621** | try-3 | peptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009969** | F53B7.7 | n/a | n/a | development | n/a | ● | ● | ● | ● | 4 |
| **WBGene00011314** | T01B7.9 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00015472** | C05D9.3 | n/a | binding | cell adhesion | membrane | ● | ● | ● | ● | 4 |
| **WBGene00003056** | lon-2 | n/a | binding | development | cell surface | ● | ● | ● | ● | 4 |
| **WBGene00001163** | efn-2 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00008776** | F13H10.5 | n/a | Catalytic | metabolism | membrane | ● | ● | ● | ● | 4 |
| **WBGene00009428** | F35E12.4 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00012073** | T27A8.1 | carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00021526** | Y41G9A.2 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00002181** | kal-1 | WAP-type (Whey Acidic Protein) 'four-disulfide core', Fibronectin type III domain (3 domains) | inhibitor | n/a | cell surface | ● | ● | ● | ● | 4 |
| **WBGene00021791** | Y51H7C.13 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00016707** | C46E10.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00010113** | F55D12.5 | Activin types I and II receptor domain | binding | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00000283** | cah-5 | carbonic anhydrase | Catalytic | metabolism | membrane | ● | ● | ● | ● | 4 |
| **WBGene00019617** | K10C2.1 | serine carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00012202** | W02B12.4 | esterase | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009434** | F35E12.10 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00044446** | C06G4.6 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00044484** | C09B9.8 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00044556** | F38G1.3 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00022827** | ZK816.4 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00010059** | F54E4.3 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00007864** | C32H11.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00010747** | K10D11.3 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00003958** | pcp-3 | lysosomal carboxypeptidase | Catalytic | metabolism | membrane | ● | ● | ● | ● | 4 |
| **WBGene00045248** | ZK180.7 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00045400** | C54D10.13 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009779** | F46C5.2 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00009798** | F46G10.4 | lipase | Catalytic | metabolism | membrane | ● | ● | ● | ○ | 3 |
| **WBGene00016354** | rig-6 | fibronectin, IG-like domains of NCAM | binding | development | membrane | ● | ○ | ● | ● | 3 |
| **WBGene00020302** | T07D1.3 | n/a | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00007097** | B0024.4 | n/a | n/a | regulation | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00001991** | hot-6 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00008233** | C50F4.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00008377** | D1054.10 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00010236** | F58B4.3 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00015713** | C12D12.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00017836** | F26F12.5 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00019988** | R09F10.5 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00016752** | C48E7.7 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00017058** | D2062.6 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00003957** | pcp-2 | lysosomal carboxypeptidase | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00000254** | bli-4 | endoprotease | Catalytic | metabolism | nucleus | ○ | ● | ● | ● | 3 |
| **WBGene00010578** | K04H8.3 | n/a | n/a | n/a | n/a | ● | ● | ● | ○ | 3 |
| **WBGene00020350** | T08B2.12 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00014135** | ZK896.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00004173** | pqn-94 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00002232** | kpc-1 | Furin like serine protease Subtilase family of serine proteases | Catalytic | metabolism | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00009909** | F49H6.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015646** | mlt-10 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00018497** | F46F5.6 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00018500** | F46F5.9 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00019260** | H34I24.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021503** | Y40D12A.2 | serine carboxypeptidase | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00008199** | C49C3.9 | n/a | n/a | defence | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00012947** | Y47H9C.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00007056** | crn-7 | n/a | Catalytic | metabolism | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021519** | Y41D4B.17 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021518** | Y41D4B.16 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00012827** | Y43F8C.5 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00012831** | Y43F8C.9 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00021732** | Y49G5B.1 | n/a | n/a | development | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021779** | Y51H7C.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021961** | Y57E12B.1 | n/a | n/a | n/a | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00006985** | zig-8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00003471** | mtd-1 | n/a | n/a | Signalling | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022711** | ZK355.1 | n/a | n/a | n/a | membrane | ● | ● | ● | ○ | 3 |
| **WBGene00022715** | ZK355.5 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00016424** | C34H4.1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00013911** | ZC482.2 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00020921** | W01C8.5 | n/a | n/a | regulation | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00001687** | gpn-1 | glypican | binding | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00021558** | Y45G5AM.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015713** | C12D12.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015805** | C15H9.11 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00044073** | tag-244 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00017762** | F23H11.7 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00014136** | ZK896.5 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00003849** | odr-2 | n/a | molecular function | Signalling | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00015328** | C02B10.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020497** | T14A8.2 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00001165** | efn-4 | n/a | n/a | regulation | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00015713** | C12D12.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00010639** | K07F5.15 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00021964** | Y57E12B.4 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00019810** | R01H2.2 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00022283** | lgc-27 | n/a | transport | transport ion | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022283** | lgc-27 | n/a | transport | transport ion | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00001989** | hot-4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00000039** | acn-1 | peptidase | Catalytic | metabolism | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00018823** | F54E2.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00017418** | F13B6.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00006611** | tre-5 | trehalase | Catalytic | metabolism | membrane | ● | ○ | ● | ● | 3 |
| **WBGene00011452** | ugt-55 | UDP-sugartransferase | Catalytic | metabolism | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00009432** | F35E12.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00011487** | T05E12.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00006404** | tag-10 | apical gut membrane protein | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00006404** | tag-10 | apical gut membrane protein | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022533** | cutl-19 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022533** | cutl-19 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00019662** | K11H12.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00013969** | ZK337.1 | Alpha-2-macroglobulin family (3 domains) | Catalytic | metabolism | extracellular | ● | ● | ● | ○ | 3 |
| **WBGene00016809** | C50D2.6 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00008369** | D1053.4 | n/a | n/a | n/a | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00010239** | F58B4.6 | n/a | n/a | n/a | n/a | ● | ● | ● | ○ | 3 |
| **WBGene00017976** | F32B5.2 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00007041** | tag-180 | calcium channel alpha-2 subunit | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00017483** | lgc-22 | n/a | transport | transport ion | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00002977** | lev-10 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00001992** | hot-7 | glycosylphosphatidylinositol (GPI)-linked signalling protein, (Ly-6 superfamily) | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00012009** | T25B9.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022645** | ZK6.11 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00007339** | C05D12.1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00004164** | pqn-83 | n/a | n/a | development | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00016354** | rig-6 | fibronectin, IG-like domains of NCAM | binding | development | membrane | ● | ● | ● | ○ | 3 |
| **WBGene00020096** | R144.6 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00000054** | acr-15 | ligand-gated ion channel subunit | receptor | transport ion | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00010660** | K08D8.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00010658** | K08D8.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00010658** | K08D8.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00010658** | K08D8.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00044138** | F31F6.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009416** | F35E2.9 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ○ | ● | ● | ● | 3 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ○ | ● | ● | ● | 3 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ○ | ● | ● | ● | 3 |
| **WBGene00044387** | C27A2.8 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00044457** | C18H7.11 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00010971** | R01E6.7 | n/a | n/a | n/a | n/a | ● | ● | ● | ○ | 3 |
| **WBGene00017193** | F07C3.2 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00020479** | T13C2.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021543** | Y43B11AR.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00013292** | Y57G11A.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020497** | T14A8.2 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00013788** | Y116A8C.8 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00013968** | ZK287.9 | n/a | n/a | n/a | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00017105** | E02H9.7 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00009913** | F49H6.12 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00013494** | Y70C5C.3 | n/a | n/a | n/a | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00016152** | pho-12 | acid phosphotase | Catalytic | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00008275** | C53B4.6 | Yeast YEA4 like protein | n/a | transport | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00010065** | F54F7.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000525** | clc-4 | n/a | n/a | n/a | membrane | ● | ● | ○ | ○ | 2 |
| **WBGene00016933** | C54G7.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00004020** | pho-1 | n/a | Catalytic | development | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00017592** | F19C7.2 | lysosomal carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ○ | ○ | 2 |
| **WBGene00007607** | C15C8.5 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00003173** | mec-9 | mechanosensory protein (mec-9) | binding | Signalling | extracellular | ○ | ● | ○ | ● | 2 |
| **WBGene00008584** | F08G5.6 | n/a | n/a | defence | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00010255** | F58E6.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00011011** | R04D3.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00011683** | phat-6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00006609** | tre-3 | trehalase | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007264** | C02F4.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00016686** | cyp-33C1 | cytochrome P450 | binding | metabolism | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00000050** | acr-11 | ligand-gated ionic channel protein | receptor | transport ion | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00018447** | F45C12.16 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00018789** | F54C1.1 | n/a | Catalytic | metabolism | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00011329** | T01D3.5 | n/a | transport | transport ion | membrane | ● | ● | ○ | ○ | 2 |
| **WBGene00011948** | T23F1.5 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00020836** | lgc-34 | ionic channel protein | transport | transport ion | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00012211** | W02D9.5 | n/a | structural | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00013882** | ZC410.5 | microfilarial antigen like | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00022751** | ZK484.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00008631** | F10A3.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00008698** | F11D11.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00010086** | F55B11.4 | n/a | binding | n/a | cytoplasm | ○ | ● | ○ | ● | 2 |
| **WBGene00010749** | K10D11.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00010750** | K10D11.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00012796** | Y43F4A.1 | zinc metallopeptidase (M8 family) | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00013915** | ZC482.7 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00017493** | F15E11.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00021222** | Y19D10A.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00002232** | kpc-1 | Furin like serine protease Subtilase family of serine proteases | Catalytic | metabolism | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00009412** | F35E2.5 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00010114** | F55D12.6 | n/a | structural | metabolism | cytoplasm | ○ | ● | ○ | ● | 2 |
| **WBGene00019067** | F58H7.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00021120** | W09G12.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00019213** | H20E11.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00010414** | H25K10.5 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00019389** | K04F1.10 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00019393** | K04F1.14 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00011592** | T07F10.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00012585** | lips-15 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00004123** | pqn-36 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00018917** | F56A4.9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00010637** | K07F5.12 | n/a | n/a | development | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00045459** | Y59A8B.26 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00021780** | scl-17 | n/a | n/a | n/a | extracellular | ○ | ● | ○ | ● | 2 |
| **WBGene00021809** | Y53G8AR.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00003959** | pcp-4 | peptidase | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00021960** | Y57E12AM.1 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00000783** | cpr-3 | cathepsin protease | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00015768** | C14C11.4 | n/a | n/a | development | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00003219** | mes-1 | tyrosine-protein kinase | Catalytic | regulation | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00000036** | ace-2 | carboxylesterase | Catalytic | Signalling | cell | ● | ○ | ○ | ● | 2 |
| **WBGene00022644** | dod-19 | n/a | n/a | development | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000862** | cwp-4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000845** | cup-4 | Acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00020487** | T13C5.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00003567** | ncx-2 | sodium/calcium exchanger protein 1 | transport | transport | membrane | ○ | ● | ○ | ● | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00003567** | ncx-2 | sodium/calcium exchanger protein 1 | transport | transport | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00016425** | C34H4.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00009645** | F42G10.1 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000799** | crn-6 | n/a | Catalytic | metabolism | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00008634** | F10A3.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00012857** | pbo-5 | Neurotransmitter-gated ion-channel | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00014125** | ZK863.8 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00006772** | unc-36 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00007340** | C05D12.2 | EGF domains | n/a | n/a | n/a | ○ | ○ | ● | ● | 2 |
| **WBGene00008964** | F19H8.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007746** | C26D10.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00012861** | Y45F3A.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00015539** | C06E7.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00017998** | F33D4.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000524** | clc-3 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00011325** | T01C3.11 | n/a | n/a | development | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00011380** | T02E1.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00012840** | grsp-1 | n/a | n/a | regulation | membrane | ● | ● | ○ | ○ | 2 |
