Reverse genetics and functional analysis of an uncharacterized cysteine-rich protein coding gene family in Caenorhabditis elegans

Chaudhuri, Poulami

Awarding institution:
King's College London

Reverse genetics and functional analysis of an uncharacterized cysteine-rich protein coding gene family in *Caenorhabditis elegans*

Poulami Chaudhuri

Thesis submitted to King's College London for the degree of Doctor of Philosophy

September 2015

Acknowledgements

I owe my gratitude and appreciation to a lot of people who have supported me throughout. Firstly, I would like to thank my first supervisor, Professor Stephen Sturzenbaum without whom I would not have been able to fulfill my dream of doing a PhD. He has been an extraordinary guide, mentor and an advisor providing encouragement, motivation and support at every step. I would like to thank my charity funder who provided me the initial support to start my project and covered the fees for my first year.

I would like to thank my collaborators, Dr. Claudia Blindauer and her entire group based in Warwick University for permitting me to finish part of my work in their lab. I would especially like to thank Dr Tanvir Imam Hassan for his time and efforts. The protein bit of my thesis would not have been possible without Claudia's guidance and support. I would like to thank Phil Cunningham from King's College London for all the help provided for the bioinformatics analysis and for his unconditional support.

I would like to thank Dr Wolfgang Maret, my second supervisor for all his advice, Dr Matthew Arno for providing me the access to the high quality instrumentation required for my project. I would like to thank Dr Collin Dolphin, for his guidance in the lab and also his thought provoking questions. I would like to thank Dr Jill Kucab from King's College London and Okasana Leszczyszyn from Warwick University for lending me their protocols (Western blot and Silver staining respectively). I would like to thank Dr Shirley Coomber and Nic Bury for providing me with teaching opportunities which allowed me to cover my living expenses.

I would like to dedicate this thesis to my parents Dr. Subir Kumar Chaudhuri and Pradipta Chaudhuri for their unconditional love and support. I would like to thank my fiancé, Anirudh Nishtala for being patient and extending his support and love

throughout this journey. I would like to thank Dr Michael Ayers, for being my agony aunt, an extraordinary friend and an inspiration. I owe my sanity and happiness to all my friends and family.

I would also like to thank Jovaras Krasauskas and Moushumi Das for exploring possible aspects of the project as part of their Master's degree. I would like to sincerely thank Yu Nie for being a true friend and a companion, Yona Essig for her friendship and all her help. I would like to thank all the past and present members of Professor Stephen Sturzenbaum's group for providing me with support, love and a great environment to work. I would like to thank Shailaja Seetharaman and Nabeel Bhatti for their help in proof reading parts of my thesis and their constant encouragement.

Lastly, I would like to thank King's College London for providing me an outstanding platform for carrying out my research.

Declaration

This thesis is a result of my own work. Where ever contributions from others have been involved, it has been clearly stated. Collaborators have been acknowledged.

ABSTRACT

Caenorhabditis elegans is a multicellular androdioecious eukaryotic organism that serves as a model for developmental and genetic studies due to its completely sequenced genome. The application of reverse genetics methodologies such as RNAi and use of fluorescent tags to investigate gene expression makes this nematode an ideal model to explore metal toxicity.

A new family *W08E(12.2-12.5)* of cysteine rich (13%) protein coding genes containing four isoforms (*W08E12.2*, *W08E12.3*, *W08E12.4* and *W08E12.5* consecutively aligned on chromosome IV were investigated. *W08E(12.2-12.5)* are ≥ 90% similar in their coding region and the promoters of *W08E12.3* and *W08E12.4* display an identity of 100%. Occurrence of such highly similar genes consecutively in the compact worm genome highlights the importance of their study. Cloning and sequencing of the genomic regions of the isoforms confirmed the presence and order of all the four genes within the worm genome. Bioinformatic screening of the isoforms predicted the presence of putative metal binding sites similar to metallothioneins within the respective promoter regions. From the bioinformatics screening it was demonstrated that the existence of highly similar *W08E(12.2-12.5)* isoforms is a rare (0.6%) event and only highly essential genes required for the nematode survival occur at that frequency within the worm.

Transgenic-GFP tagged worms showed constitutive expression of P*W08E12.3/4::GFP* in the pharyngeal region which is regulated by metals (Cd, Zn, Cu). The transgenic worm metal exposure along with the qPCR assays of the metal exposed nematodes revealed strong metal responsiveness of the gene family. Knockdown by RNAi (*W08E12.3/4/5*) was used as a tool to characterize the phenotype of the gene family. The phenotypic studies determined that the growth of the nematode is affected (drastically in the presence of metals) upon knockdown of these genes. The RNAi metal exposed worms were further screened by X-ray laser imaging at Stanford Linear Accelerator Centre (SLAC), which revealed vast Zn accumulation in the pharyngeal area of the *W08E(12.3-12.5)* knockdown worms demonstrating the role of the gene family in metal (Zn) sensing/ homeostasis/detoxification.
In order to characterize the function at the protein level, the genes were sub cloned into an intein tag fused vector but the approach failed due to internal cleavage of the protein from the tag. The S-tag based PET vector was used as an alternative to express the W08E12.3 protein. Purification of the tagged protein enabled the determination of the metal binding properties of the protein by utilizing pH titration and metal displacement assays. The assays performed with the W08E12.3 protein depicted its strong Zn binding capability which correlates with the metal accumulation data. Taken together, it can be concluded that *W08E(12.2-12.5)* gene family is regulated transcriptionally by metals and is also metal binding in nature. Future studies of the *W08E(12.2-12.5)* family might add to ongoing efforts to develop enhanced biomarkers for metal toxicosis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1: General Introduction

1.1 Heavy metals and impact of heavy metals on humans

Metallic elements are important to all organisms, acting as structural elements, stabilisers, control mechanism components or activators. Some metals such as Mn, Fe, Co, Ni, Cu, Zn, and Mo are essential to the body and their deficiency can cause impairment of biological functions. But excess of these essential metals can also be toxic to the organism. Metals such as Cd, As, Hg, Pb do not have an essential function in the body and are toxic even in moderate levels (Goyer, 1997). Some of the toxic metals inhibit the uptake of essential metals and cause detrimental effects, for example Cd does not penetrate into the foetus to a considerable degree but blocks the uptake of Zn by the placenta leading to Zn deficiency in the foetus (Apostoli et al., 2007).

Heavy metals are a subset of elements including transition metals, lanthanides, actinides and some metalloids depicting metallic properties. Any toxic metal can be termed as heavy metal irrespective of its mass or density (Singh et al., 2011). Metals are the oldest toxicants known to humans (e.g. usage of Pb began in 2000BC). Metals differ from other toxicants as they are neither created nor destroyed by humans instead they are naturally redistributed by geological and biological cycles. Geological cycles include rain water physically transporting dissolved metals or ores into streams or rivers and depositing the sediments into the soil. The biological cycle includes bio concentration of metals by plants and animals into the food chain (Goyer and Clarkson, 1996). The utility of metals by humans poses potential health concerns at least in two significant ways, first via environmental transport (human or anthropogenic contributions to air, soil, water, food) and second, via altering the

speciation or biochemical form of the elements (Beijer and Jernelov, 1986). Natural weathering of the Earth's crust, mining, soil erosion, industrial discharge, urban runoff, sewage effluents, pest or disease control agents applied to plants, air pollution fallout and others can also expose humans to heavy metals (Ming-Ho, 2005).

Heavy metal contamination in the environment has raised concerns over global public health. Their wide distribution in the environment is due to their high demand in industrial, domestic, technological, medical and agricultural applications (Tchounwou et al., 2012). Although heavy metals are naturally occurring elements in the Earth's crust human exposure and environmental contamination is mainly due to anthropogenic activities including smelting and mining, domestic, agricultural use of metals (He et al., 2005). The major route of exposure to these toxic metals is through the diet (i.e. through food and water) (Morais et al., 2012), and also by inhalation or skin penetration (**Figure 1-1**). Health impacts from ingestion of heavy metals are often long term like diminution of mental acuity, loss of motor control, organ dysfunction, cancer, chronic illnesses, concomitant suffering and ultimately death (Siegel, 2002).

Humans are rarely exposed to pure metallic forms of metals or metalloids but are instead exposed to their oxides or salts. Transition metals mostly form coordination complexes with ligands (e.g. NH_3, CO, CN^-, organic S, and N). The physiochemical properties have consequences in terms of toxicokinetics, biological effects, determined by the chemical form of the metal known as the metal speciation (Nemery, 1990).

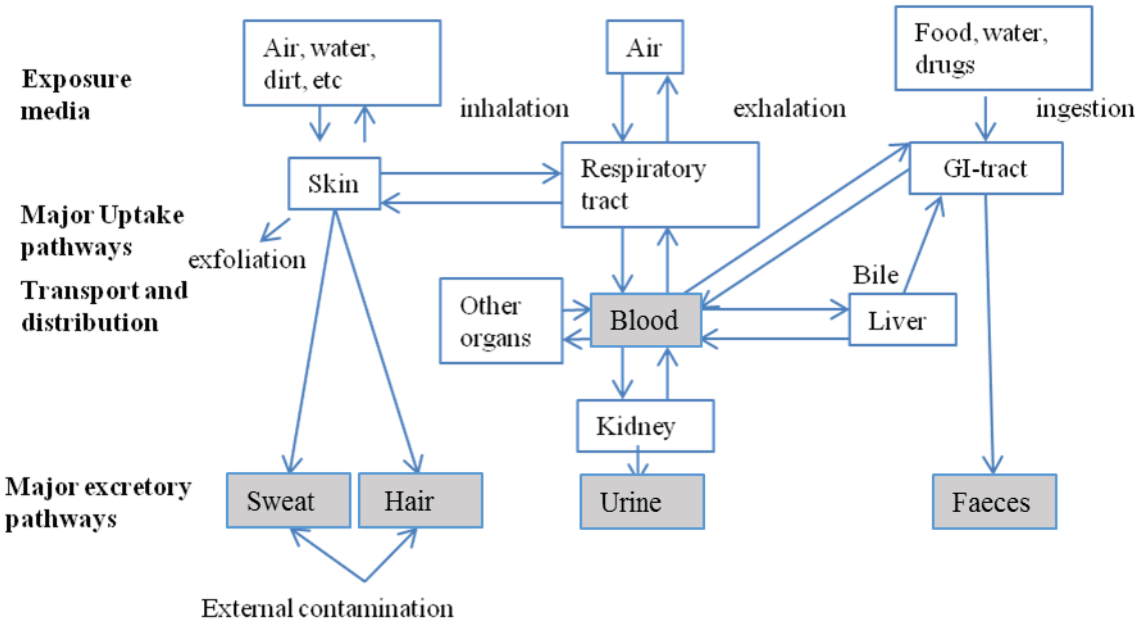


Figure 1-1: **Metabolism after exposure to metals via skin absorption, inhalation and ingestion.** The arrows describe the way metals are transported and exported while the grey boxes indicate potential biomonitors of metal contamination (adapted from Goyer and Clarkson, 1996).

One of the sources of metal exposure to humans have been the dietary uptake in the form of drugs. In India the three most prominent medical systems practiced are Ayurveda, Unani and Siddha. These involve intake of drugs constituted by plant, animal and mineral constituents. About 80% of the Indian population rely on Ayurveda. In the U.S.A. ayurvedic medicines are accessible from the internet, supermarkets and practitioners. As, ayurvedic medicines are marketed with the logo of dietary supplements they surpass the otherwise required proof of safety and efficacy. As and Hg are the most prevalent metals found in the ayurvedic medicines and since the 1970s approximately 80 cases of lead poisoning associated with ayurvedic medicine were reported (Saper et al., 2008). This serves as an example of how dietary consumption of toxic metals can lead to detrimental effects in humans (Singh et al., 2011).

Tobacco smoking is a major cause of Cd ingestion by humans (**Table 1-1**). Tobacco smoke contains prominent amounts of Cd. As absorption of Cd is far greater in the lungs than in the gastrointestinal tract, smoking contributes to be a major cause of Cd accumulation in the body (Yu, 2004, Figueroa, 2008).

Table 1-1*:* Heavy metals and their impact on human health with their permissible limits in the environment (Singh et al., 2011)

Pollutants	Major source	Effect on human health	Permissible level (mg/L)
Cd	Welding, electroplating, pesticides, fertilizers, Cd and Ni batteries nuclear fission plants	Renal dysfunction, lung cancer, bone defects, kidney damage	0.02
Zn	Metal plating, refineries, plumbing, brass manufacture	Zn fumes have corrosive effects of skin, nervous membrane damage	15
Cu	Mining, metal piping, pesticide production, chemical industry	Anaemia, stomach irritation, liver and kidney damage	0.1
Pb	Smoking, automobile emission, pesticide, smoking, coal burning, mining	Mental retardation in children, fatal infant encephalopathy, congenital paralysis	0.1

Toxicity and its threat to human health are subjected to the concentration of the metals but it is a known fact that chronic exposures of heavy metals and metalloids, even at low levels can result in drastic effects on human health (Castro-González and Méndez-Armenta, 2008). This has led to concerns in today's society about uptake of heavy

metals by humans, and mankind is keen to have a toxicant free environment (Figueroa, 2008).

1.2 Metal assessment

The amount of heavy metals exposed to humans has been defined by WHO to be the amount of substance in contact over time and space with the outer boundary of the body. To assess the amount of human exposure to contaminants present in the environment various methods can be utilised. One of the strategies being the determination of the chemical concentration scenario and the other, through the estimation of the amount of exposure of metals by use of biomarkers (Peterson, 2007).

Estimation of the amount of lead accumulation and its toxic effects over chronic exposure, by measuring lead in bones, teeth and hair serve as an example of an existing biomarker. With age, the amount of lead accumulation increases suggesting constant monitoring of lead in food is essential, otherwise chronic lead intoxication through diet might persist. Biomarkers act as early warning signals of effects at later stages in life (**Figure 1-2**), as the earlier changes in biological processes precede the effects seen at later hierarchical levels (McCarthy and Shugart, 1990).

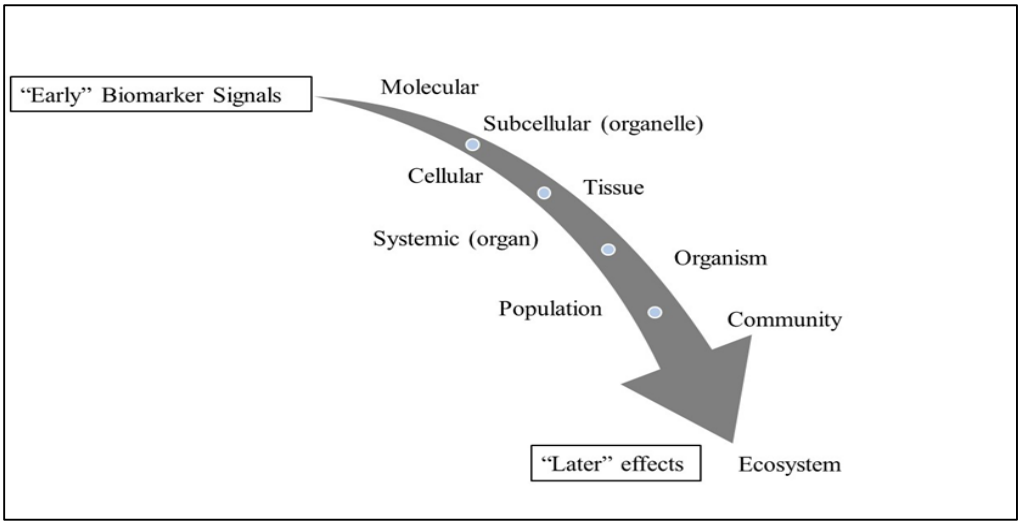


Figure 1-2: Schematic representation of sequential response order within biological systems to pollutant stress. A cascade of biological responses act as biomarkers in response to pollutant stress. Within a biological system, the responses at higher hierarchical levels are later response measures which occur when the entire environment is already impacted. The hierarchy of response pathway in a biological system begins at the molecular level and ends at the highest hierarchy i.e. the ecosystem (de Castilhos Ghisi, 2012).

Biomarkers have been described to be xenobiotically induced changes in components and processes at cellular and biochemical levels which can be measured in biological systems or samples. These changes provide an indication of the level of response incurred by the organism to the exposed contaminants defining the link between the presence of a substance and its ecological effect (Depledge et al., 1995). Biomarkers are measurements at cellular, biochemical or molecular changes in wild populations existing in contaminated source or measurement upon experimental exposure of organisms to a contaminated environment (McCarthy and Shugart, 1990). In the past few decades there has been an increase in the use of biomarkers to evaluate pollution hazards (Giesy et al., 1988, McCarthy and Shugart, 1990).

1.3 Metal homeostasis and metal binding proteins

Heavy metals (e.g. Zn, Cu) are essential to all living organisms for various physiological processes via metal dependent enzymes. These along with non-essential heavy metals (e.g. Cd, Pb, Hg) can be toxic when accumulated as they are highly reactive (Cobbett and Goldsbrough, 2002b).

Homeostasis can be disrupted by toxicity caused by metals as they generate oxidative stress, impair DNA repair mechanisms, inhibit enzyme activity, disrupt protein binding, and regular cellular functions including cell proliferation, progression of cell cycle and apoptosis (Stohs and Bagchi, 1995, Ercal et al., 2001, Beyersmann and Hartwig, 2008).

Plants and organisms have various inbuilt mechanisms to protect them from toxic metals, such as chelation and sequestering heavy metals. These particular mechanisms function by the action of specific ligands (Cobbett and Goldsbrough, 2002b).

Heavy metal stress is dealt with by exclusion, compartmentalization or synthesis of metalloproteins (Mejáre and Bülow, 2001). The two best characterised heavy metal ligands known in both plants and animals are metallothioneins (MTs) and phytochelatins (PCs). MTs are gene encoded cysteine rich proteins while PCs are enzymatically generated cysteine rich peptides (Cobbett and Goldsbrough, 2002b). Ferritin and transferrin are two other examples of specific metal binding proteins (Tamas and Martinoia, 2006).

Metal binding proteins have been a subject of major interest in the field of molecular biology of metal toxicology. Metal-protein binding interactions can be of a varied range. The protein may act as the target of toxicity for enzymes, or the protein may

have a protective role to play by reducing the activity/toxicity of the metal. The best example of a protein bearing role in protection against toxicity is metallothionein (Mejáre and Bülow, 2001).

Metallothioneins are low molecular weight proteins, abundant in thiol ligands and cysteine rich, playing a role in metal homeostasis and metal toxicity (Karin, 1985). All living organisms encode metallothioneins, which belong to a superfamily of intracellular metal-binding proteins displaying features like the archetypal MT first isolated from horse kidney and characterised by Margoshes and Vallee (Margoshes and Vallee, 1957). They were of particular interest due to their remarkable chemical structure that confers a degree of specificity, stability and dynamic behaviour almost impossible to predict from the properties of their organic and metallic ingredients (Coyle et al., 2002).

The high number of cysteines present in MTs bind to metals by mercaptide bonds. They generally comprise of two cysteine-rich metal binding domains conferring a dumbbell conformation to these metalloproteins. The MT proteins are broadly classified as Class I and Class II based on the arrangement of the cysteines. Class I MTs present in vertebrates contain 20 conserved cysteine residues based on mammalian MTs. Class II MTs do not involve a strict arrangement of the cysteines and are found in plant, fungi and non-vertebrate animals (Cobbett and Goldsbrough, 2002a).

The thiol rich ligands form the basis for high affinity binding of various essential and non-essential metals including Cd, Cu, Hg, Zn. Unlike other metalloproteins, metallothionein does not bind metals through finger or twist structure, rather it binds through two of its domains each of which is a metal binding cluster. MT binds to

several monovalent and divalent transition metals in its native form via its cysteine rich domains as it is devoid of histidine and aromatic residues in vertebrates (Chang et al., 1998). Among invertebrate MTs, snail metallothionein depicts class I MTs with no aromatic amino acids while the other studied MTs from terrestrial invertebrates differ a lot from the vertebrate MTs and might contain aromatic residues. Especially *C. elegans* MTs revealed to be distant to the other terrestrial invertebrates (Dallinger, 1996). Binding of metals by metallothionein might be transient as they are capable of releasing the metal ions rapidly. MTs perform three functions: Metal homeostasis, heavy metal detoxification and protection from oxidative stress (Martinez-Finley and Aschner, 2011).

The mammalian MTs are induced transcriptionally and biosynthesised by several factors like hormones, cytotoxic agents and metals predominantly Cd, Hg, Zn (Kaegi and Schaeffer, 1988). They are capable of interacting with the metals via complex physiological and biochemical pathways. The metallothionein has various functions including maintenance of cellular homeostasis by metal ion detoxification, scavenging Zn^{2+} from extracellular sources when intracellular Zn levels are low; it also redistributes the Zn within the cell. Zn has been proven to be the second most abundant transition metal found in organisms after iron (Dudev and Lim, 2001).

The mammalian metallothionein comprises of a β-domain and an α-domain. The β-domain consists of 9 cysteine residues sequestering either 3 Cd^{2+} or 3 Zn^{2+} or 6 Cu^{+} ions and the α-domain consists of 11 cysteine residues sequestering 4 Cd^{2+}, 4 Zn^{2+} or 6 Cu^{+} ions (Sutherland and Stillman, 2014). Metal chelation by the two domains of metallothionein is metal specific for instance Cd and Zn ions have proven to bind to the α-domain whereas Cu and Ag prefer the β-domain (Klaassen et al., 2009).

The β-domain is located at the N-terminus of the mammalian metallothionein while the α-domain is situated at the C-terminus. The binding of metals in the two domains initiates the formation of the four metal cluster followed by the three metal cluster in the α and β-domain respectively. The loss of metals has been shown to be initiated in the three metal cluster first as it is more labile in comparison to the four metal cluster. This phenomenon has been confirmed in the case of Cd binding to metallothionein where metal binding has been shown to be stronger and tighter in comparison to the three metal cluster domain (Klaassen et al., 1999).

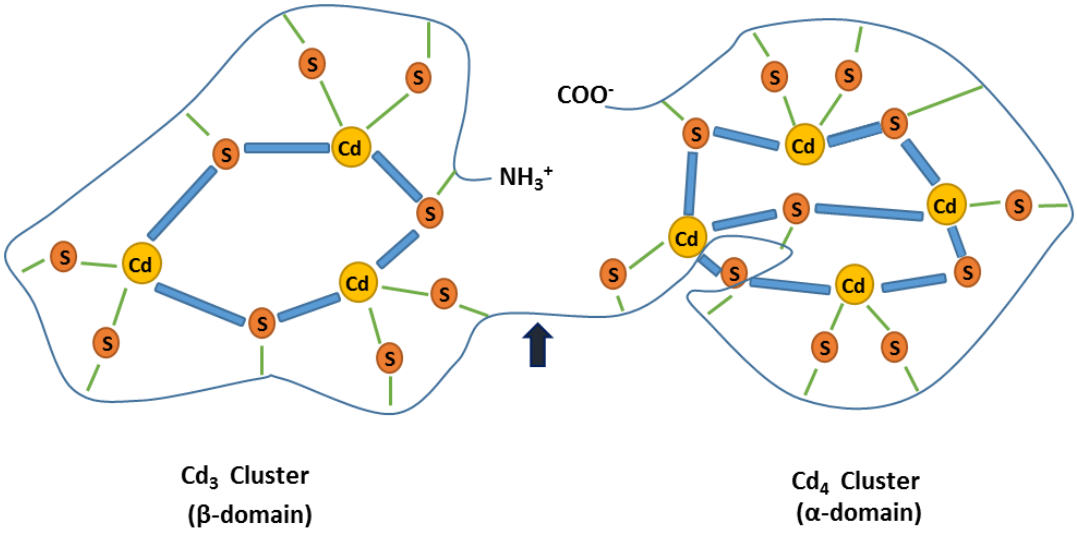


Figure 1-3: Structure of mammalian metallothionein depicting the chelation of Cd ions by the α and β-domain. The Cd atoms interact with the metallothionein domains forming a bicycle structure. The Cd 4-metal cluster is formed with 11 cysteine thiolate ligands (α- domain) and the Cd-3 metal cluster interacts with the 9 cysteine thiolate ligand forming a cyclohexane structure (β-domain) (modified and adapted from Klaassen et al., 1999).

Metal homeostasis is a system which controls the availability of the metal ions in the cells and also delivers them to their respective metalloproteins. During metal homeostasis the metal ions undergo three stages: protein mediated transfer across the bio membranes, chelation, binding to specific binding proteins (Tamas and Martinoia, 2006).

Protein metal binding sites are diverse and are attuned to their inorganic partners, but the risk of metalloproteins picking up the wrong metals when other metals are presented persists. Non-specific binding is avoided mainly due to various mechanisms involving conditions monitoring protein folding and compartmentalisation by the metallochaperons which dictate the binding of metal ions to their respective proteins (Xiao et al., 2004). The metal cofactors bind either directly to the unfolded protein assisting in folding of the protein or they bind to a well-structured cavity of a folded protein. The metal cofactors bound to the cavity of a folded protein are much more stable than the one bound to the unfolded protein for e.g. the Zn^{2+} complex in the protein cavity is much more stable than polyhedral Zn (Dudev and Lim, 2008).

Metal sensors, transporters and stores act as determinants of metal resistance. In addition to this property the sensors have shown to play a greater role in overcoming inadequate protein metal affinities to populate the number of metalloproteins with the correct metals (Waldron and Robinson, 2009) .

1.4 Cysteine domains in metal detoxification

Sulfur frequently appears in the form of cysteine, which performs a wide range of functions in proteins namely metal-binding, disulphide formation, redox-catalysis, electron donation and hydrolysis. Out of the above mentioned functions of cysteine, metal-binding, nucleophilic substitution, electron or proton transfer can be achieved only when it is present in its reduced form (Giles et al., 2003).

The sulfhydryl group of cysteine residues present in metalloproteins and peptides act as a major donor atom that binds to metal ions. Cysteine rich domains of the protein have sequestering properties towards some monovalent (Cu^{+}) and divalent metal ions (e.g. Zn^{2+} and Cd^{2+}) (Kulon et al., 2007). This metal sequestering property of the

proteins suggests that the genes encoding cysteine rich proteins might contribute towards metal homeostasis or detoxification.

Proteins containing Histidine or Cysteine rich domains bind efficiently to Zn and other divalent metal cations when they are in their reduced state (Potocki et al., 2011). Studies on the interactions of cysteine-rich sequences with metal ions are critical for understanding the possible biological function of the proteins rich in cysteines. Cysteine rich domains suggest the presence of metal binding sites or metal binding proteins.

Cysteine comprises of free –SH moieties which are the key molecules essential for the interaction with heavy metals immediately after their entry by cell transporters. Cysteine forms part of enzymes, metal binding peptides and structural proteins. It belongs to the proteins and peptides which contribute to the protective mechanisms of the cell against adverse effects caused by heavy metal ions due to the presence of a free –SH (thiol) group (Hynek et al., 2012).

1.5 *C. elegans*

Caenorhabditis elegans is saprophytic organism inhabiting soil, leaf litter and aquatic environments. The adult nematode has 959 somatic cells in the case of hermaphrodites and 1031 in adult males. The hermaphrodite produces up to 350 progeny. It has a life cycle of three days under optimal conditions. It is maintained in the lab on agar plates inoculated with *E.coli* as a food source. *C. elegans* is an androdioecious nematode (i.e. it is either a self-fertilising hermaphrodite or a male). Males occur in 0.05% of cases due to spontaneous non-disjunction in the germ line cells of the hermaphrodite. Hermaphrodites can be induced to produce males by exposure to higher temperatures.

Strains can be stored indefinitely by freezing in cryogenic solution (Jakubowski and Palczynski, 2007, Altun, 2015). It was Ellsworth C. Dougherty who introduced *C. elegans* as a model organism to Sydney Brenner in 1963. Brenner required an experimental organism suitable for genetics based studies in which the whole of the nervous system could be traced. The simplicity and consistency of the nervous system of the nematode made it a target for neurological studies rather than *Drosophila* which has a complex network of 10^5 neurons (Brenner, 1974). Sydney Brenner used the soil nematode *C. elegans* as a model organism to study development and neurobiology. Today this nematode is used to study various biological processes including cell death, signalling, cell cycle etc. The basic biology and medically relevant studies revealed conservation in molecular and cellular pathways between nematode and mammals. Further comparison disclosed that the majority of human disease genes and disease pathways are present in worms. Features like easy culturing, short generation time, rapid development (3 days) from egg to an adult, large progeny production and small size make *C. elegans* a powerful tool for research. Its transparent body permits the use of *in vivo* fluorescent markers as an easy tool to study processes such as axon growth, embryogenesis and fat metabolism within the living organism. In recent years *C. elegans* has become a tool for intense developmental and genetical studies. Its invariant cell division has permitted the tracking of the whole cell lineage from the fertilized zygote to the adult worm (Sulston et al., 1983). It is a simple multicellular eukaryote but depicts features like muscle, hypodermis (skin), intestine, reproductive system, glands, and a nervous system (Kaletta and Hengartner, 2006), features that make *C. elegans* an attractive model organism.

1.5.1 *C. elegans* life cycle

Like other nematodes *C. elegans* undergoes four larval stages (L1-L4) before reaching adulthood (**Figure 1-4**). The cycle from hatching of the egg to laying egg takes about 3 days and its life span is about 2 to 3 weeks. The end of each larval stage is marked with the shedding of a stage specific cuticle of the worm. The length of the life cycle of the nematode is temperature dependent. In the laboratory the worms are typically maintained at 20°C. The length of the life cycle can be altered by increasing or lowering the temperature to 25°C or 15°C (Klass, 1977).

During the embryo stage the proliferation and organogenesis/morphogenesis occurs. This includes cell divisions starting from one cell to 559 cells and the formation of tissues and organs. By the end of embryogenesis the main body plan of the nematode is formed. This does not alter significantly during post embryonic development.

L1 larva hatch from the egg 12 hours post fertilisation. The development of the nervous system, reproductive system, coelomocyte system begin during this phase and progress through L3 and L4 stages before transforming into an adult. At the end of L2 the worm can either enter L3 stage or might progress into a "dauer" phase. During the dauer state the worm experiences feeding arrest up to 4-6 months and its locomotion is reduced. The worm enters a dauer state under unfavourable growth conditions such as excess population density (which is detected by pheromones), the absence of sufficient food or high temperature. 1 hour after accessing food, the nematode exits dauer, resumes feeding and after 10 hour molts into the L4 larva. The L4 larva is discernible due to its vulva formation which appears as a white spot in the centre of the worm 40-48 hours post hatching. After entering the L4 stage, the worms self-fertilise, produce eggs and generate progeny by laying about 200-300 eggs over a 3 day period (Altun, 2015).

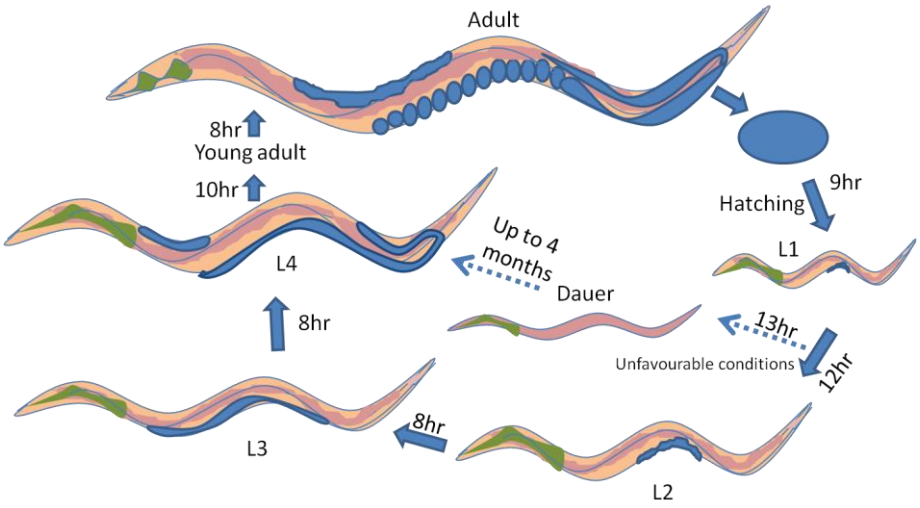


Figure 1-4: Schematic representation of *C. elegans* life cycle at 20°C. At 9 hours of post egg laying, the first L1 larva appears and progresses through L2, L3 and L4 stages before reaching adulthood. At L2 stage, if stressful conditions prevail, the larva enters dauer stage i.e. starved condition where it gains capability to remain without food intake up to 4-6 months. Once favourable conditions reappear the dauer larva molts into L4 stage and progresses to adulthood (Altun, 2015) **.**

1.5.2 *C. elegans* as a model for toxicology research

Nematodes are commonly found in aquatic and sediment environments including soil ecosystems. They participate in nutrient cycling and maintaining the quality of the environment (Barker et al., 1994). The above mentioned features supported the role of nematodes in ecotoxicological studies since the 1970s. *C. elegans* emerged as a favourable model in the 1990's as there was abundant knowledge about them for biological studies (Leung et al., 2008). *C. elegans* is extensively used in toxicity testing in aquatic media and soil (Cioci et al., 2000). As most of the environmental studies are based on aquatic medium and *C. elegans* thrive in terrestrial, benthic, and aquatic environments, ecotoxicological studies allow for a more relevant direct environmental comparison, making it an ideal biomonitor (Boyd and Williams, 2003b). Other than their ecological roles, *C. elegans* are ideal for laboratory based toxicity testing, as rapid toxicity tests with small soil volumes can be performed due to their short life cycle

and small size. They can be easily cultured in masses either on agar plates or in liquid media (Bongers, 1990). Until now, *C. elegans* has acted as an environmental biomonitor, determining the effects of chemicals on the wild type nematode survival (LC50), mutagenicity and development (Donkin and Dusenbery, 1993). It has also been suggested that *C. elegans* should substitute for vertebrates when determining mammalian acute lethality caused by metals (Williams and Dusenbery, 1988).

C. elegans is a simple organism offering tools such as reverse genetics and transgenic experiments which are more feasible and inexpensive than when conducted in more complex mammalian systems. It serves as a useful model to test the response of an *in vivo* chemical exposure on conserved pathways. It facilitates, by the use of whole organism assays, the study of functional multicellular unit (e.g. serotogenic synapse) instead of a single cell. Feeding, locomotion, lifespan, reproduction can serve as toxic end point assays such that chemical interaction with multiple target in the nematode can be investigated. Hence, *C. elegans* complements the use of both *in vitro* and *in vivo* mammalian models in toxicology (Leung et al., 2008).

C. elegans can act as a valuable toxicity model if the results obtained are predictive of the outcomes in higher eukaryotes. Research on *C. elegans* has provided sufficient evidence in this aspect both at a genetic and physiological level. Numerous physiological processes and stress responses have been observed to be conserved between the nematode and higher organisms. Bioinformatic data has showed that approximately 60% of *C. elegans* genes are homologous to humans and 12 out of 17 known signal transduction pathways are conserved between *C. elegans* and humans (Kaletta and Hengartner, 2006). Moreover, like other higher organisms, *C. elegans* have protective pathways to maintain metal homeostasis or to deal with metal toxicity via heat shock proteins (HSPs), metallothioneins (MTs), transporters and pumps (Sahu

et al., 2013) indicating that *C. elegans* has the potential to serve as a sensitive biomarker for diagnosis of metal (Cd) contamination by the use of responses to stress related gene expression. This makes it an ideal model to carry out metal toxicity research (Roh et al., 2006).

1.5.3 *C. elegans* and heavy metal toxicology/ detoxification

C. elegans has been used as a model to study toxicity and toxicological mechanisms of heavy metals such as Cd, Cu and/or Zn. Use of *C. elegans* as a model for human toxicity was first proposed in the context of heavy metals (Williams and Dusenbery, 1988). The heavy metals studies have been mostly focussed on various toxic endpoints including reproduction, life span, lethality and protein expression. There has been some focus on determining effects of metals on the nervous system by testing behaviour, reporter expression and neuronal morphology (Leung et al., 2008). Dealing with nematodes is economically more feasible in comparison to other models and the applicability of a wide range of genetic tools make it an ideal organism to study roles of specific genes and gene environment interactions. Hence the unique features of *C. elegans* make it an indispensable tool to complement mammalian models in toxicological research (Leung et al., 2008).

In conjunction with the above mentioned players in metal handling, there are genes encoding multi drug resistance associated protein (*mrp-1*) which is a homolog of the MRP gene in humans in conferring heavy metal resistance (Broeks et al., 1996).

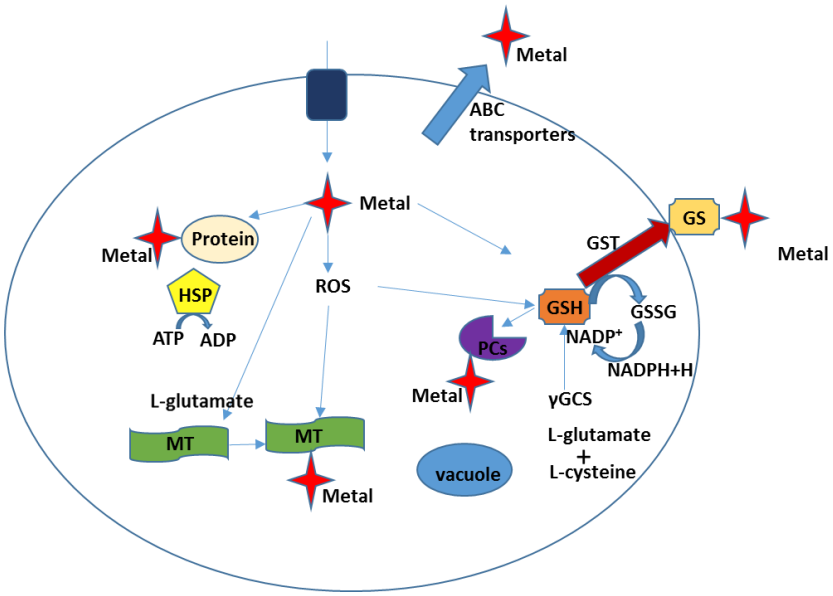


Figure 1-5: Metal detoxification systems in *C. elegans*. GSH converts to GSSG upon ROS exposure, γGCS acts as a rate limiting step for GSH synthesis. GSTs enable conjugation of GSH with the metal for excretion of the metal from the system. PCs are synthesised by GSH and act as metal chelators, whereas MTs directly bind and sequester the metal and act as antioxidants. ATP-binding cassette (ABC) transporters facilitate metal excretion and enable metal tolerance within the system. Heat shock proteins (HSPs) bind to ATP, convert them to ADP enabling binding to metals, other proteins and hence preventing metal aggregation within the system (modified and adapted from Martinez-Finley and Aschner, 2011).

Heat shock proteins (HSPs) are molecular chaperons preferentially synthesised in the presence of thermal stress but they can also be activated by other forms of stresses such as oxidative stress, heavy metal toxicity and anoxia (**Figure 1-5**). Under normal conditions HSPs have a role in protein folding and function (Lithgow and Walker, 2002). The stress induced, newly synthesised proteins maintain cellular homeostasis by enabling correct folding of stress accumulated misfolded proteins as well as the nascent proteins, thereby preventing protein aggregation or degradation of misfolded or denatured proteins. The amount of stress is directly proportional to the amount of

stress related protein (HSPs) production, which serves as a direct measure of toxicity (Gupta et al., 2010).

The ATP Binding Cassette (ABC) transporters are a widespread protein family engaged in active transportation of peptide ions, phospholipids and many other xenobiotic compounds across the biological membranes. Cytosolic metal bound chelators are transported into the vacuole with the help of these transporters. In *C. elegans* MRP-1, a member of the ABC transporter family has shown protection against Cd and As within the cells (Leslie et al., 2001).

Glutathione-S-transferases (GSTs) are found in every living species examined including plants, animals and bacteria (Eaton and Bammler, 1999). They play a major role in detoxification and excretion of many xenobiotic and physiological substances in the cell (Daulby et al., 1996). This family of multifunctional enzymes catalyse the association of reduced glutathione (GSH) with various electrophiles, enabling detoxification of exogenously and endogenously derived toxic compounds (Leiers et al., 2003). For toxic xenobiotics and carcinogens, glutathione S- conjugate formation depicts a detoxification pathway (Sugimoto, 1995). Glutathione (GSH) is a tri peptide, rich in thiol, involved in a multitude of cellular processes like detoxification of xenobiotics, defense against ROS and heavy metal sequestering. The reduced glutathione synthesises the heavy metal chelator phytochelatin (Yadav, 2010).

Phytochelatins (PCs) are a family of metal inducible thiol rich peptides which are synthesised enzymatically from GSH by PC synthase and play a prominent role in detoxification of heavy metals by complexing with metals acting as chelators. The PCs and toxic metal complexes are formed in the cytosol of the cell and eventually transported into the vacuole, protecting the organism from heavy metal toxicosis (Salt and Rauser, 1995). PCs in *C. elegans* are strongly activated in the presence of Cu but

yet play a major role in detoxification of only Cd and As. Indeed, they easily bind to Cu but do not sequester it in the vacuole, this suggests the existence of other more efficient pathways for Cu detoxification (Cobbett and Goldsbrough, 2002b). Metal chelation by PCs auto regulates biosynthesis of PCs where the activating metal is chelated by the product of the reaction, resulting in termination of the reaction (Ha et al., 1999).

Unlike PCs, metallothionein can directly bind and sequester toxicants by acting as antioxidants (Martinez-Finley and Aschner, 2011). Metallothioneins, have been identified throughout the animal kingdom, however their primary sequences are conserved only between the species related by phylogeny. There are significant differences observed between the invertebrates and it has not been possible to identify an overall consensus sequence. Irrespective of the dissimilarities, the common factor within the invertebrate metallothioneins remain the conserved short cysteine motifs which are essential for metal binding via thiolate cluster formation. This suggests they share analogous structure-function relationships. The nematode *C. elegans* metallothionein genes follow the same phenomenon (Swain et al., 2004).

C. elegans possess two metallothionein (MT) genes, *mtl-1* and *mtl-2,* which have been predicted to be the key players acting against metal toxicity. The metallothioneins in this nematode, like in other organisms, consist of many cysteines (25-30%) arranged in conserved motifs that are indispensable for thiolate cluster formation and responsible for metal binding. In nematode metallothioneins, the binding to metals takes place via tetrahedral coordination of the conserved cysteines.

Unlike other organisms the promoter region of *C. elegans* metallothionein lacks metal response elements (MREs) and the genome of the worm does not possess a metal

transcription factor (MTF-1) which is known to bind to the MREs and regulate metallothionein expression in other organisms. This suggests that the MTs of the *C. elegans* are regulated by some other, yet to be identified means (Swain et al., 2004).

C. elegans MTs have distinct functions, *mtl-1* is constitutively expressed in the lower pharyngeal bulb in the absence of metal exposure and thus may act as metal sensor, whereas *mtl-2* plays a role in metal detoxification and homeostasis as well as stress handling. Notably, *mtl-1* and *mtl-2* are both expressed in the gut region upon metal exposure. Both isoforms have distinct preferences for the metals for binding where, *mtl-1* is biased towards Zn(II) and *mtl-2* biased towards Cd(II) (Zeitoun-Ghandour et al., 2010).

The *C. elegans* MTs have two isoforms, CeMT1 and CeMT2 coding for polypeptides 75 and 63 residues in length, respectively. The isoforms display a conserved cysteine pattern (CeMT1 comprises of 19 cysteines and CeMT2 comprises of 18 cysteines) and encode aromatic amino acids (CeMT1 comprises 4 histidines, while CeMT2 comprises of 1 histidine) and a tyrosine distinguishing them from vertebrates. Both the isoforms preferentially bind to divalent metal ion, rather than Cu coordination and this binding efficiency is more pronounced for CeMT2 for Cd (Bofill et al., 2009).

CeMT-1 has 12 extra amino acids at its C-terminal in comparison to the CeMT-2. Zn is the primary constituent in these proteins like all other animal phyla, but is replaceable by Cd. CeMT-1 contains a higher amount of Zn in comparison to CeMT-2 suggesting functional differences and diversity in regulation of expression between them (You et al., 1999).

The CeMT-1 and CeMT-2 display differential binding behaviour towards different metals. CeMT-1 upon Zn binding shows optimal behaviour while CeMT-2 shows preference towards Cd coordination. CeMT-1 is constitutively expressed in the

pharyngeal region linked to Zn metabolism whereas CeMT-1 and CeMT-2 are induced upon exposure to Cd, performing a detoxification role. The presence of histidines serves critical for their coordination performance with metals, in CeMT-1 it leads to binding of the seventh Zn ion in comparison to the M(II)6-CeMT-2 complex when synthesised in the presence of supplemented Zn(II) or Cd (II). The Cu-thionine character is abolished due to the unique C-terminal histidine in CeMT-2 (Bofill et al., 2009).

1.5.4 *C. elegans* as a biomarker for metal toxicity

Understanding the mechanism of metal toxicity in mammalian systems has been rendered difficult due the complexity involved and also the reductionist approach that is inherent to the cell culture model. The nematode *C. elegans* is advantageous as a model to study toxicity as it is far less complex than mammalian systems but still has significant homology. The nematode shares 60% of human genes and encodes conserved regulatory proteins (Martinez-Finley and Aschner, 2011).

The free living nematode is a colonizer of microbe rich habitats mainly found in humus and plant decaying matter (Félix and Braendle, 2010) which confers potential to act as a bioindicator of environmental contamination. Due to the enriched presence in soil ecosytems, the convenience in handling under laboratory conditions and its sensitive stress responses, *C. elegans* is exploited in ecotoxicological studies due to its plethora of existence in soil ecosystems (Roh et al., 2009).

The completely sequenced genome of *C. elegans* makes it an ideal model to study functional relations between gene expression and the respective phenotypic response (Menzel et al., 2005, Reichert and Menzel, 2005). As natural resources (soil and water) can be used as exposure media for the nematode, it is a potential model to study

ecotoxicology (Boyd and Williams, 2003a, Peredney and Williams, 2000, Williams et al., 2000).

Nematodes are now being used as popular bioindicators of pollution stress. Due to its small body size, *in vivo* assays are possible in 96-well microplates; the transparent body allows imaging of cells in mature and developing nematodes. The different assays performed with the nematodes to estimate the amount of metal toxicity and its effects include behaviour, growth, reproduction, feeding, movement, lethality, stress related gene expression, metallothionein expression and *cdr-1* gene expression (Roh et al., 2006). The aforementioned assays can be exploited to study for both acute and chronic metal toxicity (Lund et al., 2002).

Moreover, methodologies like mutagenesis, RNAi, transgenesis, chromosomal deletions provide opportunities to manipulate and study the nematode at the molecular level unveiling information about increased sensitivity of mutant strains to metal toxicity (Jiang et al., 2009, Lund et al., 2002, Reichert and Menzel, 2005). Thus the nematode can be used in major fields of biomedical and environmental toxicology such as mechanistic toxicology, neurotoxicity and genotoxicity, high through output screening capability, environmental assessment and environmental toxicology (Leung et al., 2008) (**Table 1-2**).

Table 1-2: Different metal toxicity tests performed to assess the effects of different metals on *C .elegans* (adapted from Roh et al., 2006). The shading indicate the possible worm assay that can be performed using *C. elegans* as a model to detect toxicity conferred by the respective heavy metal.

Metals	Behaviour	Feeding	Movement	Growth	Reproduction	Lethality	Stress related gene expression	Metallothionein expression	CDR-1 gene expression
Cd									
Cu									
Zn									
Pb									
Cr									
As									
Al									

The translucent body of *C. elegans* makes it a valuable tool for researchers to visualise fluorescently labelled cells and proteins *in vivo* (Dupuy et al., 2007). This can be useful to investigate metal detoxification and is amenable for assessing the interaction between the genes and environmental factors. The role of the nematode as a biosensor of environmental risk assessment has been explored by exploiting the existing cellular systems like GSH, MTs, HSPs and other pumps and transporters involved in metal detoxification. In order to validate and develop *C. elegans* as model to understand metal toxicity, researchers are focussing on the relationship between different existing detoxifying systems (Martinez-Finley and Aschner, 2011).

1.6 Reporter Transgene

Microinjection within *C. elegans* was first carried out by Kimble and her colleagues (1982) (Mello and Fire, 1995) and later improved by Craig Mello (Mello et al., 1991). DNA is injected into the gonad of the worms and concatenates to form an extrachromosomal array which is incorporated into the nucleus. As the chromosomes in this nematode are holocentric, a part of the DNA acts like the centromere and the

extrachromosomal array is duplicated and distributed to the daughter cells. This method can be unstable during replication, and sometimes leads to mosaic expression of the transgene or over expression of the gene due to large number of copies being present, leading to possible toxic effects.

Using transposons to create in frame fusion of the transgene overcomes the above mention problems. After biolistic transformation or template directed repair, following transposon excision by homologous recombination, stable changes can be generated. *C. elegans* transposons, such as Tc1, are not used to introduce transposon mediated strand breaks because of the high copy number of the Tc1 transposon within the genome which would ultimately cause strand breaks at multiple sites.

Hence in order to introduce single copy transgenes, Mos1 transposable element from *Drosophila* is introduced into the nematode. The transposon introduces a break in the genome at a specified location where it does not interfere or disrupt the function of the neighbouring genes and also where other promoters or enhancers do not regulate expression of the transgene.

The technique is known as Mos1 excision-induced gene conversion (MosTIC). This system has demonstrated insertion of tags or caused deletions in specified genes. A large library of Mos1 inserts with defined locations in the worm genome has been created by NemaGENETAG consortium (Frøkjær-Jensen et al., 2008).

1.7 The *W08E(12.2-12.5)* Series

Based on a previous microarray experiment in the Sturzenbaum lab performed by Dr Suresh Swain, targeted to identify Cd responsive genes in *C. elegans* in a

metallothionein null background, a hitherto uncharacterized gene family was unveiled to be upregulated five-fold. The family comprised of four highly conserved genes encoding cysteine-rich proteins present consecutively on chromosome IV positioned on the forward strand at 3300824 bp to 3306638 bp in *C. elegans*. The genes are annotated *W08E12.2*, *W08E12.3*, *W08E12.4*, and *W08E12.5* based on their position in the Yeast artificial chromosome (YAC) genomic library cosmid as part of the whole genome sequencing effort.

All isoforms of the *W08E(12.2-12.5)* gene family revealed homology with uncharacterised genes present in other *Caenorhabditis* species like *C. briggsae*, *C. remanei* and also with keratin associated proteins (~65-75%) of *Homo sapiens* and *M. musculus*.

The human genome is 2.9 billion bp and the *C. elegans* genome in comparison is much smaller and compact (i.e. just over 100 million bp) with about 20,000 protein coding genes. This emphasises the importance of the presence of such highly identical genes in a compact genome and their role within the nematode. The abundance of cysteine in the proteins encoded by these isoforms hint towards heavy metal responsiveness or handling like metallothioneins (Stürzenbaum et al., 2001). Hence, the existence of four such highly conserved genes encoding cysteine rich proteins marks the importance of their functional investigation.

1.8 Aims and objectives

- This study is aimed to investigate the *W08E(12.2-12.5)* gene family at the transcriptional and translational level with an emphasis on *W08E12.3*. As

transgenic strains of nematodes can be exploited as biosensors to detect contamination by heavy metals in water and soil (Ma et al., 2009), the response of *W08E(12.2-12.5)* gene family to heavy metals at the transcriptional level will be unveiled by creation of transgenes, a single copy genome integrated transgene, P*W08E12.3/4*::GFP and an extrachromosomal array with multiple copies of the transgene, zsEx6 (P*W08E12.3/4*::GFP) worms.

- Transcriptional response of the *W08E(12.2-12.5)* will be individually quantified with the aid of qPCR.
- The phenotypic role of the *W08E(12.2-12.5)* within the worm will be investigated by exploiting reverse genetics tools like RNAi.
- The metal binding role of the *W08E(12.2-12.5)* family is aimed to be investigated by cloning the *W08E12.3* isoform in a pET29a (+) cloning vector, isolating and purifying the protein. Techniques like ICP-OES and ESI-MS would enable the estimation of the amount of metals sequestered by the W08E12.3 protein.
- Overall the study is designed to unravel the biological function of this hitherto uncharacterised family of genes and provide momentum to the field of toxicogenomics by identifying new biomarkers / rapid screening tools of toxic heavy metals.

Chapter 2: Materials and methods

2.1 Materials, reagents and solutions

Table 2-1: Source of materials, reagents and kits.

Supplier	Reagents
Promega	Wizard SV gel and PCR clean up system, Wizard plus SV miniprep DNA purification system, pGEM-T Vector System,T4 DNA ligase, GoTa qFlexi polymerase, restriction enzymes, DNA ladders, GoTaq® Long PCR Master Mix
Sigma-Aldrich	Kanamycin, Ampicillin, Nystatin, LB broth, LB agar, Cholesterol, NaOCl, $CaCl_2$, $CuCl_2$, $ZnCl_2$, qPCR primers and probes, primers, OligodTs, NP40, IPTG, L-arginine, Guanidine HCl, NaN_3, $MgSO_4$, ammonium per sulfate (APS), tetramethylethylenediamine (TEMED)
Fisher Scientific	Methanol, Iso-propanol, Chloroform, Ethidium bromide solution, 50x Tris-acetate-EDTA (TAE) buffer, NaOH, NaCl, NA_2HPO_4, K_2HPO_4, NuPAGE® LDS sample buffer (4x), NuPAGE® MES SDS running buffer (20x), Novex® Tris-Glycine SDS running buffer (10x), SeeBlue® Plus2 Pre-stained protein standard, 100 bp DNA ladder, 1Kb DNA ladder
EMD Biosciences, Inc.	BugBusterTM protein extraction reagent.
Invitrogen	DH5α competent cells, BL21 (DE3pLysS), NuPAGE system
Roche Diagnostics	Fast Start Universal Probe master mix (ROX)
BD Biosciences	Bacto Agar, Bacto Peptone
Bio-RaD Laboratories, Inc.	10% w/v SDS, Bovine Gamma Globulin standard Set

2.2 *C. elegans* maintainance

2.2.1 Nematode growth medium (NGM plates)

The nematode growth media (NGM) was prepared by autoclaving Bacto agar (1.7% w/v), Bacto peptone (0.25% w/v) and NaCl (0.03% w/v) made up to the required volume with distilled water. The autoclaved NGM was cooled to 60°C in a water bath. Under a laminar flow hood KH_2PO_4 (25 mM), $MgSO_4$ (1 mM), $CaCl_2$ (1 mM), Cholesterol (0.1% v/v) and Nystatin (0.1% v/v) were added. The molten agar was immediately dispensed into petri dishes of different sizes (e.g. for plates with a diameter 90 mm, 55 mm, 35 mm or 12 well plates, 20 mL, 10 mL, 5 mL or 2.5 mL were added respectively. Metals were added to the molten agar as desired. The plates were allowed to solidify under the hood and then stored at 4°C until required (but not beyond 3 weeks).

2.2.2 Food source

Bacteria (*Escherichia coli* OP50) was used as the food source for the nematodes. Being a Uracil auxotroph, OP50 selectively grows on NGM agar. The bacteria was provided by the *Caenorhabditis* Genetics Centre (CGC) and stored as a glycerol stock at -80˚C in a 50:50 of cells to glycerol. Under sterile conditions the bacteria were streaked on LB plates and a single isolated colony was used to make the fresh culture of OP50 in LB broth (10 mL). The freshly inoculated culture was grown overnight for about 16 hours at 37˚C at a speed of 200 rpm in a shaking incubator to facilitate growth of the bacteria. The culture was stored at 4˚C and replaced once per week.

2.2.3 M9 buffer

M9 buffer was used to wash the worms or their eggs from the NGM plates for sample collection or egg preparation respectively. M9 buffer was composed of KH_2PO_4 (22

mM), Na_2HPO_4 (42 mM), NaCl (85.5 mM) in distilled water and was autoclaved. Once the autoclaved solution cooled to room temperature $MgSO_4$ (1 mM) was added to the solution.

2.2.4 Seeding *E.coli* OP50 on NGM plates

NGM plates were spread with the OP50 suspended culture (90 mm plates – 200 µL; for 55 mm plates - 100 µL and for 12 well plates - 20 µL culture). When required, metals of appropriate concentrations were mixed with the OP50 before spreading with a plastic disposable loop on to the NGM plates. The culture was added under sterile conditions taking care that the culture is spread evenly on the plate a few mm away from the edge of the plate in order to avoid crawling of the worms onto the sides of the plates. Once the plates were inoculated they were dried under the hood and left overnight on the bench to allow the development of a bacterial lawn on the plate.

2.2.5 *C. elegans* culturing

Strains of *C. elegans* were grown on NGM agar plates with *E.coli* OP50 as their food source. The plates containing the nematodes were maintained in a 20˚C incubator. Three different methods were employed to transfer the nematodes from one plate to another i.e. worm picking, chunking and washing.

2.2.6 Worm picking, chunking and washing

Worms need to be transferred when OP50 bacteria is depleted to avoid starvation of the worms. Picking is mostly used to transfer single worms or to set up mating experiments, whereas chunking is used for routine stock transfer.

Picking of the worms required sterilizing the platinum wire pick mounted on a Pasteur pipette using a spirit lamp. Once the pick cooled down, a single worm was picked by focusing on a worm looking through a dissecting microscope. Care was taken not to

plunge the wire into the agar which would lead to cracks and promote burrowing of the worms.

Chunking of worms was performed by cutting a piece of agar containing the worms from one plate using a sterile scalpel and placing it face down onto a new OP50 seeded NGM plate. The old plate and the agar chunk were both discarded following the transfer.

Washing of the worms was performed using M9 buffer. The plates with the worms were repeatedly washed with M9 buffer using a pipette until all the worms were transferred from the plate into a 15 mL disposable plastic tube. The worms were allowed to settle down and the supernatant was removed and replaced with fresh M9 in order to remove most of the OP50 bacteria. Once the solution with worms appeared clear, the worms were counted and plated onto a new plate.

2.2.7 Egg preparation or bleaching

Egg preparation was performed mainly to synchronize the nematodes to the same stage of life cycle i.e. L1. Egg preparation also removes bacterial, fungal or yeast contamination from *C. elegans* stock plates. The bleach dissolves all worm tissues except eggs due to their strong shell covering; though over bleaching might also lead to killing of the eggs.

For egg preparation, gravid worms (mature, egg laying) were washed off the plate with M9 buffer into a 15 mL disposable plastic tube. The volume in the tube was made up to 11 mL with M9 buffer. The worms/eggs were allowed to settle for 5 minutes and then centrifuged at 1000 g for 2 minutes. The supernatant was discarded using a 1000 µL pipette leaving 2 mL containing the worm / egg pellet. Alkaline hypochlorite bleaching solution (10 mL) was prepared containing NaOH (1 M), NaOCl (4%) made

up to 10 mL using distilled water. This 10 mL bleaching solution was added to the worm/egg pellet in the tube and it was vigorously shaken for 2-2.5 minutes constantly followed by centrifugation for 1 minute at 1000 g. The supernatant was quickly removed and replaced with M9 buffer followed by centrifugation for 1 minute at 1000 g. This step was repeated four times in order to dilute the bleach. After the final wash, the supernatant was discarded and the tube was filled with 9 mL of M9 buffer. This tube was rotated overnight (or up to 48 hours) to allow the eggs to hatch into L1 worms.

2.2.8 Worm plating

The tube containing L1 larvae was removed from the rotator and the nematodes were allowed to settle by gravity for 5 minutes. The tube was centrifuged for 2 minutes at 1000 g followed by discarding all but 2 mL of the supernatant. The L1 larvae were then re-suspended and 4 µL of the supernatant was placed on a glass cover slip in order to count the number of worms/µL. The L1s were added drop-wise to the NGM OP50 spread plates under sterile conditions and dried under the laminar flow hood. The plates were then inverted and incubated at 20°C.

2.2.9 Worm freezing

L1s from a non-contaminated plate were washed off NGM plates with M9 and pelleted. The bacteria were removed from the suspension by carrying out 3-4 washes with M9 buffer. After the last wash, the volume in the suspension was reduced to about 2.5 mL. Freezing solution (100 mM NaCl, 50 mM KH_2PO_4 (pH 6), 30% (v/v) glycerol) (2.5 mL) was added to the 2.5 mL of M9 containing the pelleted worms. The solution was mixed thoroughly and aliquoted (1 mL) into five cryovial tubes. These tubes were placed in a polystyrene box, sealed with a tape and transferred to a -80˚C freezer to allow gradual freezing of the worms at the rate of 1˚C per minute. The

following day one of the tubes was removed, thawed at room temperature, plated on to a NGM agar plate and worm viability was assessed using a dissecting microscope.

2.3 *C. elegans* specific assays

2.3.1 Metal exposure and sample collection

Worms were exposed to various concentrations of Cadmium, Copper or Zinc in order to test the metal responsiveness of the genes of interest (**Figure 2-1**). The aim of metal exposure was either to monitor the effect of metals on the nematode by following its life cycle from L1 to adult stage or to expose the worms to various concentrations of the metals and then extract RNA from them.

The wild type (N2) worms and the metallothionein double knockout (*mtl-1*(tm1770); *mtl-2*(gk-125)) were exposed to a range of metal concentrations in order to extract their RNA and synthesise cDNA. The cDNA was utilised for subsequent qPCRs to quantify the metal specific responsiveness of transcripts. Exposing the nematodes to metals was achieved by adding metal ions in solution to the NGM agar prior to pouring the plates and to the OP50 solution before spreading. Three biological repeats of each condition were performed (**Table 2-2**).

GFP tagged transgenic worms were exposed to metals to evaluate changes in GFP intensity and also to determine the localization of expression (**Table 2-3**). The L4 staged worms were imaged under an inverted fluorescence Nikon microscope, using a blue laser scanning fluorescence (λ_{ex} = 450 – 490 nm); with exposure of 1/50 second and gain of 9/60 second.

Table 2-2: Metal concentrations used for nematode metal exposures for qPCR assays.

Metal	Concentration (µM) in NGM agar and OP50 solution
Cadmium Chloride ($CdCl_2$)	0, 5, 10, 20, 30
Zinc Chloride ($ZnCl_2$)	0, 50, 100, 200, 400
Copper Chloride ($CuCl_2$)	0, 50, 100, 200, 400

Table 2-3: List of metals and the concentrations used for metal exposure assays to the nematodes for quantitation of GFP intensity.

Metal	Concentration (µM) in NGM agar and OP50 solution
Cadmium Chloride ($CdCl_2$)	0, 10, 30, 50, 100, 150
Zinc Chloride ($ZnCl_2$)	100, 400, 600, 800
Copper Chloride ($CuCl_2$)	400, 600, 800

Sample collection: L4 stage worms were washed off with M9 from the plates and frozen at -80°C for RNA extraction. The tube containing the worms was then further washed with M9 by pelleting down the worms at 700 g for 1-2 minutes and replacing the supernatant with fresh M9 in order to remove the OP50 bacteria. This washing step

was repeated until the supernatant appeared clear. Once the worms settled, the M9 supernatant was discarded taking care not to disturb the worm pellet and the pellet was snap frozen in liquid nitrogen and stored at -80°C.

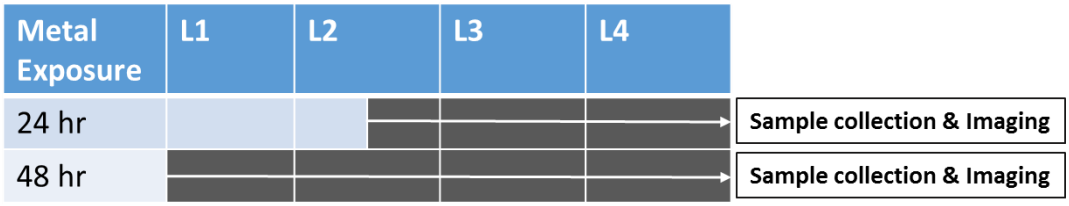


Figure 2-1: Summary of chronic (48 hours) and acute (24 hours) exposures to metals and the steps following the exposure. The worms were plated directly on the metal exposed conditions (L1-L4) for chronic exposure whereas, for acute metal exposure (24 hours) the worms were first grown until L2/L3 stage and then transferred to metal exposed conditions. The worms were collected and imaged at L4 stage for both acute and chronic exposure. The shaded region indicates metal exposed state.

2.3.2 Imaging

The transgenic GFP tagged worms were exposed to metal ions from stage L1 and imaged at stage L4. L4 worms (4-8) were picked onto a glass slide carrying a drop of M9 solution to which sodium azide (2 %) was added in order to immobilize the worms. Once the worms were immobile a cover slip was added taking care not to crush the worms. The worms were then imaged using an inverted fluorescence Nikon microscope, using a blue laser scanning fluorescence (λ_{ex}= 450-490 nm) and standard phase contrast imaging. The images were captured using a Nikon camera (Nikon UK Ltd., Kingston upon Thames, UK) at 10X magnification and analysed by dVision software or Image J software.

2.3.3 Worm plate reader

The transgenic P*W08E12.3/4*:: GFP tagged worms were exposed to metal ions from stage L1 to L4 stage, then washed off the plates with M9 into 15 mL tubes. At least

three M9 washes were carried out to remove the OP50 bacteria. 1 µL of octylphenyl-polyethylene glycol (IGEPAL) was added to the tube to prevent the worms sticking onto the pipette tips. Following a titre, the worms were aliquoted into a black 96 well microplate with three repeats of each condition. The microplate was read at 355 nm excitation and 510 nm emission wavelength. Following the plate reading, the worms were imaged using a blue laser scanning fluorescence (λ_{ex}= 450-490 nm) and standard phase contrast imaging in order to test the reproducibility between the bulk assay and single worm imaging. The worms were subjected to lysis in each well and protein content in each well was quantified using the Bradford assay. The protein concentration was used to normalise the fluorescence reading. A graph was generated based on the fluorescence reading and also based on the florescence reading/ protein content.

2.4 *C. elegans* toxicity-specific assays

2.4.1 Brood size

Worms were stage synchronised by performing egg preparation. The L1 worms were plated onto control and metal exposed NGM plates seeded with HT115 bacteria (containing either the empty RNAi vector ppD129.36 or ppD129.36 expressing *W08E(12.2-12.5*)). After two days, once the larvae reached L4 stage, single worms were picked into individual wells of a 12 well plate containing NGM supplemented according to the various conditions. Each worm was transferred each day at the same time onto a new well until the egg laying period was completed. The hatched progeny from each worm was scored once the progeny reached L4 stage. The worms were maintained at 20°C throughout the assay.

2.4.2 Volumetric growth and body length

L1 stage synchronised *C. elegans* were plated on control and metal exposed NGM plates seeded with HT115 bacteria (with RNAi (+/-) ppD129.36 vector). The change in their body area and length was recorded from day 2 of plating until day 6 (i.e. from stage L2 to adult). The growth changes were captured by taking digital pictures of the nematodes every day at the same time under the same magnification power of the microscope throughout the development period. The images were then analysed using the Image-Pro Express software. The scale for measurement was set by taking an image of a microscopic scale that was calibrated for 1000 microns.

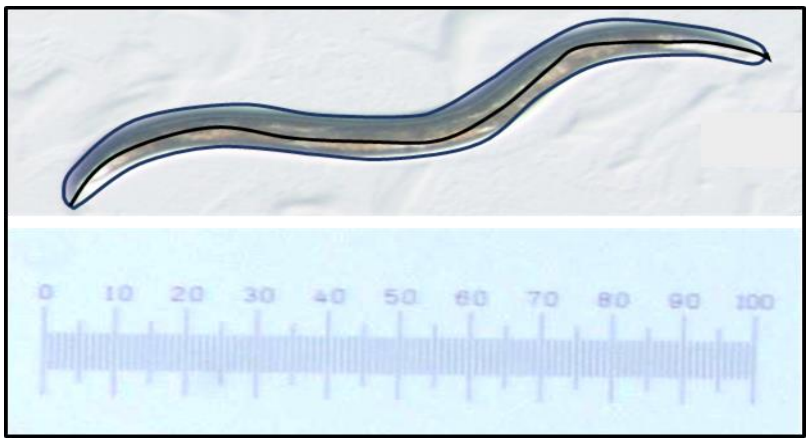


Figure 2-2: Example of body length and surface area measurement of an adult nematode using Image-Pro Express software. The black line indicates the tracing for measurement of body length whereas the blue line indicates the tracing of the body size measurement. The measurements were calibrated using a microscope graticule slide (1000 microns) as a reference within the software.

2.5 Molecular Biology methods

2.5.1 RNA extraction

Synchronized nematodes (1000-2000 worms/plate) were grown on NGM agar plates seeded with OP50 bacteria. The worms were grown to L4 larval stage (40-48 hours).

Once the nematodes reached L4 stage they were washed off the plates using M9 buffer and collected in a 15 mL tube (one tube/condition), then centrifuged at 700 g for 2 minutes. The supernatant was discarded without disturbing the pellet and replaced with fresh M9 buffer, this step was repeated in order to remove as much OP50 as possible. After 2-3 washes, the supernatant was discarded and the pellet was shock frozen in liquid nitrogen and stored at -80°C for at least 24 hours.

To extract RNA, the pellet was thawed on ice and sterile glass beads equal to the amount of the pellet were added, followed by addition of 500 µL Tri-reagent. This mixture was vortexed for 4 minutes on high speed to homogenise the pellet. Without disturbing the pellet, the supernatant was transferred into a 1.5 mL micro centrifuge tube and was allowed to stand at room temperature for 5 minutes. Chloroform (200 µL) was added to the tube and the mixture was shaken vigorously for 10 sec following incubation at room temperature for 12 minutes. The sample was then centrifuged at 12,000 g at 4°C for 15 minutes. The resulting supernatant was carefully transferred into a new 1.5 mL micro centrifuge tube without touching the interphase and 500 µL of isopropanol was added followed by inversion of the tube 5-6 times. The sample was then incubated for 12 minutes at room temperature, then centrifuged for 10 minutes at 12000 g at 4°C. The supernatant was discarded without disturbing the pellet and 1 mL of 75% ethanol was added to the tube following centrifugation at 12000 g (4°C) for 5 minutes. The supernatant was removed, the pellet was allowed to dry for 10-15 min, then resuspended in 30-50 µL RNAase-free water. The RNA concentration of each sample was determined using a nanodrop (ND-1000) Spectrophotometer and the quality of RNA was evaluated by agarose gel electrophoresis of the samples. Good quality RNA samples were reverse transcribed into cDNA by RT-PCR. The RNA

samples were stored at -80°C and were always kept on ice when used in order to prevent degradation of the RNA.

2.5.2 cDNA making

The cDNA was synthesised by reverse transcription of 1 µg of total RNA. The master mix contained M-MLV 5x reaction buffer (20% v/v), dNTPs (5 mm), oligo dT primers (100pmole/µL), RNA template (1 µg) and 1 µL reverse transcriptase enzyme (200units/ µL) made up to a volume of 20 µL using HPLC water. The reaction was placed into a PCR machine and the reverse transcription of RNA to cDNA was carried out at 42°C for 60 minutes followed by 72°C for 10 minutes and a final temperature of 4°C after completion of the reaction.

2.5.3 Genomic DNA extraction

Genomic DNA from the nematodes was extracted by performing single worm lysis. Worm lysis buffer was prepared using Tween-20 (0.1% v/v), KCl (25 mM), Tris (25 mM, pH 8.2), NP_4O (1.25 mM) and gelatine (0.005%). Worm lysis buffer (25 µL) was dispensed in a 0.2 mL PCR tube and 1.5 µL of proteinase K (1 µg/ µL) was added. 2.5 µL from the above mixture was taken into a pipette tip and a single worm was picked into a new PCR tube by the action of capillary force. A microscope was used to confirm the successful transfer of the worm from the plate into the tube. The tube with the worm was frozen at -80°C for at least 30 minutes, then loaded on to a PCR machine to carry out the lysis program (65°C: 60 minutes; 95°C: 10 minutes). Following the lysis, the solution was used to carry out nested PCR to amplify the target region of interest from the genomic DNA obtained from worm lysis.

2.5.4 Polymerase chain reaction (PCR)

Gene specific primers were designed in order to amplify the region of interest. The following conditions were followed to perform any PCR reaction:

Table 2-4: Components for a standard PCR reaction.

Component	Concentration in reaction
5x GoTaq Flexi buffer	1x
$MgCl_2$	2.5 mM
dNTP's	0.25 mM
Primers	0.1 µM each
GoTaq Polymerase	0.1 unit/µL
H_2O	Up to the reaction volume
DNA	<10 ng

The reaction mixture was aliquoted in 0.2 mL PCR tubes, centrifuged and carried out in a thermal cycler (PTC 225 Peltier Thermocycler, MJ Research, USA) programmed to include the optimal conditions based on gene specific primers. A standard protocol where X°C is the annealing temperature and Y°C the extension time of the primers would be programmed as follows: (**Table 2-5**)

Table 2-5: Temperature settings for a standard PCR reaction.

Temperature	Time
94°C	2 minutes
94°C	30 seconds
X°C	30 seconds
72°C	Y
72°C	5 minutes
4°C	∞

2.5.5 Nested PCR

Nested PCRs are performed to increase the amount of amplicon generated without a loss in the specificity. Nested PCR was performed on 1 µL genomic DNA extracted from the worms by the single worm lysis protocol.

External PCR: Combinations of primers outside the region of interest were used to carry out external PCR of the worm genome.

Internal PCR: Following external PCR, internal PCRs were set up using 1 µL of the external PCR product or its 1:10/1:100 dilutions with combinations spanning the specific region of interest with gene specific/ region specific primers (**Figure 2-3**).

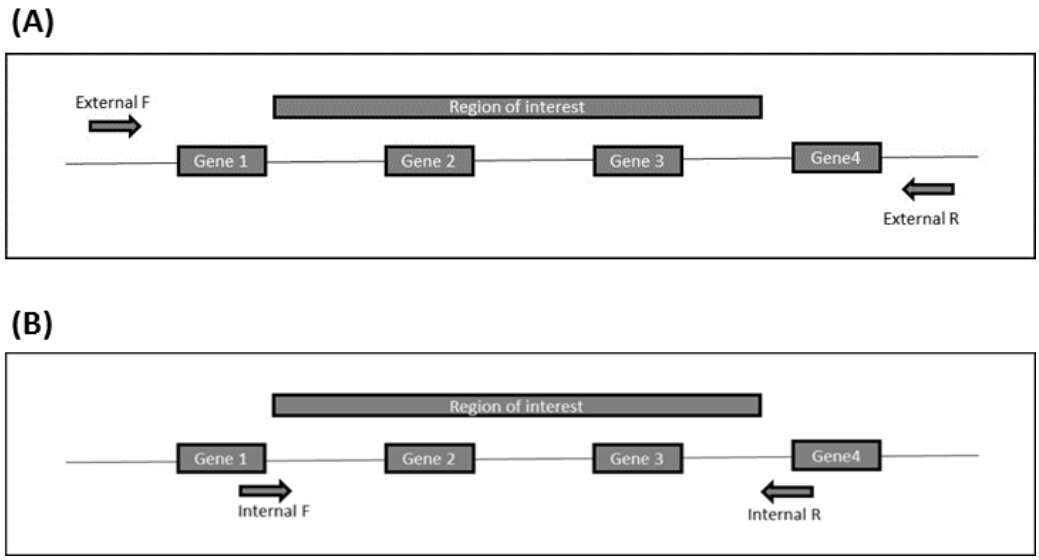


Figure 2-3: Schematic representation of an external PCR (A) step in a nested PCR and the following internal PCR (B).

After the completion of nested PCR, the PCR products were separated on an agarose gel for 40 minutes at 120V and visualized under a UV transilluminator. The bands of expected size were excised and used for cloning.

2.5.6 Agarose gel electrophoresis

For RNA/cDNA/PCR product analysis (samples smaller than 1000 bp) a 1x TAE, 1% agarose gel was prepared containing ethidium bromide (0.05 µL/ mL v/v). To visualize fragments smaller than 200 bp, a 1x TAE, 2% agarose gel was prepared. Sample: loading buffer were mixed in a ratio of 1:6 and loaded onto the gel along with 4-5 µL of either a 1 Kbp or 100 bp marker based on the sample size being analysed. The gel was run at 120V for 30-40 minutes and photographed under UV transilluminator camera (Genesnap Syngene) and reader (Synergy HT-1).

2.5.7 Molecular cloning

Molecular cloning was performed on both PCR amplified cDNA and genomic DNA fragments. The main aim to clone the cDNA copy of the four *W08E(12.2-12.5)* genes of interest was to test their primer specificity and also to clone them in a recombinant expression vector system for translational studies. The cloning of the nested PCR into a vector system would enable their sequencing.

Molecular cloning involved various steps:

PCR: Polymerase chain reactions of cDNA (*W08E12.3*) or genomic DNA with their respective primers were carried out.

Gel electrophoresis and purification: The PCR fragments were separated on an agarose gel along with a marker (100 bp marker- cDNA; 1 Kb- genomic DNA). The gel was visualized under UV light and the bands of expected sizes were excised with a disposable sterile scalpel and the gel fragments were transferred into 1.5 mL microcentrifuge tubes. The gel fragments were weighed and purified using the gel purification kit by Promega.

Ligation: Once the fragments were purified, a ligation reaction was set up utilising a suitable vector system. For example when a pGEM-T vector was used the following ligation reaction was set up in a microfuge tube:

Ligation reaction: pGEM-T vector = 50 ng

2X Rapid ligation buffer = 5 µL

T4 DNA ligase (3Weissunits/ µL) = 1 µL

PCR product = X µL

(ng of insert) $X = \frac{\text{ng of vector} \times \text{Kb of insert} \times \text{insert: vector molar ratio}}{\text{Kb of vector}}$

The ligation mix was centrifuged and was incubated at 4°C overnight.

Transformation: The ligation mix (5 µL) was used to set up a transformation of the ligated vector and insert into DH5α competent cells. The ligation mix was added to 30 µL competent cells and left on ice for 10 minutes. After 10 minutes the transformation mix was incubated at 42°C for exactly 30 seconds and was immediately returned on ice for 10 minutes. Following the incubation, 400 µL of pre-warmed LB broth was added to the transformation mix and the tubes containing the mix were incubated on a shaking incubator at 37°C for 1 hour to promote cell multiplication.

200 µL of the transformation mix containing the cells were then spread on to 55 mm LB amp agar plates and incubated at 37°C for about 18 hours.

Plasmid extraction: The following day, individual colonies were picked from the plates with a sterile pipette tip and inoculated into a 15 mL tube containing 5 mL of LB amp broth. The tubes were then incubated on 37°C shaking incubator for 24 hours. The plasmids were isolated using the Promega Mini prep kit and subjected to PCR with universal primers (e.g. SP6 and T7) or gene specific primers to confirm the

presence of the insert in the isolated plasmid. The PCR products were separated on a 1% agarose gel along with a marker and visualised under UV light. The plasmids which displayed the expected size of fragment of interest were chosen and sent for sequencing. In cases where large numbers of colonies were pre-screened, individual colonies were picked with a sterile pipette tip and inoculated into a 1.5 mL microcentrifuge tube containing 100 µL of sterile distilled water and the same tip was inoculated into the 15 mL tube containing LB amp broth. The 15 mL tube was incubated at 37°C in a shaking incubator whereas the 1.5 µL microcentrifuge tube was incubated at 100°C on a heating block for 5 minutes. The tube was then centrifuged for 1 minute at 10,000 g and the supernatant was used to set up a PCR with SP6 and T7 primers. The PCR mixture was composed of $MgCl_2$ (0.25 mM), 10% 5x GoTaq flexi buffer from Promega, the SP6 and T7 primers (0.1 µM), dNTPs (0.25 mM), GoTaq polymerase (0.1 units/µL) made up to 20 µL reaction volume with the supernatant produced. The PCR was programmed as:

95°C: 2 minutes

95°C: 45 seconds
56°C: 45 seconds
72°C: 30 seconds
} 30X

72°C: 5 minutes

The PCR products were separated on a 1% agarose gel for 40 minutes at 120V and visualized under UV light. Only the products which produced the bands of correct size were further processed**.**

2.5.8 TaqMan^{R} quantitative polymerase chain reaction (qPCR)

Quantitative PCR of each gene of interest was performed using an ABI Prism7000a platform (Applied BioSystems, Warrington, UK) qPCR machine. Each gene was quantified using *rla-1* which is an invariant ribosomal housekeeping gene in *C. elegans*. The primers and probes were designed based on the *C. elegans* genome database (www.wormbase.org) using the Universal Probe Library (UPL) assay design centre by Roche Life Sciences. Each qPCR reaction comprised of 2X ROX master mix (8.5 µL), 10x diluted cDNA (1000 ng), forward primer (0.4 µM), reverse primer (0.4 µM), probe (0.25 µM).

The standard ABI PRISM cycling conditions were used:

(50°C: 2 minutes, 95°C: 10 minutes)

95°C: 15 seconds
60°C: 60 seconds
} 40X

C_t values were determined and data analysed using the ABI PRISM software. The gene fold changes were calculated based on the 2-ΔΔ CT formula.

2.5.9 Gene silencing by RNA interference (RNAi)

The RNAi feeding procedure was followed as described Kamath et al. (Kamath and Ahringer, 2003). The RNAi clones were present in the host HT115 bacterial strain stored in the RNAi feeding library for *C. elegans*. The respective bacterial culture for RNAi feeding were grown by picking clones from the RNAi feeding library and growing them in LB/ amp (100 µg/mL) broth overnight at 30°C. Ampicillin (50 µg/mL) and Isopropylthiogalactoside (IPTG, 1 mM) were added to NGM agar just prior to pouring. Plates were poured 4-7 days before seeding to allow complete drying.

The bacterial cultures were inoculated onto the NGM (IPTG/amp) plates and grown overnight at room temperature (**Figure 2-4**).

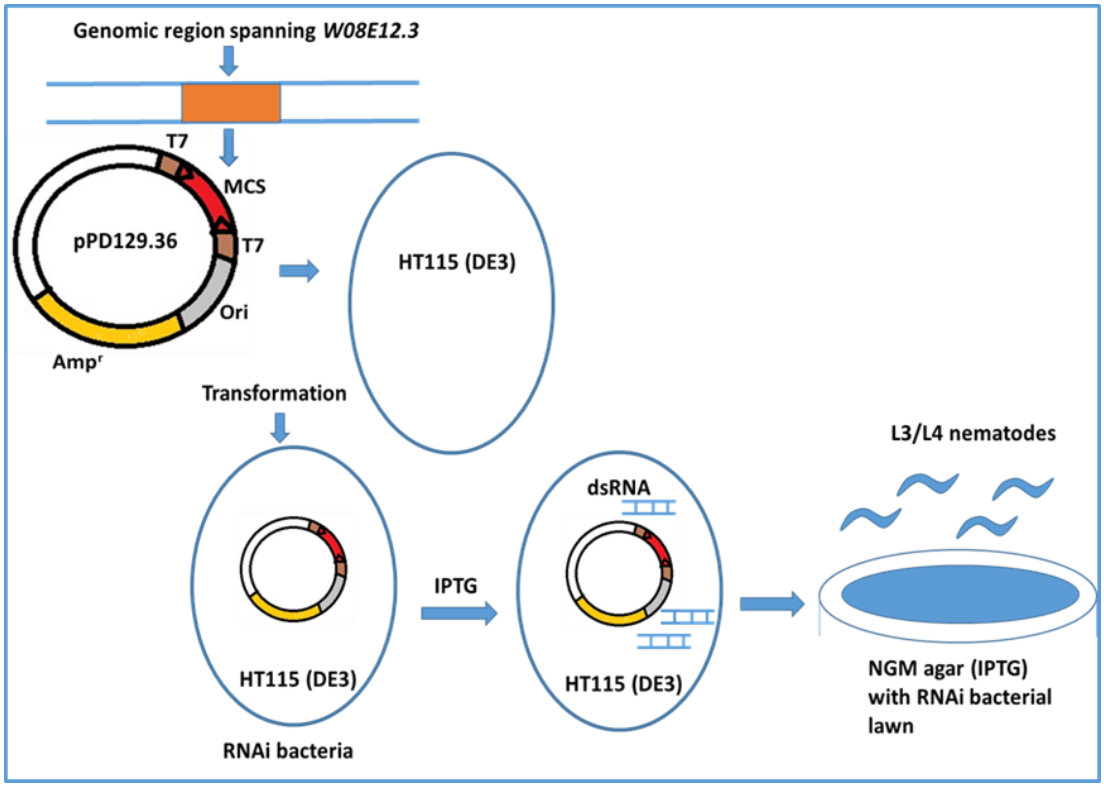


Figure 2-4: Schematic representation of the cloning of a fragment of interest into the RNAi vector (pPD129.36) and RNAi feeding process of the nematodes. The genomic fragment comprising of W08E12.3 was cloned into the multiple cloning site of pPD129.36 RNAi vector and transformed into HT115 (DE3) cells. The worms at L3/L4 stage were exposed to NGM agar containing IPTG with a lawn of the RNAi bacteria grown overnight at 37°C.

Wild type (N2) worms were synchronised by bleaching and L1 worms were plated on OP50 inoculated NGM plates. The worms (L2/L3 stage) were washed from the plates and cleaned three times with M9 buffer to remove any remaining OP50. The nematodes (200-500 worms) were plated on to NGM (IPTG/amp) RNAi bacteria inoculated plates. When the nematodes reached L4 stage, 20 worms were picked on to new NGM (IPTG/amp) inoculated plates (four repeats per condition) and were allowed to lay their first batch of eggs (F1). The adult nematodes were removed and

the F1 generation was allowed to grow at 20°C and lay eggs (F2). The F2 generation was synchronised by egg preparation. The percentage knockdown for the gene of interest was assessed by performing a PCR of the gene of interest and a house keeping gene (*rla-1*) in wild type and the RNAi exposed worms. Once the knockdown was confirmed the worms were ready for any worm assay.

2.6 Protein-Biochemistry methods

2.6.1 Recombinant expression of the *W08E12.3* protein

Plasmids containing the coding sequence of *W08E12.3* gene were transformed into (BL21 (DE3plysS)) bacteria as described in the cloning section. A single colony was picked and grown in LB broth with specific antibiotics approximately for 8 hours at 37°C. This starter culture was used to inoculate a larger volume of LB broth with antibiotics in the ratio of 1:10 dilution for large scale protein purification. The cultures were grown to OD_{600} 0.6 and then induced with Isopropylthiogalactoside (IPTG, 1 mM). After 30 minutes of induction, $ZnCl_2$ (100 µM) was added to the culture and grown for 5 hours at 37°C with shaking. After each hour 1 mL samples were collected in order to check the protein expression by SDS-PAGE and Western blotting. After induction, the cells were harvested by centrifugation at 1000 g for 20 minutes. The pelleted cells were frozen at -80°C.

2.6.2 Purification of the *W08E12.3* protein

The first attempt to purification was carried out using the IMPACT system (*W08E12.3* in pTXB vector). The protein was recombinantly expressed in the expression vector. After weighing out the cell pellets, they were lysed using BugBuster™ and the extracted cell lysate was processed based on the instructions provided by the manufacturer (NEB).