| **WBGene00007545** | C13B4.1 | n/a | Catalytic | metabolism | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00008675** | F11A5.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00009399** | F35C5.11 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00013882** | ZC410.5 | microfilarial antigen like | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00020484** | T13C5.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00008560** | pho-13 | acid phosphatase | Catalytic | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007041** | tag-180 | calcium channel alpha-2 subunit | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00009499** | F36H2.2 | n/a | transport | transport | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00004944** | sol-1 | CUB domain | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00019392** | K04F1.13 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00014666** | C05D12.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00023432** | K12B6.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000039** | acn-1 | peptidase | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00012857** | pbo-5 | Neurotransmitter-gated ion-channel | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00000048** | acr-9 | acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00017888** | acl-11 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00010064** | F54F7.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007464** | C08H9.3 | Glucosyl-transferase | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00043156** | C27F2.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00011829** | T19A6.4 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00044203** | T02E9.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00044343** | clec-77 | clec family, C-type lectin | binding | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00044452** | Y102E9.5 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ● | 2 |
| **WBGene00044683** | C36E6.8 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00020207** | T04B8.5 | n/a | n/a | transport | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00016271** | C30G4.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00022093** | Y69A2AR.22 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00011383** | T02E9.5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00014132** | ZK896.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00017260** | F08F3.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000138** | amx-2 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00045381** | F28B1.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00019009** | clec-90 | n/a | binding | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00045403** | K10H10.12 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00010994** | lgc-25 | Neurotransmitter-gated ion-channel | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00013573** | Y75B12B.11 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00010760** | K10H10.4 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00007591** | C14H10.1 | Yeast YIL023C-like protein | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00045508** | D1081.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00045482** | T03F6.9 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00013982** | ZK512.1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00004023** | pho-4 | n/a | Catalytic | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00013126** | Y52B11A.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00012718** | Y39E4B.7 | n/a | binding | n/a | membrane | ● | ○ | ○ | ● | 2 |
| **WBGene00019676** | K12D9.12 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00012445** | Y16B4A.2 | serine carboxypeptidase | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00011262** | pho-8 | histidine acid phosphatase | Catalytic | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00004017** | phg-1 | growth arrest protein extracellular domain | n/a | development | n/a | ● | ○ | ○ | ● | 2 |
| **WBGene00008277** | C53B4.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00017815** | F26B1.1 | n/a | n/a | n/a | membrane | ● | ○ | ○ | ● | 2 |
| **WBGene00022474** | Y119C1B.9 | n/a | n/a | n/a | n/a | ● | ● | ○ | ○ | 2 |
| **WBGene00017695** | fip-1 | Environmental stress | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010074** | F54G8.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015125** | B0303.3 | Acetyl-coa acetyltransferase | Catalytic | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007139** | mnp-1 | Aminopeptidase | Catalytic | metabolism | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00018133** | F37A4.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015163** | B0361.9 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009450** | ugt-58 | UDP-glucuronosyltransferase | Catalytic | metabolism | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00010314** | F59B2.12 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017127** | E04F6.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019332** | K02F3.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00020603** | T20B12.5 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000616** | col-39 | collagen | structural | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00016681** | C45G9.10 | n/a | n/a | regulation | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00001262** | emb-8 | NADPH-cytochrome P450 | binding | metabolism | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00007191** | lgc-20 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007560** | C14A4.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018048** | clec-137 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018532** | F47B7.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00008803** | lips-10 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009504** | F37B12.1 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006979** | zig-2 | IG-like C2-type domains | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009876** | F49C12.6 | n/a | transport | transport carbohydrate | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015142** | B0310.6 | n/a | n/a | Signalling | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015496** | C05E11.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016132** | C26B9.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004993** | spp-8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016596** | C42D4.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018237** | F40F4.6 | EGF-like repeats | binding | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018484** | F46C8.1 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00020043** | R13A1.5 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00020582** | T19D12.7 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00020690** | T22E5.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020738** | T23F2.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000524** | clc-3 | n/a | n/a | n/a | membrane | ● | ○ | ○ | ○ | 1 |
| **WBGene00007535** | ttr-19 | Transthyretin-like family | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00008036** | C40C9.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004060** | pmp-3 | peroxisomal membrane protein (PMP70) | binding | transport | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00008320** | C54G10.4 | mitochondrial carrier protein | transport | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009331** | F32D8.7 | Kunitz/Bovine pancreatic trypsin inhibitor domain | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009339** | F32G8.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009915** | F52A8.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011010** | R04D3.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011089** | R07B7.5 | Monooxygenase | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00005198** | srg-41 | n/a | binding | Signalling | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001730** | grl-21 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00015300** | C01F1.5 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017654** | F21C10.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017655** | F21C10.5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018333** | cyp-33E3 | n/a | binding | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006948** | wrt-2 | n/a | n/a | regulation | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019059** | F58F9.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019066** | sdz-23 | n/a | n/a | regulation | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019754** | M03E7.2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020807** | T25F10.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020826** | T26A8.3 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00000540** | cln-3.2 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015340** | C02E7.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00005602** | srj-14 | 7TM chemoreceptor, srj family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00005643** | srp-2 | serine protease inhibitor | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015577** | ugt-64 | ugt family | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00005659** | srr-8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003573** | ncx-8 | n/a | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015848** | C16C8.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007845** | C31E10.4 | n/a | n/a | n/a | n/a | ● | ○ | ○ | ○ | 1 |
| **WBGene00016430** | C35A11.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000859** | cwp-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000860** | cwp-2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016731** | C46H11.7 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00016781** | C49G7.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016782** | phat-3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017105** | E02H9.7 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00000539** | cln-3.1 | Human CLN3 protein like | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00017201** | grsp-4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001388** | far-4 | O.volvulus antigen peptide like | binding | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000055** | acr-16 | ligand-gated ion channel subunit | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009134** | F25H9.1 | Activin types I and II receptor domain | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00009136** | F25H9.3 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017880** | F28A12.3 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00017918** | F29A7.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009523** | clec-165 | receptor like | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00006570** | tig-2 | n/a | developme nt | regulation | extracellular | ○ | ○ | ○ | ● | 1 |
| **WBGene00018289** | F41E6.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00006456** | tag-83 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018720** | F53A3.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009990** | F53F4.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010135** | F55H12.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019036** | F58E1.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019077** | F59A3.4 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010514** | K02E11.5 | n/a | Catalytic | metabolism | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00000254** | bli-4 | endoprotease | Catalytic | metabolism | nucleus | ○ | ● | ○ | ○ | 1 |
| **WBGene00001990** | hot-5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019494** | K07E8.1 | n/a | structural | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011151** | R08H2.10 | n/a | structural | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000555** | cnc-1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000556** | cnc-2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000557** | cnc-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000558** | cnc-4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000559** | cnc-5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011229** | R11.4 | n/a | n/a | n/a | membrane | ● | ○ | ○ | ○ | 1 |
| **WBGene00020076** | R52.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000233** | avr-15 | glutamate-gated chloride channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020631** | T20F7.3 | n/a | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00020741** | T23F4.3 | n/a | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021145** | clec-129 | n/a | binding | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00013145** | cutl-2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000639** | col-63 | collagen | structural | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00007999** | tag-297 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00008635** | F10A3.7 | n/a | n/a | n/a | membrane | ● | ○ | ○ | ○ | 1 |
| **WBGene00009422** | F35E8.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009710** | F44G3.10 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010023** | srbc-51 | 7TM receptor, srbc family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010169** | clec-18 | CUB domain, Lectin C-type domain short and long forms (2 domains) | binding | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010798** | srbc-76 | 7TM receptor, srbc family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011020** | R05A10.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011877** | T21B4.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012362** | W09D10.4 | Protein phosphatase 2C (2 domains) | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00003091** | lys-2 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013290** | Y57G11A.2 | n/a | structural | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013494** | Y70C5C.3 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00013779** | Y116A8B.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00003765** | nlp-27 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003766** | nlp-28 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003767** | nlp-29 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00015932** | C17H12.6 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00016391** | C34B2.6 | protease | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016433** | C35B1.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017120** | E04A4.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017480** | F15B10.1 | n/a | transport | transport nucleotide | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00017485** | F15E6.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017488** | dct-7 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00017507** | F16B4.5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017539** | F17E9.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015018** | srz-85 | 7TM chemoreceptor, srz family | n/a | n/a | membrane | ● | ○ | ○ | ○ | 1 |
| **WBGene00007450** | C08F11.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007458** | C08F11.11 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00007992** | fipr-24 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010418** | H27A22.1 | guanylate cyclase | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019507** | K07H8.7 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00008492** | F01D5.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010118** | F55F3.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010127** | F55G11.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001148** | eat-20 | EGF-like domain (3 domains) | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00010662** | K08E3.2 | n/a | structural | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010993** | R03E1.2 | n/a | binding | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00012635** | Y38H8A.1 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00002109** | ins-26 | n/a | developme nt | n/a | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00013931** | clec-97 | n/a | binding | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00015821** | clec-135 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017422** | F13C5.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018506** | F46F5.15 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019214** | H20E11.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019435** | K06A9.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00019435** | K06A9.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00015449** | ugt-63 | ugt family | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007866** | C32H11.3 | n/a | n/a | Signalling | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009557** | F39B2.7 | n/a | binding | Signalling | cytoplasm | ○ | ● | ○ | ○ | 1 |
| **WBGene00009913** | F49H6.12 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00012199** | W02B8.3 | n/a | n/a | development | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00012759** | Y41C4A.1 3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013190** | Y54E2A.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003790** | npp-4 | n/a | binding | cytoskeleton | nucleus | ○ | ● | ○ | ○ | 1 |
| **WBGene00015682** | C10G8.3 | n/a | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016881** | C52E2.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00008277** | C53B4.8 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00017406** | sdz-12 | n/a | binding | n/a | cytoplasm | ○ | ● | ○ | ○ | 1 |
| **WBGene00000955** | des-2 | nicotinic acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021162** | Y5H2A.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012590** | nspe-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012591** | nspe-1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012594** | nspe-5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004164** | pqn-83 | n/a | n/a | development | membrane | ● | ○ | ○ | ○ | 1 |