In the second attempt the purification of *W08E12.3* was carried out in a pET29a vector (containing either a neutral N-terminal S-tag or no tag). The protein was recombinantly expressed in BL21 (DE3plysS) cells and the pellets were frozen at -80°C. In detail, the frozen cell pellets were resuspended in 15 mL resuspension buffer (200 mM Tris-HCl, pH 7.5; 150 mM NaCl, 1 mM DTT, 0.02% Sodium azide, 0.1% Triton x). A tablet of Protease inhibitor EDTA free cocktail (Roche) was added and the pellet was resuspended. The cell suspension was processed in a cell disruptor (20.3 PSI) and the cell lysate was centrifuged at 12,000 g for 30 minutes on a floor centrifuge.

In the first attempt, the soluble fraction was utilised to purify the W08E12.3 protein by classical chromatographic techniques but, as the protein was detected by Western blotting technique to be predominantly in the insoluble (inclusion body) fraction, the soluble proteins were removed by discarding the supernatant obtained. The pellets were then resuspended in 10 mL resuspension buffer and loaded in a cell disruptor (20.3 PSI). The resuspension was then centrifuged at 12,000 g for 15 minutes to collect the washed inclusion bodies. The supernatant was decanted and the pellet was resuspended in 5 mL of resuspension buffer containing 6 M urea and incubated on ice for 1 hour in order to dissolve the protein (1 mM Zn acetate was added during the incubation) from the inclusion bodies. To the cell suspension 10 mL of resuspension buffer (without urea) was added to reduce the overall concentration of Urea to 2 M in the solution and centrifuged at 12,000 g for 30 minutes to remove any insoluble material. The pellets from the samples after the urea wash were processed by redissolving the pellet in 5 mL (0.1 M Tris HCl buffer, 6 M Guanidine HCl, 0.1 M DTT, pH 7.5) and the suspension was left on an orbital shaker at 4°C overnight. The refolding buffer (50 mM Tris HCl, 0.1 mM Zn acetate, 0.2 M NaCl, 10 mM CaCl2, 0.02% Sodium azide, pH 7.5 and then 0.25 M L-arginine was added) was prepared to

refold the overnight solubilised inclusion bodies and the overnight dissolved 5 mL inclusion body fractions in Guanidine HCl was added to 500 mL refolding buffer drop wise in a cold room with constant stirring. The refolded protein contained some precipitated protein as well. The precipitate was collected by centrifugation (3000 g for 15-20 minutes) and the precipitate was tested on MALDI. The solubilised protein was concentrated to 3 mL, buffer exchanged into 20 mM NH_4HCO_3 by ultracentrifugation (Amicon Ultra, Millipore) and loaded on to a gel filtration column (FPLC 16/60 HiLoadTM 75 SuperdexTM prep grade, Amersham Biosciences). The eluted fractions containing the pure protein were pooled and concentrated by ultracentrifugation (Amicon Ultra, Millipore).

2.6.2.1 Gel filtration chromatography

The protein was purified by FPLC (Pharmacia Akta Purifier) mounted with a size exclusion column (HiLoad 16/60 Superdex 75, Amersham Biosciences). The solublised refolded protein from the inclusion body in ammonium bicarbonate buffer (20 mM, pH 7.8) (3-3.5 mL), was filtered through a Millipore filter (0.2 μm - 0.4 μm, Minisart®) and loaded into a super-loop. Proteins were eluted with degassed ammonium bicarbonate buffer (20 mM, pH 7.8) at a flow rate of 1 mL/minute. The elution of protein fractions (4 mL) was monitored at wavelengths of 220 nm (peptide bond and Zn-S charge transfer bonds) and 280 nm (aromatic residues). Selected FPLC fractions (**Figure 2-5A**) were analysed by Sodium Dodecyl-Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE).

2.6.2.2 Anion exchange chromatography

In contrast to the above mentioned technique, the S-tagged W08E12.3 protein isolated from inclusion bodies and refolded in presence of metals was also purified on an anion exchange column to avoid the multiple centrifugation steps involved in order overcome the minimal capacity of sample loading on Gel filtration column.

The sample was loaded into a super loop (B line) and injected to an anion exchange column (HiTrap Q XL, 5 mL, GE Healthcare) connected to an FPLC system (Pharmacia Akta Purifier). Gradient elution (buffer A: 20 mM NH_4HCO_3, pH 9.4, buffer B: 1 M NaCl, 20 mM NH_4HCO_3, pH 9.4) at a flow rate of 2 mL/minute with a gradient length of 20 column volumes was employed. In summary, the column was cleaned with 100% of buffer B and then pre-equilibrated with 100% of buffer A. The elution of protein fractions (3 mL) was monitored at an absorbance of 220 nm (peptide bond and Zn-S charge transfer) and 280 nm (aromatic residues). W08E12.3 protein eluted around 48-50% of buffer B. Selected protein fractions (**Figure 2-5B**) were analysed by SDS-PAGE.

Fraction collection pattern:

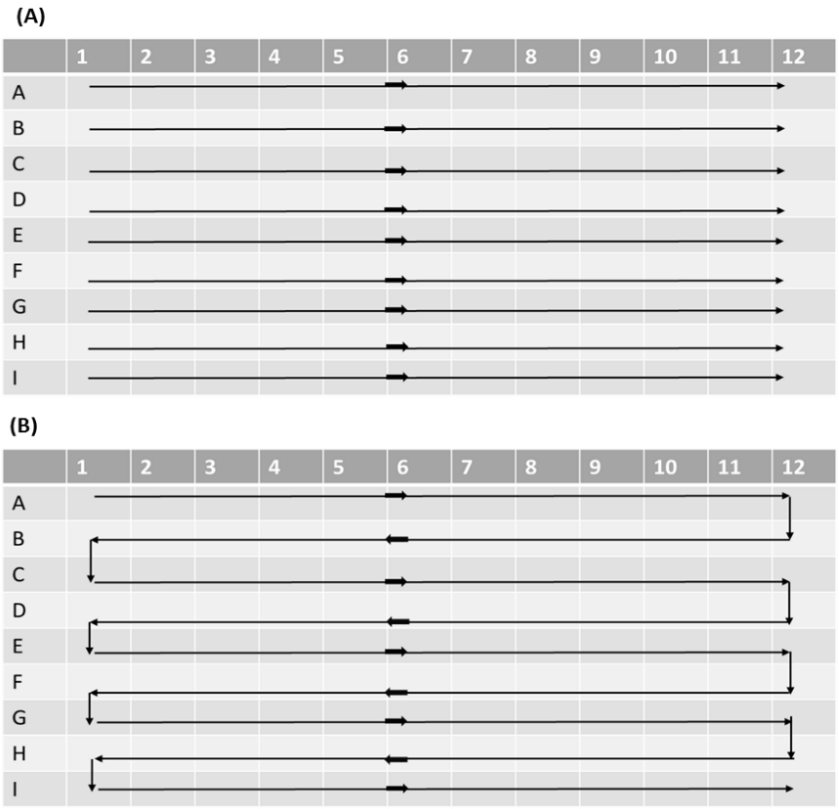


Figure 2-5: Pattern of fraction collection in a gel filtration chromatography (A) and anion exchange chromatography (B).

2.6.3 SDS-PAGE

2.6.3.1 Sample preparation

Cells were harvested from the liquid culture (1 mL) by centrifugation at 14,000x g in 1.5 mL microcentrifuge tubes. The supernatant was decanted and the pellet was allowed to dry. The pellet was resuspended with Novagen BugBuster™ reagent (5 mL for 1g cell pellet), and incubated at room temperature for up to 20 minutes, while shaken gently. Benzoase nuclease (5 µL) was added and cell debris were removed by centrifugation at 16,000 g for 20 minutes at 4°C. The supernatant (25 µL to 50 µL) was transferred to a new microcentrifuge tube and mixed with loading buffer containing β-mercaptoethanol (10 mM). The samples were boiled at 95°C for 15 minutes, then loaded onto a SDS-PAGE gel.

2.6.3.2 SDS-PAGE gel casting

SDS-Polyacrylamide gel electrophoresis was used to assess the protein induction qualitatively. SDS-PAGE gels were used to separate proteins based on their molecular size. As W08E12.3 protein is 14 KD in size a higher percentage (12%) SDS gel was chosen.

Table 2-6: Components of a 12% SDS-PAGE gel.

Resolving gel	Volume	Stacking gel	Volume
H_2O	1.6 mL	**H_2O**	1.5 mL
30% acrylamide	2 mL	**30% acrylamide**	200 µL
1.5 M Tris (pH 8.8)	1.3 mL	**1.5 M Tris (pH 6.8)**	250 µL
10% APS	50 µL	**10% APS**	20 µL
10% SDS	50 µL	**10% SDS**	20 µL
TEMED	6 µL	**TEMED**	4 µL

First the resolving gel was casted in between two glass plates of 1 mm thickness held together with a plastic holder up to the 1 cm mark below the wells of the comb. Isobutanol or isopropanol (1 mL) was added immediately on top of the resolving gel which was allowed to solidify for 15-20 minutes.

The isobutanol was washed off with distilled water (3x) and dried with a clean tissue, the stacking gel was casted and the comb was immediately placed taking care no air bubbles were inserted. The gel was left to dry for 15-20 minutes After the gel had settled, the comb was pulled out carefully and the wells were filled with 2- (*N*-morpholino) ethanesulfonic acid (MES) running buffer ready to be loaded.

2.6.3.3 Gel loading and running

The prepared samples (10 µL) were loaded into the wells alongside an Invitrogen SeeBlue® Plus2 Pre-stained Protein Standard (4-250 kDa) in order to determine the molecular weights of the proteins. The samples were separated for 1.5 hours at 70 V, until the loading dye was ~1 cm from the end of the glass plate. The gel was used either for Western blotting or coomassie/silver staining.

2.6.4 Gel staining protocols

2.6.4.1 Coomassie staining

The gel was washed three times with distilled water to remove the SDS. The gel was washed for 10 minutes each with distilled water thrice. The gel was stained with Coomassie R-250 dye overnight and then the gel was left for 3-4 hours in distilled water to destain and was then imaged.

2.6.4.2 Silver staining

Solutions:

Fixer: 60 mL 50% acetone, 1.5 mL 50% TCA, 25 µL formaldehyde (37%)

Pre-stain: 10 mg $Na_2S_2O_3$ in 60 mL distilled water (dH_2O)

Stain: 160 mg $AgNO_3$, 600 µL formaldehyde up to 60 mL with dH_2O

Developer: 1.2 g Na_2CO_3, 2.5 mg $Na_2S_2O_3$ up to 60 mL with dH_2O **(**25 µL formaldehyde added slowly with pipette while developing after adding developer to the gel)

Stopper: 1% acetic acid

Method: The Fixer added to the gel and left on a shaker (15 minutes). The gel was then washed thrice with dH_2O quickly and left in dH_2O (5 minutes) followed by three quick washes with dH_2O. The gel was then washed with 50% acetone for 5 minutes. The acetone was poured off and pre-stain was added and left for 1 minutes. The pre-stain was discarded and stain was added and left for 8 minutes. The gel was then washed with dH_2O twice and developed with the developer for 10-20 seconds. The developer was poured off and the stopping solution was immediately added. After 1-2 minutes the gel was rinsed with dH_2O and stored until ready to image.

2.6.5 Western Blotting

Table 2-7: Materials for Western blotting.

Reagent	Specification
Nitrocellulose membrane (0.2 or 0.45 µm)	SIGMA; Pre-cut to 8.5 cm (width) x 7.5 cm (height)
1X Transfer buffer	NuPAGE Transfer Buffer (Dilute 100 mL of 20X transfer buffer(Invitrogen) + 1700 mL dH_2O + 200 mL methanol
Filter Paper	Whatman 3mm; Pre-cut to 8.5 cm (width) x 7.5 cm (height)
1X TBST (TBS + 0.2% Tween): Wash buffer	Dilute 100 mL of 10X TBS stock + 900 mL dH_2O + 2 mL of Tween-20 (Sigma).
Blocking buffer (3% non-fat milk in TBST)	3 gm/ 100 mL TBST (0.2% tween), can be stored up to a week in 4° C
Primary antibody dilution buffer	1. Anti-CBD Monoclonal Antibody (NEB) in 3% non-fat milk (1:5000) with 0.1% sodium azide 2. S•Tag™ Monoclonal Antibody (Novagen) in 3% non-fat milk (1:5000) with 0.1% sodium azide
Secondary antibody dilution buffer	1. Anti-rabbit/HRP (Immunostar Biorad) (1:20,000) in 3% non-fat milk without sodium azide. 2. Polyclonal-anti-goat-anti-mouseimmunoglobulin (Dakocytomation) in 3% non-fat milk (1:10,000) without sodium azide
Autoradiography films	GE healthcare
Kodak developing solution	Sigma; dilute 220 mL stock + 790 mL dH_2O (Protect from light)
Kodak fixer solution	Sigma; dilute 220 mL stock + 790 mL dH_2O (Protect from light)
Mini Trans-Blot module	Biorad
West-Pico Chemiluminiscent substrate	Pierce (Thermo)

Method: After the SDS-PAGE had finished, the gel was removed from the cassette, the stacking gel was removed and the resolving gel was placed in transfer buffer until the blot was ready to be assembled. In a large plastic container, transfer buffer was poured until all the components of the blot were immersed. The components of the blot module were arranged layer by layer into a sandwich (**Figure 2-6**).

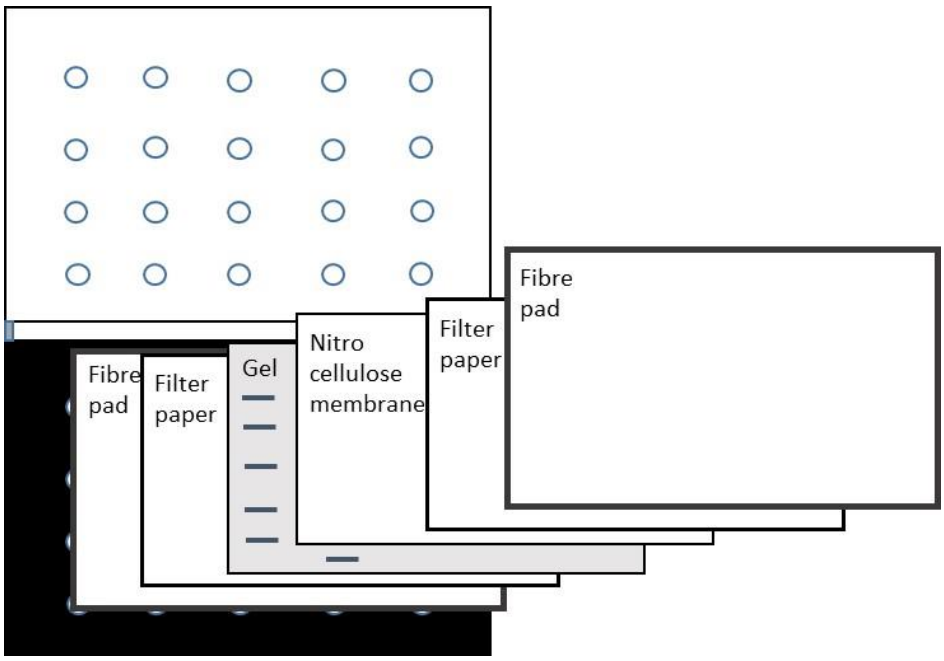


Figure 2-6: Schematic representation of the sandwich assembly for the transfer of the proteins from gel to the membrane during Western blotting.

The sandwich was placed into the blotting module and was topped up with transfer buffer. The tank around the blot module was filled with transfer buffer (3/4 full) and the transfer was performed in a 4°C room for 1 hour at 100V with continuous soft stirring. After the transfer, the sandwich was disassembled, discarding the gel and filter paper. The blot / membrane was placed in 3% non-fat milk (dissolved in TBST) and blocked for 1 hour at room temperature or overnight at 4°C. The blocking solution was discarded and the primary antibody dilution buffer was added. The blot was incubated in the primary antibody buffer overnight at 4°C. The primary antibody was discarded and the blot was washed thrice for 10 minutes each with TBST at room temperature with constant shaking. The secondary antibody buffer was then added to the blot and incubated for an hour at room temperature on a rotator. The secondary antibody was discarded and the blot was washed twice with TBST for 10 minutes each and once for 20 minutes. During the last TBST wash, the chemiluminiscent detection was prepared. A PVDC (polyvinylidene chloride) wrap was laid on the benchtop

smoothing out the wrinkles. A working solution of SuperSignal West Pico substrate was prepared so that there was 3 mL per blot (mix 1 part Lumino / Enhancer to 1 part stable Peroxide in a conical tube covered in foil). After the last wash the blot was washed with the West Pico substrate working solution for 4 minutes. The blot was then placed on the PVDC wrap with the protein side facing up and the excess liquid was drained off and the wrinkles were smoothed out with a paper towel. For the development of the Western on the autoradiography film, the blot was placed into a film cassette and development was carried out in a dark room. In the dark room, the blot was exposed to the film for 10 seconds, 1 minute and 5 minutes. The film was processed using Kodak developer and fixer until the bands were sufficiently dark but not saturated. For the electronic development of the blot, the blot from was transferred to the gel doc system. The first image was taken in presence of light to obtain an image of the ladder that could be later merged with the developed blot to indicate the sizes of the proteins in the blot. The images of the developed samples on the blot were taken in the dark as a set of five images, selecting the standard set up for development of Westerns where each new image would be superimposed with the previous image to enhance the visualisation of the protein bands developed on the blot. The final image was merged with the image containing the ladder to determine the overall impression of the developed protein bands and their sizes.

2.7 Quantification of protein concentration

2.7.1 Cysteine assay

The basis of the cysteine assay is the reaction between DTNB (5, 5'-dithiobis 2-nitrobenzoic acid) and thiols present in the protein resulting in disulphide and TNB (thiobenzoic acid) which is monitored by change in absorbance at 412 nm. The

concentration of the protein can be assessed by plotting the value on a scatchard plot created by using the cysteine standards. The protein was in Tris (0.1 M) buffer, to which DTNB (2.5 mM) was added and the reaction (5 µL protein + 200 µL DTNB + 195 µL water) was incubated for 10-15 minutes before the absorbance at 412 nm was recorded.

2.7.2 Bradford assay

The Bradford assay is a protein determination method based on the binding of Coomassie Brilliant Blue G-250 dye to proteins. On binding to proteins the dye changes its colour to blue which is detected at 595 nm in the assay using a spectrophotometer or microplate reader.

The quick start Bradford protein assay kit was used for the quantification of the amount of protein to be loaded on a SDS-PAGE gel or to quantify the total protein content in the nematodes aliquoted in each well while performing worm bulk assays using a fluorescent plate reader.

Bovine gamma globulin standards (2, 1.5, 1, 0.75, 0.5, 0.25, 0.125 mg/mL) provided by the supplier were used to obtain the standard curve. The standards and samples (5 µL of standard/sample + 250 µL of 1X dye reagent) were aliquoted into a 96 well microplate and was incubated for 15 minutes at room temperature. The plate was read at 595 nm in a microplate reader. A standard curve was plotted using the absorbance at 595 nm of the standards against their respective concentrations and the protein concentration of the unknown samples were determined from the curve.

2.7.3 BCA (Bicinchoninic acid) assay

Bicinchoninic acid (BCA) assay is a detergent free protein assay utilising bicinchoninic acid as a reagent to colorimetrically and quantitatively detect the amount

of protein. It is based on the biuret reaction i.e. reduction of Cu^{+2} to Cu^{+1} by proteins in alkaline conditions. The Cu^{+1} is selectively detected by the bicinchoninic acid leading to a purple coloured product which absorbs at 562 nm. The absorbance at 562 nm linearly increases with increasing protein concentrations (20-20,000 µg/mL). The Pierce™ BCA Protein Assay Kit was used in order to determine the concentration of the purified protein. The instructions were followed based on the manufacturer's guidelines.

A set of albumin standards were prepared of concentrations ranging from 0-2000 µg/mL in the buffer of choice. The standards and the unknown samples to be measured were mixed with the working reagent and incubated for 30 minutes at 37ºC. The samples were then read at 562 nm on a spectrophotometer to determine the concentrations. The concentration of the unknown protein samples were determined based on the standard curve that was plotted using the absorbance (562 nm) of standards against their respective protein concentrations.

2.7.4 ICP-OES (Inductively Coupled Plasma Optical Emission Spectroscopy)

Inductively Coupled Plasma-Optical Emission Spectroscopy (ICP-OES) (Perkin-Elmer Optima 5300 DV, Model S10) was used to determine the S, Zn, Cd and Cu content of the proteins by measuring S at 180.669 nm and 181.975 nm, Zn at 206.200 nm and 213.857 nm, Cd at 228.802 nm and 214.440 nm and Cu at 327.393 nm and 324.752 nm. Plasma operating conditions were, Argon (Ar) flow rate 13.0 L/minute, auxiliary gas flow rate 0.2 L/minute, nebuliser flow rate 0.8 L/minute and RF power at 1300 W (**Figure 2-7**).

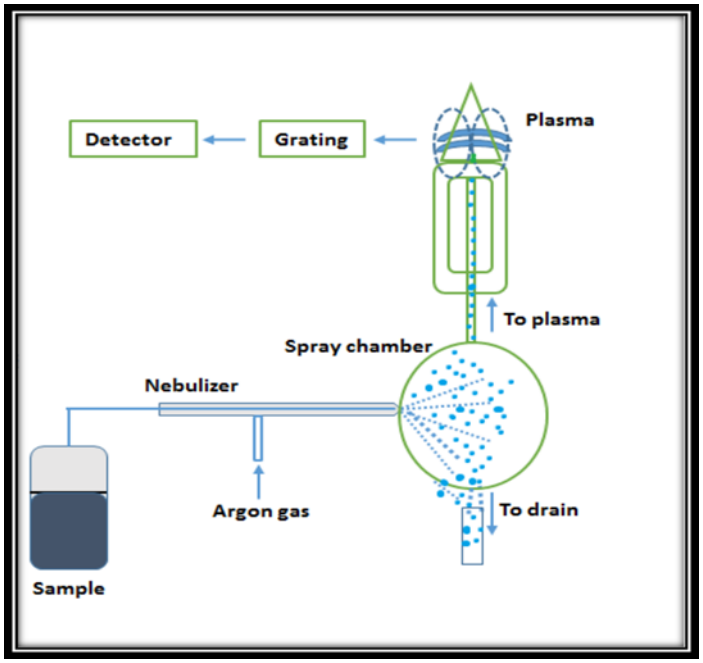


Figure 2-7: Schematic representation of sample introduction into the ICP-OES. The liquid sample is first loaded into the nebulizer that converts it into aerosols. The nebulizer sucks up the sample due to the pressure generated by the argon gas. The spray chamber separates the heavier droplets into the drain and the finer droplet aerosols in to the plasma. Plasma emits radiations from the samples that are analysed to detect metals (Skoog and West, 1980).

The samples for ICP-OES were prepared in 0.1 M nitric acid in the ratio of 1 part of sample: 3 parts of ultrapure nitric acid (prepared from 72% distilled HNO_3) in order to remove possible traces of labile sulfide ions. The ICP-OES instrument was calibrated using S, Zn, Cd and Cu 1000 p.p.m. standards (Fisher Scientific). All samples were analysed using three replicate readings, with a washing time of 60 seconds for each sample. Data were analysed using WinLab 32 software (Perkin-Elmer).

The values in mg/L were divided by dilution factor followed by 1000 to obtain units in µg/L. The result was then divided by the appropriate atomic masses which gives the concentration in µmol/L of the given elements in the sample. Based on the sulfur content the concentration of S-tagged W08E12.3 protein was determined as it

comprises of 22 sulfhydryl groups i.e. 19 cysteine, 3 methionine residues. Using ICP-OES based on the metal standards, the metal: sulfur (based on which the metal: protein) ratios were calculated and the concentration of the protein was also determined.

2.7.5 ESI-MS (Electrospray Ionisation Mass Spectrometry)

Electrospray ionisation mass spectrometry (ESI-MS) is used to identify and characterise a single species in solution. Based on the main charge on the protein, the stoichiometry of the protein or protein complex can be studied by the mass by charge (m/z) ratio. The intensities of the molecular peaks provides information about the abundance and stoichiometric stability of the each species present in the sample. The protein sample was desalted in 10 mM NH_4HCO_3 using a PD-10 sephadex G-25 column and the ESI-MS was performed at basic pH to obtain the metallated or the native spectra of the protein and acidic pH to obtain the metal devoid (apo) form of the protein. For the metallated form the sample was made up of 40 µM protein in 10 mM ammonium bicarbonate and 10% methanol, and the apo form was prepared in the same way but with the addition of 2 % formic acid. The sample was then loaded onto the ESI-MS with the help of a syringe.

Acquisition parameters

Mass spectra were recorded on a Bruker Daltonics MicroTOF fitted with an electrospray ionization source operating in positive mode. The samples were injected into the spectrometer by a syringe pump with a flow rate of 240 µL/hour. Other parameters were fixed as: temperature 195°C, nebulizer 0.6 Bar, dry gas 4.5 l/minute, capillary exit 100 V, skimmer1 50 V, skimer2 25.2 V, hexapole1 24.2 V, hexapole2 22.4 V, hexapole RF 450 V, transfer time 81 µs and detector TOF 2300 V.

The data were obtained over 0.4-2 minutes for the range of 500-5000 m/z. The experimental data were then smoothed and deconvoluted using the data analysis software Bruker compass data analysis v 4 provided by Bruker Daltonics.

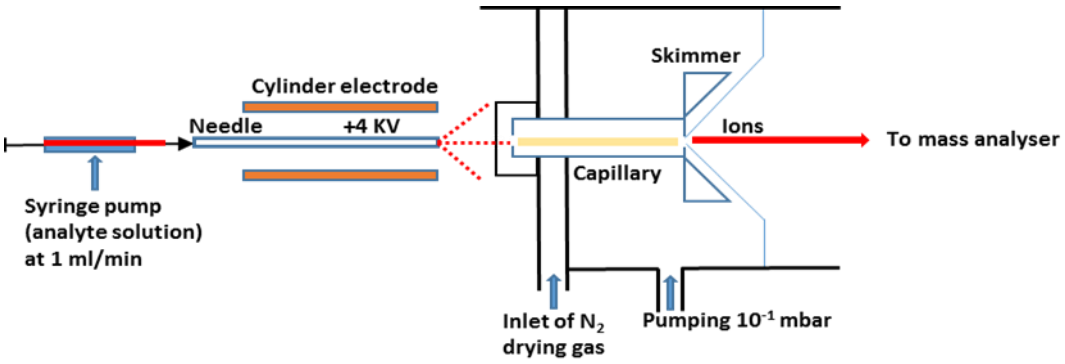


Figure 2-8: ESI-MS technique from sample loading and processing for mass analysis. The sample in solution is loaded with a syringe pump onto a capillary tube after it has been charged by the electric current (4 Kv) provided by the cylindrical electrodes. With the help of the electric field a dipolar meniscus is formed at the edge of the capillary tube. Due to excess charges present, the meniscus is destabilised producing a jet of droplets with excessive charged ions. The charged ions become smaller capable of producing gas phase ions as the solvent is evaporated by the inert N_2 gas. The gas phase ions travel into the analyser due to varying pressures and electric fields. As the gas phase ions have multiple charges, the mass to charge ratio is reduced to few thousands from tens of thousands making the data manageable by the mass analyser. Once, the ions come in contact with the detector based on the m/z ratio, signals are generated and displayed graphically (Ho et al., 2003).

2.8 pH titration

Metals are released from a protein at acidic pH. pH titrations can therefore determine if metals are bound to the proteins and at what pH they reach their apo (demetallated) form. Protein was equilibrated in 10 mM NH_4HCO_3 buffer and a concentration of 5-10 µM was needed to carry out the experiment.

pH titrations of zinc-loaded proteins were performed in the range of 200-300 nm at a scan rate of 300 nm/minute at room temperature. Respective protein samples (1.5 mL, 5-10 μM, 10 mM NH_4HCO_3 buffer, pH 7.92) were titrated with HCl (1-5 μL, 0.01-5 M). The dilution effect due to addition of HCl was negligible. The pH of the samples was measured before and after the UV-spectrum was recorded.

The pH titrations was initiated by measuring the pH of the protein in buffer which was reduced uniformly in small fractions until all the metals were released from the protein. Each measurement was taken after 10 minutes of equilibration time. After each pH change the UV-Vis spectroscopic data was collected. The pH readings were taken at each point both before and after the UV spectrometric measurement. In order to determine the pH of half metal displacement, the absorbance at 230 nm for each pH measurement was plotted.

2.9 Metal exchange assay (UV-Vis data)

The protein was equilibrated in 10 mM NH_4HCO_3 buffer and a concentration of 5 μM was required for the assay. If the protein was devoid of Zn and was oxidised, the protein was first treated with 1 mM Zn acetate and 2 mM TCEP/ DTT overnight. The metal exchange assay was based on the efficiency of Cd to replace Zn bound to the protein. The protein in a total volume of 700 μL of 10 mM NH_4HCO_3 was transferred into a quartz cuvette. The UV-Vis data of the Zn bound protein was first recorded and then 2 mM $CdCl_2$ was added progressively from 1 equivalent to 10-12 equivalent relative to the protein at a time interval of 10 minutes and the UV-Vis data was recorded for each equivalent increase of the Cd to the protein in the range of 200-300 nm at a scan rate of 300 nm/minute at room temperature. Dilution effect for addition

of cadmium was negligible. In order to monitor if the Cd ions displaced the Zn the absorbance at 254 nm for each molar equivalent addition of Cd was plotted.

2.10 X-ray fluorescence imaging at SLAC (Stanford Linear Accelerator Centre)

The Stanford Linear Accelerator Centre (SLAC) beam line accelerator in Stanford, California USA, was used to detect metal accumulation within single nematodes. The high energy electron beam enables the detection of metal accumulation within the worm body. The worms were exposed to metals (150 µM Zn, 30 µM Cd) from L1-late adult stage and imaged at L4 or later stages, N2 (wild type) worms were used as the control. The worms were then examined under two different beamlines i.e. for Zn and Cd accumulation. The accumulation of the metals within the worm body was tested by processing the images taken at SLAC with the aid of custom designed microprobe analysis software (SMAK).

The metal accumulation in the worms was tested by X-ray fluorescence imaging at the Stanford Synchrotron Radiation Laboratory (SSRL). The nematodes were immobilised by adding sodium azide on a plastic slide at room temperature. Zn and Cd were measured at different beam lines. 10 KeV incident energy was used at beam line 2-3, selected with a Si(111) double crystal monochromator for Zn measurements. The X-ray beam was focussed at 2 x 2µm spot size using a Rh coated Kirkpatrick-Baez (KB) mirror pair (Xradia Inc). The worms were rastered continuously across the beam and the elements of interest and the sample were monitored at each pixel by recording the intensity of the fluorescence lines using silicon drift Vortex detector (Hitachi) equipped with Xspress3 electronics (Quantum Detectors) with dwell time/ pixel: 75 ms.

Cd measurements were performed at lower incident energy at beam line 14-3, 3.575 KeV. The X-ray beam was focussed at 5 x 5µm spot size using Kirkpatrick-Baez (KB) mirror pair coated with Ni. Calibrating with standard x-ray fluorescence concentration thin film standards (MicroMatter) enabled detection of the concentrations of the elements.

Chapter 3: Genomic characterization of the *W08E(12.2-12.5)* gene family

3.1 Introduction

The importance of studying the genomic architecture and the arrangement of genes across the genome along with their evolution pattern has attracted increasing attention over recent years. It has come to light that the three dimensional organisation of the genes within the genome and the architecture of the genome is not random. The availability of genomic sequences of eukaryotes has indicated that beyond tandem arrays, eukaryotes comprise of co-expressed gene clusters that are related by function which could be either cellular or by fate due to recent duplication events. This suggests that the genomic architecture is organised in a way that certain gene families are proximal to each other. The proposition of genome organisation linked to gene regulation is contradictory to the ideology that genes are distributed homogenously across the chromosomes within the genome (Kosak and Groudine, 2004). This phenomenon has been proven to exist in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, mouse and human. In *Schizosaccharomyces pombe* highly expressed genes, G2 co-regulated genes and few functionally related genes are clustered together (Tanizawa et al., 2010), whereas in humans transcriptionally active GC-rich clusters were identified (Lieberman-Aiden et al., 2009). Hence, it was hypothesized that the genes with shared functions and levels of expression exist in proximity to facilitate co-transcription via shared transcription factors and ensure optimal chromatin remodelling (Diament et al., 2014).

The *C. elegans* genome is just over 100 Mb in size comprising of 20,000 protein coding genes. The entire genome has been sequenced into a composite of 2527 cosmids, 257 YACs, 113 fosmids and 44 PCR products (Consortium, 1998). The importance of studying the *W08E(12.2-12.5)* gene cluster arises from the information accessed from WormBase indicating that four copies of these highly similar genes are present consecutively on Chromosome 1V.

3.2 Results

3.2.1 Investigating the probability of occurrence of similar genes in a cluster

A script in PERL language was written in collaboration with Phil Cunningham, a senior bioinformatics officer at King's College London, in order to use the bioinformatics program (UVA FASTA SERVER) to perform a thorough search on all five autosomes and the X chromosome to define the frequency of four or more highly similar genes aligned next to each other on the worm genome. The criteria set for the screen were:

- The chains of genes (i.e. transcripts) identified are related by sequence similarity (cut off E value ≤ 0.0001).
- Each chain comprises of at least four genes.
- The gap in between two genes consecutive to each other in a chain is less than 2000bp.

The identified gene chains on each chromosome were further subjected to Gene Ontology in order to identify whether clusters of genes had similar functions and also to assess if clusters in proximity shared a common function.

In total 116 chains of genes were identified across the whole genome which fulfilled the criterion set by the bioinformatics program.

The number of protein coding genes within each chromosome is similar, with the exception of chromosome V, which has a greater number of genes due to its bigger size. The frequency of similar gene chains is highest (~1.3%) on the chromosome II (**Table 3-1**).

Table 3-1: Identified gene clusters in respective chromosomes and their frequency of distribution within each chromosome.

Chromosome	Length (Mb)	No. of protein coding genes	Chain of 4	Chain of 5	Chain of 6	Chain of 7	Chain of 8	Chain of 9	Chain of 10	Chain of 12	Total Chains	Frequency (no. of chains/total no. of protein coding genes)
1	15.08	2920	1	1	1	1	0	0	0	0	4	0.0013
2	15.28	3552	24	12	6	3	0	0	1	0	46	0.0129
3	13.78	2699	5	2	0	1	0	0	0	0	8	0.0030
4	17.49	3348	8	7	1	2	2	0	0	1	21	0.0062
5	20.92	5149	12	11	2	1	0	1	0	0	27	0.0052
X	17.72	2844	3	4	2	1	0	0	0	0	10	0.0035
Total	100.27	20512	53	37	12	9	2	1	1	1	116	0.0056

The frequency of occurrence of a family of genes which form a gene chain based on sequence similarity across the entire genome is calculated to be 0.6%, which highlights that the existence of similar genes in a cluster on the worm genome is a rare event. Moreover, the search included genes that were not necessarily protein coding in nature and also pseudogenes.

The existence of the *W08E(12.2-12.5)* isoforms on chromosome IV is therefore a significant and rare event. Therefore, the occurrence of four highly identical genes coding for cysteine rich proteins makes the *W08E(12.2-12.5)* gene family important to study. This is particularly true as the *C. elegans* genome is compact and gene duplication or the occurrence of identical/similar isoforms are typically lower than in more complex genomes of higher organisms.

It can be inferred that the most commonly occurring gene clusters are of the serpentine receptor family and the F-box protein family. Others include C-type lectins, MATH domains, Glutathione S transferase, Nematode specific peptides and the Cytochrome P450 family (**Figure 3-1**). A detailed description along with the other gene clusters that appeared on the worm genome can be found in the appendix **(9.1)**.

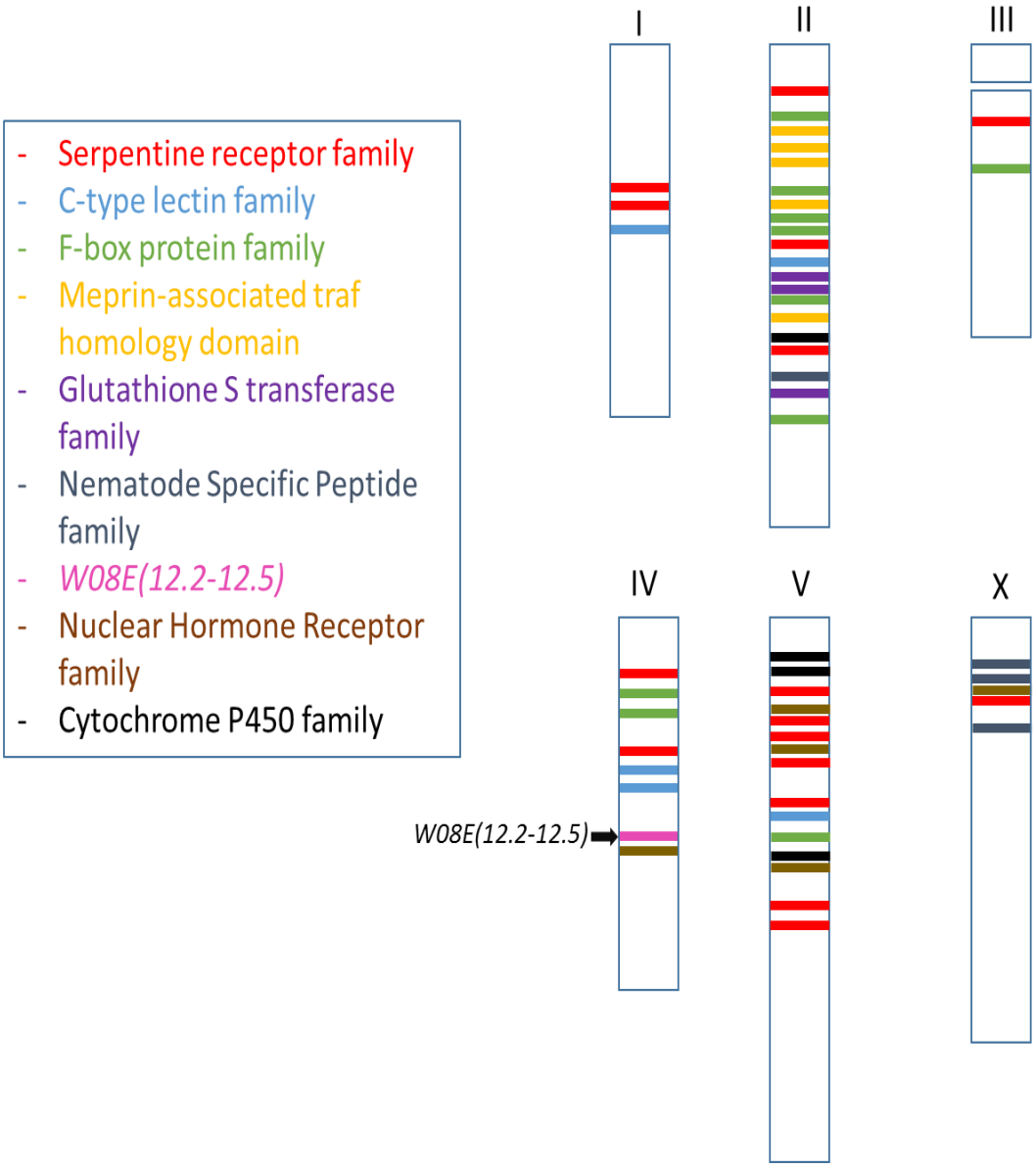


Figure 3-1: Diagrammatic representation of distribution of most commonly occurring gene chains (similar gene clusters) across each *C. elegans* chromosome identified by bioinformatics screening. All the chromosomes have been labelled with the most commonly occurring gene chains based on relative distance between them.

Table 3-2: List of the most common gene chains occurring across the entire worm genome and their functions.

Family	Function
Serpentine receptor family	• The serpentine receptor family are part of the chemoreceptor family. • They encompass a wide range of functions, including autocrine, paracrine and endocrine mechanisms (Troemel, 1999).
F-box protein	• This protein family facilitates protein-protein interaction. • It has also been shown to be involved in controlling apoptosis (Chiorazzi et al., 2013).
C-type lectins	• Protein family with an affinity for carbohydrates in presence of $Ca^{(2+)}$. • In the genome of Caenorhabditis elegans, there are almost 300 genes encoding proteins that contain C-type lectin-like domains (CTLDs). • Their product however have never shown to have carbohydrate binding activity (Takeuchi et al., 2008).
Meprin associated traf domains	• They have been recruited for their functions in apoptosis from the ubiquitin based pathways (Aravind et al., 2001).
Glutathione S transferase	• Facilitates Conjugation of reduced glutathione to a wide number of hydraphobic electrophiles both exogenous and endogenous. • It is mainly involved in dealing with oxidative stress (Leiers et al., 2003).
Nematode specific peptide family	• Families are predicted to code for small secreted unrelated proteins that are unrelated to any known proteins outside of nematodes. • They might be possibly involved in reproduction, larval development, egg hatching, growth of the nematode but their exact function is still unknown (Thomas, 2006).
Nuclear hormone receptor	• They are regulated by lipophilic hormones involved in metazoan metabolism, development and homeostasis. • *C. elegans* harbours 284 of these receptors in its genome. These affect life history traits such as sex determination, molting, developmental timing, diapause, and life span. • They also influence neural development, axon outgrowth and neuronal identity. • They are involved in lipid and xenobiotic metabolism as well (Antebi, 2006).
W08E(12.2-12.5) gene family	• Uncharacterised family of genes encoding cysteine rich proteins.
Cytochrome P450	• Gene families within C. elegans like CYP34 and CYP35 subfamilies may represent xenobiotic metabolising genes in the parasite (Laing et al., 2015).

3.2.2 Promoter analysis

In order to understand the regulation of *W08E(12.2-12.5)*, an investigation of the promoters of the isoforms was deemed important. Using GCG, TESS, JASPAR based software and the TransFAC database, putative transcription binding sites positioned up to 1000 bp promoter region upstream of each of the *W08E(12.2-12.5)* isoform were identified and then sorted to screen for putative metal binding or metal responding loci.

GCG Analysis revealed the following transcription binding sites to be present in all four genes: Ad2MLP_US.3:TATAAA, AP-1_CS3:TGANTMA, CaFA_site, EcR-consensus (2): KNTCANTNNMM, EF-1A_CS:RNMGGAWGT, GATA-1_CS2:WGATAMS, H2A_conserved_US:YCATTC, MtBF_CS, PEA3_CS:AGGAAR, TATA-box-CS: TATAWAW.

The JASPAR CORE database was chosen to screen the promoters of *W08E(12.2-12.5)* isoforms as JASPAR contains a curated, non-redundant set of profiles, derived from published collections of experimentally defined transcription factor binding sites for eukaryotes. As the *W08E(12.2-12.5)* isoforms code for cysteine rich proteins, suggesting metal binding or metal response, the promoter region of 1000 bp upstream of each isoform was screened through the JASPAR CORE database against a pre-existing matrix containing a list of Zn binding or coordinating transcription factors. The metallothionein (MT) isoforms were also subjected to the screening in order to identify if *W08E(12.2-12.5)* isoforms share any common transcription factor binding sites which might suggest a possible linkage between the two families. The transcription factors found in the promoter region of *W08E(12.2-12.5)* isoforms and metallothionein isoforms based on the preselected matrix are zinc coordinating in nature. ELT-3, CHE-1 and EOR-1 originate from the BetaBetaAlpha-zinc finger family. Whereas, ELT-3 is from the GATA family; DAF-12 from hormone nuclear receptor family and MAB-3 from DM family.

BLMP-1, CHE-1 and ELT-3 were identified to be the most abundant sites in *W08E(12.2-12.5)* and also in MT's (**Table 3-3**). The occurrence of Zn coordinating transcription factor binding sites in the *W08E(12.2-12.5)* gene family is similar to that of metallothionein isoforms suggests a role in metal binding or a function which is

similar to that of the metallothioneins. More detailed information about the existing transcription binding sites in the promoter region of *W08E(12.2-12.5)* isoforms can be found in the appendix section (**9.2**).