| **WBGene00021509** | Y41D4A.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00003763** | nlp-25 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001133** | eat-2 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004374** | rme-2 | LDL-like receptor | binding | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00021919** | cutl-25 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022070** | Y67D8C.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001636** | gly-11 | Glycosyl transferases | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00003566** | ncx-1 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013775** | Y116A8A.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00008652** | F10D11.6 | n/a | binding | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018381** | F43C11.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010893** | cutl-9 | cuticulin 1 | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011500** | T05G5.5 | Hypothetical protein A (T. aquaticus) | binding | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021981** | lgc-26 | ion channel protein | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006950** | wrt-4 | Hint module | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004372** | rig-5 | Drosophila amalgam protein like | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010245** | F58D5.6 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00000053** | acr-14 | neuronal acetylcholine receptor protein | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022447** | Y110A2AL.12 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021400** | Y38C1AA.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001046** | dnj-28 | n/a | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00006131** | str-69 | 7TM chemoreceptor, str family | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00013828** | Y116F11B.13 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00003575** | ncx-10 | n/a | n/a | transport | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00020921** | W01C8.5 | n/a | n/a | regulation | n/a | ● | ○ | ○ | ○ | 1 |
| **WBGene00009774** | F46B6.9 | n/a | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00000560** | cnc-6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021847** | Y54F10AL.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00022336** | Y82E9BR.3 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006592** | dpy-31 | Zinc metalloprotease | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00001819** | haf-9 | transporter protein | transport | transport | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00014091** | ZK822.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021325** | Y34B4A.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021874** | clec-81 | clec family, C-type lectin | binding | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00022100** | Y69A2AR.31 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015578** | C07G3.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015940** | C18A3.2 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001148** | eat-20 | EGF-like domain (3 domains) | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00017381** | ddr-2 | tyrosine kinase | Catalytic | regulation | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000046** | acr-7 | Acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022532** | ZC155.4 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018009** | F33D11.12 | n/a | binding | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001814** | haf-4 | ABC transporter | transport | transport | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00000061** | lgc-11 | Acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021160** | Y4C6B.6 | n/a | Catalytic | metabolism | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00020063** | R13D11.10 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012293** | W06A7.4 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00016133** | C26B9.3 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016335** | C33C12.3 | Glucosyl-ceramidase | Catalytic | metabolism | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00017294** | F09E10.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017299** | F09F7.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022580** | ZC262.3 | N-CAM IG domain | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022642** | ZK6.7 | lipase | n/a | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018272** | F41C3.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019127** | cgt-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019127** | cgt-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00002977** | lev-10 | n/a | n/a | n/a | membrane | ● | ○ | ○ | ○ | 1 |
| **WBGene00021384** | Y37F4.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00022033** | Y65B4BL.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00022106** | lgc-46 | ion channel protein | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021941** | lgc-33 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00016642** | C44C1.2 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00020413** | T10E9.3 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000232** | avr-14 | n/a | transport | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019127** | cgt-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022077** | Y69A2AR.6 | n/a | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000058** | acr-19 | nicotinic acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016329** | osr-1 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015055** | B0222.3 | phosphate permease | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00001479** | fmo-4 | flavin-containing monoxygenase | Catalytic | metabolism | membrane ER | ○ | ● | ○ | ○ | 1 |
| **WBGene00018206** | ugt-61 | ugt family | Catalytic | metabolism | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00015786** | C15B12.4 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00000988** | dhs-25 | short-chain alcohol dehydrogenase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004017** | phg-1 | growth arrest protein extracellular domain | n/a | development | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00018411** | F44B9.10 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019401** | nuo-4 | NADH dehydrogenase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021448** | Y39D8A.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015619** | C08G9.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019848** | R03G5.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016336** | C33C12.4 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018928** | F56B3.2 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010342** | F59F5.3 | tyrosine-protein kinase | Catalytic | regulation | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010597** | K06A4.7 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00001476** | fmo-1 | flavin-containing monoxygenase | Catalytic | metabolism | membrane ER | ○ | ● | ○ | ○ | 1 |
| **WBGene00013225** | Y56A3A.2 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007807** | C29F3.7 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018716** | F52H2.4 | n/a | transport | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018977** | F56E10.3 | n/a | binding | cytoskeleton | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00020984** | W03D8.1 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00013574** | Y76A2B.2 | Leucine Rich Repeat (2 copies) (2 domains) | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00008320** | C54G10.4 | mitochondrial carrier protein | transport | transport | mitochondria | ○ | ● | ○ | ○ | 1 |
| **WBGene00009971** | F53C11.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010573** | K04H4.2 | Chitin-binding motifs | binding | metabolism | extracellular | ○ | ○ | ○ | ● | 1 |
| **WBGene00012761** | Y41C4A.1 8 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00009406** | F35C11.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007178** | B0457.2 | elastin precusor | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016193** | C28H8.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017312** | F09G2.3 | permease | transport | transport ion | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00000055** | acr-16 | ligand-gated ion channel subunit | transport | transport ion | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00010901** | M28.10 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00018112** | F36H9.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000254** | bli-4 | endoprotease | Catalytic | metabolism | nucleus | ○ | ● | ○ | ○ | 1 |
| **WBGene00020693** | T22E7.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022506** | ZC21.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015303** | rga-6 | n/a | n/a | Signalling | cytoplasm | ○ | ● | ○ | ○ | 1 |
| **WBGene00017886** | F28B3.5 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020479** | T13C2.3 | n/a | n/a | n/a | n/a | ● | ○ | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00003519** | nac-3 | Yeast ORF YCR37C | transport | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017071** | D2096.3 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012718** | Y39E4B.7 | n/a | binding | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00001512** | gab-1 | GABA receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022677** | ZK180.3 | n/a | transport | transport lipid | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007325** | C05C9.1 | LBP / BPI / CETP family | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007402** | ugt-60 | UDP-glucuronosyl-transferase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001406** | fce-2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010959** | MTCE.11 | NADH dehydrogenase ND1 | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011084** | srsx-21 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00001692** | grd-3 | n/a | binding | regulation | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00002975** | lev-8 | acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013684** | Y105E8A.27 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00016732** | phat-1 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00019845** | R03G5.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007497** | C09G9.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016174** | C27H5.4 | n/a | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00008340** | C55A6.11 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009528** | F38A6.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010723** | cpg-7 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00010954** | clec-189 | n/a | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004264** | qua-1 | hedgehog-like protein | Catalytic | metabolism | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00012152** | cnc-10 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012488** | clec-105 | n/a | binding | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00012603** | nspe-6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012604** | nspe-2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012371** | W09G3.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007055** | tag-196 | cysteine protease and a protease inhibitor | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00005957** | srx-66 | 7TM chemoreceptor, srx family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016641** | C44C1.1 | n/a | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00017580** | lgc-4 | member of the ligand-gated ionic channels family | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017890** | F28B4.1 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00018250** | F40H3.2 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00010350** | H01G02.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010350** | H01G02.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010745** | dod-17 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00005120** | srd-42 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00011354** | lgc-13 | nitcotinic acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013657** | Y105C5B.18 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00013493** | clec-9 | CUB domain, Lectin C-type domain short and long forms | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00020420** | T10E10.3 | G-protein receptor | n/a | Signalling | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020863** | T27E4.5 | n/a | inhibitor | n/a | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00003525** | nas-6 | Zinc-binding metalloprotease domain | Catalytic | metabolism | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00009204** | F28C6.4 | Yeast YJ10 like | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009204** | F28C6.4 | Yeast YJ10 like | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011965** | T23G7.2 | n/a | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022443** | Y110A2AL.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00022891** | ZK1290.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012315** | immt-2 | [031110 dl] Modified prediction based on EST data, correct splice donor from exon 2 | n/a | n/a | mitochondria | ○ | ● | ○ | ○ | 1 |
| **WBGene00013013** | clec-145 | n/a | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013412** | Y64G10A.2 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00006027** | srx-136 | 7TM receptor, srx family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044067** | hke-4.1 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00005898** | srx-7 | 7TM chemoreceptor, srx family | n/a | Signalling | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006006** | srx-115 | 7TM chemoreceptor, srx family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010940** | M163.8 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017389** | lgc-38 | gamma-aminobutyric acid receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00002098** | ins-15 | ins family | development | n/a | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00016663** | C45E1.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018865** | F55A12.6 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00011355** | lgc-14 | nicotinic acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012915** | lgc-35 | n/a | transport | transport ion | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00021582** | clec-71 | clec family, C-type lectin | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016079** | C24H12.10 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00005297** | srh-76 | 7TM chemoreceptor, srh family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000062** | acr-23 | channel protein | receptor | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00014669** | C06G8.3 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016417** | C34F11.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00002132** | inx-10 | type-1 membrane protein | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00021095** | mlt-8 | n/a | n/a | development | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00004164** | pqn-83 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006952** | wrt-6 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015284** | C01B10.10 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016721** | C46G7.1 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004890** | smp-2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00008583** | ugt-65 | UDP-glucuronyl-transferase like | Catalytic | metabolism | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00023504** | F26F2.8 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00003762** | nlp-24 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009882** | vha-17 | n/a | transport | transport ion | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00013601** | Y87G2A.13 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00013601** | Y87G2A.13 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010354** | cyp-31A2 | Cytochrome P450 | binding | metabolism | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00007070** | ugt-49 | UDP-glucuronosyl-transferase | Catalytic | metabolism | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00012200** | W02B8.4 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044074** | W02B8.6 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00008595** | clec-56 | C-type lectin domain | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015315** | srbc-29 | 7TM chemoreceptor, srbc family | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00015316** | srbc-30 | 7TM chemoreceptor, srbc family | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00007954** | C35C5.2 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011971** | T23G11.6 | Rat insulin-like growth factor binding protein complex acid labile chain like | binding | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00012847** | srxa-15 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044176** | C30G7.2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044189** | F36D3.14 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018880** | acc-3 | Ligand-gated ionic channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044152** | W04G3.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00044287** | F21H12.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00044301** | lgc-28 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044292** | F56D6.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044423** | F53F10.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006494** | hke-4.2 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044436** | Y47G6A.31 | n/a | binding | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044560** | C36C9.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044472** | dct-8 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00044411** | R12B2.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044548** | cnc-9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017399** | lgc-51 | ligand-gated ionic channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00021626** | Y47D7A.14 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044637** | Y47D7A.16 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003572** | ncx-7 | Na/Ca, K antiporter | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015476** | C05D9.9 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00016152** | pho-12 | acid phosphotase | Catalytic | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00001587** | ggr-2 | Glycine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020760** | T24C4.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020657** | lgc-53 | ligand-gated ionic channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011102** | R07E3.1 | cysteine proteinase | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010655** | K08D8.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00012789** | Y43D4A.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00005625** | srj-42 | 7tm receptor protein | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00044756** | F58F12.4 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00044754** | Y119C1B.12 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00022474** | Y119C1B.9 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00022255** | Y73B6BL.36 | n/a | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018226** | F40B5.2 | n/a | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018226** | F40B5.2 | n/a | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003566** | ncx-1 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021586** | clec-75 | n/a | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013471** | clec-242 | n/a | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010418** | H27A22.1 | guanylate cyclase | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044900** | cnc-11 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013451** | Y67A10A.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00044922** | Y43C5A.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044988** | W01A8.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010748** | K10D11.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012542** | Y37A1B.9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044801** | ZC262.9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011360** | lgc-18 | nicotinic acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012814** | Y43F8B.3 | Kunitz/Bovine pancreatic trypsin inhibitor domain | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001455** | flp-12 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007807** | C29F3.7 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017306** | F09F9.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00009331** | F32D8.7 | Kunitz/Bovine pancreatic trypsin inhibitor domain | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011927** | T22C8.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012814** | Y43F8B.3 | Kunitz/Bovine pancreatic trypsin inhibitor domain | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00045397** | Y54G2A.52 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00045455** | F26G1.11 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00045251** | F54F7.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019069** | lgc-30 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011328** | T01D3.3 | Von Willebrand factor-like Copper/zinc superoxide dismutases (SODC) | binding | metabolism | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00043066** | acr-25 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013351** | Y59A8B.19 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010027** | F54B11.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00014669** | C06G8.3 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018978** | sdz-22 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00045488** | F57B1.9 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00045486** | K05F6.12 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011121** | R07E5.17 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00045494** | ZK662.6 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00004372** | rig-5 | Drosophila amalgam protein like | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013487** | tag-336 | EGF-like domain | Catalytic | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017569** | F18E9.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018278** | F41C6.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00017885** | F28B3.4 | n/a | binding | n/a | nucleus | ○ | ○ | ○ | ● | 1 |
| **WBGene00009762** | F46B3.9 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00006624** | try-6 | peptidase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019746** | M03A1.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012612** | Y38H6A.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012391** | Y6B3B.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013060** | Y51A2A.4 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012215** | W02D9.9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044399** | F11F1.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00008583** | ugt-65 | UDP-glucuronyl-transferase like | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00008277** | C53B4.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00020642** | T20H9.6 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |

## Appendix 2.

All *C. elegans* proteins that also have *C. briggsae* orthologues with GPI predictions, of which there are 382 proteins. 201 of these have predictions in two programs or more. Columns 1 to 3 contain the Wormbase gene ID, gene name and a brief description of the protein. Columns 4 to 6 contain the GO terms for the proteins where available. Columns 7 to 10 contain the programs with which the protein was predicted, with ● indicating a positive prediction and ○ a negative. Column 11 denotes the number of prediction programs that gave the protein a positive result.

| Wormbase gene ID | gene name | brief description | Molecular function | Biological process | Cellular component | Big PI | GPI SOM | Frag Anchor | Pred GPI | No. of hits |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009700** | F44F4.1 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00017969** | F32A5.3 | Serine carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00015803** | C15H9.9 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00004370** | rig-3 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ● | ● | ● | ● | 4 |
| **WBGene00013969** | ZK337.1 | Alpha-2-macroglobulin family (3 domains) | Catalytic | metabolism | extracellular | ● | ● | ● | ● | 4 |
| **WBGene00007299** | C04F12.5 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018787** | cutl-20 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00021452** | Y39F10A.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009679** | F44D12.2 | n/a | binding | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00022246** | acp-7 | n/a | Catalytic | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00000038** | ace-4 | Acetylcholine-esterase | Catalytic | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00006869** | vab-2 | n/a | binding | Signalling | anchored | ● | ● | ● | ● | 4 |
| **WBGene00007911** | C34B7.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00020195** | T03G6.3 | plasma cell membrane protein and phosphor-diesterase I (weak) | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00000037** | ace-3 | n/a | Catalytic | Signalling | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018576** | F47G3.1 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00001988** | hot-3 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00016979** | C56G2.4 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00019320** | K02E10.6 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00018984** | F56F10.1 | peptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00017594** | F19C7.4 | lysosomal carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00007722** | C25D7.15 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00006621** | try-3 | peptidase | Catalytic | metabolism | n/a | ● | ● | ● | ● | 4 |
| **WBGene00011314** | T01B7.9 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00015472** | C05D9.3 | n/a | binding | cell adhesion | membrane | ● | ● | ● | ● | 4 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00003056** | lon-2 | n/a | binding | development | cell surface | ● | ● | ● | ● | 4 |
| **WBGene00001163** | efn-2 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00008776** | F13H10.5 | n/a | Catalytic | metabolism | membrane | ● | ● | ● | ● | 4 |
| **WBGene00009428** | F35E12.4 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00021526** | Y41G9A.2 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00002181** | kal-1 | WAP-type (Whey Acidic Protein) 'four-disulfide core', Fibronectin type III domain (3 domains) | inhibitor | n/a | cell surface | ● | ● | ● | ● | 4 |
| **WBGene00021791** | Y51H7C.13 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00012202** | W02B12.4 | esterase | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00044484** | C09B9.8 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00044556** | F38G1.3 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00045400** | C54D10.13 | n/a | n/a | n/a | n/a | ● | ● | ● | ● | 4 |
| **WBGene00009779** | F46C5.2 | n/a | n/a | n/a | membrane | ● | ● | ● | ● | 4 |
| **WBGene00016627** | C44B7.5 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020302** | T07D1.3 | n/a | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00001991** | hot-6 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00008233** | C50F4.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015713** | C12D12.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00017836** | F26F12.5 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00019988** | R09F10.5 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00016752** | C48E7.7 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00008509** | F01G10.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00003957** | pcp-2 | lysosomal carboxypeptidase | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00001581** | gfi-1 | n/a | n/a | regulation | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00019017** | F57F4.4 | n/a | n/a | regulation | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00019660** | K11H12.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00019663** | K11H12.7 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020350** | T08B2.12 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00014135** | ZK896.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00004173** | pqn-94 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020995** | W03F8.6 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00012439** | Y12A6A.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00019260** | H34I24.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021503** | Y40D12A.2 | serine carboxypeptidase | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00012947** | Y47H9C.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021519** | Y41D4B.17 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021518** | Y41D4B.16 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021779** | Y51H7C.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00003471** | mtd-1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00003956** | pcp-1 | n/a | Catalytic | metabolism | n/a | ○ | ● | ● | ● | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00016424** | C34H4.1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00013911** | ZC482.2 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00020921** | W01C8.5 | n/a | n/a | regulation | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00001687** | gpn-1 | glypican | binding | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00021558** | Y45G5AM.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015713** | C12D12.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015805** | C15H9.11 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00044073** | tag-244 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00014136** | ZK896.5 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00015328** | C02B10.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020497** | T14A8.2 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00001165** | efn-4 | n/a | n/a | regulation | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00015713** | C12D12.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022283** | lgc-27 | n/a | transport | transport ion | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022283** | lgc-27 | n/a | transport | transport ion | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00018823** | F54E2.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00011879** | pho-7 | histidine acid phosphatase | Catalytic | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00009432** | F35E12.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00011487** | T05E12.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022645** | ZK6.11 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00019662** | K11H12.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00013969** | ZK337.1 | Alpha-2-macroglobulin family (3 domains) | Catalytic | metabolism | extracellular | ○ | ● | ● | ● | 3 |
| **WBGene00016809** | C50D2.6 | n/a | n/a | n/a | n/a | ● | ● | ● | ○ | 3 |
| **WBGene00008369** | D1053.4 | n/a | n/a | n/a | membrane | ● | ● | ○ | ● | 3 |
| **WBGene00010239** | F58B4.6 | n/a | n/a | n/a | n/a | ● | ● | ○ | ● | 3 |
| **WBGene00012073** | T27A8.1 | carboxypeptidase | Catalytic | metabolism | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00007041** | tag-180 | calcium channel alpha-2 subunit | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00017483** | lgc-22 | n/a | transport | transport ion | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00001992** | hot-7 | glycosylphosphatidylinositol (GPI)-linked signalling protein, (Ly-6 superfamily) | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00000283** | cah-5 | carbonic anhydrase | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00012009** | T25B9.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00022645** | ZK6.11 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00007339** | C05D12.1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00016354** | rig-6 | fibronectin, IG-like domains of NCAM | binding | development | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00000054** | acr-15 | ligand-gated ion channel subunit | receptor | transport ion | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00044138** | F31F6.8 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00009416** | F35E2.9 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ○ | ● | ● | ● | 3 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ○ | ● | ● | ● | 3 |
| **WBGene00009431** | dct-17 | n/a | binding | metabolism | cytoplasm | ○ | ● | ● | ● | 3 |
| **WBGene00010971** | R01E6.7 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00021543** | Y43B11AR.1 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00010059** | F54E4.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00013292** | Y57G11A.4 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00020497** | T14A8.2 | n/a | n/a | n/a | n/a | ○ | ● | ● | ● | 3 |
| **WBGene00003958** | pcp-3 | lysosomal carboxypeptidase | Catalytic | metabolism | membrane | ○ | ● | ● | ● | 3 |
| **WBGene00008275** | C53B4.6 | Yeast YEA4 like protein | n/a | transport | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00000525** | clc-4 | n/a | n/a | n/a | membrane | ● | ● | ○ | ○ | 2 |
| **WBGene00016354** | rig-6 | fibronectin, IG-like domains of NCAM | binding | development | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00004020** | pho-1 | n/a | Catalytic | development | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00017592** | F19C7.2 | lysosomal carboxypeptidase | Catalytic | metabolism | n/a | ● | ● | ○ | ○ | 2 |
| **WBGene00007097** | B0024.4 | n/a | n/a | regulation | n/a | ○ | ○ | ● | ● | 2 |
| **WBGene00007607** | C15C8.5 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00003173** | mec-9 | mechanosensory protein (mec-9) | binding | Signalling | extracellular | ○ | ● | ○ | ● | 2 |
| **WBGene00010236** | F58B4.3 | n/a | n/a | n/a | n/a | ● | ● | ○ | ○ | 2 |
| **WBGene00011683** | phat-6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00006609** | tre-3 | trehalase | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000050** | acr-11 | ligand-gated ionic channel protein | receptor | transport ion | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00018789** | F54C1.1 | n/a | Catalytic | metabolism | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00010578** | K04H8.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00011329** | T01D3.5 | n/a | transport | transport ion | membrane | ● | ● | ○ | ○ | 2 |
| **WBGene00020836** | lgc-34 | ionic channel protein | transport | transport ion | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00013882** | ZC410.5 | microfilarial antigen like | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00022751** | ZK484.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00008698** | F11D11.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00010086** | F55B11.4 | n/a | binding | n/a | cytoplasm | ○ | ● | ○ | ● | 2 |
| **WBGene00013915** | ZC482.7 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00017493** | F15E11.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00021222** | Y19D10A.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00019067** | F58H7.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00019213** | H20E11.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00008199** | C49C3.9 | n/a | n/a | defence | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00007056** | crn-7 | n/a | Catalytic | metabolism | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00011592** | T07F10.6 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00012585** | lips-15 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00012827** | Y43F8C.5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00021732** | Y49G5B.1 | n/a | n/a | development | n/a | ○ | ● | ○ | ● | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00018917** | F56A4.9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00010637** | K07F5.12 | n/a | n/a | development | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00021809** | Y53G8AR.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00003959** | pcp-4 | peptidase | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000783** | cpr-3 | cathepsin protease | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00015768** | C14C11.4 | n/a | n/a | development | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00000036** | ace-2 | carboxylesterase | Catalytic | Signalling | cell | ● | ○ | ○ | ● | 2 |
| **WBGene00022644** | dod-19 | n/a | n/a | development | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000845** | cup-4 | Acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00000039** | acn-1 | peptidase | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00003567** | ncx-2 | sodium/calcium exchanger protein 1 | transport | transport | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00003567** | ncx-2 | sodium/calcium exchanger protein 1 | transport | transport | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00016425** | C34H4.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007652** | C17G1.5 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00014125** | ZK863.8 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00006772** | unc-36 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00007340** | C05D12.2 | EGF domains | n/a | n/a | n/a | ○ | ○ | ● | ● | 2 |