Table 3-3: Identified transcription factor binding sites in the 1000 bp promoter region upstream of the *W08E(12.2-12.5)* isoforms, *mtl-1*, and *mtl-2*.

Transcription factor	*W08E12.2*	*W08E12.3*	*W08E12.4*	*W08E12.5*	*mtl-1*	*mtl-2*
CHE-1	21	16	16	18	13	19
ELT-3	10	5	5	3	9	6
BLMP-1	15	15	15	4	42	23
MAB-3	1	1	1	0	1	0
DAF-12	2	0	0	3	2	0
EOR-1	0	0	0	1	8	2

3.2.3 Confirming the existence of the *W08E(12.2-12.5)* series within the worm genome

Following the bioinformatic analysis, the physical existence of the four isoforms of *W08E(12.2-12.5)* had to be confirmed. Hence the first step in characterization of the *W08E(12.2-12.5)* gene family was to confirm the presence of all four genes consecutively within the worm genome as suggested by WormBase (www.wormbase.org).

3.2.3.1 *W08E(12.2-12.5)* gene series alignment

As per the data available on the WormBase database, *W08E(12.2-12.5)* are consecutively aligned on chromosome IV of the nematode. The annotation of these genes is based on their position on the yeast artificial chromosome (YAC) library in which they were cloned during the whole genome sequencing of the worm.

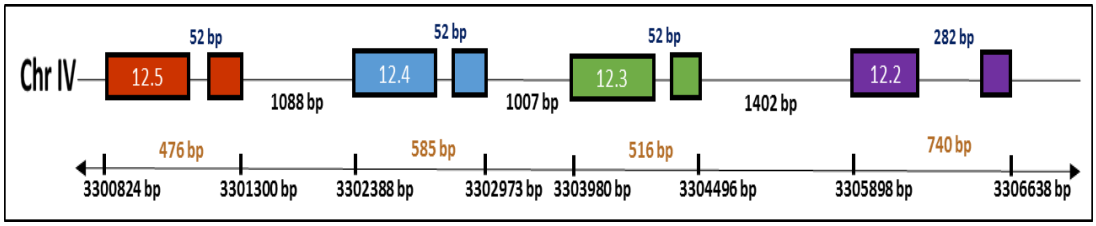
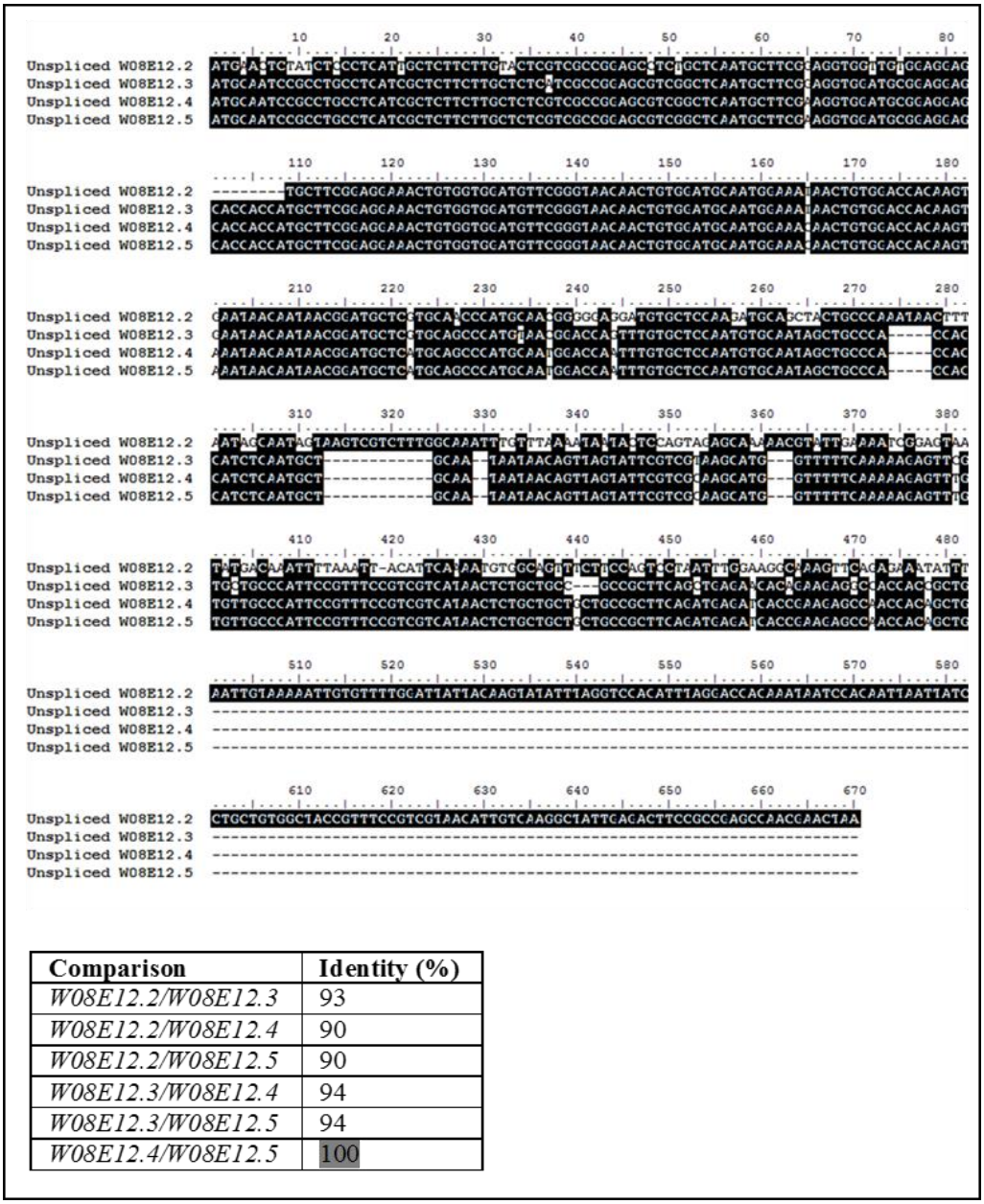


Figure 3-2: Schematic diagram of a section of the YAC cosmid *W08E12* depicting the four homologous isoforms *W08E12.2, W08E12.3, W08E12.4* and *W08E12.5*.

All the isoforms have an identical intron and exon structure and are evenly spaced with the exception of *W08E12.2* where the two exons are distanced out in comparison to the other three genes (**Figure 3-2**).

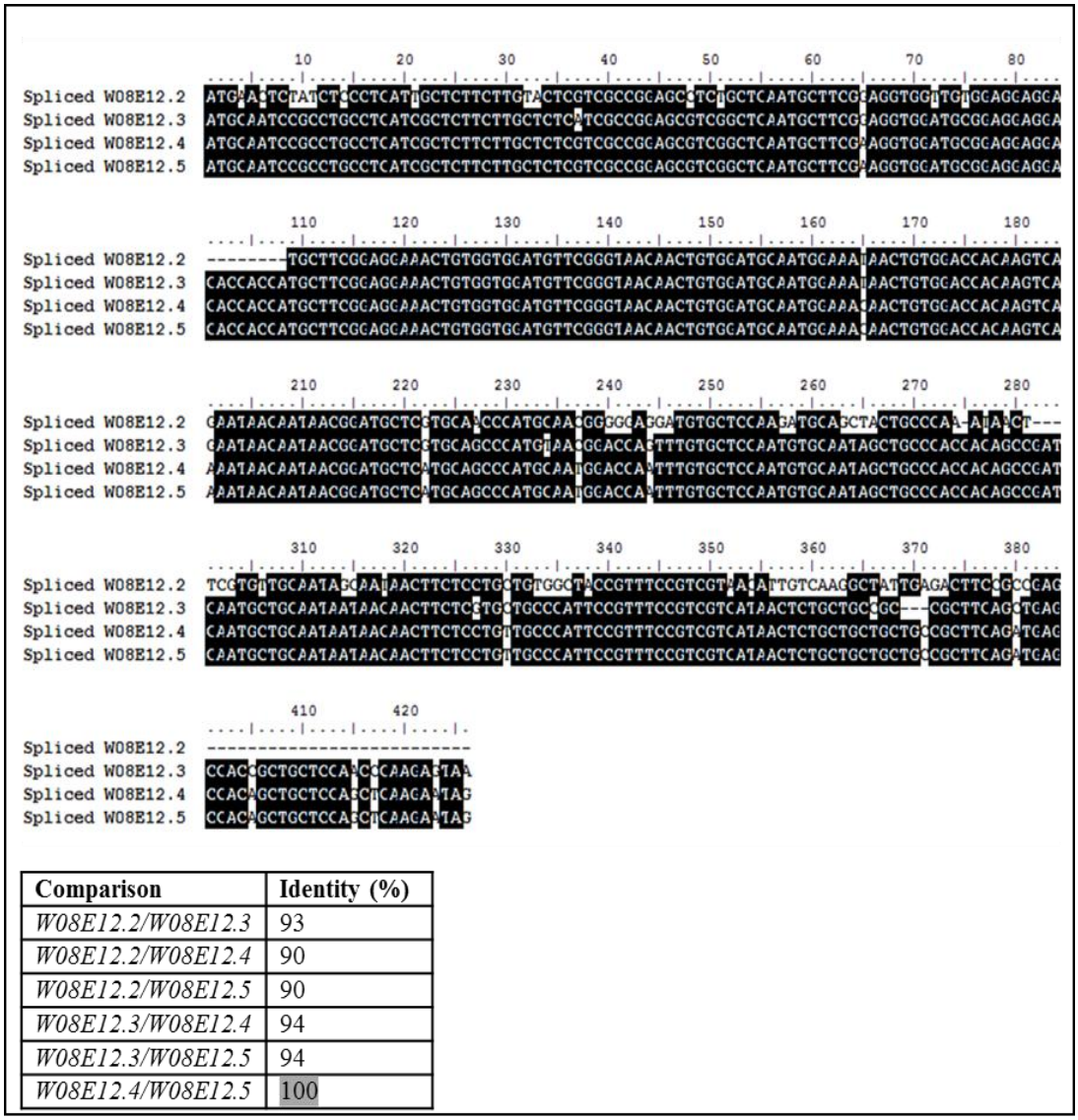
The four isoforms of the *W08E(12.2-12.5)* gene family were subjected to alignment following BLAST using the ClustalW alignment program in order to define the percentage similarity between the isoforms.

The nucleotide sequences of the coding region (spliced and unspliced) (**Figure 3-3**, **Figure 3-4**) and their promoter regions were aligned using the BioEdit software and were subjected to BLAST (Madden, 2013) which revealed that all the four genes depicted high sequence similarity to each other with *W08E12.4* and *W08E12.5* being 100% identical (**Figure 3-4**). When 500 bp of the promoter region of the four genes was compared it was observed that the promoters of *W08E12.3* and *W08E12.4* are 100% identical. The amino acid sequences revealed that the *W08E(12.2-12.5)* series are cysteine rich (18-19 cysteine out of 141 aa) (**Figure 3-6**).



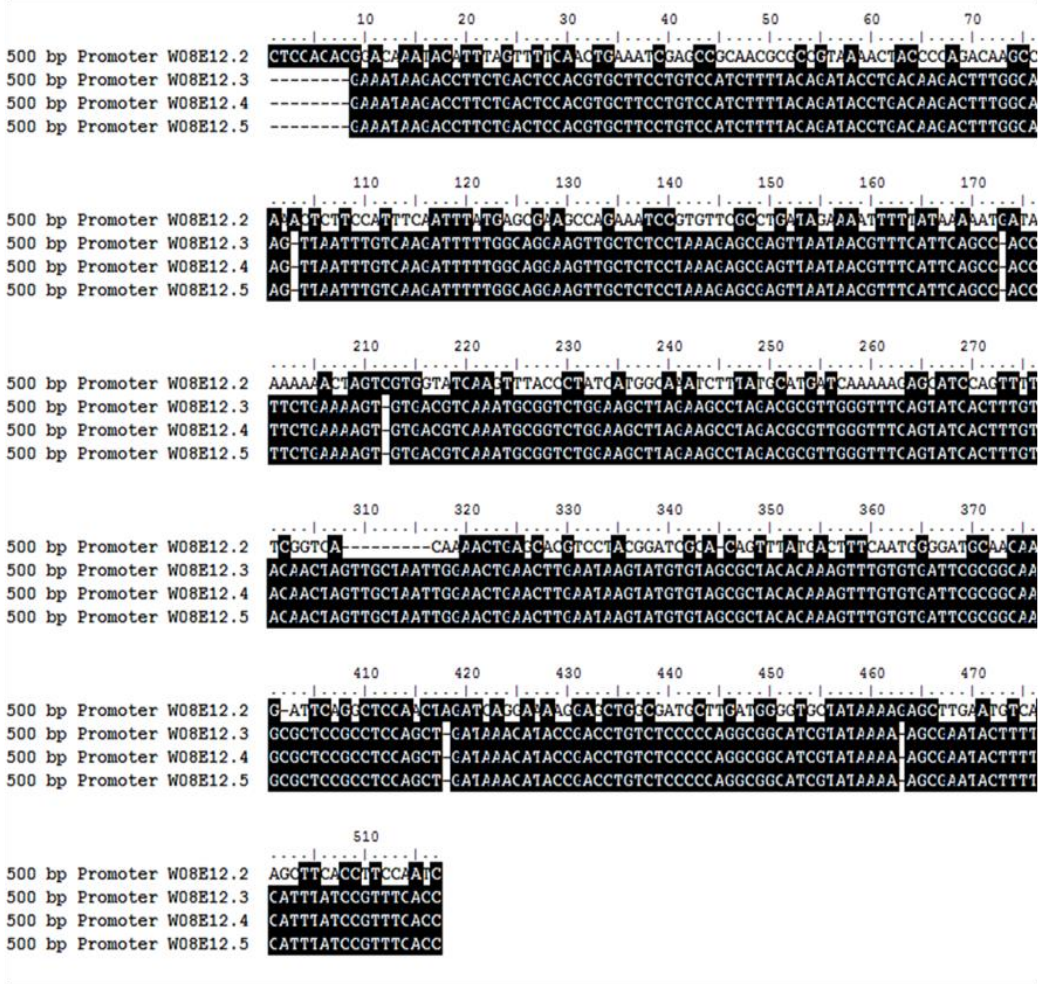
Comparison	Identity (%)
W08E12.2/W08E12.3	93
W08E12.2/W08E12.4	90
W08E12.2/W08E12.5	90
W08E12.3/W08E12.4	94
W08E12.3/W08E12.5	94
W08E12.4/W08E12.5	100

Figure 3-3: ClustalW alignment of the unspliced coding regions of the *W08E(12.2-12.5)* isoforms. Identical nucleotides are shown in black. The table depicts the BLAST results revealing the sequence similarity between the unspliced coding regions of the *W08E(12.2-12.5)* isoforms.



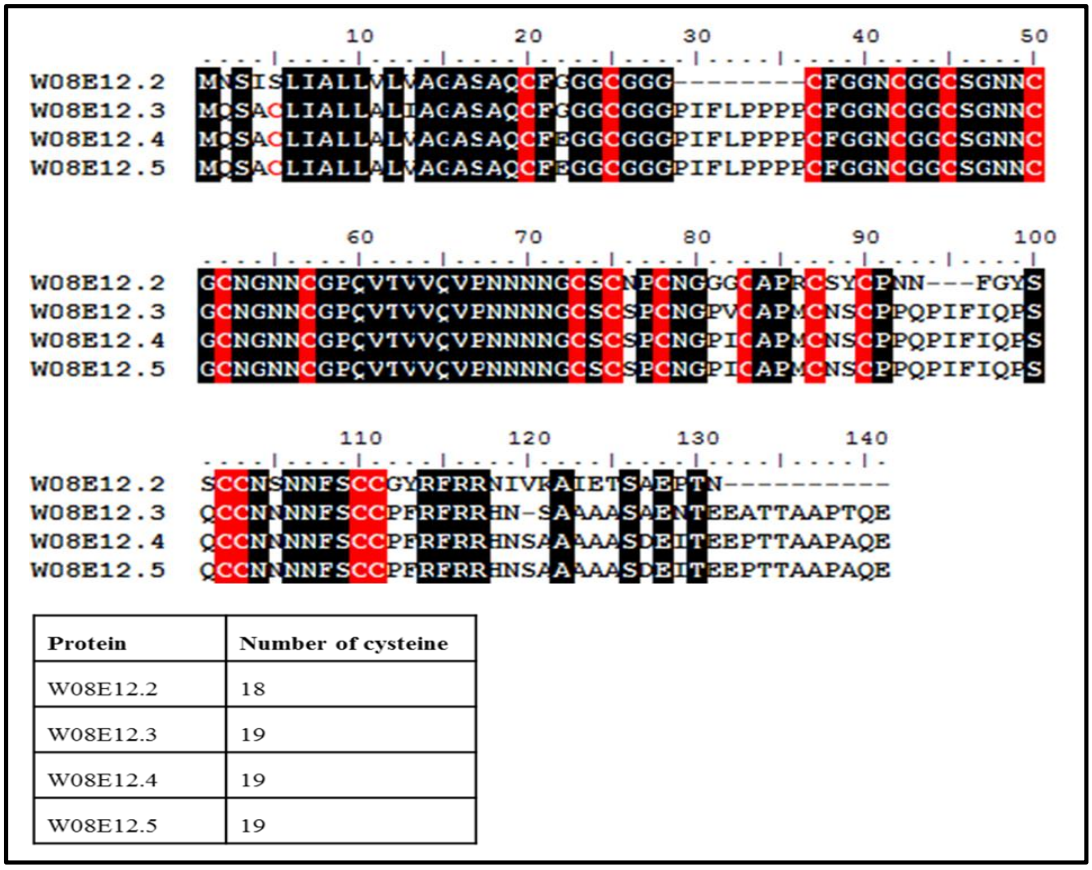
Comparison	Identity (%)
W08E12.2/W08E12.3	93
W08E12.2/W08E12.4	90
W08E12.2/W08E12.5	90
W08E12.3/W08E12.4	94
W08E12.3/W08E12.5	94
W08E12.4/W08E12.5	100

Figure 3-4: ClustalW alignment of spliced coding regions of the *W08E(12.2-12.5)* isoforms. Identical nucleotides are shown in black. The table depicts the BLAST results revealing the sequence similarity between the spliced coding regions of the *W08E(12.2-12.5)* isoforms.



Comparison	Identity (%)
W08E12.2/W08E12.3	42
W08E12.2/W08E12.4	42
W08E12.2/W08E12.5	40
W08E12.3/W08E12.4	100
W08E12.3/W08E12.5	92
W08E12.4/W08E12.5	92

Figure 3-5: **ClustalW alignment of the 500 bp promoter regions of the *W08E(12.2-12.5)* isoforms. Identical nucleotides are shown in black. The table depicts the BLAST results revealing the sequence similarity between the promoter regions of the *W08E(12.2-12.5)* isoforms.**



Protein	Number of cysteine
W08E12.2	18
W08E12.3	19
W08E12.4	19
W08E12.5	19

Figure 3-6: ClustalW alignment of amino acid sequences of the *W08E(12.2-12.5)* isoforms. Identical residues are shown in black. The cysteines are highlighted in red. The table depicts the number of cysteine residues in each protein isoform of the *W08E(12.2-12.5)* family.

3.2.3.2 Validating the order and presence of all four *W08E(12.2-12.5)* genes

Information regarding the order of *W08E12.2*, *WO8E12.3*, *W08E12.4*, and *W08E12.5* on the YAC cosmid is based on the information derived from the WormBase database. In order to cross validate this information, the whole genomic fragment comprising all four genes along with their upstream promoter sequences were attempted to be cloned and sequenced.

Single worm lysis was followed by a series of nested PCRs to amplify the whole genomic fragment consisting of all four genes or partial fragments. Different

combinations of primers were used to carry out the nested PCRs. First the external PCRs were performed and then the products of the external PCRs were used to conduct the internal PCRs. The following PCRs were set up (**Figure 3-7**):

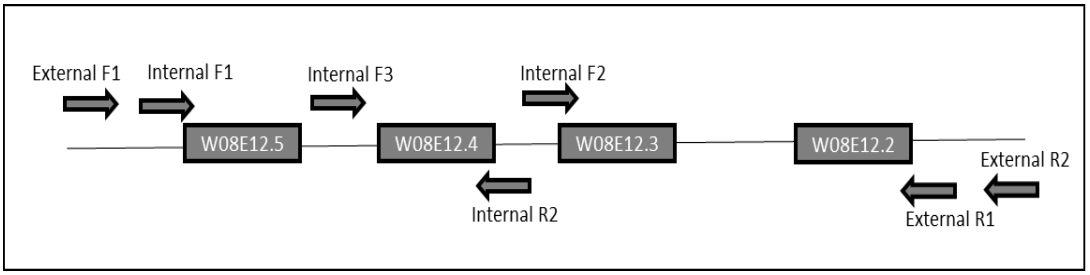


Figure 3-7: Schematic representation of the PCRs set up using different combination of primers in order to amplify the genomic fragments comprising the *W08E(12.2-12.5)* isoforms.

The gel electrophoresis results of the nested PCR products Int F2/ Ext R1, Int F1/Ext R2 followed by an internal PCR with Int F2/Ext R1, displayed a band of 2816 bp; Int F1/Ext R2 followed by an internal PCR with IntF3/Ext R1, nested PCR products produced expected bands of 3083 bp; Ext F1/Ext R1 followed by an internal PCR with Int F1/Int R2 displayed a band of expected size i.e. 1676 bp, whereas the other PCR combinations resulted in smears on the gel. The smears appeared most likely due to background noise amplification of the genomic DNA (**Figure 3-8**). All bands obtained from the nested PCR products were gel purified and attempted to be cloned into pGEM-T vector.

In the initial screening, which was performed by extracting the plasmids via disrupting the *E.coli* cells at high temperature and subjecting them to PCR with their respective primers, it was observed that only the plasmids amplified with Int F2/Ext R1; Ext F1/Ext R1 followed by an internal PCR with Int F1/Int R2 primers displayed the product of the expected size i.e. 2816 bp; 1676 bp respectively (data not shown) and were sent to GATC for sequencing.

The sequencing reactions confirmed successful cloning and sequence identification of these two fragments which validates the genomic locus of *W08E12.2, W08E12.3* and *W08E12.5* and also confirm their order of presence in the worm genome (**Figure 3-8**)

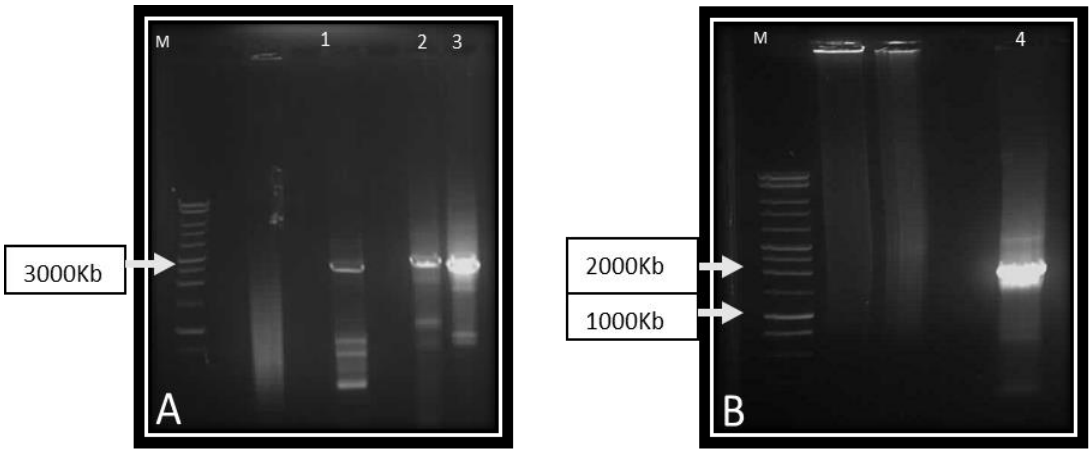


Figure 3-8: Evaluation of the XLPCR products by gel electrophoresis. The XLPCR products were separated against a 1Kb marker (Promega) on a 0.8% agarose gel. A: (1) Int F2/ Ext R1, (2) Int F1/Ext R2 (Int F2/Ext R1), (3) Int F1/Ext R2 (IntF3/ Ext R1), B: (4) Ext F1/Ext R1 (Int F1/Int R2).

Proving the presence of *W08E12.4* within the worm genome was a challenge as it shares 100% sequence identity in the coding region with *W08E12.5* and in the promoter region with *W08E12.3* (**Figure 3-9A**). Finally, a PCR reaction was set up to amplify the entire genomic region from *W08E12.3* to *W08E12.5* using a high fidelity Go Taq Long PCR master mix (**Figure 3-9B**).

Amplification with the high fidelity long PCR mastermix yielded a genomic fragment of 3835 bp using 1F/ ExtR primer combination on the worm genomic DNA (**Figure 3-10**). This fragment confirms the presence of *W08E12.4* along with the other isoforms because if *W08E12.4* did not exist then the PCR product would have been of 2660 bp instead of 3835 bp (**Figure 3-9B**). Hence, the existence and order of presence of all

the four isoforms of the *W08E(12.2-12.5)* family within the worm genome could be confirmed.

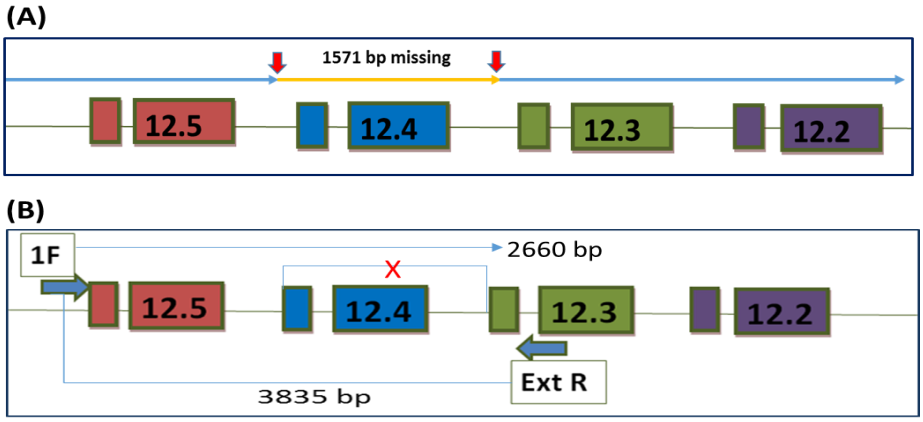


Figure 3-9: (A) Schematic diagram depicting the sequenced regions of the genomic fragment comprising of the *W08E(12.2-12.5)* isoforms and also highlighting the 1517 bp region involving *W08E12.4* gene that could not be sequenced. (B) Schematic representation of extra-long PCR set up using Go Taq Long PCR master mix from Promega with primers spanning the entire genomic region from *W08E12.5* until *W08E12.3*. In presence of *W08E12.4*, expected PCR product of 1F/ExtR would be 3835 bp, while in its absence in the worm genome, the expected PCR product would be 2660 bp.

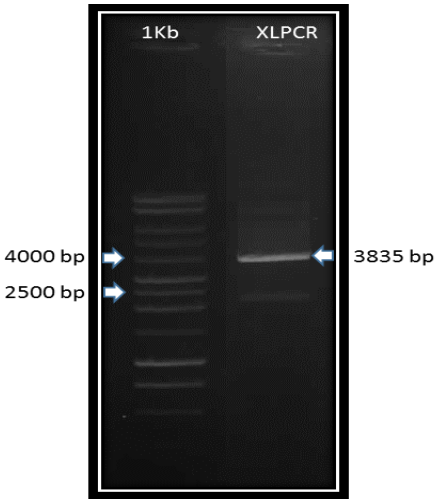


Figure 3-10: Gel electrophoresis of XLPCR product (IF/Ext R) against a 1Kb Promega marker on a 0.8% agarose gel. The Go Taq Long PCR mix was used to amplify the genomic region spanning *W08E12.3*, *W08E12.4* and *W08E12.5*. The PCR amplified genomic product of 1F/Ext R was separated on a 0.8% agarose gel to test if the PCR yielded the product of expected size.

3.3 Discussion

3.3.1 Investigating the probability of occurrence of similar genes in a cluster

A whole genome screening was performed in order to verify the occurrence of clusters of genes across the entire worm genome. The identified genes were subjected to Gene Ontology analysis to investigate if chains coded for similar functions were in close proximity to each other. The Gene Ontology data revealed the most frequent gene chains across the entire worm genome were the serpentine receptor family and the F-box protein family. The serpentine receptors are part of the chemoreceptor family, and a series of disparate G protein coupled receptors were found to be candidate olfactory receptors in *C. elegans* (Troemel et al., 1995). The genes were classified into small gene families and given names starting with sr for serpentine receptor (*sra*, *srb*, *srd*, *sre*, and *sro*). Their probable chemosensory function was based on transgene expression patterns in one or more known pairs of chemosensory neurons (Robertson and Thomas, 2006). *C. elegans* can sense touch, temperature, and light, but its responses to chemicals are the most diverse responses in its behavioural routine. The nematode feeds on bacteria which produce chemicals stimulating life cycle events like defecation, reproduction, feeding, chemotaxis (Bargmann and Horvitz, 1991), and the avoidance of toxic compounds (Culotti and Russell, 1978). Pheromones contribute to the mating success between males and hermaphrodites (Liu and Sternberg, 1995) and also controls the development of dauer stage, an alternative larval stage (Golden and Riddle, 1984). The nematode comprises of only 14 types of chemosensory neurons but responds to a large number of chemicals as each neuron detects varied stimuli (Troemel, 1999). The existence of a multitude of serpentine receptors throughout the worm genome reconfirms that chemoperception is an essential sense to the soil

nematode. The clustering of chemosensory genes across the worm genome especially across chromosome V, II and IV has been confirmed by others (Barnes et al., 1995).

The F-box protein facilitates protein-protein interaction. An interesting finding that *C. elegans* has 326 F-box proteins, and is the fourth most common protein domain in the nematode (Kipreos and Pagano, 2000). The notion that closely related genes cluster in the *C. elegans* genome has been suggested to be due to gene duplication events (Chen et al., 2005). Overall it has been proven that the worm genome is not organised in a random pattern, quite the contrary, its architecture is designed in a way that similar genes and similar gene chains cluster together within the genome.

3.3.2 Promoter analysis

Predicting metal binding sites in proteins by using the sequence as the source of information helps in predicting the structure and function of the protein. Metal binding domains are mostly cysteine or histidine rich motifs as they have shown to participate in the binding to transition metals. Hence, identification of such motifs in the sequence may indicate the presence of metal binding properties of the protein (Passerini et al., 2006).

The 500 bp promoter region of all the four genes were analysed in order to pinpoint metal responsive transcription factor binding sites to explore *in silico*, metal responsiveness of the *W08E(12.2-12.5)* series.

Manual screening of metal responsive binding sites revealed sites such as **MRE-CS2** (metal response element), **MBF-I_CS** (metal binding factor) in the promoter region of *W08E12.2* at position 285. **MRE_CS2** is also positioned in the promoter region of metallothionein genes and drives its expression in other organisms but not in *C. elegans*. **EBV-ZRE5** (zinc response element), **NF-E1_CS1** (zinc finger transcription

factor) binding sites were identified in the promoter region of *W08E12.2* and *W08E12.5*. **GATA** elements like **GATA-1_CS2** which have shown to play a role in transcription of metallothionein genes *mtl-1* and *mtl*-2 (Moilanen et al., 1999) were confirmed to be present in the promoter region of all four *W08E(12.2-12.5)* genes. **GR-MT-IIA** which is a metallothionein gene specific transcription factor (Burcelin et al., 1995) was present in the promoter region of *W08E12.5* at position 23. Stress response elements such as **HIF** (hypoxia inducible factor), **HRE** (hypoxia response element) were also identified in the promoter region of all four genes. The afore mentioned transcription factor binding sites suggest putative metal responsive or metal binding roles of the *W08E(12.2-12.5)* series.

In order to highlight the transcription factors involved in metal binding, JASPAR CORE database with the selection of Zn binding transcription factor matrix was used. The promoter regions of *mtl-1* and *mtl-2* were also screened with *W08E(12.2-12.5)* isoforms. The results signify that the promoters of *W08E(12.2-12.5)* isoforms contain numerous Zn coordinating transcription factors including **CHE-1**, **ELT-3** and **EOR-1**, all of which are common to metallothioneins. Overall, the results of promoter analysis suggest that the *W08E(12.2-12.5)* series are likely to be transcriptionally regulated in a manner similar to metallothionein and might be involved in metal homeostasis or detoxification.

3.3.3 Confirming the existence of the *W08E(12.2-12.5)* series within the worm genome

W08E(12.2-12.5) series is an uncharacterised gene family and the sequence information of the isoforms was obtained from the WormBase database. Aligning the amino acid sequence and nucleotide sequence of the coding region as well as the

upstream promoter region revealed that all four isoforms are highly similar, conserved and cysteine rich. Whilst *W08E12.4* and *W08E12.5* are 100% identical in their coding regions as well as non–coding intron regions their promoters differ. In contrast, the promoter region (500-1000 bp sequence upstream to the genes) of *W08E12.3* and *W08E12.4* are 100% identical to each other but not the coding region. The high level of similarity between the four genes suggests a recent duplication event. However *W08E12.2* is comparatively different to the other three genes when its entire genomic sequence is considered, suggesting its evolution to be a result of an earlier divergence.

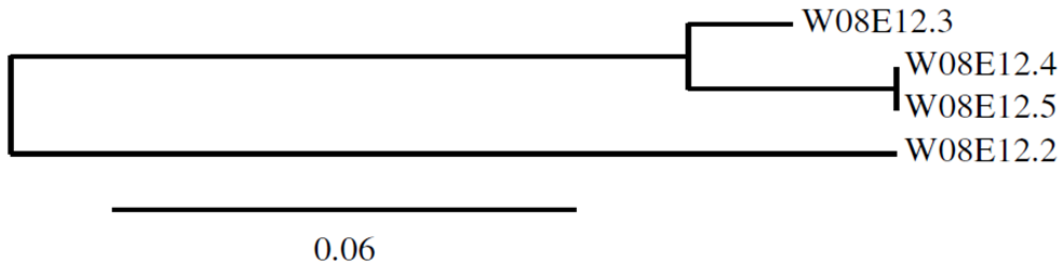


Figure 3-11 : Cladogram of *W08E(12.2-12.5)* gene family (generated using www.phylogeny.fr).

The amino acid sequence of the *W08E(12.2-12.5)* series contains 13% cysteines, a proportion that is lower than in metallothioneins which typically have a cysteine content of about 30%. However, the metallothionein protein is smaller than the *W08E(12.2-12.5)* isoforms and the absolute cysteine number is similar, where MTL-1 comprises of 19cys/75a.a., MTL-2 comprises of 18cys/63a.a. whereas the W08E12.2, W08E12.3, W08E12.4 and W08E12.5, code for 18cys/120a.a., 19cys/140a.a., 19cys/141a.a., 19cys/141a.a. respectively. Cysteines are rare in most of proteins and their presence indicate roles in metal binding or reactive oxygen (ROS) sequestering. If metal binding of the of *W08E(12.2-12.5)* series resembles MTs their secondary and tertiary structures will certainly differ.

Using nested PCRs, extra-long PCRs and cloning, the existence and the order of all the four isoforms within the worm genome was confirmed. The experimental results matched with the information provided by the WormBase database i.e. *W08E12.2*, *W08E12.3*, *W08E12.4* and *W08E12.5* are all present on chromosome IV and are in consecution to each other.

3.4 Conclusion

The existence and order of *W08E(12.2-12.5)* isoforms on the worm genome was confirmed. The occurrence of similar cluster of genes in proximity supports the notion that the architecture of the *C. elegans* genome is not randomly organised. The amino acid sequence analysis of the four isoforms revealed a high proportion of cysteine residues suggesting metal binding properties of the proteins. The promoter analysis of the *W08E(12.2-12.5)* genes indicated the presence of putative Zn coordinating sites also found in metallothioneins, thereby suggesting that the *W08E(12.2-12.5)* family may be involved in metal binding, metal homeostasis and /or metal detoxification.

Chapter 4: Localization and characterization of *W08E(12.3/12.4)*

4.1 Introduction

4.1.1 Creating single copy transgenic strain

In 1991, Craig Mello and his colleagues developed a method to introduce DNA into *C. elegans* where the DNA is injected into the gonad of the worms and concatenates to form an extrachromosomal array which is incorporated into the nucleus. As the chromosomes in *C. elegans* are holocentric, a part of the DNA acts like the centromere and the extrachromosomal array is duplicated and distributed to the daughter cells.

An alternative to the extra-chromosomal array is the creation of in frame transgenes. These are integrated into the genome of the worm by means of transposons, either by biolistic transformation or template directed repair, following transposon excision by homologous recombination (Berezikov et al., 2004). *C. elegans* transposons, such as Tc1, are not used to introduce transposon mediated strand breaks because of the high copy number of the Tc1 transposon within the genome which could ultimately cause strand breaks at multiple sites (Barrett et al., 2004).

Hence in order to introduce single copy transgenes, the Mos1 transposable element from *Drosophila* was introduced into the nematode (Bessereau et al., 2001). The transposon introduces a break in the genome at a specified location where it does not interfere or disrupt the function of the neighbouring genes and also where other promoters or enhancers do not regulate expression of the transgene (Robert and Bessereau, 2007).

The technique is known as Mos1 excision-induced gene conversion (MosTIC). This system has facilitated the insertion of tags or caused deletions in specified genes. A large library of Mos1 inserts with defined locations in the worm genome has been created by NemaGENETAG consortium (Duverger et al., 2007).

Transgenic cop-136(P*W08E12.3/4*::GFP) worms

To allow the characterization of the *W08E(12.2-12.5)* gene family, worms with a single copy inserted transgene were created. Among the four isoforms, the promoter region of *W08E12.3* and *W08E12.4* were selected, as both isoforms share identical promoters and hence the use of P*W08E12.3/4*::GFP transgene will identify the location of expression and the functional aspects of both *W08E12.3* and *W08E12.4* isoforms. The transgenic strain was constructed by KnudraTransgenics, a company based in Utah, U.S.A. The transgenic worms were created by adapting the Mos1 mediated single copy insertion technique (MosTIC), to insert a single copy at a specified locus which drives expression in a broad range of tissues at endogenous levels (Frokjaer-Jensen et al., 2008).

Prior to MosTIC transposition an extra chromosomal array was created containing homologous chromosomal DNA from flanking sequence ends of the Mos1 element. In order to identify the transgene insertion into the chromosome, a positive marker was also integrated in between the flanking sequences of the extra chromosomal array. The positive marker causes a rescue phenotype of *unc-119*, a strain that is small and almost paralysed and the injection of an *unc-119* wild type gene along with the transgene will rescue the severe phenotype and allows the selection of transgene integration (Frokjaer-Jensen et al., 2008).

In this strategy, the Mos 1 element introduces a double stranded break. While making the extrachromosomal array a sequence of Mos1 transposase was integrated under a heat shock promoter. Heat shock activates production of transposase which excises the transposon and produces a double stranded break which is repaired by gene conversion using the extra chromosomal array containing the transgene as a template. In order to distinguish between worms containing the extra chromosomal array from the ones having the transgene fused to the genome, a negative selection marker was used namely the gene expressing a red fluorescent proteins and TWK-18(gf), an activated K+ channel that leads to muscle paralysis at elevated temperatures. These negative markers are incorporated into the transgene arrays but not copied into target site integration. Hence the worms carrying the array with negative markers are active at 15°C but paralyzed at 25°C. Only the worms which did not lose the array do not get paralyzed (Frokjaer-Jensen et al., 2008). The above mentioned method was applied to generate a transgenic strain with an extrachromosomal array of the **P*W08E12.3/4*::GFP** transgene.

4.1.2 Creating an extrachromosomal array with multiple copies of the transgene

Transgenic strains of *C. elegans* are used for various applications. They are generally created by injecting DNA into the cytoplasm of the gonad in a syncytial hermaphrodite (Berkowitz et al., 2008). The DNA fragment containing the transgene of interest is coinjected with plasmids containing coinjection markers (Hobert, 2002, Evans, 2006). The markers are also established transgenes which either code for fluorescence or a specific phenotype. The successful transgenic lines are picked from the F2 generation of the injected worms. However partial and variable transmission rates (10-100%) of extrachromosomal arrays in the next generation can be observed in the transgenic

lines. The segregation of the extrachromosomal arrays during cell divisions decide their fate in being transmitted into both the germ line and somatic cell line (Mariol et al., 2013).

This method can be unstable during replication (Frokjaer-Jensen et al., 2008), and sometimes leads to mosaic expression of the transgene or over expression of the gene due to large number of copies being present, resulting in possible dominant negative and/or toxic effects (Thellmann et al., 2003). With the extrachromosomal arrays there is a risk of drift of expression over several generations which might arise eventually due to structural changes in the array or heritable silencing of the array (Hammarlund et al., 2007, Sha and Fire, 2005).

In the current study, in order to create an extrachromosomal overexpressing transgenic strain, copies of P*W08E12.3/4*::GFP were microinjected into the gonad of nematodes. The F2 generation of the injected worms were screened for successful transformants and maintained as single lines. Screening was performed in order to detect the nematodes with high GFP intensity using an inverted fluorescence Nikon microscope (Nikon Eclipse FE2000-S).

4.2 Results

4.2.1 Location of expression of cop-136(P*W08E12.3/4*::GFP) within the nematode

The transgenic strain cop-136(P*W08E12.3/4*::GFP) created by KnudraTransgenics, was grown up to L4 stage and visualised under an inverted fluorescence Nikon microscope, using a blue laser scanning fluorescence (λ_{ex}= 450-490 nm) with exposure time 1.5 seconds and gain 9.6 seconds.

Within the worm the expression of *W08E12.3* and *W08E12.4* is concentrated in the pharynx and a faded fluorescence signal around the intestinal region can be observed (**Figure 4-1**). The pharynx of the nematode is divided into the corpus, isthumus and terminal bulb area (**Figure 4-2A**). The expression of the *W08E12.3/4* is predominantly in the procorpus and the isthumus area. There is significant expression of the transgene in the posterior terminal bulb area in the dorsal epithelial pm8 cells present above the pharyngeal-intestinal valve (**Figure 4-2B**). The faded fluorescence signal observed in the outer intestinal rings of the nematode can be attributed to auto fluorescence generated due to high exposure and gain settings of the microscope.

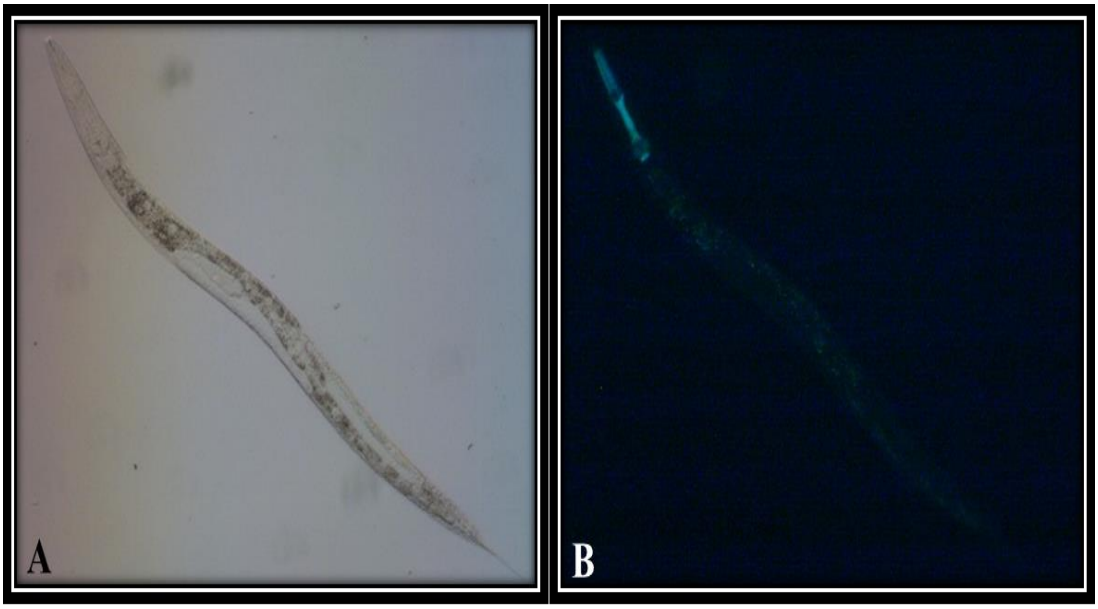


Figure 4-1: (A) cop-136(P*W08E12.3/4*::GFP) worm imaged by phase contrast imaging at 10X magnification (B) cop-136(P*W08E12.3/4*::GFP) worm imaged by inverted fluorescence microscopy (Nikon), using a blue laser scanning fluorescence (λ_{ex}= 450-490 nm) at 10X magnification under UV light.

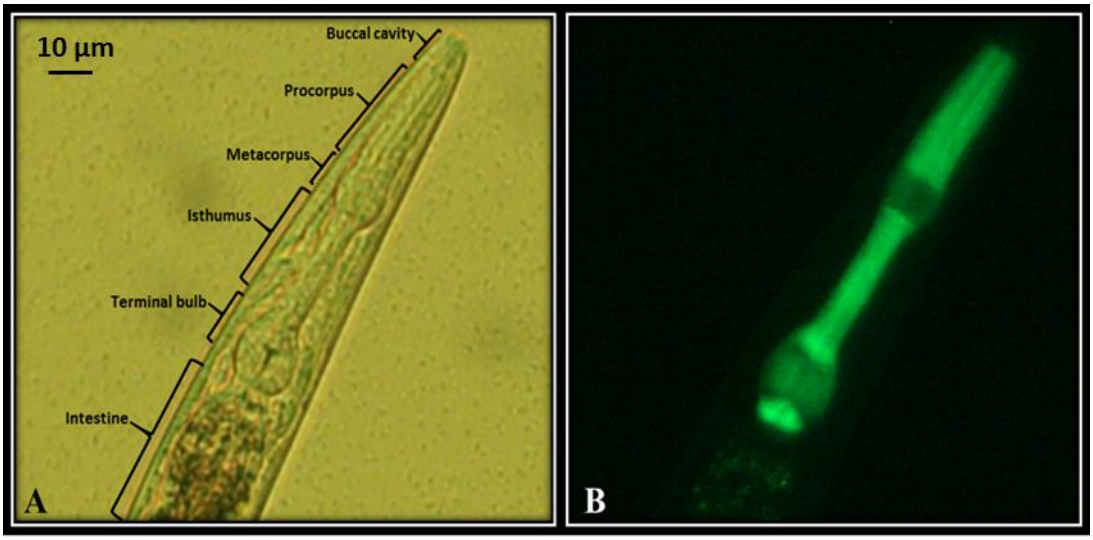


Figure 4-2: (A) cop-136(P*W08E12.3/4*::GFP) worm imaged by differential interference contrast (DIC) imaging at 40X magnification with description of the pharynx (B) cop-136(P*W08E12.3/4*::GFP) worm imaged by inverted fluorescence microscopy (Nikon), using a blue laser scanning fluorescence (λ_{ex}= 450-490 nm) at 40X magnification under UV light (scale: 10 µm).

4.2.2 Expression studies of cop-136(P*W08E12.3/4*::GFP) within the nematode on exposure to metals

As the *W08E(12.2-12.5)* family displayed high cysteine content (13%) which is a trait of metal binding proteins the cop-136(P*W08E12.3/4*::GFP) transgenic strain was exposed to defined concentrations of Zn, Cd and Cu at L1 stage in order to test the responsiveness of *W08E12.3* and *W08E12.4* isoforms from L1 stage to late adults to heavy metals. It was observed that the GFP expression was most defined at L4 stage and was negligible at L1-L3 stage, hence all images were taken at L4 stage using a blue laser fluorescence (λ_{ex}= 450-490 nm) and the fluorescence intensity of each worm was quantified using the ImageJ software. The entire pharynx of the nematode was outlined, the fluorescence within that region measured, and the mean fluorescence determined.

Alternatively, metal responsiveness of the transgenic worms were measured by a bulk assay where the fluorescence of the metal exposed L4 transgenic worms was measured using a fluorescence plate reader.

4.2.2.1 Bulk assay

In order to determine the optimal number of worms required per well of the 96 well black microplate to obtain a considerable fluorescence signal, a test run was performed using the cop-136(P*W08E12.3/4*::GFP). At L4 stage the worms were washed off the plates and a titre provided an estimate of the total number of worms remaining after the washing step. Defined worm counts were aliquoted in repeats (n=3) on the plate and the absorbance measured at 355 nm excitation and 510 nm emission using a monochromator fluorescent plate reader.

The wells with 10,000 worms had the highest fluorescence (1.78 a.u.) in comparison to the other worm counts used in this assay (**Figure 4-3**). As the purpose of using the fluorescence plate reader was to record the fluorescence change in cop-136(P*W08E12.3/4*::GFP) worms on exposure to heavy metals like Cd, Cu, Zn, the optimal worm count for the plate reader assays was deemed to be 5000 worms as this number will accommodate an increase or decrease in the fluorescence of the cop-136(P*W08E12.3/4*::GFP) on exposure to metals.

Once the worm count/well was optimised, the cop-136(P*W08E12.3/4*::GFP) worms were exposed to heavy metals (Cd, Cu and Zn). Their GFP expression was followed from L1 stage to late adults. Worms were subjected to chronic metal exposure from L1-L4 stage (48 hours) but also more acutely from late L2-L4 stage (24 hours).

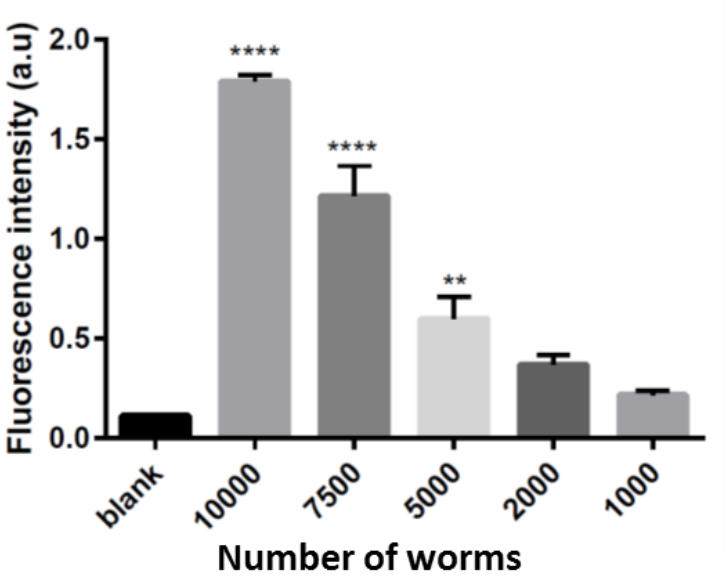


Figure 4-3: Fluorescence quantification (355 nm excitation and 510 nm emission) of defined numbers of cop-136(P*W08E12.3/4*::GFP). (The statistical analysis was performed using one-way ANOVA and Dunnett's multiple comparison test (n=3); **** $P < 0.0001$, ** $P < 0.005$, error bars: SD. The significance was determined in comparison to the blank i.e. the wells without worms on the plate).

In order to quantify the fluorescence of the cop-136(P*W08E12.3/4*::GFP) strain, the concentration of the metal was limited to doses that did not affect the development significantly. In order to detect the changes in the fluorescence of cop-136(P*W08E12.3/4*::GFP) strain on exposure to Cd a dose range was set up with a maximum dose to 50 µM Cd.

There was a significant increase in fluorescence in worms exposed to 10 µM Cd compared to the control (unexposed) worms (**Figure 4-4**). Overall, it can be stated that there was a significant increase in the fluorescence of the cop-136(P*W08E12.3/4*::GFP) strain on exposure to 10 µM Cd for 48 hours but fluorescence decreased in a dose dependent manner at higher concentrations.

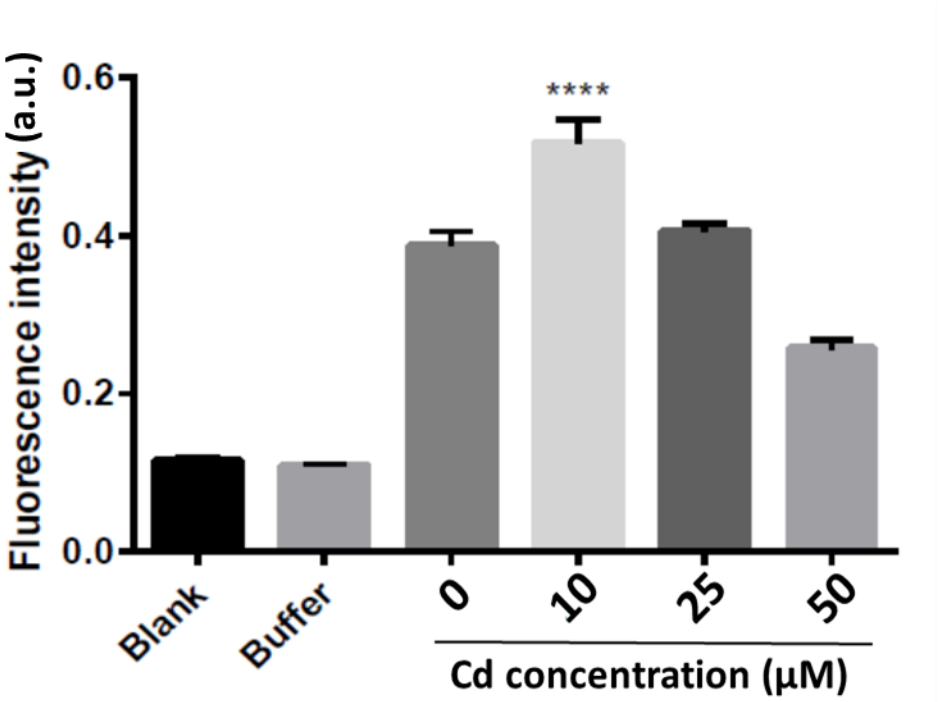


Figure 4-4: Fluorescence quantification of cop-136 (P*W08E12.3/4*::GFP) strain on exposure to different concentrations ranging from 0 µM- 50 µM cadmium for 48 hours. The transgenic worms were exposed to Cd from L1-L4 stage (48 hours). The fluorescence of the nematodes was measured at 355 nm excitation, 510 nm emission monochromator of a fluorescence plate reader; n = 3 x 5000 worms / condition. Blank was the fluorescence emitted by empty wells and the buffer control was the wells filled with the M9 buffer. (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

In order to test the responsiveness of the cop-136(P*W08E12.3/4*::GFP) strain to other heavy metals, the worms were exposed from L1 to L4 stage (i.e. for 48 hours) to Zn or Cu. Chronic exposure to Zn changed the fluorescence intensities of the transgenic worms in comparison to that of the control (unexposed). The worms exposed to 100 µM Zn displayed the highest fluorescence. The fluorescence displayed by 0 µM, 400 µM and 600 µM Zn exposed worms were statistically no different to controls (**Figure 4-5A**). The response of cop-136(P*W08E12.3/4*::GFP) to Cu was negative, as overall fluorescence decreased significantly in worms challenged with 400 µM or 600 µM Cu

(**Figure 4-5B**). Altogether, it can be stated that the cop-136(P*W08E12.3/4*::GFP) strain is metal responsive in nature.

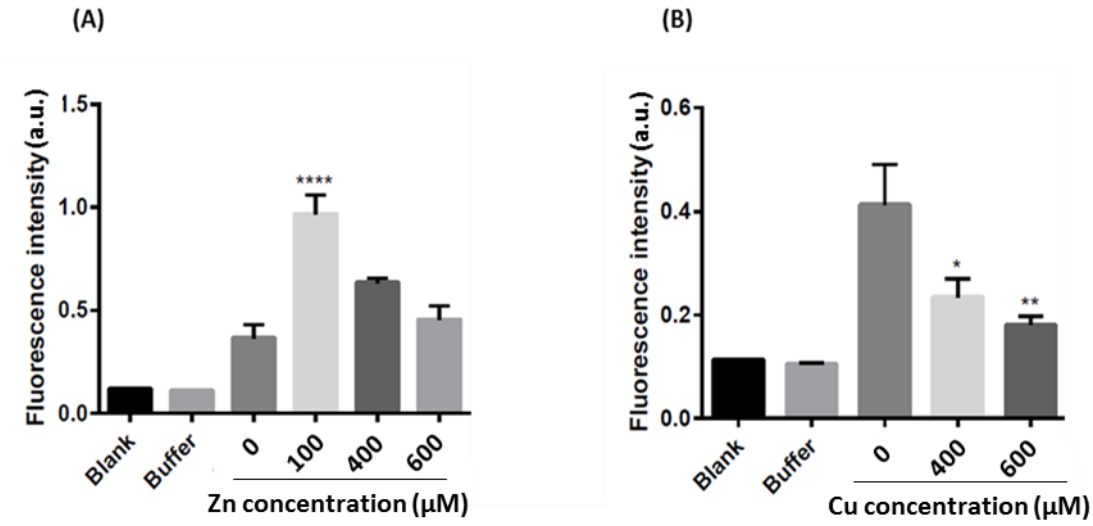


Figure 4-5: Fluorescence quantification of cop-136(P*W08E12.3/4*::GFP) on exposure to different concentrations of metals for 48 hours. The transgenic worms were exposed to either Zn (A) or Cu (B) from L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes was measured at 355 nm excitation, 510 nm emission monochromator using a fluorescence plate reader; n = 3 x 5000 worms / condition. Blank was the fluorescence emitted by empty wells and the buffer control was the wells filled with the M9 buffer. (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test, **** $P < 0.0001$; * $P < 0.05$; ** $P < 0.005$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

The concept of testing the *W08E(12.2-12.5)* isoforms for metal responsiveness arose from the fact that the proteins coded by them are cysteine rich (13%). The number of cysteines within these proteins is similar to that of metallothioneins and also the location of the P*W08E12.3/4*::GFP transgene and *mtl-1* overlap i.e. in the pharynx. Taken together this suggests that the *W08E(12.2-12.5)* family might be functionally linked to metallothioneins. In order to test the relationship between metallothioneins and the *W08E(12.2-12.5)* family, the cop-136 (P*W08E12.3/4*::GFP) strain was crossed with the metallothionein double knockout strain (*mtl-1(tm1770);mtl-2(gk125)*).

Following a series (four) of back crossing the selected mutants with the desired genotype i.e. P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*)* were screened to produce a population containing homogenous worms containing the P*W08E12.3/4*::GFP transgene in a metallothionein double knockout background (**Figure 4-6**). These worms were exposed to defined Cd concentrations.

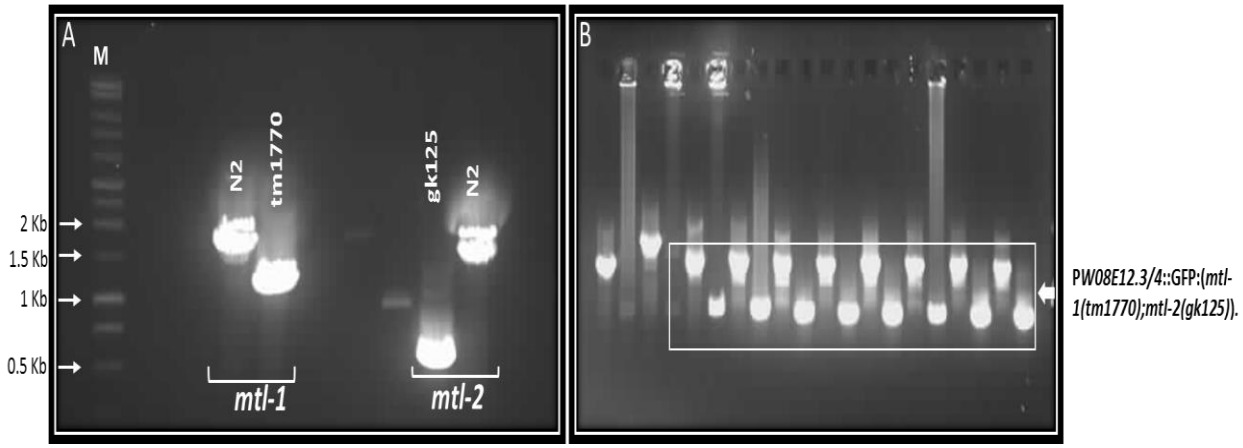


Figure 4-6: (A) Nested PCR products of N2 (wild type) vs. the metallothionein double knockout (*mtl-1 (tm1770);mtl-2(gk125))* on a 0.8% agarose gel against a 1Kb ladder. (B) Nested PCR screening for GFP tagged double knock out transgene cop-136(P*W08E12.3/4*::GFP);*mtl-1(*tm1770*);mtl-2(*gk125*)* on a 0.8% gel. The box represents the successful GFP tagged double knock out transgenic worms, while the upper panel of fig (B) represents the worms from failed crosses lacking the metallothionein double knockout (*mtl-1(*tm1770*);mtl-2(*gk125*)*) genotype.

The strain cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*)*) was exposed to 10 µM and 30 µM Cd respectively for 48 hours i.e. from L1-L4 stage. There was a decrease in fluorescence on exposure to Cd in a dose dependent manner (**Figure 4-7**). Whilst cop-136(P*W08E12.3/4*::GFP) responded positively when challenged with 10 µM Cd, this was lost upon deletion of the metallothioneins. However, it is evident that the worms exposed to 30 µM Cd display significantly less fluorescence in comparison to the unexposed worms. Hence, it appears that cop-136(P*W08E12.3/4*::GFP;*mtl-1*

*(*tm1770*);mtl-2(*gk125*))* is following the same trend as the cop-136(P*W08E12.3/4*::GFP) strain at least upon exposure to 30 µM Cd. In order to validate the result obtained, the experiment was repeated by exposing simultaneously the cop-136(P*W08E12.3/4*::GFP) strain and the mutant cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*))* to 30 µM Cd.

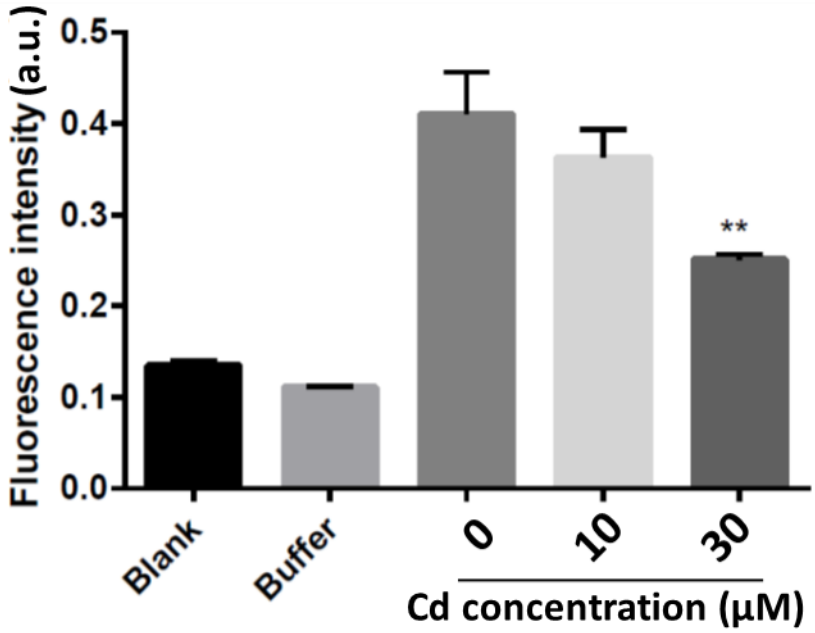


Figure 4-7: Fluorescence quantification of cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2 (*gk125*)) strain on exposure to different concentrations of Cadmium (0-30 µM) for 48 hours. The transgenic worms were exposed to Cd from L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes was measured at 355 nm excitation, 510 nm emission monochromator of a fluorescence plate reader; n = 3 x 5000 worms / condition. Blank was the fluorescence emitted by empty wells and the buffer control was the wells filled with the M9 buffer. (The statistical analysis was performed by means of a one-way ANOVA and Dunnett's multiple comparison test, ** $P < 0.005$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

This confirmed that on exposure to 30 µM Cd both cop-136 (P*W08E12.3/4*::GFP) strain and cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*))* display a

decrease in overall fluorescence (**Figure 4-8**). This reduction in fluorescence may be attributed to the toxicity that is caused by Cd.

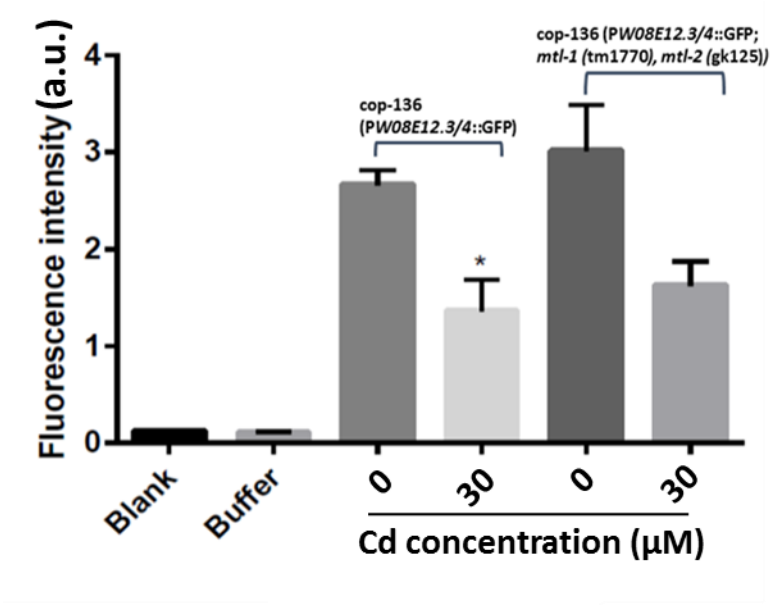


Figure 4-8: Fluorescence quantification of cop-136(P*W08E12.3/4*::GFP) strain and cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*))* strain on exposure to 30 µM Cd. The transgenic worms were exposed to Cd from L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes was measured at 355 nm excitation, 510 nm emission monochromator of a fluorescence plate reader; n = 3 x 5000 worms / condition. Blank was the fluorescence emitted by empty wells and the buffer control was the wells filled with the M9 buffer. (The statistical analysis was performed by means of a one-way ANOVA and Dunnett's multiple comparison test, * $P < 0.05$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

4.2.2.2 Single worm imaging

The transgenic P*W08E12.3/4*::GFP tagged worms were exposed to heavy metals from stage L1 stage. They were imaged at stage L4 as the GFP expression was most defined at that particular stage. For imaging, L4 worms (4-8) were picked onto a glass slide

carrying a drop of M9 solution to which sodium azide (2%) was added in order to immobilize the worms. Once the worms were immobile a cover slip was placed on top to cover the drop taking care not to crush the worms. The worms (n=15) were then imaged using an inverted fluorescence microscope (Nikon), using a blue laser scanning fluorescence (λ_{ex}= 450-490 nm) and standard phase contrast imaging. The images were captured using a Nikon camera (Nikon UK Ltd., Kingston upon Thames, UK) and analysed by ImageJ software.

On exposure to 30 µM Cd, the cop-136(P*W08E12.3/4*::GFP) strain shows a significant decrease in fluorescence intensity in comparison to the control (unexposed) worms (**Figure 4-9**). This trend is comparable to the trend observed in case of the bulk assay (**Figure 4-8**). Hence, this data reconfirms that on exposure to 30 µM Cd cop-136(P*W08E12.3/4*::GFP) strain is characterized by a reduction in overall fluorescence.

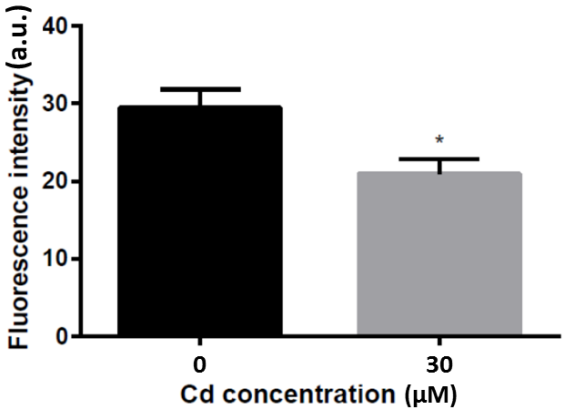


Figure 4-9: Fluorescence quantification of cop-136(P*W08E12.3/4*::GFP) strain on exposure to 0 µM and 30 µM Cd. The transgenic worms were exposed to Cd from L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes (n=15) was imaged with an inverted fluorescence microscope (Nikon), using a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence. (The statistical analysis was performed using one-way ANOVA and by means of an unpaired t-test; * $P < 0.05$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

Upon exposure to high doses of Zn and Cu (400 µM – 600 µM) cop-136(P*W08E12.3/4*::GFP) showed an overall dose dependent decrease in fluorescence (**Figure 4-10**). Although, the decrease in the fluorescence intensity was not statistically significant when compared to the unexposed worms. The reason behind the insignificance was attributed to the amount of fluctuation in the fluorescence signal of the cop-136(P*W08E12.3/4*::GFP) within each condition.

The effect of Cd on cop-136(P*W08E12.3/4*::GFP) strain was also tested in a metallothionein null background. The cop-136(P*W08E12.3/4*::GFP) strain was fed with RNAi bacteria directed against *mtl-1* and *mtl-2*.

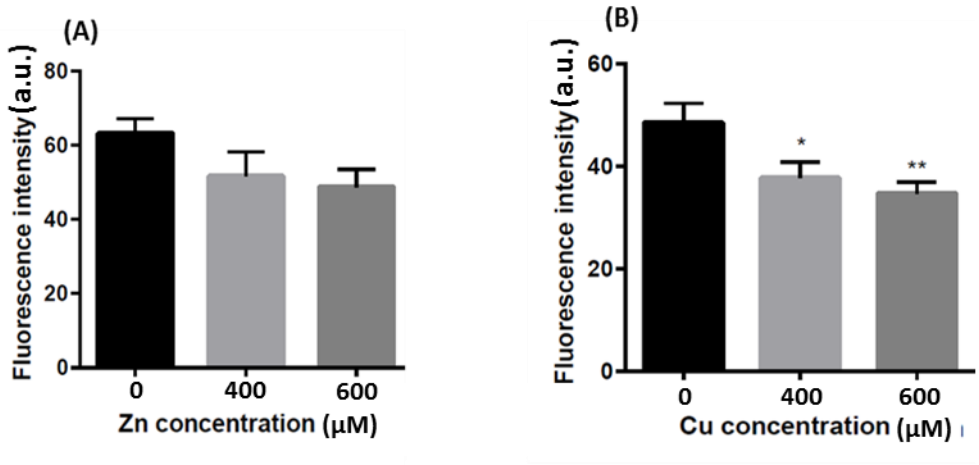


Figure 4-10: Fluorescence quantification of cop-136(P*W08E12.3/4*::GFP) strain on exposure to defined concentrations of heavy metals Zn or Cu for 24 hours. The transgenic worms were exposed to Zn (A) or Cu (B) from late L2-L4 stage i.e. for 24 hours. The fluorescence of the nematodes (n=15) was imaged with an inverted fluorescence microscope (Nikon), using a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence. (The statistical analysis was performed by means of a one-way ANOVA and Dunnett's multiple comparison test; * $P < 0.05$, ** $P < 0.005$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

The knock down of *mtl-1* and *mtl-2* within the cop-136(P*W08E12.3/4*::GFP) strain lead to a significant decrease in the fluorescence signal (**Figure 4-11**) which is

comparable to the cop-136 (P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*))* bulk assay data (However the baseline was higher in comparison to the wild type background).

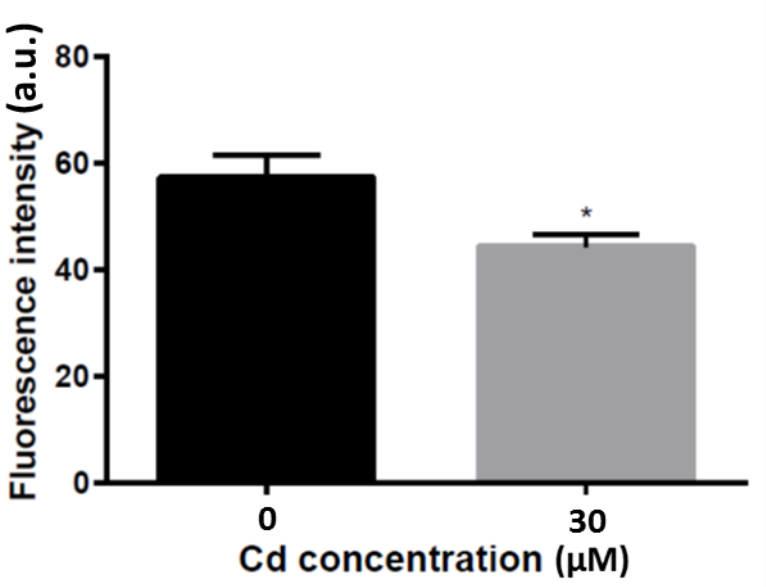


Figure 4-11: Fluorescence quantification of cop-136(P*W08E12.3/4*::GFP), in a RNAi mediated metallothionein knock down background on exposure to 30 µM Cd. The transgenic knock down worms were exposed to Cd from late L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes (n=15) was imaged using an inverted fluorescence microscope (Nikon), with a blue laser ((λ_{ex}= 450-490 nm) (The statistical analysis was performed using an unpaired t-test; * $P < 0.05$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

In order to test if the transgene signal fluctuated, the cop-136(P*W08E12.3/4*::GFP) strain was screened over 3-4 generations and the fluorescence between the progeny from the same parent was recorded.

Although the offspring from the same parent of a single copy transgenic worm would be expected to have similar fluorescence, significant degree of fluctuation was observed within the same progeny of cop-136(P*W08E12.3/4*::GFP) (**Figure 4-12**). The instability in the fluorescence signal of the cop-136(P*W08E12.3/4*::GFP) was

reasoned to be due to extremely low fluorescence signal generated by the single copy transgene cop-136(P*W08E12.3/4*::GFP).

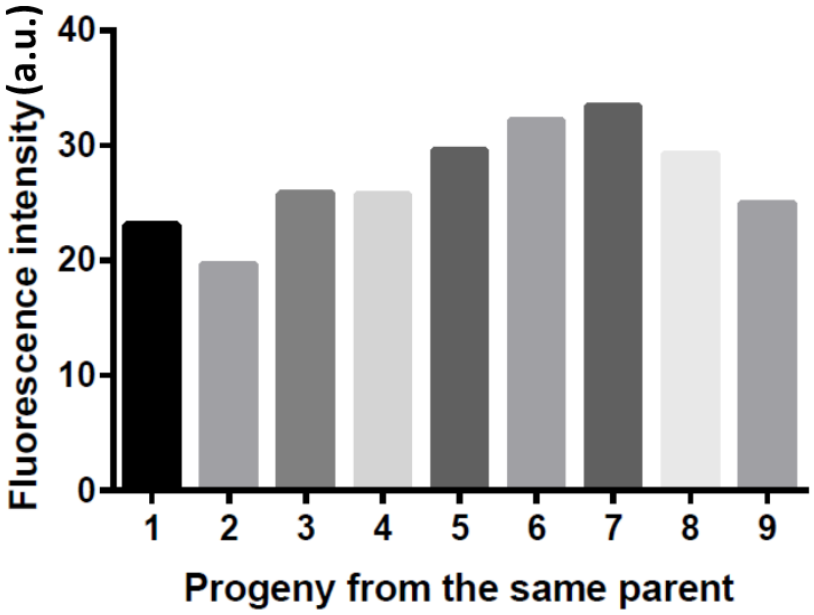


Figure 4-12: Fluorescence quantification of individual worms of cop-136(P*W08E12.3/4*::GFP) strain generated from the same parent. The progeny of a single parent were separated into individual wells of the 12 well plate. The isolated offspring were then imaged at L4 stage using an inverted fluorescence microscope (Nikon), by means of a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence.

In order to overcome the afore mentioned problem of low fluorescence in individual worms due to a single copy transgene, a transgenic strain zsEx6(P*W08E12.3/4*::GFP) comprising of an extrachromosomal array of P*W08E12.3/4*::GFP was generated.

The expression of the P*W08E12.3/4*::GFP transgene was reconfirmed to be confined to the pharynx of the nematode (**Figure 4-13**). The zsEx6(P*W08E12.3/4*::GFP) lines **(Figure 4-13B)** displayed a much stronger fluorescence signal in comparison to the single copy transgenic worm cop-136 (P*W08E12.3/4*::GFP) (**Figure 4-13A**). Individual worms depicted partial expression of the transgene i.e. displaying mosaicism (**Figure 4-13C**). In order to overcome the interference in the experiments

due to mosaic expression of the transgene, the lines were manually screened to select only the non-mosaics. While screening for the non-mosaics it was observed that along with the worms there was some detectable fluorescence being depicted by the eggs laid by them on the NGM plate.

Figure 4-13: Comparison of fluorescence signal between the cop-136(P*W08E12.3/4*::GFP) and the different lines zsEx6(P*W08E12.3/4*::GFP) strain. Multiple copies of P*W08E12.3*::GFP transgene was injected into the gonad of the nematode. The transgenic worms were imaged under an inverted fluorescence Nikon microscope, using a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence. Exposure: 1/60 seconds; Gain : 9/60 seconds. (A) Worm with a single genome integrated copy of P*W08E12.3/4*::GFP (B) Highly fluorescent worm with multiple copies of the transgene (C) Line depicting mosaic transgenic expression.

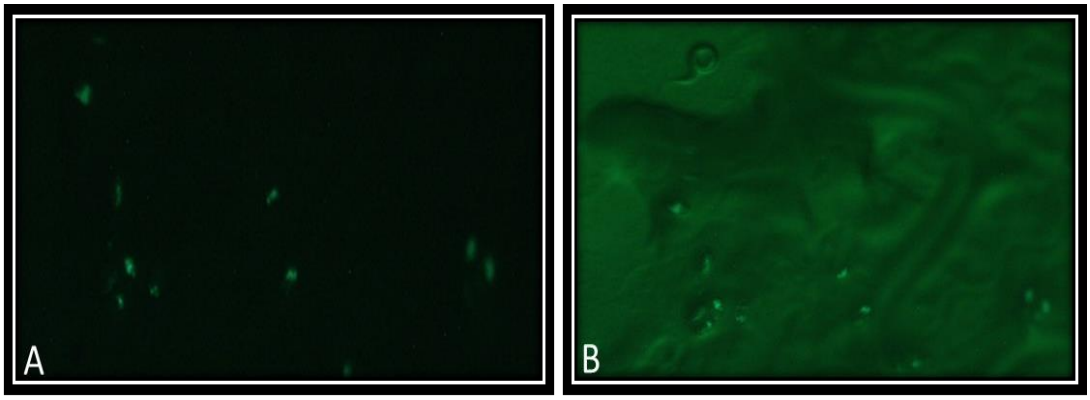


Figure 4-14: Fluorescence observed within the eggs of the zsEx6(P*W08E12.3/4*::GFP) strain using an inverted fluorescence microscope. When the worms were being screened for non-mosaics it was observed that the eggs of a certain stage comprising of the embryo depict fluorescence. (A) Image taken at lower exposure setting to avoid auto fluorescence. (B) Image taken at higher exposure settings in order to confirm that the fluorescence was defined to the eggs.

The nematode embryo expresses the transgene at a defined location (**Figure 4-14**). Hence, this suggests that the transgenic expression starts from the embryonic stage and carries on until adulthood.

The zsEx6(P*W08E12.3/4*::GFP) transgenic strain was exposed to 30 µM Cd to test if it produced comparable results to that of the cop-136(P*W08E12.3/4*::GFP) strain. On exposure to 30 µM Cd, the zsEx6(P*W08E12.3/4*::GFP) strain depicted a significant decrease in fluorescence in comparison to the control (unexposed) worms (**Figure 4-15**). This data is comparable to the single worm imaging data of cop-136(P*W08E12.3/4*::GFP) strain. The change in fluorescence is statistically highly significant in case of the zsEx6(P*W08E12.3/4*::GFP) implying less fluctuation between worms and the fluorescence signal of zsEx6(P*W08E12.3/4*::GFP) in comparison to cop-136(P*W08E12.3/4*::GFP) strain.

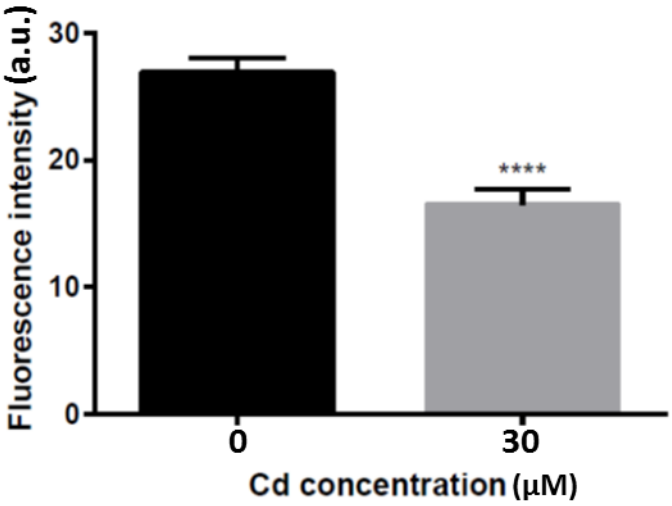


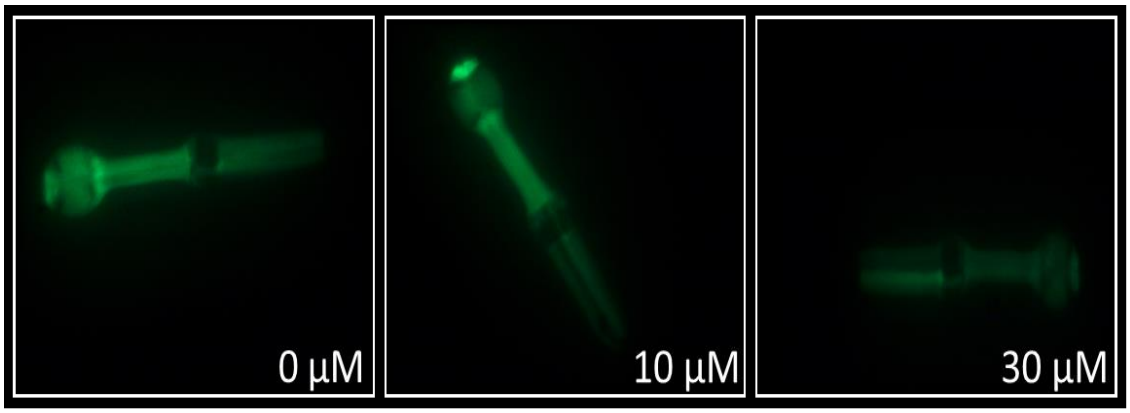
Figure 4-15: Fluorescence quantification of zsEx6(P*W08E12.3/4*::GFP) strain on exposure to 0 µM and 30 µM Cd. The transgenic worms were exposed to Cd from late L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes (n=15) was imaged using an inverted fluorescence microscope (Nikon), with a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence. (The statistical analysis was performed using a unpaired t-test; **** $P < 0.0001$; error bars: SD. The significance was determined in comparison to the unexposed (0 µM Cd) worms).

The response of P*W08E12.3/4*::GFP on exposure to Cd in a metallothionein knock down background was tested in zsEx6(P*W08E12.3/4*::GFP) to verify the results obtained from the cop-136(P*W08E12.3/4*::GFP) strain.

It is evident from the images of L4 zsEx6(P*W08E12.3/4*::GFP) worms that the transgenic worms in a metallothionein knock down background (**Figure 4-16B**) fluoresce higher than the wild type background (**Figure 4-16A**). The zsEx6(P*W08E12.3/4*::GFP) worms upon exposure to 10 µM Cd display a slight increase in fluorescence within the terminal bulb of the pharynx whereas, at 30 µM Cd there is an obvious decrease in the fluorescence intensity. By observing the zsEx6(P*W08E12.3/4*::GFP) worms in a metallothionein knock down back ground it appears that upon exposure to 10 µM Cd there was increased fluorescence in the pm8 cells of the terminal pharyngeal bulb in comparison to the control (unexposed) worm. When the zsEx6(P*W08E12.3/4*::GFP) (*mtl-1/mtl-2*)) knock down strain was exposed to 30 µM Cd a significant drop in the overall fluorescence in the pharyngeal area of the worm was observed (**Figure 4-16**).

Upon quantification of the fluorescence using ImageJ software (from **Figure 4-16**), it was observed that both zsEx6(P*W08E12.3/4*::GFP) strain and zsEx6(P*W08E12.3/4*::GFP) (*mtl-1/mtl-2*)) knockdown strain follow a similar trend, where fluorescence was comparable to worms exposed to 10 µM Cd but significantly reduced upon exposure to 30 µM Cd. However, the overall fluorescence signal displayed by the worms in a metallothionein knock down background was significantly higher than that of the zsEx6(P*W08E12.3/4*::GFP) strain in the case of 0 µM Cd and 10 µM Cd (**Figure 4-17**).

(A) zsEx6(P*W08E12.3/4*::GFP) in wild type background



(B) zsEx6(P*W08E12.3/4*::GFP) in metallothionein double knockdown background

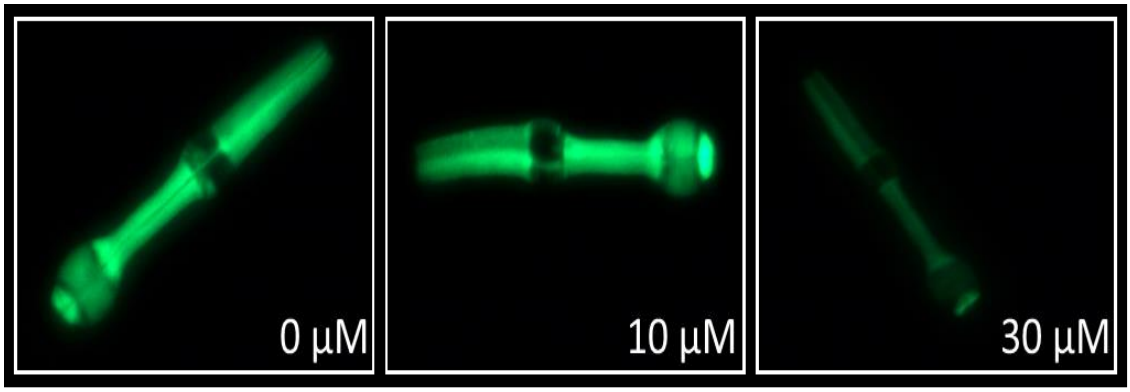


Figure 4-16: (A) Imaging L4 zsEx6(P*W08E12.3/4*::GFP) worms (B) Imaging of zsEx6(P*W08E12.3/4*::GFP) in metallothionein double knock down worms, using an inverted fluorescence microscope (Nikon), with a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence upon exposure to defined concentrations of Cd (0 µM, 10 µM, 30 µM).