| **WBGene00008964** | F19H8.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00012861** | Y45F3A.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00015539** | C06E7.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00011498** | T05G5.1 | Caldesmon-like repeats | n/a | n/a | membrane | ○ | ○ | ● | ● | 2 |
| **WBGene00011380** | T02E1.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007545** | C13B4.1 | n/a | Catalytic | metabolism | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00013882** | ZC410.5 | microfilarial antigen like | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00008560** | pho-13 | acid phosphatase | Catalytic | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00007041** | tag-180 | calcium channel alpha-2 subunit | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00009499** | F36H2.2 | n/a | transport | transport | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00004944** | sol-1 | CUB domain | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00019392** | K04F1.13 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00014666** | C05D12.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00023432** | K12B6.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00000039** | acn-1 | peptidase | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00000048** | acr-9 | acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00017888** | acl-11 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00020096** | R144.6 | n/a | n/a | n/a | membrane | ● | ● | ○ | ○ | 2 |
| **WBGene00006942** | wrk-1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00043156** | C27F2.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00010660** | K08D8.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00011829** | T19A6.4 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00044457** | C18H7.11 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00044452** | Y102E9.5 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ● | 2 |
| **WBGene00017193** | F07C3.2 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00044683** | C36E6.8 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00016271** | C30G4.6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00011383** | T02E9.5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00014132** | ZK896.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00010747** | K10D11.3 | n/a | n/a | n/a | n/a | ○ | ● | ● | ○ | 2 |
| **WBGene00045248** | ZK180.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00000138** | amx-2 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00045403** | K10H10.12 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00010994** | lgc-25 | Neurotransmitter-gated ion-channel | transport | transport ion | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00045482** | T03F6.9 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00013982** | ZK512.1 | n/a | n/a | n/a | membrane | ○ | ● | ● | ○ | 2 |
| **WBGene00004023** | pho-4 | n/a | Catalytic | n/a | n/a | ○ | ● | ○ | ● | 2 |
| **WBGene00013126** | Y52B11A.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ● | 2 |
| **WBGene00017815** | F26B1.1 | n/a | n/a | n/a | membrane | ● | ○ | ○ | ● | 2 |
| **WBGene00017695** | fip-1 | Environmental stress | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015125** | B0303.3 | Acetyl-coa acetyltransferase | Catalytic | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015163** | B0361.9 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009450** | ugt-58 | UDP-glucuronosyl-transferase | Catalytic | metabolism | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00010314** | F59B2.12 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017127** | E04F6.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019332** | K02F3.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000616** | col-39 | collagen | structural | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00007191** | lgc-20 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009798** | F46G10.4 | lipase | Catalytic | metabolism | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00018532** | F47B7.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015142** | B0310.6 | n/a | n/a | Signalling | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004993** | spp-8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016596** | C42D4.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00020582** | T19D12.7 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00020690** | T22E5.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020738** | T23F2.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009331** | F32D8.7 | Kunitz/Bovine pancreatic trypsin inhibitor domain | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009339** | F32G8.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009915** | F52A8.1 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011011** | R04D3.3 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00001730** | grl-21 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015300** | C01F1.5 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017654** | F21C10.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006948** | wrt-2 | n/a | n/a | regulation | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000540** | cln-3.2 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00015340** | C02E7.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007264** | C02F4.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00005643** | srp-2 | serine protease inhibitor | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015848** | C16C8.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001388** | far-4 | O.volvulus antigen peptide like | binding | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000055** | acr-16 | ligand-gated ion channel subunit | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009134** | F25H9.1 | Activin types I and II receptor domain | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00009136** | F25H9.3 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00017880** | F28A12.3 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00018289** | F41E6.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009990** | F53F4.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010135** | F55H12.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019077** | F59A3.4 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001990** | hot-5 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00000556** | cnc-2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000557** | cnc-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000558** | cnc-4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000559** | cnc-5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020076** | R52.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007999** | tag-297 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010169** | clec-18 | CUB domain, Lectin C-type domain short and long forms (2 domains) | binding | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010750** | K10D11.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011020** | R05A10.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012362** | W09D10.4 | Protein phosphatase 2C (2 domains) | Catalytic | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00003765** | nlp-27 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003767** | nlp-29 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00016433** | C35B1.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017480** | F15B10.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00017485** | F15E6.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007992** | fipr-24 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010118** | F55F3.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010127** | F55G11.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012635** | Y38H8A.1 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00009557** | F39B2.7 | n/a | binding | Signalling | cytoplasm | ○ | ● | ○ | ○ | 1 |
| **WBGene00010005** | cnc-7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013190** | Y54E2A.5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003790** | npp-4 | n/a | binding | cytoskeleton | nucleus | ○ | ● | ○ | ○ | 1 |
| **WBGene00000955** | des-2 | nicotinic acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012594** | nspe-5 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003763** | nlp-25 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001133** | eat-2 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003566** | ncx-1 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013775** | Y116A8A.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00008652** | F10D11.6 | n/a | binding | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004123** | pqn-36 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00018381** | F43C11.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021981** | lgc-26 | ion channel protein | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004372** | rig-5 | Drosophila amalgam protein like | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021960** | Y57E12AM.1 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00003575** | ncx-10 | n/a | n/a | transport | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00009774** | F46B6.9 | n/a | n/a | development | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00000560** | cnc-6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021847** | Y54F10AL.1 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00022336** | Y82E9BR.3 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006592** | dpy-31 | Zinc metalloprotease | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001819** | haf-9 | transporter protein | transport | transport | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00021325** | Y34B4A.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015578** | C07G3.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00015940** | C18A3.2 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00000046** | acr-7 | Acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001814** | haf-4 | ABC transporter | transport | transport | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00000061** | lgc-11 | Acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021160** | Y4C6B.6 | n/a | Catalytic | metabolism | membrane lysosome | ○ | ● | ○ | ○ | 1 |
| **WBGene00012293** | W06A7.4 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00016133** | C26B9.3 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019127** | cgt-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00019127** | cgt-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021384** | Y37F4.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00021941** | lgc-33 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00019127** | cgt-3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020487** | T13C5.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00022077** | Y69A2AR. | n/a | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 6 | | | | | | | | | |
| **WBGene00016329** | osr-1 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001479** | fmo-4 | flavin-containing monoxygenase | Catalytic | metabolism | membrane ER | ○ | ● | ○ | ○ | 1 |
| **WBGene00000988** | dhs-25 | short-chain alcohol dehydrogenase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018411** | F44B9.10 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00001476** | fmo-1 | flavin-containing monoxygenase | Catalytic | metabolism | membrane ER | ○ | ● | ○ | ○ | 1 |
| **WBGene00013225** | Y56A3A.2 | n/a | Catalytic | metabolism | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007807** | C29F3.7 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00012857** | pbo-5 | Neurotransmitter-gated ion-channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009971** | F53C11.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009406** | F35C11.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007178** | B0457.2 | elastin precusor | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00001512** | gab-1 | GABA receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007325** | C05C9.1 | LBP / BPI / CETP family | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00007402** | ugt-60 | UDP-glucuronosyl-transferase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011084** | srsx-21 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00013684** | Y105E8A.27 | n/a | n/a | n/a | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00009528** | F38A6.4 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009969** | F53B7.7 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00010954** | clec-189 | n/a | binding | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00004264** | qua-1 | hedgehog-like protein | Catalytic | metabolism | extracellular | ○ | ● | ○ | ○ | 1 |
| **WBGene00012152** | cnc-10 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012603** | nspe-6 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012840** | grsp-1 | n/a | n/a | regulation | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00008675** | F11A5.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010350** | H01G02.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00010350** | H01G02.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013657** | Y105C5B.18 | n/a | n/a | n/a | n/a | ○ | ○ | ○ | ● | 1 |
| **WBGene00020484** | T13C5.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00022443** | Y110A2AL.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013412** | Y64G10A.2 | n/a | n/a | development | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00006027** | srx-136 | 7TM receptor, srx family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044067** | hke-4.1 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006006** | srx-115 | 7TM chemoreceptor, srx family | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017389** | lgc-38 | gamma-aminobutyric acid receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00014669** | C06G8.3 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021095** | mlt-8 | n/a | n/a | development | n/a | ○ | ○ | ○ | ● | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **WBGene00006952** | wrt-6 | n/a | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00016721** | C46G7.1 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012857** | pbo-5 | Neurotransmitter-gated ion-channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004890** | smp-2 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003762** | nlp-24 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00009882** | vha-17 | n/a | transport | transport ion | membrane | ○ | ○ | ○ | ● | 1 |
| **WBGene00010064** | F54F7.2 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00013601** | Y87G2A.13 | n/a | n/a | n/a | membrane | ○ | ○ | ● | ○ | 1 |
| **WBGene00013601** | Y87G2A.13 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00018880** | acc-3 | Ligand-gated ionic channel | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044152** | W04G3.10 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00044287** | F21H12.7 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00044292** | F56D6.8 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00006494** | hke-4.2 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044548** | cnc-9 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00021626** | Y47D7A.14 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003572** | ncx-7 | Na/Ca, K antiporter | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00020760** | T24C4.4 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00022474** | Y119C1B.9 | n/a | n/a | n/a | n/a | ○ | ○ | ● | ○ | 1 |
| **WBGene00022255** | Y73B6BL.36 | n/a | n/a | transport | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018226** | F40B5.2 | n/a | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00018226** | F40B5.2 | n/a | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00003566** | ncx-1 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044900** | cnc-11 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00044922** | Y43C5A.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00017260** | F08F3.1 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00011360** | lgc-18 | nicotininc acetylcholine receptor | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007807** | C29F3.7 | n/a | n/a | regulation | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00009331** | F32D8.7 | Kunitz/Bovine pancreatic trypsin inhibitor domain | Catalytic | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011927** | T22C8.6 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00045397** | Y54G2A.52 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00013573** | Y75B12B.11 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00045251** | F54F7.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019069** | lgc-30 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00043066** | acr-25 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00007591** | C14H10.1 | Yeast YIL023C-like protein | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00014669** | C06G8.3 | n/a | transport | transport ion | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00011121** | R07E5.17 | n/a | n/a | development | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004372** | rig-5 | Drosophila amalgam protein | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | like | | | | | | | | |
| **WBGene00017569** | F18E9.3 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00006624** | try-6 | peptidase | Catalytic | metabolism | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00019746** | M03A1.3 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00012391** | Y6B3B.7 | n/a | n/a | n/a | membrane | ○ | ● | ○ | ○ | 1 |
| **WBGene00004017** | phg-1 | growth arrest protein extracellular domain | n/a | development | n/a | ● | ○ | ○ | ○ | 1 |
| **WBGene00008277** | C53B4.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00008277** | C53B4.8 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |
| **WBGene00022474** | Y119C1B.9 | n/a | n/a | n/a | n/a | ○ | ● | ○ | ○ | 1 |