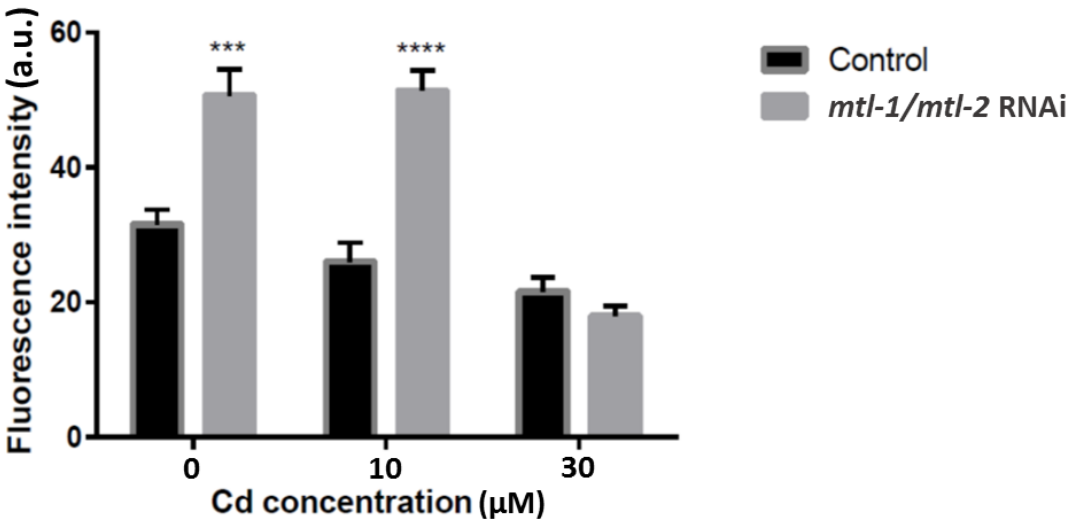


Figure 4-17: Fluorescence quantification of zsEx6(P*W08E12.3/4*::GFP) strain and zsEx6(P*W08E12.3/4*::GFP) in a metallothionein double knockdown background on exposure to 10 µM and 30 µM Cd. The transgenic worms were exposed to Cd from L1-L4 stage i.e. for 48 hours. The fluorescence of the nematodes (n=15) was imaged using an inverted fluorescence microscope (Nikon), with a blue laser ((λ_{ex}= 450-490 nm) scanning fluorescence. (The statistical analysis was performed using two-way ANOVA and Bonferroni's multiple comparison test; *** P < 0.005; **** P <0.0001; error bars: SD). The significance was determined in comparison to the unexposed (0 µM Cd) worms).

4.3 Discussion

One of the major advantages of using *C. elegans* as a model organism is that it has a transparent body due to which fluorescent markers can be used to study mechanisms like embryogenesis, axon growth, and fat metabolism within the living organism (Kaletta and Hengartner, 2006). *C. elegans* has proved itself to be a potent model to study toxicity and toxicological mechanisms of various heavy metals like As, Cd, Zn, Cu, Pb by either facilitating toxicity end point studies or via tracking reporter transgenic expression within the worm (Leung et al., 2008). This trait of *C. elegans* being transparent was exploited in the current study to trail the expression of

W08E12.3 and *W08E12.4* gene within the worm by following a reporter (GFP) gene expression. In order to study the location of expression of *W08E12.3* and *W08E12.4* within the worm, a transgenic strain cop-136(P*W08E12.3/4*::GFP) was created by Knudra Transgenics. Out of the four isoforms this promoter region was chosen to create the transgene as it is identical in *W08E12.3* and *W08E12.4*. In cop-136 (P*W08E12.3/4*::GFP) the promoter region upstream of *W08E12.3/4* gene was fused with a GFP and microinjected into the worm to create a single copy integrated array within the worm genome. Hence, by tracking the expression of P*W08E12.3/4*::GFP transgene both *W08E12.3* and *W08E12.4* can be characterized. By tracing the GFP expression within the nematode the site of expression of *W08E12.3/4* gene was found to be confined to the pharyngeal region. Within the pharynx there was significant expression of the transgene in the posterior terminal bulb area, especially in the dorsal epithelial pm8 cells present above the pharyngeal-intestinal valve (Rasmussen et al., 2008). The P*W08E12.3/4*::GFP transgene also showed expression in the embryonic stage of the nematode. This is similar to the expression of P*mtl-2*-GFP in the final stages of embryogenesis of the nematode upon exposure to Cd (Swain et al., 2004).

The location of P*W08E12.3/4*::GFP transgene was found to be overlapping with the site of expression of *mtl-1*, which is constitutively expressed in the second pharyngeal bulb in the nematode acting as a metal sensor. Viable GFP expressing transgenes helped in defining the function of metallothioneins in response to cadmium (Swain et al., 2004). Given the overlapping expression, suggests a possible functional similarity between *mtl-1* and *W08E12.3/4*. Hence, metal responsiveness of the cop-136(P*W08E12.3/4*::GFP) strain was tested by exposing the transgenic worms to defined concentrations of heavy metals like Cd, Zn and Cu.

First, the response of cop-136(P*W08E12.3/4*::GFP) strain to heavy metals was tested via a bulk assay where 5000 worms were used to generate a fluorescence signal that could be detected by a fluorescent plate reader. The bulk assays revealed that the fluorescence by cop-136(P*W08E12.3/4*::GFP) strain was up regulated upon acute and chronic exposure to 10 µM Cd whereas at higher doses of Cd the fluorescence decreased in comparison to the unexposed worms. A similar trend was observed in Zn exposed animals, but fluorescence decreased upon exposure to Cu.

The reduction in fluorescence in response to higher doses of heavy metals can be due to the toxic effects caused to the nematodes. No upregulation to Cu exposure could be due to other existing The Cu detoxifying mechanisms in the worm is abolished due to the presence of the unique C-terminal histidine in CeMT-2 (Bofill et al., 2009) which possibly also suggests the presence of a non-selectivity towards Cu in *W08E12.3/4* (which also contain histidine residues). Although the down regulation of the fluorescence upon Cu exposure is predominantly due to the high Cu concentrations resulting in toxic effects, leading to the shrinkage of body size of the worms. It was observed that high concentrations of heavy metals caused delay or halt the development of the nematodes which might influence the overall fluorescence signal generated by the exposed nematode. Concentrations of Cd were confined to 10 µM and 30 µM as the higher concentrations (50 µM) caused immense toxicity.

When the mutant cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*))* strain was created to test the responsiveness of *W08E12.3/4* to heavy metals in absence of metallothioneins, it was observed that upon exposure to Cd (10 µM, 30 µM) the fluorescence of the worms decreased. However, under control (metal devoid) conditions the cop-136(P*W08E12.3/4*::GFP;*mtl-1(*tm1770*);mtl-2(*gk125*))* appeared to

fluoresce more than that of the cop-136(P*W08E12.3/4*::GFP) wild type transgene, again suggesting an interplay between metallothioneins and *W08E12.3/4.*

Single worm imaging was carried out to quantify the fluorescence of individual worms exposed to heavy metals and determine if the data generated replicate the bulk assay. The metal responsiveness of P*W08E12.3/4*::GFP to Cd seemed to yield similar results to that of the bulk assay i.e. there was a significant reduction in signal upon exposure to 30 µM Cd or higher concentrations and the fluorescence signal reduced when cop-136(P*W08E12.3/4*::GFP) was exposed to 400-600 µM Cu and Zn. The *mtl-1/2* RNAi cop-136(P*W08E12.3/4*::GFP) strain depicted a significant reduction in fluorescence when exposed to 30 µM Cd.

The fluctuations in the fluorescence signal by cop-136(P*W08E12.3/4*::GFP) prompted the creation of a transgenic worm strain zsEx6(P*W08E12.3/4*::GFP) containing an extrachromosomal array of P*W08E12.3/4*::GFP. Overcoming the issues of mosaicism by selecting specifically the non-mosaics, it was observed that zsEx6(P*W08E12.3/4*::GFP) strain displayed significantly stronger fluorescence signal in comparison to the cop-136(P*W08E12.3/4*::GFP) strain. Multiple copies of the transgene in comparison to a single integrated copy within the worm was the reason behind the high fluorescence of zsEx6(P*W08E12.3/4*::GFP). When zsEx6(P*W08E12.3/4*::GFP) was exposed to Cd it reconfirmed the data generated by cop-136(P*W08E12.3/4*::GFP) strain, namely the significant drop in the fluorescence upon exposure to 30 µM Cd. The zsEx6(P*W08E12.3/4*::GFP) strain further confirmed that upon exposure to 30 µM Cd in a metallothionein null background there was a significant drop in fluorescence generated by P*W08E12.3/4*::GFP. The zsEx6(P*W08E12.3/4*::GFP) metallothionein double knockdown strain also proved that the fluorescence depicted by P*W08E12.3/4*::GFP in metallothionein null background

(0 µM Cd, 10 µM Cd) is significantly higher than that generated in presence of metallothioneins.

Overall, the transgenic strains cop-136(P*W08E12.3/4*::GFP) and zsEx6(P*W08E12.3/4*::GFP) confirm that upon exposure to lower doses of Cd *W08E12.3/4* might be upregulated at specific doses (10 µM Cd) but higher concentration induce toxic effects to the nematode and hence, reduces the fluorescence intensity. Zinc exposure enhances the expression of *W08E12.3/4* at doses up to 100 µM but again higher doses do confer toxicity to the worms. The *W08E12.3/4* is downregulated in the presence of Cu (400 µM- 600 µM). When worms were exposed to 600µM Cu, 600µM Zn or 30 µM Cd, growth was stunted and they appeared to be starved i.e. slimmer and shorter, which was reflected in a reduced GFP signal. The other reason for the reduction of the GFP signal upon exposure to 30 µM Cd could be the unavailability of free Cd to induce the P*W08E12.3/4*::GFP promoter as observed in snail HpCdMT isoform upon quantitation with real time-PCR where, at the beginning of Cd exposure, the concentration of free Cd ions to interact with gene regulatory elements is higher in comparison to later stages where the majority of the Cd ions are sequestered by the CdMT protein (Höckner et al., 2011). It is conceivable that the stress mechanisms in the worm compensates the toxic effects caused by low concentrations of Cu, Zn and Cd but beyond a threshold heavy metal toxicity is apparent. The stress effects were more prominent in the worms exposed to highest condition of heavy metals implying that metal concentration is directly proportional to the toxicity of the metal.

In the absence of metallothioneins, the *W08E12.3/4* seemed to respond to metals similar to the wild type worms but the baseline expression of *W08E12.3/4* appeared to be significantly higher. This suggested that the *W08E(12.2-12.5)* series might have a

role in metal detoxification or metal homeostasis in absence of metallothioneins. Given that the expression of *mtl-1* and *W08E12.3/4* co-localize, suggests that there might be a functional relationship between the two genes. In the absence of metallothioneins, *W08E12.3/4* are upregulated probably to compensate for the role requiring metallothioneins.

4.4 Conclusion

By exploiting the transgenic strains cop-136(P*W08E12.3/4*::GFP) and zsEx6(P*W08E12.3/4*::GFP), it can be confirmed that the expression of *W08E12.3/4* takes place in the pharynx of the nematode and are metal responsive. The similarity with metallothioneins in the site of expression and upregulation of *W08E12.3/4* in absence of metallothioneins suggest a possible relationship between the two. Further investigations into the transcriptional regulation of the *W08E(12.2-12.5)* gene family by qPCRs will aid the more detailed characterization of this putative metal responsive gene family within the nematode.

Chapter 5: Quantitation of *W08E(12.2 -12.5)* and unravelling their role in metal handling

5.1 Introduction

In the previous chapter, the metal responsiveness of the *W08E(12.2-12.5)* family was evaluated based on transgenic strains, cop-136(P*W08E12.3/4*::GFP) and zsEx6(P*W08E12.3/4*::GFP). The promoter activity of the two genes (*W08E12.3/4)* provided an overview of the possible metal responsive role. The quantitation of transcriptional regulation of the *W08E(12.2-12.5)* in the presence of heavy metals by quantitative polymerase chain reaction (qPCR) is ideally suited to assess the amplitude of metal responsiveness. Ideally, the gene expression levels of individual isoforms would facilitate transcriptional characterization of the entire gene family. But, the high degree of identity between the individual isoforms of *W08E(12.2-12.5)* family poses a challenge for designing gene specific primer-probe combinations for the subsequent exploitation by qPCR.

The metallothioneins present in *C. elegans* are a prominent response pathway involved in chelating metal ions within the worm. The metallothioneins have been described as multipurpose proteins with a role in detoxifying Cd and Cu, the homeostasis of Zn and the scavenging of free radicals (Zeitoun-Ghandour et al., 2010) . *W08E12.3/4* isoforms displayed a possible relationship with metallothioneins making it interesting to further investigate the transcriptional response of the *W08E(12.2-12.5)* family to heavy metals. In order to further investigate the relationship between the two families, the fold change in gene expression of the *W08E(12.2-12.5)* isoforms were quantified by qPCR in the presence and absence of the metallothioneins.

Phenotypic studies followed by reverse genetic tools like RNAi were employed to determine the importance of the *W08E(12.2-12.5)* isoforms within the nematode. For evaluating the role of the *W08E(12.2-12.5)* family in the development of the nematode, and also to investigate their role in stress responsiveness, the phenotypic effect on growth and brood size was verified upon exposure to heavy metals in a *W08E(12.3-12.5)* knockdown background. RNAi of the *W08E(12.3-12.5)* isoforms was also exploited to test the accumulation of metals within the worm in the presence and absence of the isoforms using the X-ray laser generated by the linear accelerator at SLAC (Stanford Linear Accelerator Centre). Overall, it is aimed to unveil a comprehensive overview of the transcriptional characterization of the *W08E(12.2-12.5)* family.

5.2 Results

5.2.1 Quantitative real time polymerase chain reaction (qPCR)

Real-time PCR is the gold standard for detecting and quantifying genes, specifically due to its extensive dynamic range, high sensitivity of the assay and negligible carry over contamination risk (Yao et al., 2006). qPCR quantifies specific PCR amplicons (i.e. target genes) by following fluorescent reporters which are proportional to the amount of amplicons generated. The fluorescence signals can arise from ds-DNA binding fluorescent dyes (e.g. SYBR green) or sequence specific pairs of oligonucleotides labelled with specific fluorophores (Wilhelm et al., 2003). In the current study TaqMan probes were used, which are linear oligonucleotides fused to the fluorogenic dye 6- carboxyfluorescein (FAM) at the 5' end and the quencher tetramethylrhodamine (TAMRA) at the 3' end. Once the hybridisation between the dye and the specific target amplicon takes place, the probe is degraded by the

exonuclease activity of Taq polymerase, releasing the fluorescence from the reporter dye (Yao et al., 2006). The probe does not fluoresce when it is intact, therefore fluorescence is only emitted at 518 nm when the probe binds specifically to the amplicon of the target gene (Wilhelm et al., 2003).

The most challenging part was to design probe and primer combination which are specific to individual *W08E(12.3-12.5)* isoforms. Due to their high sequence similarity, there was a high possibility of cross reactivity between the isoforms. Following numerous trials it was concluded that although it was possible to design specific probe and primer combinations for *W08E12.2*, the other isoforms (*W08E12.3*, *W08E12.4* and *W08E12.5*) would have to be quantified together as one unit.

As described in the methods section, the nematodes were grown (in presence or absence of Cd) to L4 stage and total RNA was extracted from them. The extracted mRNA was transcribed into cDNA which was used to quantify the expression of *W08E(12.3-12.5)* isoforms after being normalised to the house keeping gene *rla-1*.

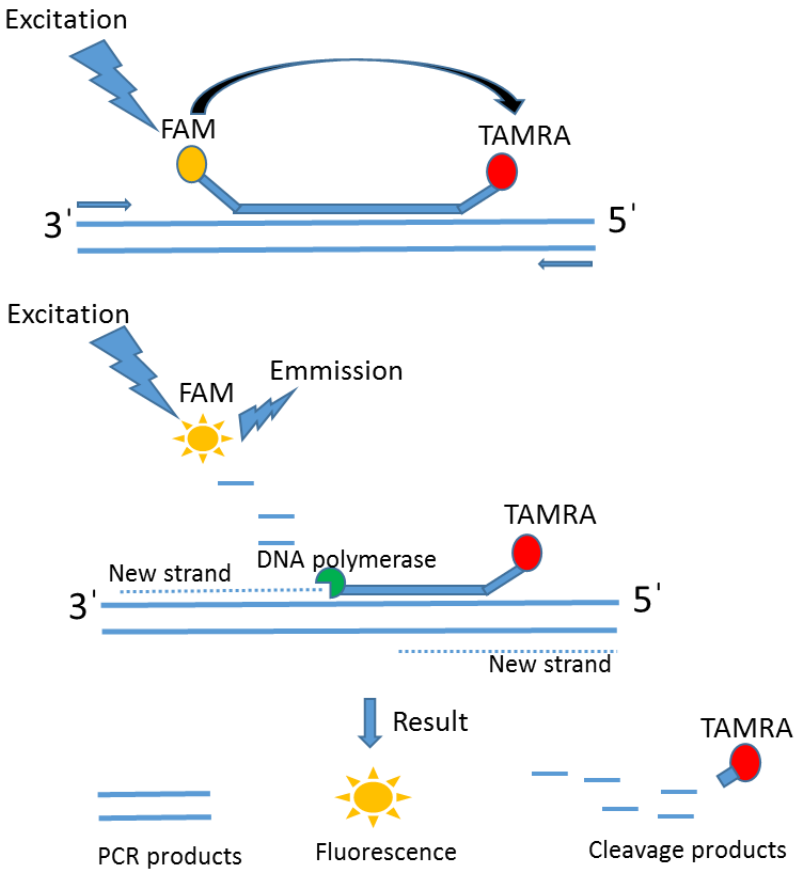


Figure 5-1: Steps followed in qPCR with Taqman probes. First the probe complementary to the specific sequence attaches to the DNA template. As the fluorescence and quencher are intact at this point no fluorescence is detected because the quencher prevents fluorescence emission when excited. When amplification takes place, the DNA polymerase comes in proximity to the probe and it degrades the probe by hydrolysing the fluorophore end of the probe. As the hydrolysis separates the quencher from the fluorophore, fluorescence is emitted by the fluorophore. The amplification of the DNA is monitored by the changes in the fluorescence emission (adapted from (Bustin, 2000)).

The primers and probes specific for the *W08E(12.2-12.5)* were designed using the Universal Probe Library (UPL) assay design software, which claims to deliver intron spanning, sequence specific primer-probe combinations. The primers and probes used for quantitation of *W08E(12.2-12.5)* are listed in **Table 5-1**.

Table 5-1: List of primer/ probe combinations along with their sequences, used to quantify the respective genes by qPCR.

Primer ID	Primer sequence (5ˈto 3ˈ)
rla-1 forward	ACGTCGAGTTCGAGCCATA
rla-1 reverse	CCGGAAGAGACAGAAGTGA
W08E12.2 forward	CGTGTTGCAATAGCAATAAC
W08E12.2 reverse	TTACGACGGAAACGGTAGC
W08E12.3-5 forward	TTTGTGCTCCAATGTGCAA
W08E12.3-5 reverse	ATGGGCARCASGAGAAGTTG
Probe ID	Probe sequence/ number
rla-1 probe	GGCCAGGA (Probe 162)
W08E12.2 probe	TTCTCCTG (Probe 2)
W08E(12.3-12.5) probe	GCTGCCA(Probe 44)

The wild type (N2) worms were exposed to Cd and the average fold change in the expression of the isoforms *W08E12.2* and (*W08E12.3-12.5)* was quantified using the above mentioned primer-probe combinations (**Table 5-1**). A statistically robust, fivefold increase in the *W08E12.2* expression was observed when the worms were exposed to 20 µM Cd. The relative fold change in the expression of *W08E(12.3-12.5)* to Cd followed a dose dependent pattern, with peak response at 10 µM Cd. The upregulation of the *W08E(12.3-12.5)* expression upon Cd exposure was less than *W08E12.2*, deemed to be statistically significant at the two higher doses but minimal when compared to the *W08E12.2* (**Figure 5-2**).

The *W08E(12.2-12.5)* expression upon Cd exposure was tested in a metallothionein null background in order to assess any change in their response in comparison to the wild type (N2). The *W08E12.2* in the absence of metallothioneins displayed a significant increase in its expression upon exposure to 10 µM Cd and 20 µM Cd. The *W08E12.2* attained paramount amplification at 10 µM Cd. Biological repeats of the experiment were performed and they revealed the same response trend, although the

overall amplitude of expression of *W08E12.2* between the repeats varied significantly (**Appendix, Figure 9-1**).

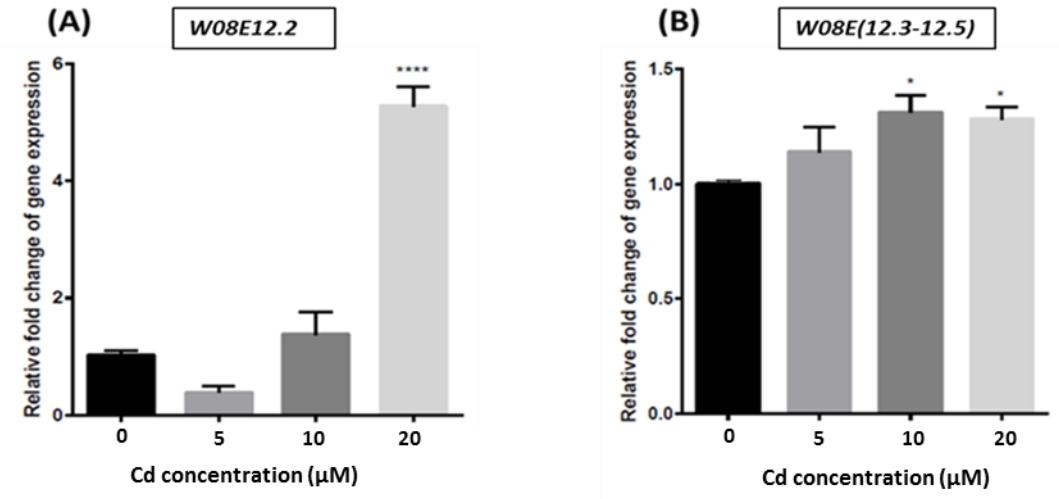


Figure 5-2: Differential *W08E(12.2-12.5)* expression in cDNA samples from (N2) nematodes exposed to defined cadmium concentrations for 48 hours. Nematodes were exposed to Cd (0-20 µM) from L1-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA was synthesised. Average fold change in expression of *W08E12.2* (A) and *W08E(12.3-12.5)* (B) is presented. Fold changes of the respective isoforms were calculated following the normalization to the invariant control (*rla-1*) normalization. (The statistical analysis was performed using one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$, * $P < 0.05$, error bars: SEM. The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

The *W08E(12.3-12.5)* displayed a dose dependent increase in the expression fold upon exposure to Cd in a metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) background. The response peaked at 10 µM Cd (**Figure 5-3B**). The biological repeats performed for this experiment also displayed maximal gene upregulation at 10 µM Cd, although a definitive pattern of expression could not be deduced due to the non-uniformity between the repeats (**Appendix, Figure 9-2**).

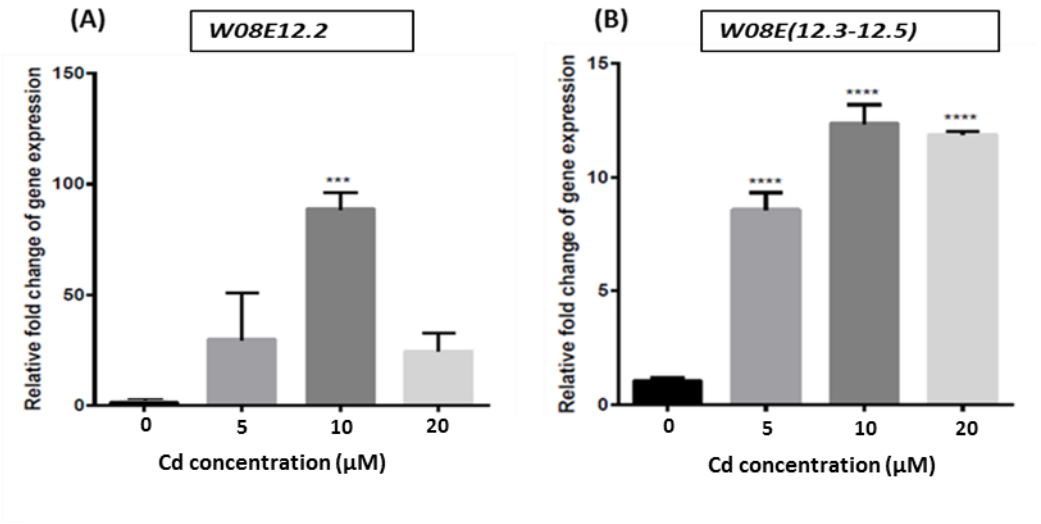


Figure 5-3: Differential *W08E(12.2-12.5)* expression in cDNA samples from metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) worms exposed to defined cadmium concentrations for 48 hours. Nematodes were exposed to Cd (0-20 µM) from L1-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA (1000 ng/µL) was synthesised. Average fold change in expression of *W08E12.2* (A) and *W08E(12.3-12.5)* (B) when exposed to defined doses of Cd is presented. Fold changes of the respective isoforms were calculated following the normalization to the invariant control (*rla-1*) normalization. (The statistical analysis was performed using one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$, *** $P < 0.001$, error bars: SEM. The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

Acute (24 hours) Cd exposure of the metallothionein double knockout (*mtl-1(tm1770); mtl-2(gk125)*) worms showed an upregulation of *W08E12.2* and *W08E(12.3-12.5)* isoforms (although there was variability in the pattern of expression between the biological repeats (**Appendix, Figure 9-3**). Due to the differences between the biological repeats it was difficult to draw a conclusion about the expression of *W08E(12.2-12.5)*. As *W08E12.3, W08E12.4* and *W08E12.5* was quantified together, it is possible that cross reactivity between the probes and primers of the entire *W08E(12.2-12.5)* family interfered with the quantification. Hence, the cross reactivity of all probes and primers was evaluated.

Plasmids containing individual cDNA clones of *W08E12.2*, *W08E12.3* and *W08E12.4/5* were used to test the specificity of various primer-probe combinations. The Ct values for each qPCR are presented in a matrix against each primer-probe combination used for the test specific for each *W08E(12.2-12.5)* isoform (**Figure 5-4A**).

The primer-probe combinations that yielded a Ct value < 20 (green) are ideal for specific and efficient amplification of the respective *W08E(12.2-12.5)* isoform. Based on the matrix, the primer-probe combinations used for performing the qPCRs on *W08E(12.2-12.5)* before the specificity test displayed non specificity and cross reactivity, hence a new set of probe-primer combination was set up (**Figure 5-4**).

The matrix based ideal primer-probe combination were used to characterize the *W08E(12.2-12.5)* isoforms. It was observed that upon exposure to Cd the wild type (N2) nematodes displayed an upregulation of *W08E12.2*, the maximal increase in the fold change was at 10 µM Cd (**Figure 5-5A, B**). The biological repeats of *W08E12.2* depicted variable intensities in expression, but they both confirmed a statistically robust increase in amplification at 10 µM Cd.

The *W08E12.3* expression followed a dose dependent upregulation in response to Cd, and the peak in the expression profile was attained at 10 µM Cd. The biological repeats confirmed the same pattern of expression of *W08E12.3* upon exposure to Cd, however the magnitude of fold change in expression was different by threefold (**Figure 5-5C, D**). An upregulation of *W08E12.4*/5 was observed upon Cd exposure (**Figure 5-5E**) which was non reproducible in its biological repeat (**Figure 5-5F**). Therefore, based on the biological repeats, the quantitation of *W08E12.4* and *W08E12.5* in the wild type upon Cd exposure revealed no specific pattern of expression.

(A)

Key	cDNA	Combination of probe/primers used before the specificity test	Combinations designed to be specific by UPL	Ideal Combination of probe/primers after specificity test
Ct value > 40				
40 > Ct value > 35				
35 > Ct value > 30	*W08E12.2*	12.2 F/ 12.2 R; # 2	12.2 F/ 12.2 R; # 2	12.2 F/ 12.2 R; # 2
30 > Ct value > 25	*W08E12.3*	12.3 F/ 12.3 R; # 160	12.3/4/5 F/R; #44	12.3/4/5 F/R; #44
25 > Ct value > 20	*W08E(12.4-12.5)*		12.4 F/ 12.4 R; #74	12.4 F/ 12.4 R; #44
Ct value < 20	*W08E(12.3-12.5)*	12.3/4/5 F/R; #44	12.3/4/5 F/R; #44	

(B)

cDNA / Primer & Probe	*W08E12.2*	*W08E12.3*	*W08E12.4/5*
12.2 F/R #2	14.5	29.2	29.6
12.2 F/R #44			
12.2 F/R #160			
12.2 F/R #74			
12.3 F/R #2			
12.3 F/R #44	23.3	8.5	20.9
12.3 F/R #160	39		22.8
12.3 F/R #74	24.1	8.4	20.9
12.4 F/R #2			
12.4 F/R #44	29.4	31.8	14.9
12.4 F/R #160			
12.4 F/R #74	27.5	22.2	11
12.3/4/5 F/R #2	25.3		9.5
12.3/4/5 F/R #44	26.2	10.1	
12.3/4/5 F/R #160			
12.3/4/5 F/R #74			

Figure 5-4: Matrix representation of the different cDNA and primer/probe combinations based on the extent of specificity and efficiency of each combination to amplify respective isoforms based on the Ct values (B). The key represents the colour codes used in the matrix based on the Ct values (A). The combinations of primers and probes used to characterise the *W08E(12.2-12.5)* isoforms before the specificity test, the suggested combinations meant to be specific by the Universal Probe Library assay design centre and the specific combinations identified on the matrix have also been listed in the table (B) underneath the matrix.

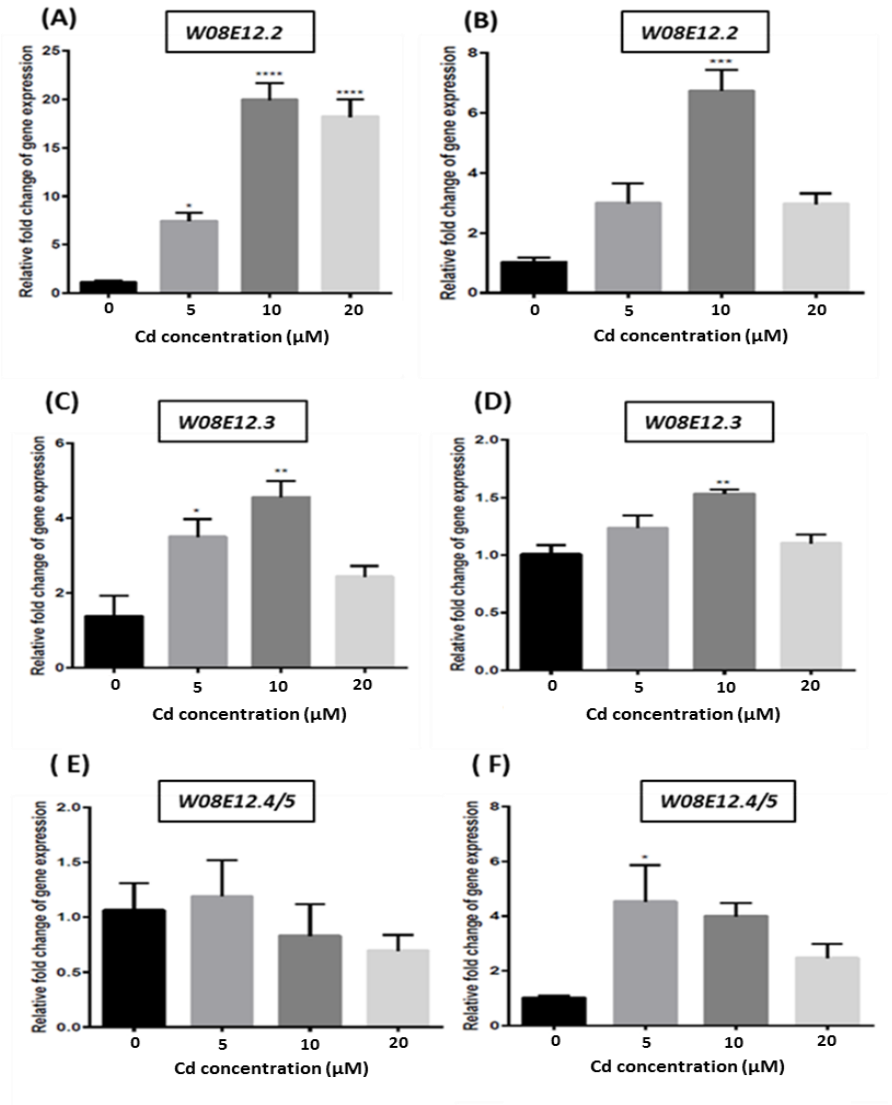


Figure 5-5: Expression of *W08E(12.2-12.5)* in cDNA samples derived from (N2) nematodes exposed to defined cadmium concentrations for 48 hours, amplified with new primer-probe combination. Nematodes were exposed to Cd (0-20 µM) from L1-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA (1000 ng/µL) was synthesised. Average fold change in expression of *W08E12.2* (A, B biological repeats), *W08E12.3* (C, D biological repeats) and *W08E12.4/5* (E, F biological repeats) when exposed to defined doses of Cd are presented. Fold changes of the respective isoforms were calculated following the normalization to the invariant control (*rla-1*). (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.005$, * $P < 0.01$; error bars: SEM. The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

The *W08E12.2* expression in response to Cd in a metallothionein null background was dose dependent, displaying a gradual increase in amplification from 0-10 µM Cd followed by a significant drop at 20 µM Cd. Although the intensity of expression between the biological repeats were different, they were reproducible in the pattern of expression and therefore confirmed the expression profile of *W08E12.2* in metallothionein knockout worms upon exposure to Cd (**Figure 5-6A, B**).

The *W08E12.3* expression in response to Cd in the metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) worms followed the same pattern as *W08E12.2*. The biological repeats although slightly variable at certain doses (10 and 20 µM Cd) displayed similar response to Cd with the peak increase in fold change to be fivefold in magnitude (**Figure 5-6C, D**).

The biological repeats confirmed an overall upregulation of *W08E(12.4-12.5)* expression when exposed to Cd in a metallothionein null background. Although, there was a variation observed in the intensity of *W08E(12.4-12.5)* expression in the 20 µM Cd dose between the biological repeats, an overall upregulation of 4-fold in the amplification upon 5 µM and 10 µM Cd exposure was observed in both (**Figure 5-6E, F**).

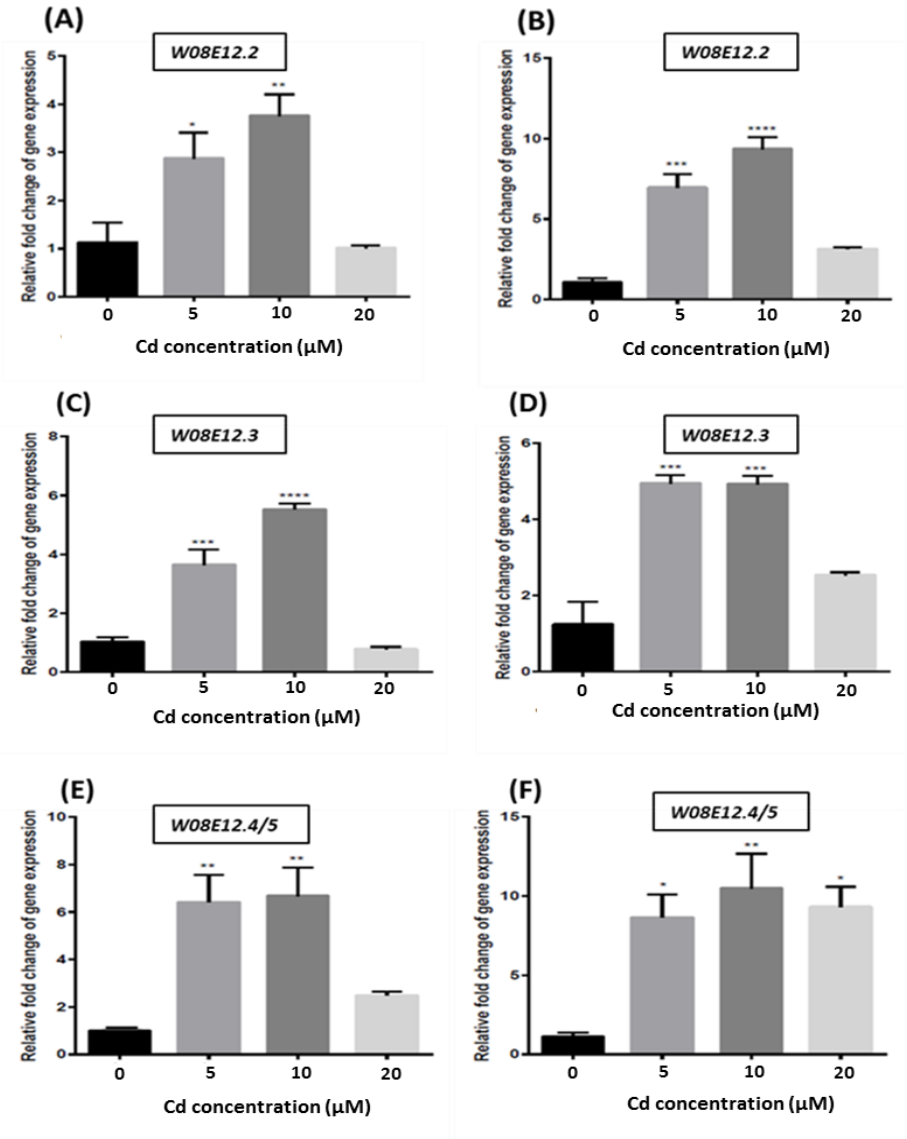


Figure 5-6: Differential expression of *W08E(12.2-12.5)* in cDNA samples derived from metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) worms exposed to defined cadmium concentrations for 48 hours amplified with new primer-probe combination. Nematodes were exposed to Cd (0-20 µM) from L1-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA (1000 ng/µL) was synthesised. Average fold change in expression of *W08E12.2* (A, B biological repeats), *W08E12.3* (C, D biological repeats) and *W08E12.4/5* (E, F biological repeats) when exposed to defined doses of Cd has been represented. Fold changes of the respective isoforms were calculated following the normalization to the invariant control (*rla-1*) normalization. (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$, *** $P < 0.001$, ** $P < 0.005$, * $P < 0.01$; error bars: SEM. The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

5.2.2 Knockdown of *W08E(12.2-12.5)* by RNA interference (RNAi)

Given that the entire genome of the worm has been sequenced, this information can be used to understand the functional aspects of specific genes, for example by applying reverse genetics tools such as the gene specific knockdown by RNAi. Andrew Fire discovered that antisense-RNA can hybridise with mRNA to inhibit translation leading to silencing of eukaryotic endogenous genes (Fire et al., 1991), and subsequently that injection of sense RNA could also cause the same effect (Guo and Kemphues, 1995).

In the current study, RNAi of *W08E(12.2-12.5)* was applied as a tool to identify the loss of function phenotype within the nematode. The worms were fed with *W08E(12.2-12.5)* RNAi bacteria and the knock down worms were exposed to heavy metals. Generally to perform a gene knock down by feeding within the nematode, the feeding bacteria for a particular gene is selected from the commercially sourced RNAi library which covers 19,762 protein coding genes. But, unfortunately the *W08E12.2* and *W08E12.3* isoforms seemed to be absent in the library when the specific RNAi clones were tested with the recommended primers spanning the genomic region of the respective *W08E(12.3-12.5)* isoforms (**Figure 5-7**). Hence, a bespoke clone for the *W08E12.3* isoform was created by amplifying the genomic region of the *W08E12.3* isoform and cloning it into the RNAi vector pPD129.36. The cloned vector was transformed into competent HT115 bacteria and fed to the worms for the knock down of the specific genes within the nematode.

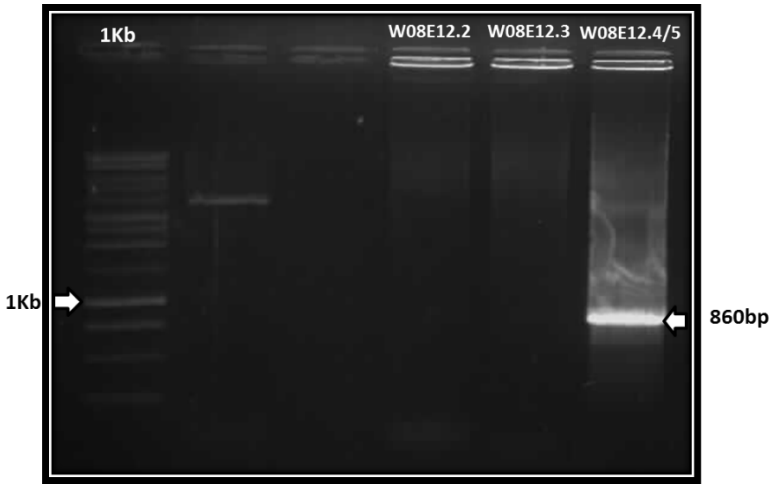


Figure 5-7: Agarose gel (0.8%) electrophoresis of PCR amplified fragments using gene specific primers of the RNAi clones. Clones for the respective *W08E(12.2-12.5)* RNAi bacteria were picked from the RNAi library and screened by amplifying the isolated plasmids of the respective isoforms with gene specific primers.

In detail, the primers used to amplify the *W08E12.3* region from the worm genome were tagged with *EcoR1* recognition sites extensions, to allow the amplified region to be digested with *EcoR1* and therefore ligated into the pPD129.36 vector which also has an *EcoR1* restriction site (**Figure 5-8**).

The PCR amplified *W08E12.3* from the genomic region after the restriction digestion with *EcoR1* yielded a genomic product (848 bp) with sticky ends compatible with the *EcoR1* digested vector pPD129.36.

Ligating the *EcoR1* digested *W08E12.3* fragment into the pPD129.36 was followed by transformation into the HT115 (DE3) cells. The successful transformants were screened with *W08E12.3* specific primers. A fragment of 848 bp was obtained upon amplification of the isolated plasmids with the gene specific primers, then confirming the successful construction of *W08E12.3* genomic RNAi clone (**Figure 5-9**). The cDNA clones for the *W08E12.2*, *W08E12.3* and *W08E12.4/5* were also constructed following the same protocol (data not shown).

(A)

cgaatacttttacattagaaccatcagctcttacatttatccgtttcaccatgcaatccgcctgcctcatcgctc
ttcttgctctcatcgccggagcgtcggctcaatgcttcggaggtggatgcggaggaggaccaatcttccttccac
caccaccatgcttcggaggaaactgtggtggatgttcgggtaacaactgtggatgcaatggaaataactgtggac
cacaagtcactgtagtccaggtcccgaataacaataacggatgctcgtgcagcccatgtaacggaccagtttgtg
ctccaatgtgcaatagctgcccaccacagccgatcttcattcagccatctcaatgctgcaataataacagttagt
attcgtcgtaagcatggtttttcaaaaagagttcgctactttcagacttctcgtgctgcccattccgtttccgtc
gtcataactctgctgccgccgcttcagctgagaacacagaagaggccaccaccgctgctccaacccaagagtaaa
tctctacgatcgtctgattacttgatccaatttttgtatatacgaaacctagatttatacgtaataataaacacg
tttgaatttatattttatttctgaacagttaggtggttcctgtggcataacagaaagttggtaatggagaaacaa
gactagacttttccacaatacttcttttcaagcaaacggttcttcggacgttgtagtcaaacaggaagatgag
atcagcgaaaaaccagattcaacatacaaacttcgtttgatcagagcacaagctttgtaggggacaatttaaaat
gtcgccctaaaatttatgcctataagttgcgtagcaggcactacaggctgatc

(B)

EcoR1 F: GGG GAATTC CCCATCAGCTCTTACATTTATCCG

EcoR1 R: CTATAAGTTGCGTAGCAGGCACTCGAATTC CCC

Figure 5-8: The genomic sequence encoding *W08E12.3* (A). The yellow highlighted region depicts the unspliced sequence of the *W08E12.3* gene. The red highlights show the start and end point of the region that was cloned into the pPD129.36 vector. The primers used to amplify the *W08E12.3* genomic region with the extensions including *EcoR1* recognition sites are depicted in red (B).