## Appendix 3.

Results for all of the *C. elegans* proteins identified with LC MS/MS. A total of 287 proteins were identified with the MASCOT program. Gene name, public name and Wormbase ID were taken from Wormbase. The score and query matched refers to the MASCOT output for the total score of the protein and the number of peptides assigned to the protein by the program, respectively. Unique peptides refer to the number of unique statistically significant peptides assigned to each protein after manual curation (for an in-depth description of MASCOT output see Figure 5.7). The main band refers to the gel band (from the 1st dimension of separation with SDS-PAGE, Figure 5.6) from which the protein identification score is the highest, and other bands refers to the gel bands that also contain identifications for the protein.

| Gene name | Wormbase gene ID | Score | Query matched | Unique peptides | Main band | Other bands |
|---|---|---|---|---|---|---|
| **K10C2.1** | WBGene00019617 | 783 | 18 | 12 | 4 | 1,2,3,5,7,8 |
| **F54F11.2** | WBGene00010070 | 734 | 16 | 11 | 3 | 4,5,6 |
| **Y16B4A.2** | WBGene00012445 | 597 | 16 | 8 | 2 | 1,4,7 |
| **pcp-3** | WBGene00003958 | 593 | 14 | 8 | 4 | 1,2,3,5 |
| **C29F3.7** | WBGene00007807 | 594 | 13 | 7 | 4 | 1,2,3 |
| **lec-2** | WBGene00002265 | 487 | 12 | 7 | 9 | 2,7,8 |
| **lec-4** | WBGene00002267 | 399 | 8 | 6 | 7 | 1,2,8,9 |
| **ZK6.11** | WBGene00022645 | 395 | 10 | 6 | 5 | 1,2,3,4,6 |
| **F32A5.3** | WBGene00017969 | 373 | 10 | 6 | 7 | 4 |
| **pho-1** | WBGene00004020 | 348 | 9 | 6 | 5 | 4 |
| **pcp-2** | WBGene00003957 | 449 | 9 | 5 | 3 | 4 |
| **Y41D4B.16** | WBGene00021518 | 347 | 6 | 5 | 3 | 2 |
| **lec-1** | WBGene00002264 | 340 | 9 | 5 | 9 | 1,2,3,4,5,7,8,10 |
| **F21D5.3** | WBGene00009008 | 338 | 6 | 5 | 3 | 1 |
| **F57F4.4** | WBGene00019017 | 337 | 9 | 5 | 1 | none |
| **R05G6.7** | WBGene00019900 | 319 | 8 | 5 | 8 | 1,2,3,4,5,7,9,10 |
| **F56F10.1** | WBGene00018984 | 358 | 6 | 4 | 4 | 2,5 |
| **pcp-4** | WBGene00003959 | 354 | 6 | 4 | 5 | 3,4 |
| **dod-19** | WBGene00022644 | 333 | 13 | 4 | 5 | 1,2,3,4,7 |
| **C26B9.5** | WBGene00016134 | 252 | 8 | 4 | 5 | 4 |
| **lec-5** | WBGene00002268 | 262 | 8 | 3 | 7 | none |
| **vps-32.1** | WBGene00016961 | 255 | 6 | 3 | 9 | none |
| **T19D12.4** | WBGene00020579 | 249 | 5 | 3 | 5 | none |
| **tag-10** | WBGene00006404 | 245 | 4 | 3 | 3 | 4 |
| **stl-1** | WBGene00006061 | 226 | 3 | 3 | 7 | 9 |
| **F54E2.1** | WBGene00018823 | 210 | 6 | 3 | 5 | 2,4 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **K08D8.6** | WBGene00010660 | 200 | 5 | 3 | 4 | none |
| **gfi-1** | WBGene00001581 | 181 | 5 | 3 | 2 | 3 |
| **act-4** | WBGene00000066 | 169 | 4 | 3 | 8 | none |
| **dct-17** | WBGene00009431 | 225 | 4 | 2 | 3 | none |
| **lec-2** | WBGene00002265 | 187 | 4 | 2 | 10 | 1,4,5 |
| **eft-4** | WBGene00001169 | 162 | 4 | 2 | 9 | 1,5 |
| **vha-19** | WBGene00021952 | 148 | 2 | 2 | 2 | 3,5,7,9 |
| **act-4** | WBGene00000066 | 145 | 3 | 2 | 3 | none |
| **F53C11.1** | WBGene00009971 | 144 | 3 | 2 | 4 | none |
| **daf-21** | WBGene00000915 | 140 | 2 | 2 | 9 | none |
| **B0024.4** | WBGene00007097 | 138 | 3 | 2 | 4 | 5 |
| **Y54G2A.18** | WBGene00021883 | 134 | 2 | 2 | 2 | 1,6,7,8,10 |
| **vha-1** | WBGene00006910 | 130 | 3 | 2 | 1 | none |
| **F35E12.10** | WBGene00009434 | 125 | 3 | 2 | 5 | 1 |
| **T25B6.2** | WBGene00020788 | 107 | 6 | 2 | 5 | none |
| **Y40D12A.2** | WBGene00021503 | 107 | 4 | 2 | 9 | none |
| **tre-3** | WBGene00006609 | 89 | 3 | 2 | 3 | none |
| **npp-21** | WBGene00019940 | 85 | 2 | 2 | 8 | 2,3,5 |
| **Y12A6A.1** | WBGene00012439 | 79 | 3 | 2 | 10 | none |
| **vha-4** | WBGene00006913 | 181 | 2 | 1 | 1 | 2 |
| **pcp-1** | WBGene00003956 | 132 | 2 | 1 | 4 | none |
| **C18H7.11** | WBGene00044457 | 121 | 2 | 1 | 4 | none |
| **vha-2** | WBGene00006911 | 108 | 2 | 1 | 10 | 1,2,3,4,5,7,8,9 |
| **C12D12.1** | WBGene00015713 | 89 | 3 | 1 | 1 | none |
| **crn-6** | WBGene00000799 | 89 | 1 | 1 | 5 | none |
| **C02B10.3** | WBGene00015328 | 80 | 1 | 1 | 9 | none |
| **atp-2** | WBGene00000229 | 79 | 2 | 1 | 3 | 1,4 |
| **C18E9.6** | WBGene00007686 | 77 | 1 | 1 | 5 | 1,9 |
| **M116.5** | WBGene00019792 | 75 | 3 | 1 | 6 | 2,7,8,10 |
| **Y47H9C.1** | WBGene00012947 | 75 | 1 | 1 | 5 | none |
| **C48E7.1** | WBGene00016749 | 74 | 1 | 1 | 9 | none |
| **phb-1** | WBGene00004014 | 74 | 1 | 1 | 9 | none |
| **dnc-1** | WBGene00001017 | 67 | 3 | 1 | 6 | 7,9,10 |
| **nurf-1** | WBGene00009180 | 66 | 1 | 1 | 3 | none |
| **vha-16** | WBGene00016258 | 65 | 2 | 1 | 1 | 3,5,7,9 |
| **F23F12.8** | WBGene00017754 | 63 | 3 | 1 | 9 | none |
| **unc-54** | WBGene00006789 | 63 | 1 | 1 | 5 | none |
| **Y46D2A.2** | WBGene00021590 | 60 | 2 | 1 | 4 | none |
| **R02F2.9** | WBGene00019838 | 60 | 2 | 1 | 9 | 1,5 |
| **K04H4.2** | WBGene00010573 | 59 | 3 | 1 | 3 | 1,2,4,5 |
| **Y32H12A.8** | WBGene00021316 | 59 | 2 | 1 | 6 | none |
| **ftt-2** | WBGene00001502 | 58 | 1 | 1 | 8 | none |

| | | | | | | |
|---|---|---|---|---|---|---|
| **C03F11.3** | WBGene00015389 | 57 | 2 | 1 | 3 | none |
| **F23C8.6** | WBGene00017735 | 54 | 2 | 1 | 9 | none |
| **ZK896.4** | WBGene00014135 | 54 | 1 | 1 | 4 | none |
| **Y51A2D.15** | WBGene00013082 | 52 | 2 | 1 | 2 | 5,7,10 |
| **rpl-38** | WBGene00004452 | 52 | 2 | 1 | 7 | 2,5,10 |
| **lmp-1** | WBGene00003053 | 52 | 1 | 1 | 7 | none |
| **W03F9.10** | WBGene00021004 | 52 | 1 | 1 | 7 | 2,3 |
| **srab-6** | WBGene00016479 | 50 | 2 | 1 | 7 | none |
| **hmg-12** | WBGene00001977 | 49 | 1 | 1 | 6 | none |
| **F22E12.1** | WBGene00009058 | 48 | 2 | 1 | 6 | 3,7,8,10 |
| **F56E10.3** | WBGene00018977 | 48 | 2 | 1 | 6 | 8,10 |
| **K02H11.9** | WBGene00019350 | 48 | 2 | 1 | 6 | 7,8,10 |
| **Y54E10A.6** | WBGene00021828 | 48 | 2 | 1 | 6 | 7,8,10 |
| **hecd-1** | WBGene00016405 | 48 | 2 | 1 | 7 | 2,3,5,6 |
| **lys-1** | WBGene00003090 | 47 | 1 | 1 | 9 | none |
| **rpt-6** | WBGene00004506 | 46 | 2 | 1 | 6 | 7,8,10 |
| **T24C12.4** | WBGene00020766 | 46 | 1 | 1 | 1 | 5,9 |
| **C34H4.2** | WBGene00016425 | 46 | 1 | 1 | 2 | none |
| **Y105E8B.9** | WBGene00013693 | 46 | 1 | 1 | 2 | 9 |
| **aman-3** | WBGene00018594 | 45 | 1 | 1 | 3 | none |
| **T06D4.3** | WBGene00020292 | 45 | 1 | 1 | 8 | 2 |
| **C09E7.4** | WBGene00015638 | 44 | 2 | 1 | 10 | 2,5,6,7,8 |
| **C32E8.11** | WBGene00016326 | 44 | 1 | 1 | 4 | none |
| **F41G4.7** | WBGene00018310 | 44 | 1 | 1 | 4 | none |
| **pqn-38** | WBGene00004125 | 44 | 1 | 1 | 6 | none |
| **ajm-1** | WBGene00000100 | 44 | 1 | 1 | 9 | none |
| **R148.3** | WBGene00020102 | 44 | 1 | 1 | 9 | none |
| **hint-3** | WBGene00016150 | 42 | 2 | 1 | 2 | 5,7,10 |
| **Y54E10A.12** | WBGene00021832 | 42 | 1 | 1 | 2 | 5,10 |
| **col-171** | WBGene00000744 | 42 | 1 | 1 | 2 | 1 |
| **MTCE.16** | WBGene00010961 | 42 | 1 | 1 | 4 | none |
| **myo-2** | WBGene00003514 | 42 | 1 | 1 | 6 | 2,3,7 |
| **ZK973.1** | WBGene00022830 | 42 | 1 | 1 | 6 | none |
| **R03E1.2** | WBGene00010993 | 42 | 1 | 1 | 9 | 1,2 |
| **aptf-1** | WBGene00019424 | 42 | 1 | 1 | 10 | none |
| **H12I13.2** | WBGene00019191 | 41 | 2 | 1 | 6 | 9 |
| **F55F8.2** | WBGene00018890 | 41 | 1 | 1 | 1 | none |
| **F37A4.6** | WBGene00018136 | 41 | 1 | 1 | 5 | none |
| **lin-3** | WBGene00002992 | 40 | 2 | 1 | 5 | 2,6,7,8,10 |
| **F58H7.1** | WBGene00019067 | 40 | 2 | 1 | 5 | 2,10 |
| **H35N09.2** | WBGene00019266 | 40 | 2 | 1 | 7 | 8 |
| **F36H9.5** | WBGene00018113 | 40 | 1 | 1 | 2 | 3 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **M05B5.1** | WBGene00010869 | 40 | 1 | 1 | 2 | 3 |
| **T18D3.1** | WBGene00011820 | 40 | 1 | 1 | 2 | 5,7 |
| **Y73C8C.8** | WBGene00022265 | 40 | 1 | 1 | 5 | none |
| **F13D2.1** | WBGene00008735 | 40 | 1 | 1 | 6 | none |
| **F59A2.2** | WBGene00010302 | 40 | 1 | 1 | 6 | none |
| **CD4.8** | WBGene00016993 | 40 | 1 | 1 | 7 | none |
| **rfc-2** | WBGene00004338 | 39 | 2 | 1 | 1 | 2,4,5,8 |
| **C17G10.1** | WBGene00015915 | 39 | 2 | 1 | 2 | 1 |
| **D1037.1** | WBGene00017025 | 39 | 1 | 1 | 1 | none |
| **F22B5.10** | WBGene00009045 | 39 | 1 | 1 | 2 | none |
| **Y17G7B.17** | WBGene00012468 | 39 | 1 | 1 | 3 | none |
| **F54D5.11** | WBGene00010054 | 39 | 1 | 1 | 4 | none |
| **lev-8** | WBGene00002975 | 39 | 1 | 1 | 8 | 6 |
| **F54D10.3** | WBGene00018804 | 39 | 1 | 1 | 8 | none |
| **T09B4.4** | WBGene00020378 | 39 | 1 | 1 | 8 | none |
| **R166.2** | WBGene00011302 | 39 | 1 | 1 | 9 | none |
| **Y50D4C.2** | WBGene00021747 | 38 | 3 | 1 | 5 | none |
| **C44E4.4** | WBGene00016653 | 38 | 2 | 1 | 9 | 5 |
| **F26D2.16** | WBGene00009154 | 38 | 1 | 1 | 2 | 1,9 |
| **bicd-1** | WBGene00016611 | 38 | 1 | 1 | 3 | none |
| **dct-16** | WBGene00012615 | 38 | 1 | 1 | 3 | none |
| **cuc-1** | WBGene00000835 | 38 | 1 | 1 | 3 | none |
| **otpl-5** | WBGene00018478 | 38 | 1 | 1 | 6 | none |
| **F55C10.4** | WBGene00010108 | 38 | 1 | 1 | 6 | none |
| **T05H10.6** | WBGene00011510 | 38 | 1 | 1 | 6 | none |
| **unc-83** | WBGene00006815 | 38 | 1 | 1 | 6 | none |
| **mig-22** | WBGene00003253 | 38 | 1 | 1 | 8 | none |
| **F15E11.12** | WBGene00017498 | 38 | 1 | 1 | 10 | none |
| **Y41E3.8** | WBGene00012766 | 38 | 1 | 1 | 10 | none |
| **clp-1** | WBGene00000542 | 37 | 9 | 1 | 2 | 3 |
| **tag-273** | WBGene00013289 | 37 | 2 | 1 | 5 | 2,10 |
| **ZK795.2** | WBGene00014082 | 37 | 2 | 1 | 8 | none |
| **K09D9.9** | WBGene00019567 | 37 | 1 | 1 | 2 | none |
| **Y47G6A.17** | WBGene00021643 | 37 | 1 | 1 | 2 | 3 |
| **mig-1** | WBGene00003238 | 37 | 1 | 1 | 4 | 5 |
| **grl-12** | WBGene00001721 | 37 | 1 | 1 | 6 | 5 |
| **F31F4.1** | WBGene00017957 | 37 | 1 | 1 | 6 | none |
| **R74.6** | WBGene00011280 | 37 | 1 | 1 | 6 | none |
| **pink-1** | WBGene00017137 | 37 | 1 | 1 | 7 | none |
| **ntl-3** | WBGene00003826 | 37 | 1 | 1 | 7 | none |
| **C34C6.2** | WBGene00007915 | 37 | 1 | 1 | 8 | 6 |
| **tag-65** | WBGene00006442 | 37 | 1 | 1 | 8 | none |

| | | | | | | |
|---|---|---|---|---|---|---|
| **C32E12.4** | WBGene00016330 | 37 | 1 | 1 | 9 | none |
| **ins-14** | WBGene00002097 | 37 | 1 | 1 | 10 | none |
| **F44F4.10** | WBGene00009705 | 37 | 1 | 1 | 10 | none |
| **cyh-1** | WBGene00021714 | 36 | 2 | 1 | 7 | none |
| **egl-30** | WBGene00001196 | 36 | 1 | 1 | 1 | none |
| **F17H10.3** | WBGene00008927 | 36 | 1 | 1 | 6 | none |
| **ech-8** | WBGene00001157 | 36 | 1 | 1 | 7 | none |
| **zif-1** | WBGene00006977 | 36 | 1 | 1 | 9 | none |
| **ucr-2.1** | WBGene00012158 | 36 | 1 | 1 | 10 | none |
| **ugt-33** | WBGene00007946 | 35 | 2 | 1 | 2 | none |
| **C24A3.1** | WBGene00016032 | 35 | 2 | 1 | 5 | none |
| **srw-42** | WBGene00005789 | 35 | 2 | 1 | 9 | none |
| **ZK484.5** | WBGene00022751 | 35 | 2 | 1 | 9 | none |
| **fbxc-21** | WBGene00019042 | 35 | 1 | 1 | 1 | none |
| **Y34B4A.8** | WBGene00021324 | 35 | 1 | 1 | 2 | none |
| **ugt-42** | WBGene00017959 | 35 | 1 | 1 | 3 | none |
| **glb-24** | WBGene00011287 | 35 | 1 | 1 | 3 | none |
| **Y105E8A.23** | WBGene00013680 | 35 | 1 | 1 | 4 | none |
| **Y40C7B.1** | WBGene00021498 | 35 | 1 | 1 | 4 | none |
| **Y50F7A.2** | WBGene00021760 | 35 | 1 | 1 | 4 | none |
| **usp-14** | WBGene00006856 | 35 | 1 | 1 | 5 | none |
| **F19F10.5** | WBGene00017601 | 35 | 1 | 1 | 5 | none |
| **fcd-2** | WBGene00012767 | 35 | 1 | 1 | 6 | none |
| **T22F7.5** | WBGene00020704 | 34 | 2 | 1 | 3 | none |
| **ugt-64** | WBGene00015577 | 34 | 2 | 1 | 6 | none |
| **lsl-1** | WBGene00009937 | 34 | 2 | 1 | 7 | none |
| **C47F8.6** | WBGene00008162 | 34 | 1 | 1 | 1 | none |
| **srj-16** | WBGene00005604 | 34 | 1 | 1 | 2 | none |
| **mppa-1** | WBGene00022159 | 34 | 1 | 1 | 3 | none |
| **Y56A3A.31** | WBGene00013243 | 34 | 1 | 1 | 6 | none |
| **Y51A2D.7** | WBGene00013075 | 34 | 1 | 1 | 7 | none |
| **gon-4** | WBGene00001653 | 34 | 1 | 1 | 8 | none |