Late L2 or early L3 worms were fed with either *W08E12.3* or a mixture of *W08E12.3/4/5* RNAi bacteria and transferred to new plates once they reach L4 stage. The early adult worms were transferred to new RNAi plates and left for 4-6 hours to allow eggs to be laid. The parents were then killed and the F2 generation was used to perform all the RNAi based experiments.

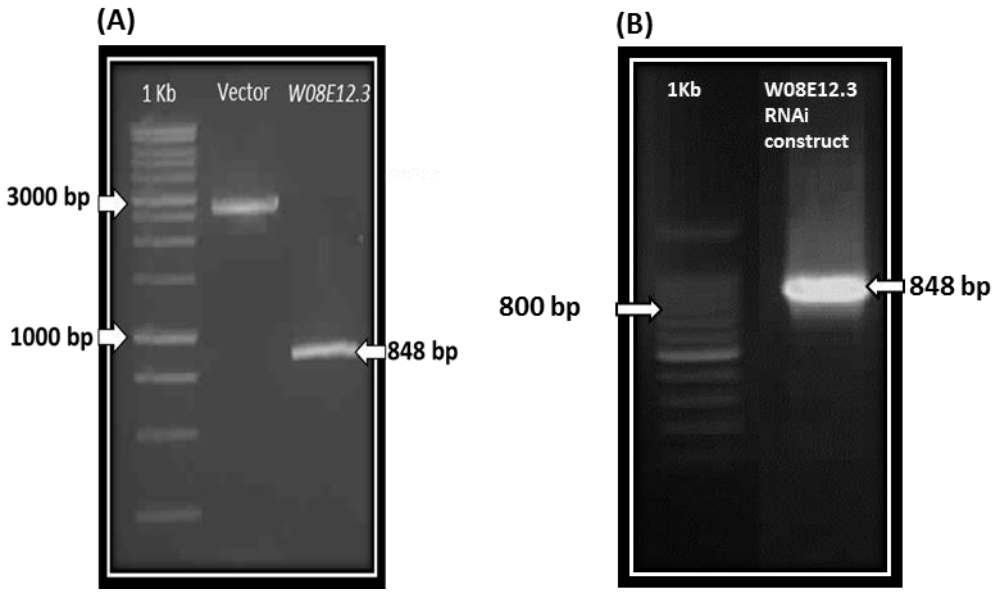


Figure 5-9: Amplification of the *EcoR1* digested pPD129.36 vector, *W08E12.3* insert and the amplified region of *W08E12.3* from the pPD129.36 cloned vector. (A) Gel electrophoresis of the *EcoR1* digested pPD129.36 vector and the *W08E12.3* region amplified with *W08E12.3* specific primers against a 1Kb marker on a 0.8% agarose gel. The *W08E12.3* region was amplified from the worm genome by using gene specific primers with *EcoR1* extensions. The amplified fragment and the vector pPD129.36 were subjected to digestion with *EcoR1*. **(B) Gel electrophoresis of the *W08E12.3* region amplified from the pPD129.36 vector with *W08E12.3* specific primers against a 1Kb marker on a 0.8% agarose gel.** The *W08E12.3* region was amplified from the cloned pPD129.36 construct with gene specific primers in order to test the success of *W08E12.3* cloning into the pPD129.36 plasmid for construction of the RNAi feeding clone.

The cDNA samples were prepared from the RNAi exposed and control (unexposed) worms in order to test the knockdown efficiency. The cDNA samples were subjected to PCRs with the gene specific primers and the results were normalized based on the amplification of the house keeping gene *rla-1* for all the samples. The *W08E12.3* amplification in the *W08E12.3* RNAi cDNA sample was effectively reduced in comparison to the unexposed samples, whereas in the *W08E12.4* RNAi cDNA samples the knockdown was less significant, albeit still lower than the control. (**Figure 5-10**).

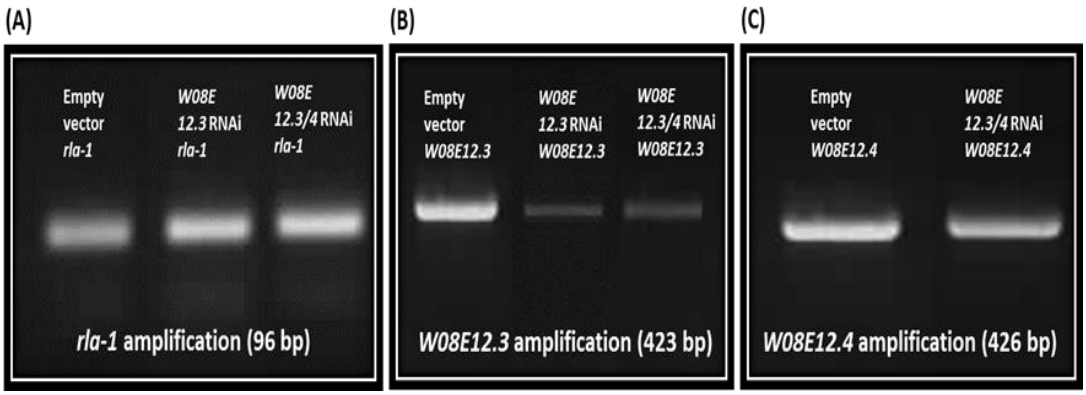


Figure 5-10: Evaluation of knockdown efficiency of *W08E12.3* and *W08E12.4* following RNAi. Amplification of *rla-1* (A), *W08E12.3* (B), *W08E12.4/5* (C) in control worms (no knockdown) compared to worms subjected to the RNAi of *W08E12.3* and the *W08E12.3/4/5*. The control worms were exposed to bacteria with the empty RNAi (ppD129.36) vector, whereas the RNAi worms were exposed to either *W08E12.3* or both *W08E12.3/4* RNAi vector. The worms were exposed to the standard RNAi feeding protocol and tested for the relative knockdown of each specific gene using *rla-1* as the house keeping gene control.

5.2.3 Growth Assay

One of the main advantages of using *C. elegans* as a model organism lies in its application in toxicology research with the ability to use quantitative assays like development, growth and brood size. Growth assays provide a detailed view of the effect of toxins on the different larval stages from L1 to adulthood. It has been shown before that tracking development from L1 through 48 hours post hatch enables an indirect measurement of neurotoxicity (Boyd et al., 2010).

The growth and development of wild type *C. elegans* in control conditions follows a sigmoidal pattern when body length is plotted against time (Jager et al., 2005). But when growth is plotted for worms exposed to certain toxins, which affect the nematodes growth, the sigmoidal growth curve can change.

Based on the growth assay performed, growth of the *W08E(12.3-12.5)* knockdown worms resembled the wild type up to 72 hours post L1s. Exposure to Cd (30 µM) affected the growth drastically of both control and the knockdown worms although the effect was more prominent in case of the *W08E(12.3-12.5)* knockdown worms which displayed a severe reduction in length and volumetric surface area 24-120 hours post L1s (**Figure 5-11**). The length and surface area of the Zn exposed control and knockdown worms was comparable to the unexposed worms up to 72 hours post L1s. Thereafter growth retardation was observed in *W08E(12.3-12.5)* knockdown worms upon exposure to Zn or Cd (**Figure 5-11**).

Overall it can be stated that the *W08E(12.3-12.5)* knockdown affected the growth rate of the worms and this effect was more pronounced in the presence of heavy metals (30 µM Cd, 200 µM Zn). This suggests that *W08E(12.3-12.5)* protects the nematode from metal induced toxicity (**Figure 5-11**).

5.2.4 Brood size (number of offspring)

Brood size is routinely used as a quantitative measure in toxicological research. The assay is carried out by transferring the worm onto a new plate every day from the time the egg laying starts until it ends. Once, the eggs hatch, the progeny from each worm per day is counted.

The reproductive toxicity caused due to the knockdown of *W08E(12.3-12.5)* was tested by exposing the wild type (subjected to the RNAi empty vector or the *W08E(12.3-12.5)* RNAi vector) to heavy metals (30 µM Cd, 200 µM Zn). Based on the brood size assay performed, the number of progeny produced per day was significantly lower upon exposure to heavy metals, especially upon exposure to 30 µM

Cd for both wild type and knockdown worms (**Figure 5-12B**). The total number of viable offspring produced by the knockdown worms resembled the wild type, although it is evident that Cd (30 µM) caused the *W08E(12.3-12.5)* knockdown worms to lay eggs for a longer period than the wild type worms (**Figure 5-12A**).

Overall it can be stated that Cd (30 µM) induced reproductive toxicity in both the wild type and the *W08E(12.3-12.5)* knockdown worms resulting in a reduction in brood size in comparison to the unexposed worms. As the difference between the brood size of the wild type and the *W08E(12.3-12.5)* knockdown worms was negligible, it suggests that the silencing of *W08E(12.3-12.5)* in the nematode has no drastic effect on the brood size.

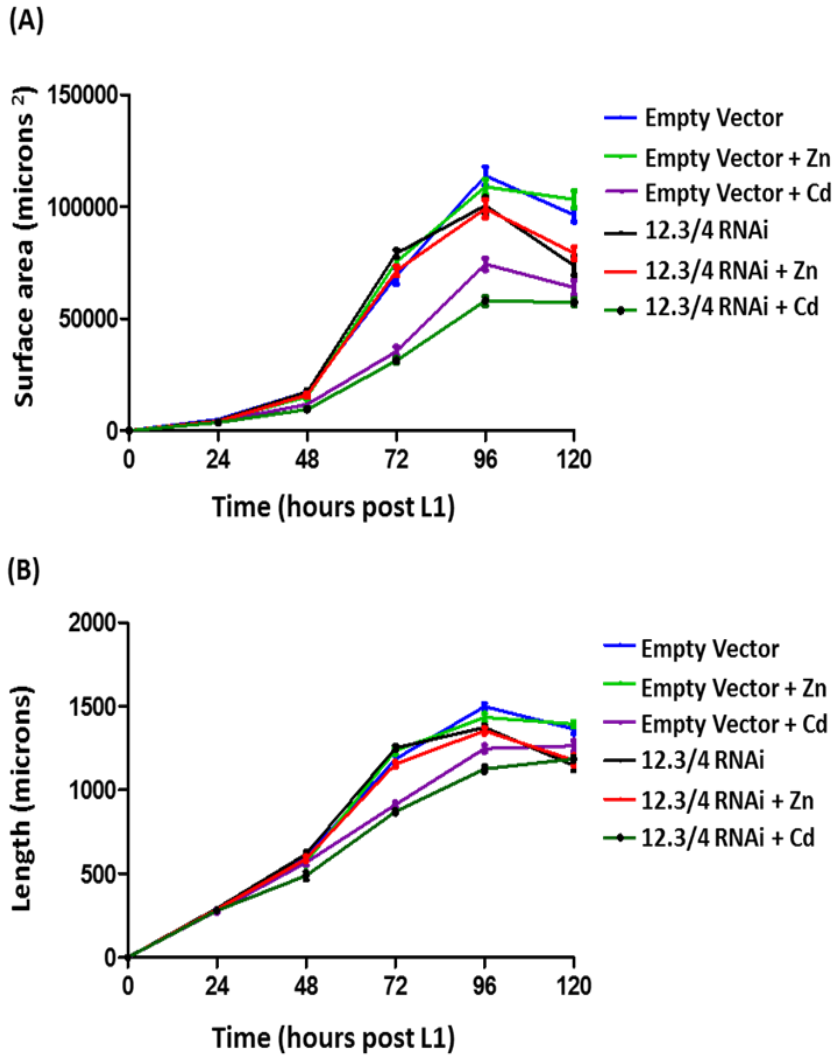


Figure 5-11: Developmental growth recorded over 120 hours post L1 stage of wild type subjected to the empty (control) RNAi vector or the *W08E(12.3-12.5)* vector in the absence and presence of heavy metals (30 µM Cd, 200 µM Zn). The volumetric area of the worm bodies (A) and the length of the worm body (B) has been recorded against time. Growth of the nematodes was followed from the L2 – late adult i.e. from day 2 to day 6 post plating the nematodes on the culture plates. The values correspond to means (± SEM) of 10 worms per strain, per condition.

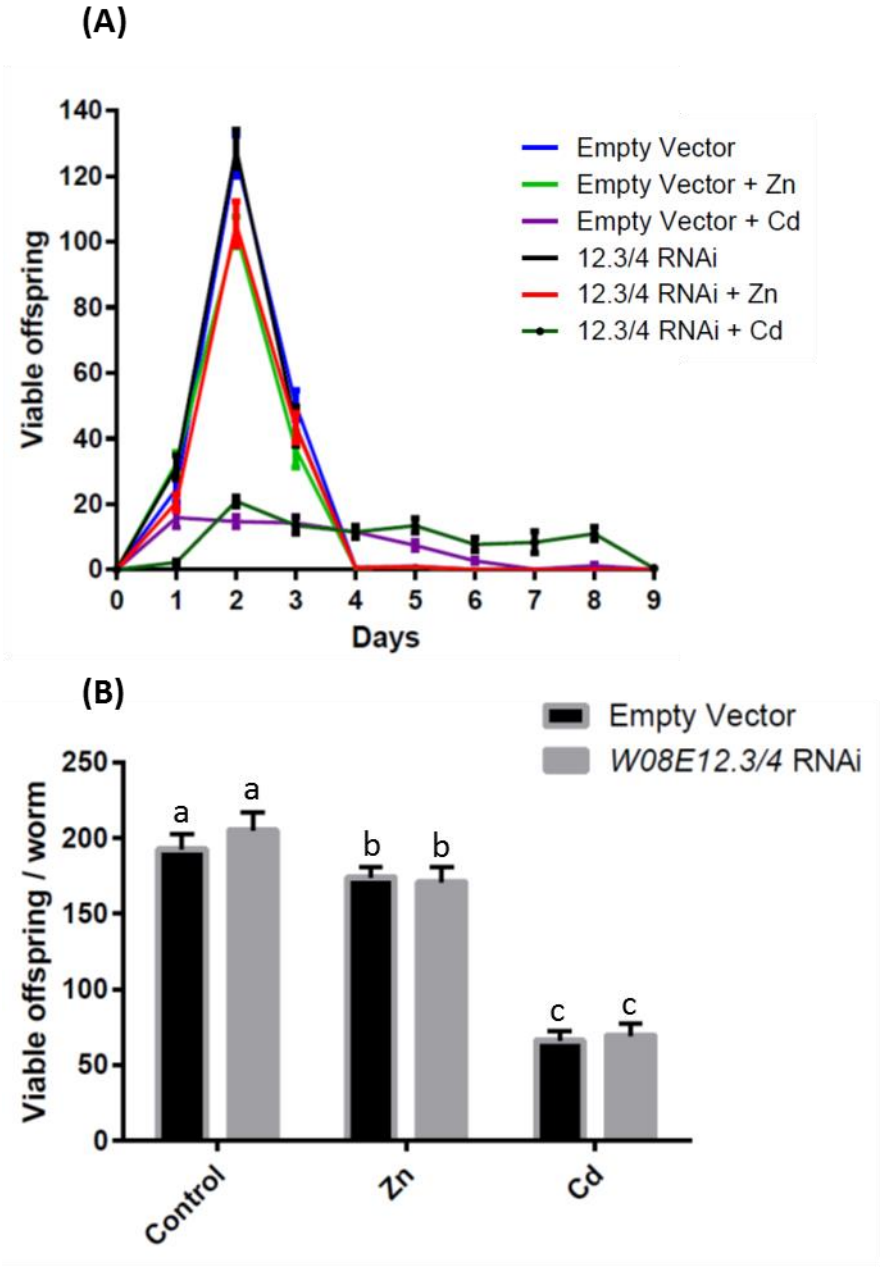


Figure 5-12: (A) Brood size recorded over 9 days from day 1 of egg laying of wild type subjected to the empty RNAi vector or the *W08E(12.3-12.5)* vector in the absence (control) and presence of heavy metals (30 µM Cd, 200 µM Zn). (B) Total number of eggs produced over 9 days by wild type worms subjected to the empty RNAi vector and the *W08E(12.3-12.5)* knock down worms in the absence (control) and presence of heavy metals (30 µM Cd, 200 µM Zn). Values plotted are means calculated from 12 replicates per condition (± SEM). The mean difference of each metal exposure is significantly different to the respective metal-devoid condition, but no RNAi mediated effect on brood size was observed. Different letters refer to a statistical difference between conditions ($P < 0.001$).

5.2.5 X-ray fluorescence imaging of *W08E(12.3-12.5)* knockdown worms

The Stanford linear accelerator at SLAC (Stanford linear accelerator Centre) uses a polarised electron beam and stands to be the world's longest linear accelerator (3 km) of electrons and positrons. The accelerator boosts the kinetic energy of charged sub atomic particles/ions via oscillation electric potentials by the linear beam line. The linear accelerator (Linac) provides X-ray laser pulses to detect matter at the smallest scale. Linacs are capable of accelerating heavy ions to energies beyond the ring-type accelerators (synchrotrons) which compromise the strength of magnetic field in order to maintain the curved path of the ions (Alley et al., 1995).

With the aid of SLAC it is in principle possible to visualise the accumulation of different metals within the nematode body. In a proof of concept experiment, control (N2) and the *W08E12.3/4* RNAi worms were exposed to a mixed dose of Zn (150 µM) and Cd (30 µM) from L1 stage until late adults. The worms were then examined under two different beamlines optimized to allow the concentration of Zn or Cd to be visualized at a 5 micron resolution by Yona Essig (King's College London) and processed with a custom designed microprobe analysis software (SMAK) (**Figure 5-13**).

All worms were exposed to Cd and Zn resulting which defined the baseline accumulation within a wild type worm (**Figure 5-13A, Figure 5-13D**). Upon knockdown of *W08E(12.3-12.5),* Cd levels did not change markedly however a striking accumulation of Zn was observed in the upper pharyngeal region, especially in the procorpus and metacorpus area. This finding was consistent in both biological repeats (**Figure 5-13B, C**).

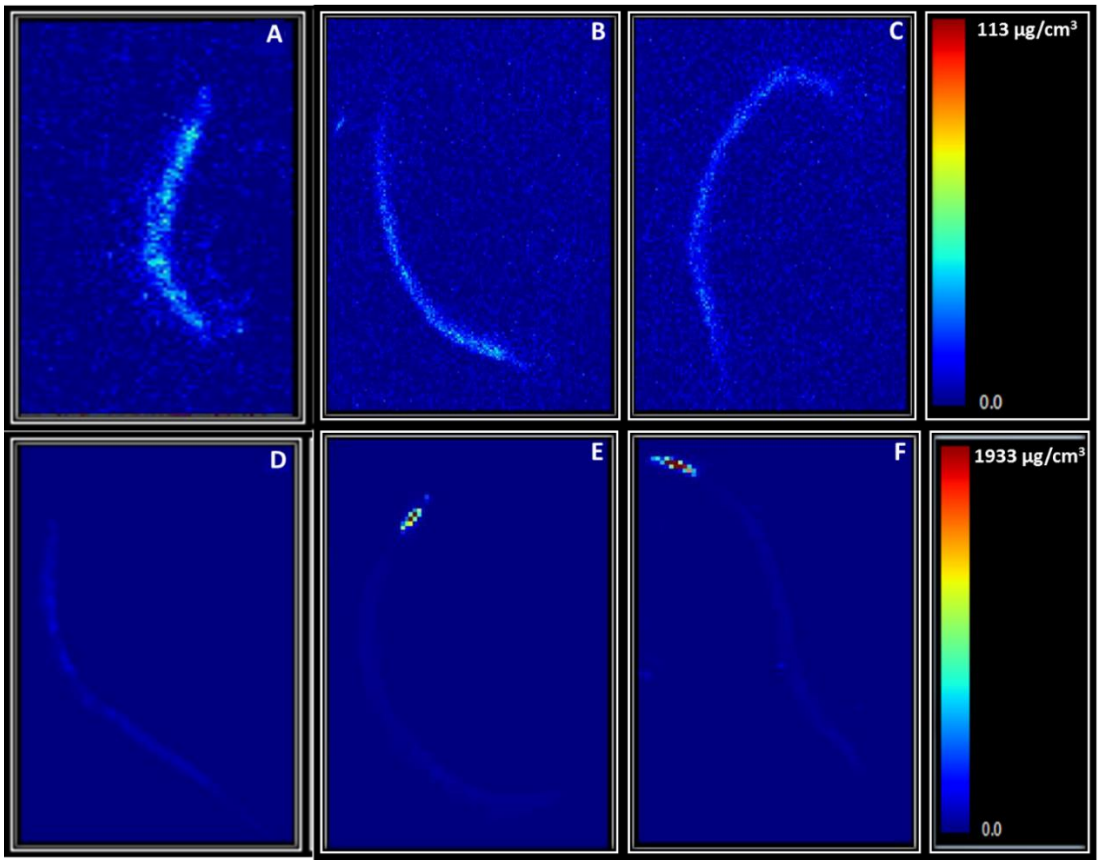


Figure 5-13: X-ray laser images processed using SMAK of Control (N2) and *W08E12.3/4* RNAi worms exposed to a mixture of Zn and Cd. The worms were exposed to a mixture of Zn (150 µM) and Cd (30 µM) from L1 until adult and imaged using the X-ray laser produced by the linear accelerator at SLAC. The images were processed using an analysis software known as Sam's Microprobe Analysis kit (SMAK) and the Zn and Cd accumulation within the nematodes was quantified normalising it with the respective metal standards with the heat map scale on the right. Heat map of the mixed metal exposed worms tested under Cd filter (beamline 14.3) (A, B. C) and Zn filter (beamline 2.4) (D, E, and F). (A, D- N2 control worms; B, C, E, F- *W08E12.3/4* RNAi worms)

5.3 Discussion

The completely sequenced genome of *C. elegans* allows the use of reverse genetic tools to unveil the function of uncharacterised genes. The location of *W08E(12.3/12.4)* within the worm was traced and their responsiveness to heavy metals was investigated

by means of GFP tagged transgene (**chapter 4**), but it was deemed important to quantify the expression of the entire *W08E(12.2-12.5)* family in response to heavy metals by an alternative approach namely qPCR.

The *W08E12.2* expression was upregulated in wild type worms exposed to 20 µM Cd and *W08E(12.3-12.5)* at the two highest doses (10 µM, 20 µM). The overall expression of *W08E12.2* was higher than *W08E(12.3-12.5)* upon exposure to Cd. This confirms the results obtained using the transgene and further substantiates that *W08E(12.2-12.5*) is metal responsive in nature.

The *W08E(12.2-12.5)* expression was also tested in a metallothionein null background in order to assess the presence of a functional dependency between the two families. When the metallothionein double knock out worms were subjected to acute (24 hours) Cd exposure, the *W08E12.2* followed a dose dependent increase in expression, while the cumulative response of *W08E(12.3-12.5)*, though increasing at 10 µM Cd, followed no strict pattern. As the amplitude of expression varied between the biological repeats, the reproducibility of the data has to be questioned and only general observations/deductions can be derived from this initial experiment.

In consequence it was decided to test the specificity of the probes and primers designed by the Universal Probe Library assay design software. Based on the matrix design, a cross reactivity between the isoforms was unveiled, hence new primer-probe combinations were designed to quantify each of the isoforms, namely *W08E12.2*, *W08E12.3* and *W08E12.4/12.5*.

The new primer probe combinations revealed that in the wild type (N2) Cd exposed (48 hours) worms, the *W08E12.2* and *W08E12.3* displayed a significant increase in expression at 10 µM Cd, however the *W08E12.4/*quantification did not yield a

conclusive outcome due to inconsistent results between the biological repeats. This indicated that the high similarity between the genes still plays a role in cross reactivity of the primers and probes.

When the expression of the isoforms was tested in a metallothionein null background, the *W08E12.2, W08E12.3 and W08E12.4/5* displayed dose dependent patterns with a gradual increase in response to Cd (0-10 µM Cd). The results signify a strong response of *W08E(12.2-12.5)* to 10 µM Cd when metallothioneins are absent. A higher dose (30 µM) of Cd was not included in the tests as the worms suffered toxic effects leading to larval arrest. Based on the quantification of the *W08E(12.2-12.5)* isoforms upon exposure to metals, it was observed that the gene response was in general stronger for *W08E(12.3-12.5)* in absence of metallothioneins, suggesting a possible additive effect between *W08E(12.3-12.5)* isoforms and metallothioneins.

A valuable reverse genetic tool applicable to *C. elegans* is the RNAi technique, which was optimised to knockdown the *W08E(12.3-12.5)* isoforms in order to investigate their role within the worm in response to heavy metals (Zn and Cd). Although there was negligible difference observed in the brood size, the overall growth was significantly restricted in worms subjected to the knockdown of *W08E(12.3-12.5)* especially in response to Cd, indicated that these isoforms might play a role in the development and stress response pathway of the nematode.

The *W08E(12.3-12.5)* knockdown worms accumulated vast amounts of Zn in the pharyngeal corpus region in comparison to the rest of the worm body. Interestingly, the P*W08E12.3/4*::GFP was also expressed in the pharyngeal region of the worm. This suggested a possible link between Zn sensing or Zn homeostasis/ detoxification and the *W08E12.3/4* isoforms.

5.4 Conclusion

Taken together from the quantitative transcriptional analysis by qPCR it can be deduced that *W08E(12.2-12.5)* are upregulated upon Cd induction. The knock down of the gene family by RNAi followed by phenotypic developmental studies revealed the crucial role of *W08E(12.2-12.5)* in handling metal stress. Based on the X-ray fluorescence imaging, Zn accumulation was predominantly observed in the pharyngeal region of the *W08E(12.3-12.5)* knock down worms indicating their significance in a possible Zn sensing pathway.

Overall, it can be concluded that in the absence of the *W08E(12.3-12.5)* isoforms, a major Zn homeostasis pathway or a Zn sensing neuronal pathway is disrupted. Therefore suggesting that, *W08E(12.3-12.5)* might play a major role in Zn sensing or homeostasis and detoxification pathway within the nematode.

Chapter 6: Assessing the metal binding role of the W08E12.3 protein

6.1 Introduction

The *W08E(12.3-12.5)* isoforms have been characterised at the gene level in order to confirm their metal responsive nature (**chapter 4 & 5**) followed by reverse genetics approach which unveiled heavy Zn accumulation in the pharyngeal region of the worm upon the knock down of the *W08E(12.3-12.5)* isoforms (**chapter 5**). Given that the gene expression is linked to the expression of the proteins, it was deemed important to study the *W08E(12.3-12.5)* family at the protein level. Out of the four isoforms *W08E12.3* was most characterised at the gene level, therefore was selected for further investigation.

Heterologous protein expression is largely carried out in the gram negative bacteria *E. coli* because of its rapid growth on inexpensive substrates, well explored genetics and the availability of wide range of expression vectors (Baneyx, 1999). In order to identify the unique biological activity of a protein, isolation and purification is essential. The W08E12.3 protein was expressed in *E. coli* as a fusion protein to improve its solubility in the cell cytoplasm and facilitate its purification. An affinity tag based approach was adapted to facilitate the purification of W08E12.3, as affinity tags enhance stability and solubility of the fused protein. They can also lead to a purity of up to 99% (Terpe, 2003). Various approaches like mass spectrometry, ICP-OES, pH titration can be applied to understand the structure and function of the protein.

6.2 Results

6.2.1 Purification of the W08E12.3 protein

6.2.1.1 The Intein Mediated Purification with an Affinity Chitin binding Tag (IMPACT) system

The IMPACT system includes an intein mediated inducible, self-cleavable tag to allow the purification of the protein of interest. It is designed to exploit the protein splicing mechanism to separate the affinity chitin binding domain (intein) from the target protein in one chromatographic step without the use of any exogenous proteases. It provides the choice of vectors to facilitate a C-terminal or N-terminal fusion of the intein tag to the protein (Cui et al., 2006). The above mentioned features of the IMPACT system made it the first choice of recombinant protein expression system in comparison to the other existing tagged systems like GST, His, S-tag etc.

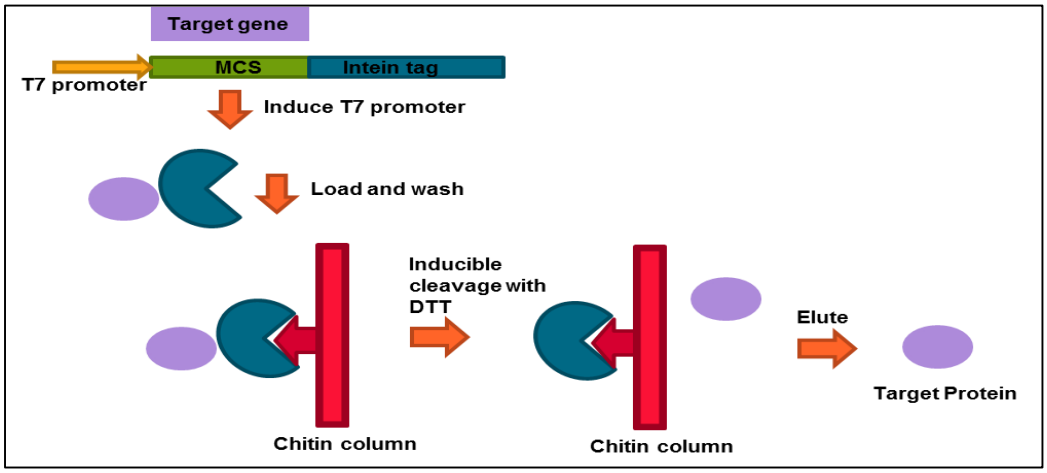


Figure 6-1: Schematic representation of intein mediated purification with an affinity chitin binding tag system to purify a fusion protein.

The C- terminal (pTXB vector) fused intein tag based system (**Figure 6-1**) was used to purify the W08E12.3 protein expressed in *E. coli* (DE3) cells. The protein expression was initially optimised by inducing the cells (containing the pTXB vector

with the *W08E12.3* cDNA insert) with IPTG over various time points and assessed on a SDS-PAGE by Western blotting using antibodies against the intein tag.

The induction test of the W08E12.3 protein on the silver stained gel (**Figure 6-2**) and the Western blot confirmed the expression of the intein fused W08E12.3 protein (43 kDa) in the induced samples. However, in addition to the intein tagged protein, another significant band was observed both on the silver stained gel and the Western blot which was confirmed to be the intein tag (27 kDa) on its own. This confirms that the tag within the *E. coli* cells is, during the induction prone to self-cleavage.

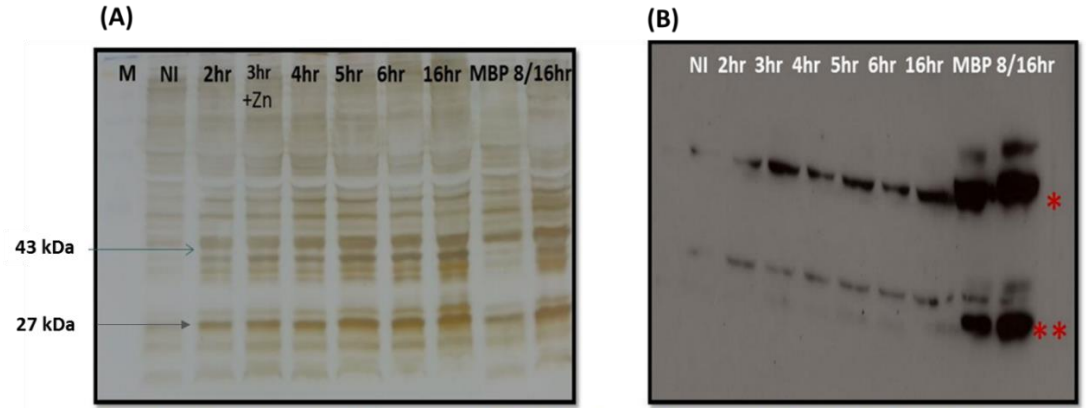


Figure 6-2: Total cell extracts of the IPTG induced protein samples separated by SDS-PAGE (4-12%) and silver stained (A) and probed with an intein tag specific secondary antibodies on a Western blot (B). The *E. coli* cells with the pTXB vector carrying the *W08E12.3* cDNA were induced with 1 mM IPTG over different time points. The cells were then centrifuged and lysed by sonication to extract the total protein. The protein was then boiled with Laemmli's buffer and loaded on to the SDS-PAGE gel. One gel was stained with silver and the other probed for the intein tag using a Western blotting procedure (* W08E12.3 protein fused with intein tag; ** the intein tag) (M-marker, NI- non-induced sample, MBP- maltose binding protein).

The W08E12.3 intein tagged protein was attempted to be purified using the chitin based column followed by protein splicing with dithiothreitol (DTT) and elution with Tris buffer.

After the purification of W08E12.3 protein using the IMPACT system, the SDS-PAGE gel revealed that there was negligible amounts of the tag fused protein in the samples before cleavage and no visible protein in the elute (**Figure 6-3**). The DTT based cleavage of the tag from the target protein was continued for 3 days in order to facilitate maximum cleavage but again no substantial amounts of the W08E12.3 protein was in the elute (**Figure 6-3**). Overall from the induction studies and the SDS-PAGE gel of the purified samples, it can be concluded that the intein tag self-cleaved from the target protein (W08E12.3) within the *E.coli* cells, likely due to the internal pH conditions of the cell or due to the mis-folding of the W08E12.3 tag fused protein within the cell. Hence, most of the W08E12.3 protein was lost even before the purification process was initiated.

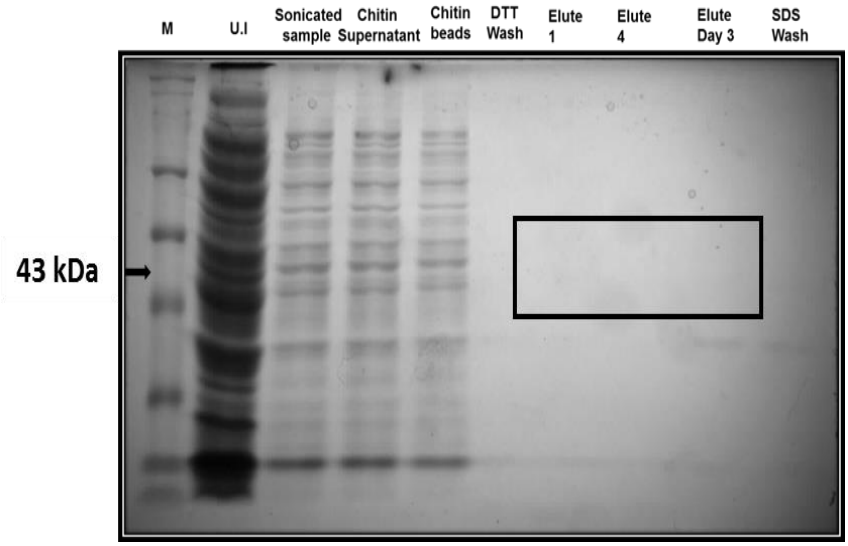


Figure 6-3: SDS-PAGE (4-12%) of samples processed through the IMPACT purification system. Samples were collected at each step of the purification of W08E12.3 protein and they were boiled with Laemmli's buffer and loaded on to the SDS-PAGE gel (M-marker, U.I- Uninduced).

6.2.1.2 S-tag based purification of protein

The S-tag based purification involves an N-terminal fusion peptide tag based system and enables the detection of the recombinant protein by homogenous assays or colorimetric detection in Western blots. There are four cationic, three anionic, five non polar and three uncharged polar residues present in the tag (Terpe, 2003). The tagged protein binds to S-protein coated agarose beads to separate it from the other bacterial proteins, and is cleaved by the action of thrombin. The purification involves a single chromatographic step.

An S-tag based system was chosen for the expression and purification of W08E12.3 due to the small neutral tag and also the fact that many metallothioneins have been expressed and their metal stoichiometry studied with the tag being uncleaved (Chan et al., 2002). The S-tagged system provided an assurance that if tag cleavage arises as an issue, the W08E12.3 protein could be studied with the S-tag extension. The pET29a (+) vector was used in order to clone *W08E12.3* cDNA and express the protein.

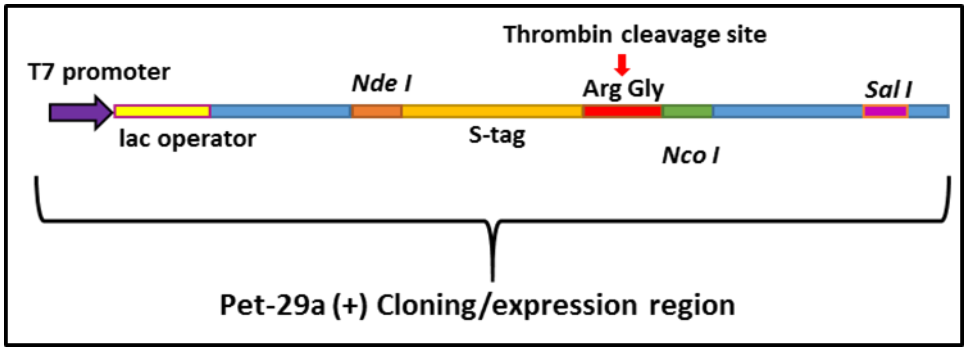


Figure 6-4: The nucleotide sequence with the corresponding amino acid sequence of cloning region of the pET29a (+) vector (modified from (Raines et al., 2000)).

Two constructs were created for the protein expression of W08E12.3, one with and one without the tag. To clone the S-tagged version of the protein, W08E12.3 was

inserted into the *NcoI* and *SalI* flanking sites, whereas for the untagged version *NdeI* and *SalI* restriction sites were used (**Figure 6-5**).

The constructs were sent for sequencing which revealed that two additional amino acids, namely a methionein and a glutamate, were inserted between the S-tag and the start codon. As the extra amino acids were deemed to be non interfering, this construct was decided to be used for protein studies of W08E12.3.

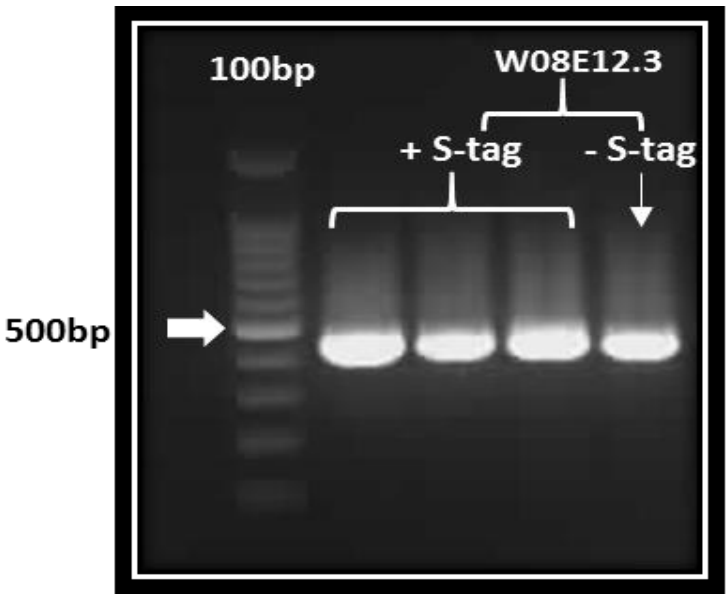


Figure 6-5: Agarose gel electrophoresis of the PCR amplified products of pET29a (+/- tag) vector containing the *W08E12.3* cDNA insert. The transformed cells were screened for the succesful constructs i.e. the pET29a (+) vector with the *W08E12.3* cDNA insert by subjecting the isolated plasmids to amplification with the *W08E12.3* cDNA specific primers.

The sequenced constructs (**Figure 6-6**) were used to study the expression pattern of the W08E12.3 protein. The induction of the protein expression was performed using 1 mM IPTG and the induced samples were collected every one hour over a period of 0-5 hr with an overnight sample in order to detect if there was a time dependent expression profile of the W08E12.3 protein. The expression studies were carried out

in the presence of metals and in absence of metals to investigate if metals provide any extra stability in the protein additional stability in the protein expression of W08E12.3. The induction studies of the S-tagged W08E12.3 protein revealed that the pET29a (+) expressed the W08E12.3 protein stably over time (**Figure 6-7**). No difference was observed between the metal induced and metal devoid W08E12.3 expression, suggesting that metals did not change the course of W08E12.3 expression but might play a role in keeping the expressed recombinant protein unoxidised while performing the large scale purification.

Figure 6-6: **W08E12.3 S-tagged construct sequenced with T7 promoter and compared to the theoretical sequence and mass.** The sequence of the tag is given in red and the yellow highlights show the extra amino acids introduced by the cloning strategy.

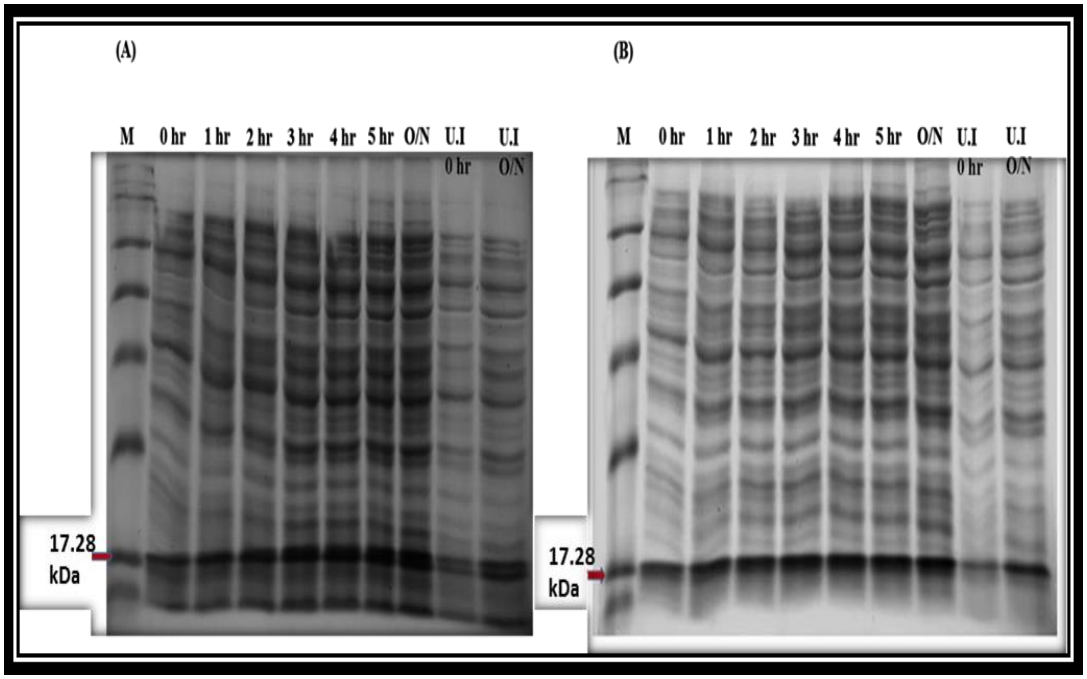


Figure 6-7: Coomassie stained SDS-PAGE (4-12%) of samples induced with 1 mM IPTG over a period of 0-5 hours and overnight. The bacteria containing of the pET29a vector (+ S-tag) with the *W08E12.3* insert were induced with IPTG and collected every hour. One batch of the cells were induced in presence of 100 µM Zn (A) and the other batch was induced in absence of any metals (B). The collected samples were lysed using sonication and the total protein was loaded onto a SDS-PAGE gel. The Bradford assay was used to ensure equal loading of samples and the samples were boiled with Laemmli's buffer before loading onto the SDS-PAGE (M- marker, U.I- uninduced sample, O/N-sample induced overnight).