| **kel-8** | WBGene00020952 | 34 | 1 | 1 | 8 | none |
| **catp-7** | WBGene00022010 | 33 | 3 | 1 | 2 | 1,4,5,6,9,10 |
| **ztf-1** | WBGene00018833 | 33 | 2 | 1 | 6 | none |
| **W08G11.1** | WBGene00012346 | 33 | 1 | 1 | 2 | 9 |
| **dpy-22** | WBGene00001081 | 33 | 1 | 1 | 4 | none |
| **ZK355.5** | WBGene00022715 | 33 | 1 | 1 | 6 | none |
| **syg-2** | WBGene00007750 | 32 | 1 | 1 | 3 | none |
| **K09H9.5** | WBGene00019597 | 31 | 1 | 1 | 1 | none |
| **K10E9.1** | WBGene00019634 | 74 | 3 | 0 | 3 | 1,5 |
| **cogc-4** | WBGene00021784 | 55 | 2 | 0 | 2 | none |

| | | | | | | |
|---|---|---|---|---|---|---|
| **nhr-141** | WBGene00017787 | 54 | 2 | 0 | 8 | none |
| **pde-5** | WBGene00016328 | 51 | 2 | 0 | 3 | none |
| **acdh-11** | WBGene00012860 | 51 | 2 | 0 | 6 | none |
| **gei-6** | WBGene00001563 | 50 | 3 | 0 | 4 | none |
| **F39C12.1** | WBGene00018193 | 50 | 2 | 0 | 3 | none |
| **cdh-12** | WBGene00022103 | 50 | 2 | 0 | 5 | none |
| **W05B2.4** | WBGene00012272 | 50 | 2 | 0 | 8 | none |
| **B0524.4** | WBGene00015243 | 50 | 2 | 0 | 10 | none |
| **ztf-4** | WBGene00020399 | 49 | 2 | 0 | 4 | none |
| **acp-7** | WBGene00022246 | 49 | 2 | 0 | 7 | 5 |
| **unc-89** | WBGene00006820 | 49 | 2 | 0 | 9 | 1,5 |
| **ZK970.1** | WBGene00014171 | 49 | 2 | 0 | 10 | none |
| **psa-1** | WBGene00004203 | 48 | 2 | 0 | 5 | none |
| **qui-1** | WBGene00004265 | 48 | 2 | 0 | 5 | none |
| **map-2** | WBGene00003130 | 48 | 2 | 0 | 8 | 2 |
| **unc-68** | WBGene00006801 | 47 | 3 | 0 | 1 | 5 |
| **twk-30** | WBGene00006682 | 47 | 2 | 0 | 4 | none |
| **sdc-2** | WBGene00004746 | 47 | 2 | 0 | 6 | none |
| **prp-8** | WBGene00004187 | 47 | 2 | 0 | 7 | none |
| **F55F10.1** | WBGene00018898 | 46 | 4 | 0 | 5 | none |
| **C34C12.2** | WBGene00007921 | 46 | 2 | 0 | 4 | none |
| **C55A6.3** | WBGene00008332 | 46 | 2 | 0 | 4 | none |
| **R06C7.5** | WBGene00011064 | 46 | 2 | 0 | 6 | none |
| **cpna-2** | WBGene00015061 | 46 | 2 | 0 | 8 | none |
| **nsy-1** | WBGene00003822 | 45 | 2 | 0 | 6 | none |
| **C49F5.6** | WBGene00008210 | 45 | 2 | 0 | 10 | none |
| **rabs-5** | WBGene00021538 | 44 | 2 | 0 | 2 | none |
| **sdc-3** | WBGene00004747 | 43 | 2 | 0 | 3 | none |
| **larp-1** | WBGene00020097 | 43 | 2 | 0 | 3 | none |
| **lpd-3** | WBGene00003060 | 42 | 2 | 0 | 6 | none |
| **puf-5** | WBGene00004241 | 42 | 2 | 0 | 8 | none |
| **sdc-2** | WBGene00004746 | 41 | 2 | 0 | 4 | none |
| **ZK402.5** | WBGene00022731 | 41 | 2 | 0 | 6 | none |
| **Y110A7A.9** | WBGene00022459 | 41 | 2 | 0 | 10 | none |
| **C05C10.2** | WBGene00007329 | 40 | 2 | 0 | 6 | none |
| **B0207.5** | WBGene00015028 | 37 | 3 | 0 | 3 | none |
| **anc-1** | WBGene00000140 | 37 | 3 | 0 | 5 | none |
| **sma-1** | WBGene00004855 | 36 | 2 | 0 | 7 | none |
| **F59E12.9** | WBGene00019124 | 35 | 2 | 0 | 1 | none |
| **flr-1** | WBGene00001465 | 35 | 2 | 0 | 5 | none |
| **grl-14** | WBGene00001723 | 35 | 2 | 0 | 5 | none |
| **C07E3.3** | WBGene00007414 | 35 | 1 | 0 | 8 | 10 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Y57A10A.8 | WBGene00013253 | 34 | 2 | 0 | 3 | none |
| noah-1 | WBGene00016422 | 34 | 2 | 0 | 6 | none |
| col-76 | WBGene00000652 | 34 | 2 | 0 | 6 | none |
| F56H1.3 | WBGene00018994 | 34 | 2 | 0 | 8 | none |
| lfi-1 | WBGene00022500 | 34 | 2 | 0 | 9 | none |
| Y55F3BL.1 | WBGene00021935 | 34 | 1 | 0 | 2 | 5 |
| T12A2.8 | WBGene00020442 | 34 | 1 | 0 | 3 | none |
| F15D4.6 | WBGene00008863 | 34 | 1 | 0 | 6 | none |
| F31C3.2 | WBGene00009284 | 34 | 1 | 0 | 6 | none |
| lin-35 | WBGene00003020 | 34 | 1 | 0 | 7 | none |
| F26D11.2 | WBGene00017819 | 34 | 1 | 0 | 8 | none |
| Y51H7C.5 | WBGene00021783 | 34 | 1 | 0 | 9 | none |
| tag-233 | WBGene00044071 | 34 | 1 | 0 | 10 | none |
| K06A5.8 | WBGene00019434 | 34 | 1 | 0 | 10 | none |
| C49C8.3 | WBGene00016767 | 33 | 2 | 0 | 2 | none |
| duox-2 | WBGene00018771 | 33 | 2 | 0 | 3 | none |
| B0284.2 | WBGene00007132 | 33 | 2 | 0 | 9 | none |
| T28A11.20 | WBGene00020882 | 33 | 1 | 0 | 1 | 9 |
| T05A1.3 | WBGene00011454 | 33 | 1 | 0 | 3 | none |
| Y6B3B.3 | WBGene00012388 | 33 | 1 | 0 | 3 | none |
| C39F7.1 | WBGene00016538 | 33 | 1 | 0 | 4 | none |
| C47A4.1 | WBGene00008122 | 33 | 1 | 0 | 4 | none |
| gcy-6 | WBGene00001533 | 33 | 1 | 0 | 6 | none |
| twk-5 | WBGene00006660 | 33 | 1 | 0 | 6 | none |
| F47G9.4 | WBGene00009831 | 33 | 1 | 0 | 6 | none |
| try-2 | WBGene00006620 | 33 | 1 | 0 | 7 | none |
| T24B8.7 | WBGene00011980 | 33 | 1 | 0 | 7 | none |
| Y104H12D.2 | WBGene00022426 | 33 | 1 | 0 | 8 | 3 |
| Y38C1AB.4 | WBGene00021406 | 33 | 1 | 0 | 9 | none |
| F26F12.3 | WBGene00017834 | 33 | 1 | 0 | 10 | none |
| T27A8.3 | WBGene00012075 | 33 | 1 | 0 | 10 | none |
| ZK945.4 | WBGene00014166 | 32 | 3 | 0 | 2 | none |
| W02B12.10 | WBGene00012205 | 32 | 2 | 0 | 1 | 5 |
| Y56A3A.30 | WBGene00013242 | 32 | 2 | 0 | 1 | 10 |
| B0412.3 | WBGene00015173 | 32 | 2 | 0 | 5 | none |
| ric-4 | WBGene00004364 | 32 | 1 | 0 | 2 | none |
| glb-14 | WBGene00008996 | 32 | 1 | 0 | 3 | none |
| fbxb-34 | WBGene00021092 | 32 | 1 | 0 | 3 | none |
| asg-2 | WBGene00000210 | 32 | 1 | 0 | 4 | none |
| Y24D9B.1 | WBGene00021287 | 32 | 1 | 0 | 4 | none |
| C42C1.8 | WBGene00016586 | 32 | 1 | 0 | 5 | none |
| H11L12.1 | WBGene00019189 | 32 | 1 | 0 | 7 | none |

| | | | | | | |
|---|---|---|---|---|---|---|
| **K09F6.9** | WBGene00019592 | 32 | 1 | 0 | 7 | none |
| **M7.9** | WBGene00010885 | 32 | 1 | 0 | 8 | none |
| **Y37H9A.3** | WBGene00012578 | 32 | 1 | 0 | 8 | none |
| **Y71G12B.5** | WBGene00022145 | 32 | 1 | 0 | 8 | none |
| **tlf-1** | WBGene00006577 | 32 | 1 | 0 | 9 | none |
| **K03E5.2** | WBGene00019361 | 32 | 1 | 0 | 9 | none |
| **sulp-5** | WBGene00010789 | 32 | 1 | 0 | 9 | none |
| **xnd-1** | WBGene00001514 | 31 | 2 | 0 | 1 | none |
| **rsp-5** | WBGene00004702 | 31 | 2 | 0 | 1 | none |
| **Y6E2A.5** | WBGene00012399 | 31 | 1 | 0 | 1 | none |