There was a strong band of 17 kDa present in the uninduced sample signifying leaky expression of the W08E12.3 protein. A Western blot was performed using S-tag antibodies in order to investigate if this band was corresponding to the W08E12.3 protein or any other bacterial background protein.

Western blotting of the IPTG induced cell lysates confirmed the expression of the S-tagged W08E12.3 protein (**Figure 6-8**). In the Western blot, a very faint signal was observed in the uninduced samples, confirming that the 17 kDa protein band observed

in the uninduced samples co-localized with a background *E.coli* protein, affirming the tight regulation of the S-tagged W08E12.3 protein by the T7 promoter.

The Western blot also demonstrated that the expression differed based on the induction time. Upon induction for 4 hours, the maximum expression of W08E12.3 was achieved whereas an overnight induction lead to the degradation of the protein. It was confirmed that pET29a (+) is a suitable expression vector for W08E12.3 induction as leaky expression of the recombinant protein observed was negligible and a substantial level of induced expression of W08E12.3 protein could be achieved.

Based on the induction study of the W08E12.3 protein, a large scale purification was attempted. A 1 litre culture of the DE3 (plysS) cells containing the W08E12.3 S-tagged construct was grown and induced with 1 mM IPTG and 100 µM Zn once the OD_{600} reached 0.6. The cells were centrifuged after 4 hours of induction and the pellet was frozen at -80 ºC until further use. The frozen pellets were thawed and processed using a cell disruptor and the soluble fraction of the protein was subjected to S-agarose binding. The W08E12.3 protein was cleaved from the S-tag using thrombin (10 units/mg of protein) and the eluted fraction along with the samples generated in the process of purification were seperated by SDS-PAGE followed by Western blotting using a S-tag antibody.

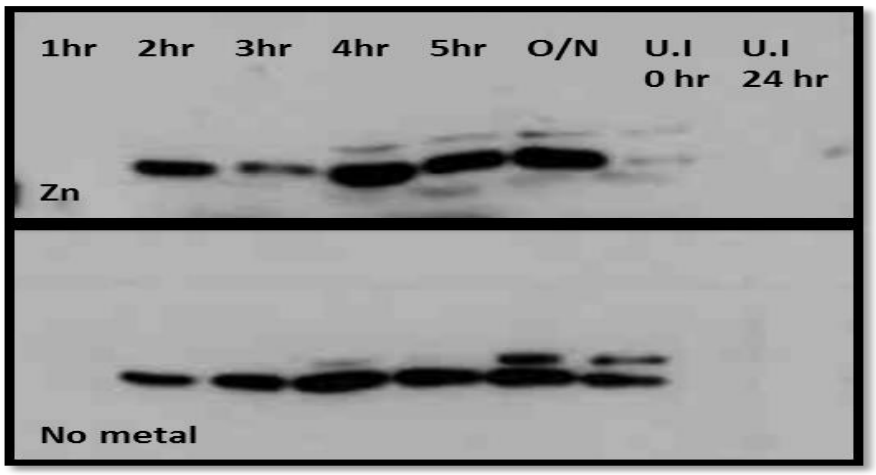


Figure 6-8: Western blotting of the Coomassie stained SDS-PAGE (4-12%) gel of samples induced with 1 mM IPTG over a period of 0-5 hours, overnight at 37°C with Zn (0.1 mM) (A) and without a metal supplement (B) using an anti S-tag antibody. The SDS-PAGE gel was blotted onto a nitrocellulose membrane at 100V for 1 hour. The blot was then probed with a S-tag monoclonal primary antibody and polyclonal goat-anti-mouse immunoglobulins/HRP secondary antibody (U.I- uninduced, O/N- sample induced overnight).

The W08E12.3 protein yield using the S-tag based large scale purification was exceptionally poor (**Figure 6-9**). The Western blot of the thrombin cleaved pure protein fraction retained a faint band signifying the incomplete cleavage by thrombin. The majority of the protein remained in the pellets after two batches of sonication and in the whole cell lysates. Urea based extraction of the protein from these pellets displayed a significant amount of W08E12.3 protein. Overall this suggests that the protein accumulated in the inclusion bodies i.e. the insoluble fraction of the cell lysate.

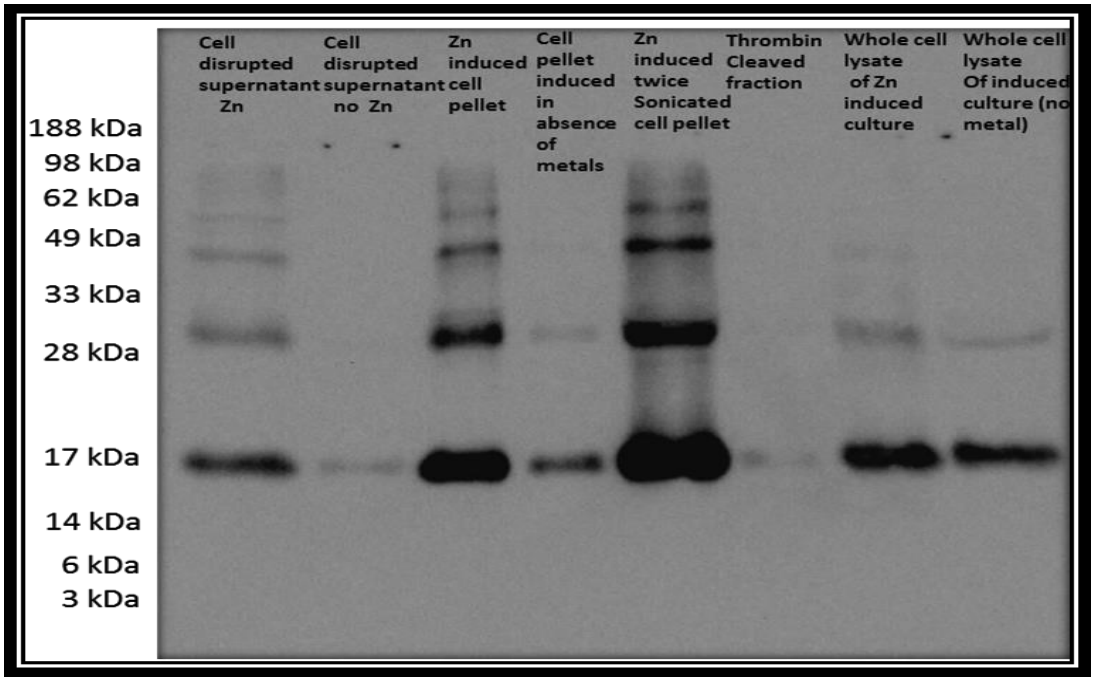


Figure 6-9: Western blotting of the SDS-PAGE (4-12%) gel of samples generated in the process of large scale purification of the W08E12.3 protein using an S-tag system using anti S-tag antibodies. The different fraction of samples obtained while purifying the W08E12.3 protein were boiled with Laemmli's buffer and loaded on to the SDS-PAGE gel. Lane 1: supernatant of Zn exposed culture after sonication; Lane 2: supernatant of culture (in the absence of Zn) after sonication, Lane 3: Pellet of the Zn exposed culture following sonication but then extracted using urea, Lane 4: same as lane 4 but extracted using urea, Lane 5: Culture twice sonicated, then pellet extracted using urea, Lane 7: The concentrated cleaved S-tagged protein, Lane 8-9: Whole cell lysates with and without Zn.

Separation and probing of the soluble and insoluble fractions of the cell lysates on a SDS-PAGE followed by a Western blot revealed that the W08E12.3 protein was predominantly present in the insoluble fraction as inclusion bodies (**Figure 6-10**). The optimal conditions for carrying out a large scale purification were also identified, as the cells in LB broth produced better yield of protein in comparison to the super optimal broth and the 1 mM IPTG lead to optimal induction of W08E12.3 in comparison to the other concentrations that were tested. Hence, for the large scale purification of W08E12.3, the cells were grown in LB broth at 37°C and induced with

1 mM IPTG for 4 hours. To extract the protein from the cells, a method was designed based on an existing technique developed to purify human matrix metalloproteinase 7 from inclusion bodies (Oneda and Inouye, 1999).

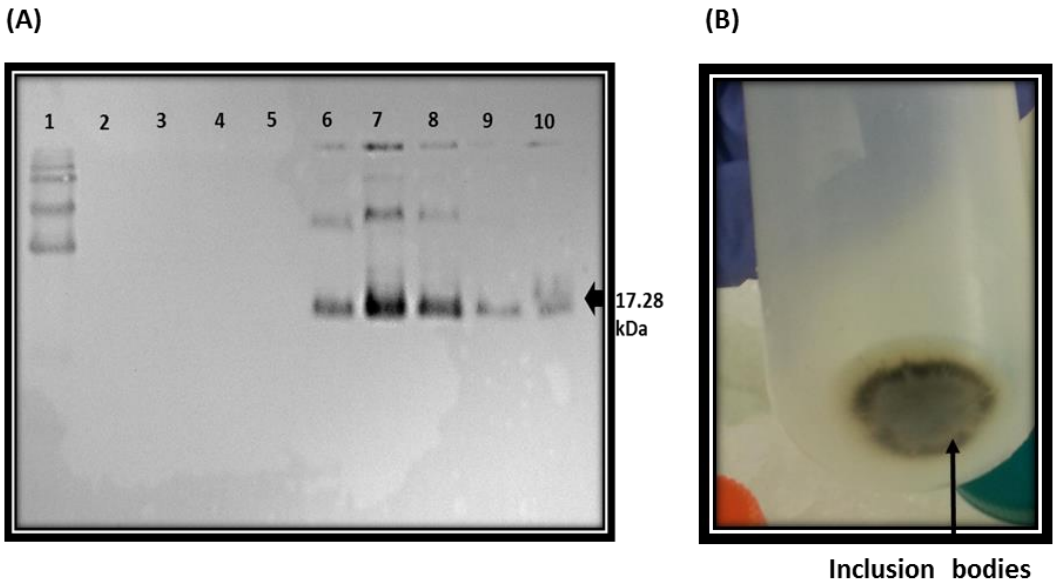


Figure 6-10: Western blotting of the soluble and insoluble cell lysate fractions of samples induced with different IPTG concentrations and different growth media (A), pelleted inclusion bodies after the soluble fraction of the cell lysate is removed (B). Lane 1-Rainbow marker, Lane 2-4: soluble fractions of cells induced with 0.5 mM, 1 mM, 5 mM IPTG respectively grown in LB broth, Lane 5-8: Insoluble fractions or inclusion body fractions of cells induced with 0.5 mM, 1 mM, 5 mM IPTG, Lane 9: Inclusion body fraction of cells grown in super optimal broth with 1 mM IPTG, Lane 10: Inclusion body fraction from the uninduced cells.

The untagged version of the protein was attempted to be purified using the classical chromatographic methods (**Figure 6-11**) but this was not successful. It was very difficult to track the protein of interest as the only source of information available was the mass of the protein. The expression profile of the untagged W08E12.3 protein could not be traced on the SDS-PAGE gels after the anion exchange chromatography, followed by gel filtration chromatography as very low levels of W08E12.3 protein was obtained with a lot of other background proteins (**Figure 6-11B**) and hence it was

assumed that the S-tagged version of W08E12.3 was more stable in expression and easier to purify than the untagged W08E12.3. The S-tag, after W08E12.3 purification was attempted to be cleaved with the aid of Thrombin but the cleavage was unsuccessful even after numerous attempts and always resulted in protein loss. Given that the 15 a.a. long S-tag (**Figure 6-12**) is a neutral charged tag, it was assumed that it will not interfere with the protein function. Hence the purification of W08E12.3 was continued with the S-tag version.

In order to identify the optimal amount of metal required to stabilise the S-tagged W08E12.3 protein expression in the BL21 DE3(pLysS) cells, a growth profile of the expression host bacteria was investigated in the presence of Cd and Zn respectively upon induction with 1 mM IPTG. The cells reached the stationary growth phase at 5 hours in presence of both Zn and Cd while being induced at 3.5-4 hours (0.7 $O.D_{600}$) i.e. at the late log phase. Zn and Cd retarded the growth of the cells after 5 hr, although the effect was more drastic in the case of 0.5 mM Zn (**Figure 6-13A**). The reduction in growth of the cells once the protein is induced within the cells signify that the S-tagged W08E12.3 protein was toxic to the cells. The toxicity was partly conferred by the metals added for induction, Zn deemed to be more toxic than Cd as its concentration was higher by five fold in comparison. The growth was recovered after 10 hr of metal induction in case of both Cd and Zn confirming that the effect of toxicity on the cells faded over time (**Figure 6-13B**). The growth curve emphasises that the harvesting of the cells should be performed after 4-5 hours of IPTG induction as the cells are in their stationary phase.

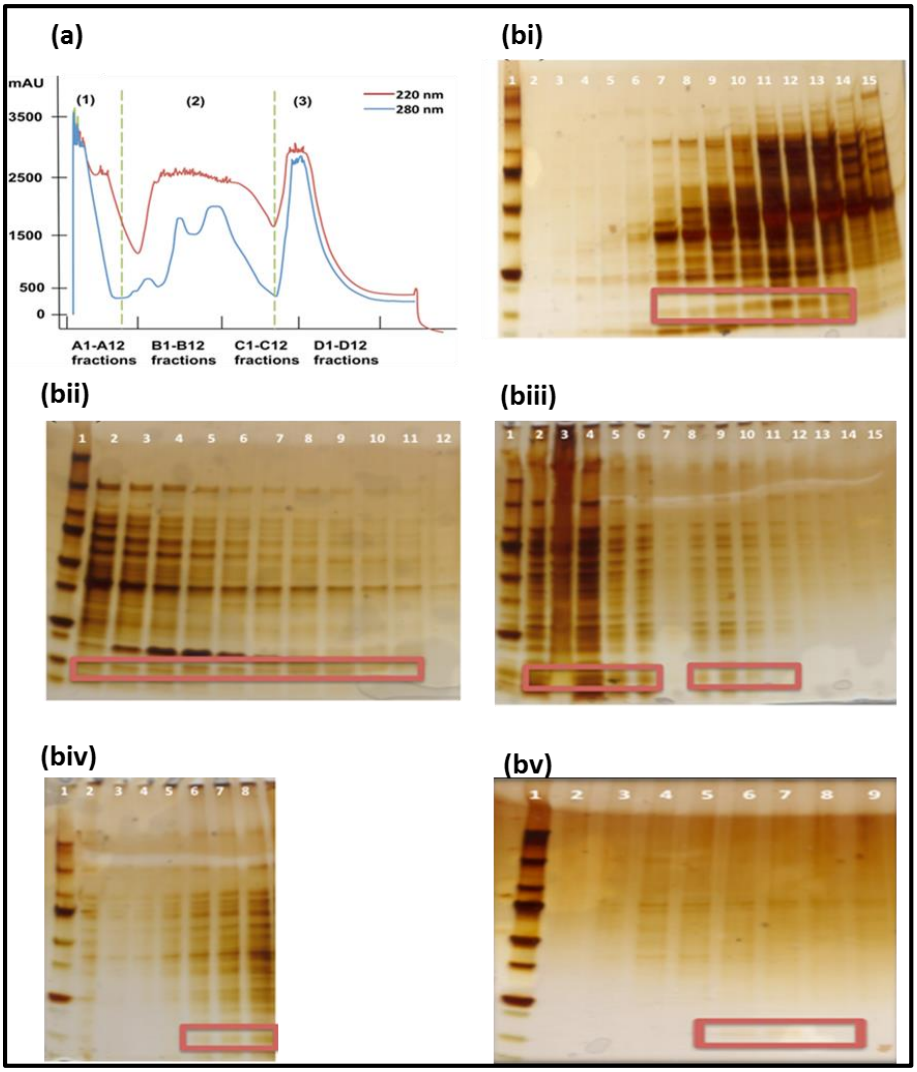


Figure 6-11: Anion exchange chromatography to identify and extract the W08E12.3 protein. (A) The cell disrupted suspension was loaded onto the anion exchange chromatography (column used: HiTrap Q XL, 5 mL, GE Healthcare). The gradient elution and flow rate was set to 2 mL/minute, all the fractions were collected. The UNICORN 5.11 software was used to monitor the chromatography by measuring the absorption at 220 nm and 280 nm representing peptide bonds and aromatic compounds respectively and generating a chromatogram with three major peaks, (1), (2) and (3). (B) Each individual fraction was concentrated and analysed on a 15% SDS-PAGE gel and silver stained (bi) Lane 1: Rainbow marker. Lane 2-15: Fractions b1-b12 and c12-c11 respectively, (bii) Lane 2: Rainbow marker. Lane 2-12: Fractions c10-c1 and d3 respectively, (biii) Lane 1: Rainbow marker. Lane 2-15: Fractions a1-a12 and d5-d6 respectively, (biv) Lane 1: Rainbow marker. Lane 2-7: Fractions from d7-d12 respectively, (bv) Concentrated samples from anion exchange (peak 2, peak 3) chromatography further purified on gel filtration chromatography (chromatogram not shown) and concentrated and tested on 15% SDS-gel.(red box: indicates 14 KDa band) .

A

LysGluThrAlaAlaAlaLysPheGluArgGlnHisMetAspSer

\+ − 0 0 0 0 + 0 − + 0 0/+ 0 − 0

B

AAAGAAACCGCTGCTGCTAAATTCGAACGCCAGCACATGGACAGC

LysGluThrAlaAlaAlaLysPheGluArgGlnHisMetAspSer

pET

Figure 6-12: The amino acid sequence with the net charge (A) and the nucleotide sequence (B) of the S-tag in the pET29a (+) vector.

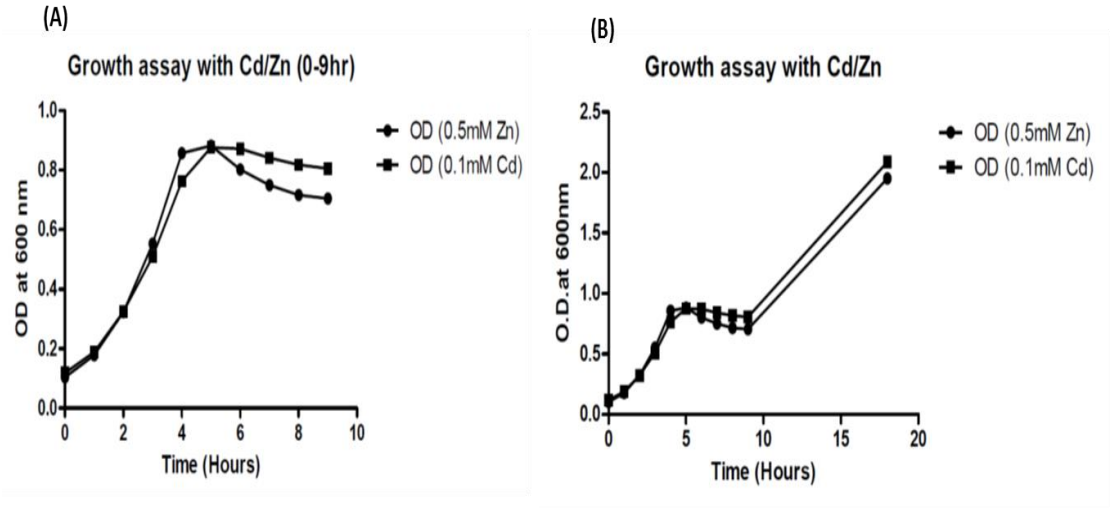


Figure 6-13: Growth curve of BL21 DE3(pLysS) with the pET29a (+ W08E12.3) construct over 24 hours induced with IPTG in the presence of Zn or Cd. The cells were grown in LB media at 37 °C until the $O.D_{600}$= 0.7 and then IPTG (1 mM) induction was started with the addition of Zn (0.5 mM) and Cd (0.1 mM) respectively for 0-9 hr (A) and overnight (B).

Although Cd (0.1 mM) appeared to be less toxic to the cells in comparison to Zn (0.5 mM) (**Figure 6-13**) there was no significant difference observed between the two metal exposures (as the cells resumed normal growth over night, the experiment was not repeated), Western blot detection of the protein induction (**Figure 6-14**) revealed that the expression increased over time (0-4 hours) for the Zn supplemented samples, whereas there was no significant change in expression between the 1 hour induced

sample and overnight sample in the presence of Cd. Hence, it was decided that for the large scale purification of the S-tagged W08E12.3 protein Zn would be added instead of Cd but at a lower dose (0.4 mM) in order to express the protein stably and also to avoid metal related cell toxicity.

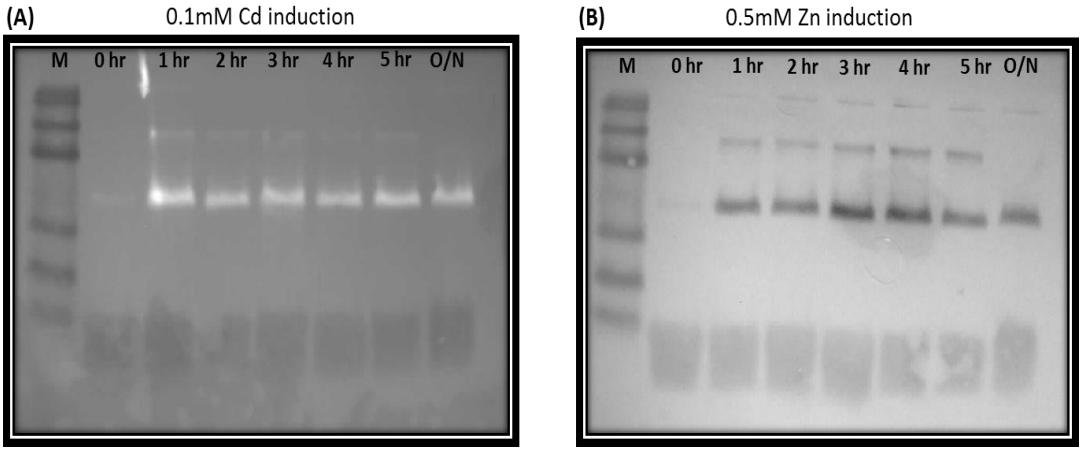


Figure 6-14: Western blotting of the insoluble fractions of the cell lysates from the BL21 DE3(pLysS) containing the pET29a (+ W08E12.3) construct supplemented with metals (Cd and Zn). The cells were grown in LB media at 37°C until the $O.D_{600}$ = 0.6 and then IPTG (1 mM) mediated induction was initiated with the addition of either (A) Cd (0.1 mM) or (B) Zn (0.5 mM) respectively for 0-5 hours and overnight. The insoluble fraction of cell lysate was separated from the soluble fraction by centrifugation and Western blotted with an anti-S tag antibody.

6.2.2 Gel Filtration chromatography

The expressed S-tagged W08E12.3 protein was purified from the inclusion bodies as described in the methods section. The extracted protein from the inclusion bodies was refolded in the presence of Zn (0.1 mM) at pH 8. The presence of protein was monitored by SDS-PAGE at each step, and visualised via Western blotting and silver staining to assess the purification of the S-tagged W08E12.3 protein. The pure protein, buffer exchanged into 10 mM NH_4HCO_3, was loaded onto a gel filtration column (FPLC 16/60 HiLoad Superdex 75 prep grade, Amersham Biosciences). The gel filtration fractions were collected based on the peaks and separated by SDS-PAGE.

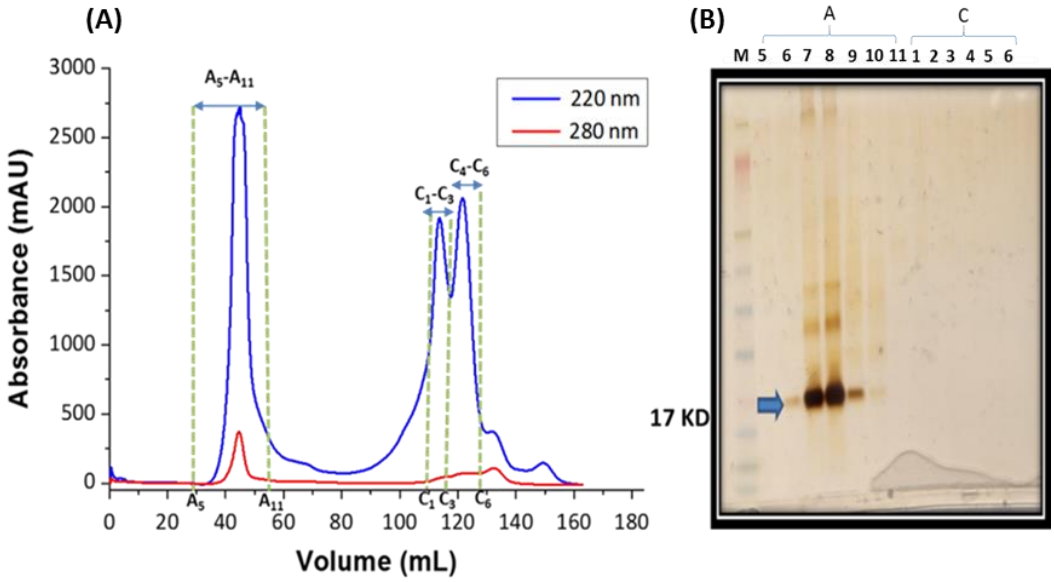


Figure 6-15: Fig: Gel filtration chromatography to purify the inclusion body purified and refolded S-tagged W08E12.3 protein (A) and SDS-PAGE of the isolated peaks from the chromatogram (B). Gel filtration chromatography (column used: HiLoad 16/60 Superdex 75, Amersham Biosciences) was applied to purify the W08E12.3 protein. Isocratic elution and flow rate of 1 mL/minute were set for fraction elution. To monitor the chromatographic process, the UNICORN 5.11 software program was used to generate chromatograms showing absorption at 220 nm and 280 nm representing peptide bonds and aromatic side-chains, respectively. The major peaks are separated and highlighted with lines. The samples were collected for the separated peaks and run on an SDS-PAGE gel which was later silver stained. Lane 1: rainbow marker, Lanes 2-7: fraction A_6-A_{11}, Lanes 8-13: fraction C_1-C_6 respectively.

Three major peaks were observed on the chromatogram after the gel filtration of the inclusion body purified protein (**Figure 6-15A**). When the fractions from individual peaks identified on the gel filtration chromatogram were analysed on SDS PAGE (15%), a strong 17.8 kDa band was observed for peak (A_6-A_{11}) especially for fractions A8 and A9 suggesting that the S-tagged W08E12.3 protein elutes in peak (A_6-A_{11}) (**Figure 6-15B**). The other fractions from peak (C_1-C_3) and peak (C_4-C_6) did not yield any detectable amount of S-tagged protein. The gel filtration fractions

from peak (A_6-A_{11}) were pooled, concentrated and desalted on a PD10 gel filtration column. The desalted protein was checked on a Western blot using S-tagged antibodies and then subjected to protein analysis.

A further expression and purification was carried out to generate more S-tagged W08E12.3 protein in order to perform the protein based assays, but instead of gel filtration chromatography (pH 8), the viability of anion exchange chromatography (pH 9.6) was explored, as this separation method is much faster, and avoids the multiple centrifugation steps involved prior to gel filtration. In this instance, the extracted protein from the inclusion bodies was divided into two equal fractions and refolded in the presence of Zn (0.1 mM) and Cd (0.1 mM), respectively.

The refolded protein was injected directly onto the anion exchange column and the purified samples were eluted in 10 mM NH_4HCO_3 using a NaCl gradient. The protein refolded in the presence of Zn (0.1 mM) and Cd (0.1 mM) eluted at 0.48-0.50 M NaCl at 40-78 mL elution volume (B_{11}-C_2) (**Figure 6-16**). The other peaks did not contain any S-tag bearing proteins, as judged by SDS-PAGE and Western blotting. The protein fractions of interest were pooled, desalted and checked via Western blot analysis using an anti S-tag antibody.

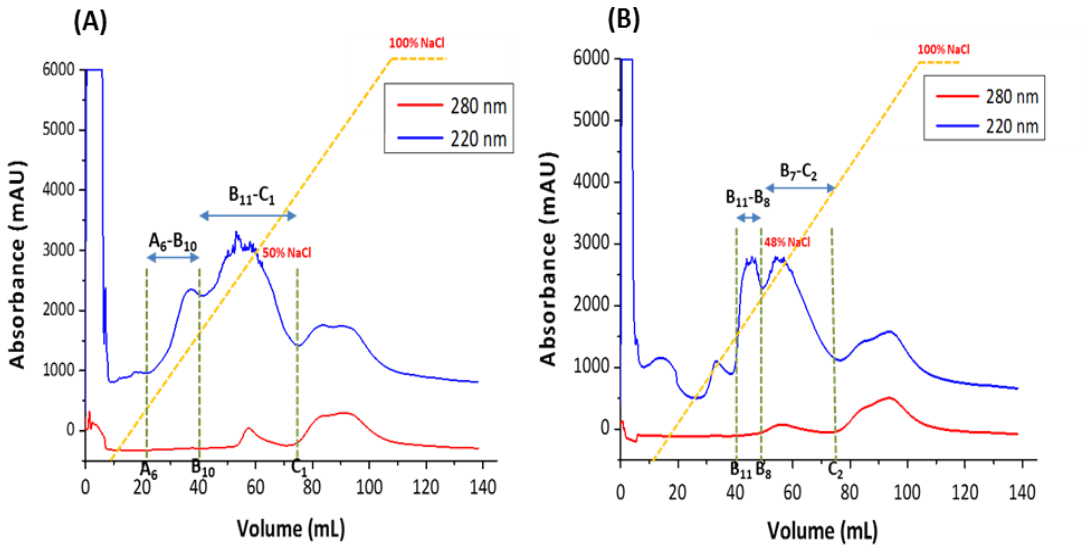


Figure 6-16: Anion exchange chromatography of the protein extracted from inclusion bodies and refolded in the presence of metals, 0.4 mM Zn (A), 0.1 mM Cd (B) respectively. The IPTG (1 mM) induced W08E12.3 protein stabilized using zinc (0.4 mM) was extracted from the inclusion bodies. The protein was split into two fractions and refolded individually with Zn (0.1 mM) and Cd (0.1 mM) respectively. The refolded protein was purified using anion exchange chromatography (column used: HiTrap Q XL, 5 mL, GE Healthcare). The gradient elution and the flow rate was adjusted to 2 mL/minute, the fractions were collected including the flow through. The UNICORN 5.11 software was used to monitor the chromatography process by tracing absorption at 220 nm (peptide bonds) and 280 nm (aromatic residues). Fractions from each peak were pooled together, concentrated and analysed on a 15% SDS-PAGE gel.

The purified protein from the gel filtration (**Figure 6-15**) column and anion column chromatography (**Figure 6-16A**) was subjected to Western blotting in order to probe for S-tagged W08E12.3 protein. Western blotting confirmed that the purified protein was the S-tagged W08E12.3 protein (**Figure 6-17**). From the Western blots it can also be deduced that the purified W08E12.3 protein was present in its dimeric form (35.6 kDa) in addition to its monomeric structure. The purified protein was analysed using ICP-OES, ESI-MS for metal detection, pH titration and metal displacement assay.

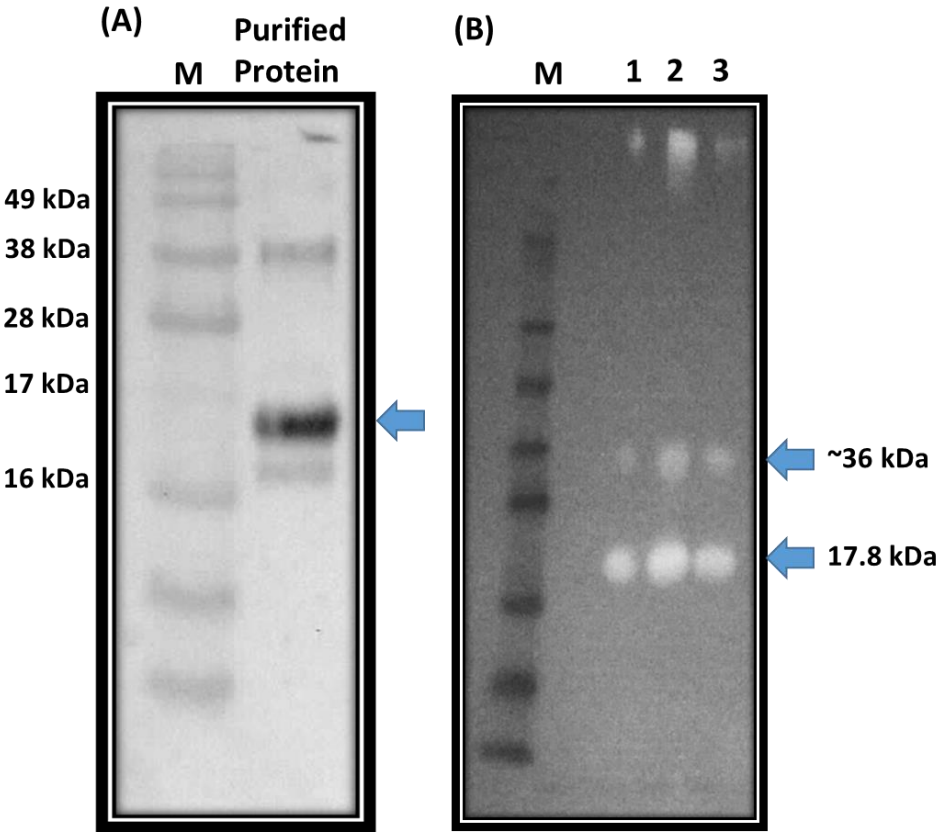


Figure 6-17: Western blotting of the gel purified S-tagged W08E12.3 protein from the insoluble fractions of the cell lysates from the BL21 DE3(pLysS) probed with an anti-S tag antibody. (A) The W08E12.3 protein refolded in Zn and purified by means of gel filtration, (B) W08E12.3 protein refolded in the presence of Zn and Cd respectively and purified using anion exchange chromatography (Lane 1: Zn refolded diluted protein, Lane 2: Concentrated protein refolded in presence of Zn, Lane 3: Concentrated protein refolded in presence of Cd).

6.2.3 ICP-OES

Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES) is an analytical technique designed to measure the amount of metals and some non-metals (e.g. P and S) in a particular sample. Samples introduced into ICP must be in solution. Fine aerosols of the samples are generated by a nebuliser and are atomised and

ionised by an Argon plasma. This process generates excited atoms and ions, which emit element specific radiation in the UV-visible range. The intensity of the emitted radiation can be measured, and with adequate calibration, metal and for e.g. sulfur contents of a sample can be quantified (Skoog and West, 1980).

Once a metal binding protein is isolated there are three major aspects that need to be determined, namely the correct identity of the protein, its concentration, and the number of metals bound to the protein. The confirmation of the protein identity can be achieved by means of a Western blotting assay (for tagged proteins) or by measuring its mass using mass spectrometry. The concentration, especially in case of cysteine-rich proteins, can be obtained by determining the thiol content in the protein by DTNB 5,5′-dithiobis (2-nitrobenzoic acid) (DTNB) (Ellman, 1959). In the case of metal-bound thiols (e.g. like in metallothioneins), it is necessary to incubate the sample with ethylene diamine tetra acetic acid (EDTA), as metal-bound thiols tend to be unreactive in this assay. Alternatively, ICP-OES can be used to measure the sulfur content. ICP-OES remains the most reliable method to quantify both the protein concentration and also the metal:protein ratio simultaneously if the reactions are carried out in acidified conditions (Blindauer and Leszczyszyn, 2010).

All fractions generated in the process of purification of W08E12.3 were prepared for ICP-OES analysis to quantify the concentration of the W08E12.3 protein and also the amount of metal in each of the samples. Typically the samples were diluted fourfold in 0.1 M nitric acid and were then subjected to ICP-OES.

The ICP-OES instrument was calibrated using S, Zn, Cd and Cu 1000 p.p.m. standards (Fisher Scientific), diluted accurately to concentrations matching those expected for the protein samples. Based on the measured sulfur content, the

concentration of S-tagged W08E12.3 protein was determined as it comprises of 22 sulfhydryl groups i.e. 19 cysteine and 3 methionine residues, which in turn allowed the metal:sulfur (and hence metal:protein) ratios to be calculated (**Table 6-1**). It was found that the S-tagged W08E12.3 protein (purified using gel filtration chromatography) displayed a protein: metal ratio of 1:6.97 Zn (**Figure 6-18**) (the S values on the ICP-OES were negative for sample A10 because of absence of an internal standard which could calibrate the instrument better) whereas the anion column chromatography showed ratios of 1:4.5 Zn and 1:3 Cd. It was observed from the ICP-OES data that the amount of protein obtained after refolding in presence of Cd was much lower than the purified protein obtained after refolding in presence of Zn. This correlates with the S-tagged W08E12.3 protein induction data, where upon Western blotting it was revealed that the Cd supplemented samples produced comparatively lower amounts of protein (**Figure 6-14**).

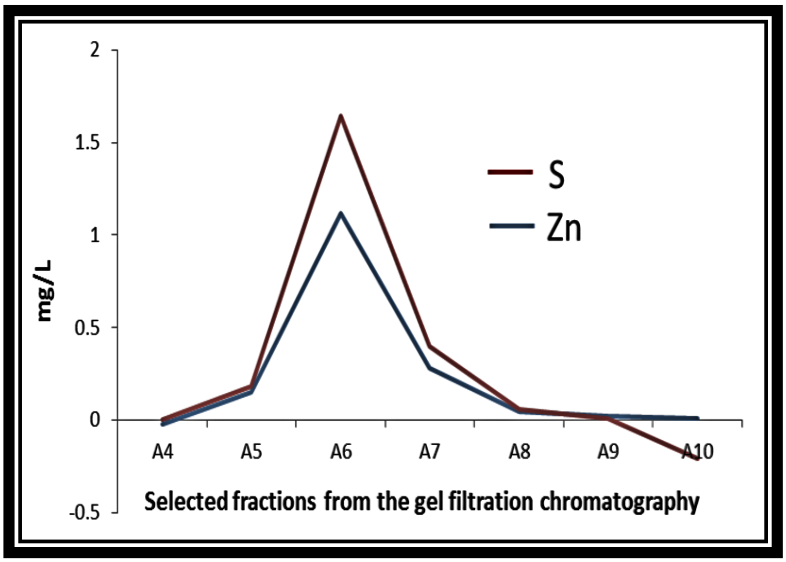


Figure 6-18: ICP-OES data of the individual fractions eluted from the gel filtration chromatography from peak A_4-A_{10} (Figure 6-15) of the S-tagged W08E12.3 Zn (6.5) bound pure protein. The graph depicts the concentration of S and Zn over the collected gel filtration

chromatography samples (peak A_4-A_{10}) based on the outcome of the W08E12.3 protein analysis on ICP-OES.

About 30 mg (1/50th of the total pellet) of the inclusion body (used for anion chromatography) was dissolved in 600 µL of 1 M nitric acid and then diluted 10 fold for ICP-OES sample preparation. This fraction was also analysed on ICP-OES in order to identify the amount of metal that is bound to the protein within the cell inside the inclusion bodies. A protein to Zn ratio of 1:5 was revealed by the ICP-OES analysis. Based on the ICP-OES analysis the S-tagged W08E12.3 protein purified (by anion exchange chromatography) from the inclusion bodies and refolded in the presence of Zn depicted a protein to Zn ratio of 1:4.5.

Table 6-1: The metal concentration, protein concentration and the protein:metal ratio present in the purified S-tagged W08E12.3 fractions derived by ICP-OES.

Sample name	Metal Concentration (µM)	Protein concentration (µM)	Protein: metal
Gel filtration column purified S-tagged W0812.3	244 (Zn)	35	1:6.85(±0.2) (Zn)
Anion exchange column purified S-tagged W0812.3	142 (Zn)	32	1:4.5(±0.3) (Zn)
Anion exchange column purified S-tagged W0812.3	24 (Cd)	8	1:2.90(±0.2) (Cd)
S-tagged W08E12.3 in the inclusion bodies	245 (Zn)	48	1:5.01(±0.1) (Zn)

Hence, in this case the amount of metal (Zn) bound to the protein before denaturation from the inclusion bodies and the amount of metal (Zn) bound after refolding the protein in the presence of metals remains almost the same i.e. 4.5-5 Zn(II) per mole of the S-tagged W08E12.3 protein. But the protein within the inclusion bodies is most likely in the unfolded or misfolded or partially folded form hence, possibly not reflecting the full potential of the S-tagged W08E12.3 protein in metal binding. The difference between the Zn binding ability of the S-tagged W08E12.3 protein in the two purification system highlights that the protein is capable of binding up to ~7 Zn(II) when properly folded, also the presence of Zn bound forms of the protein within the inclusion bodies suggests its strong potential to sequester Zn (**Table 6-1**).

6.2.4 ESI-MS

Electrospray ionisation mass spectrometry (ESI-MS) is typically used for structural study or quantification of metabolites in complex biological samples (Ho et al., 2003). Mass spectra can be directly acquired from a solution, with minimal fragmentation or alteration of the species in the sample. On the basis of the main charge on the protein, the stoichiometry of the protein or protein complex can be studied by the mass to charge (m/z) ratio. The intensities of the molecular peaks provide the abundance and stoichiometry of each species present in the sample. ESI-MS can also be utilised to study the stoichiometry of metal- ligand complexes in very small/ dilute samples (10^{-6} M) (Di Marco and Bombi, 2006). In the Blindauer group "native ESI-MS" is majorly used to study speciation of metalloproteins (Leszczyszyn and Blindauer, 2010).

The 4.5 Zn(II) and 6.5 Zn(II) bound S-tagged W08E12.3 protein was metal depleted by exposure to low pH (3.8) and analysed by positive ESI-MS in order to determine the apo (metal devoid) mass of the protein. The ESI-MS spectra revealed various

peaks indicating the different charged states of the same purified S-tagged W08E12.3 protein (**Figure 6-19A**).

The raw ESI-MS data depicts the relative abundance of each charged state of the S-tagged W08E12.3 to the tallest peak. The peaks indicate the multiple mass assessment of the same protein (Banerjee and Mazumdar, 2012). The deconvoluted ESI-MS spectra (**Figure 6-19C**) reveals the mass (17,279.18 Da) of the identified S-tagged W08E12.3 protein in its neutral charge state which is very close to the theoretical mass (17,280.5 Da). This confirmed that the 22 sulfur containing groups (19 cysteines and 3 methionines) in the S-tagged W08E12.3 protein were in their reduced state, or else the mass of the S-tagged protein would have been 22 Da less than the expected mass (as each cysteine or methionine would have one less H atom) (Zeitoun-Ghandour et al., 2011).

The metallated form of the S-tagged W08E12.3 protein (both, the gel purified and anion exchange chromatography purified protein) was attempted to be analysed by ESI-MS at pH 7.4 but this did not yield any expected metal-bound species, hence the ESI-MS data are not shown. It may be assumed that metallated W08E12.3 protein does not ionise well at neutral pH.

6.2.5 pH titration

For proteins containing multiple metals (for example with metallothioneins) it is complicated to define the metal affinity, as there may be multiple binding affinities (Blindauer and Leszczyszyn, 2010). However, a relatively simple way of establishing metallothionein-like character for a protein involves the determination of its pH of half displacement, as for Zn-MTs, this value is typically below pH 5, whereas other proteins containing Zn-Cys sites demetallated at much higher pH.

As the W08E12.3 protein was shown to bind 4.5-6.5 Zn(II), the affinity of the Zn to the protein was tested by its competition against hydrogen ions upon decreasing the pH i.e. by calculating the pH of half metal displacement by following the absorbance of the Zn-S Ligand to Metal Charge Transfer (LMCT) bands at constant absorbance (ca.) 230 nm (Szunyogh et al., 2015).

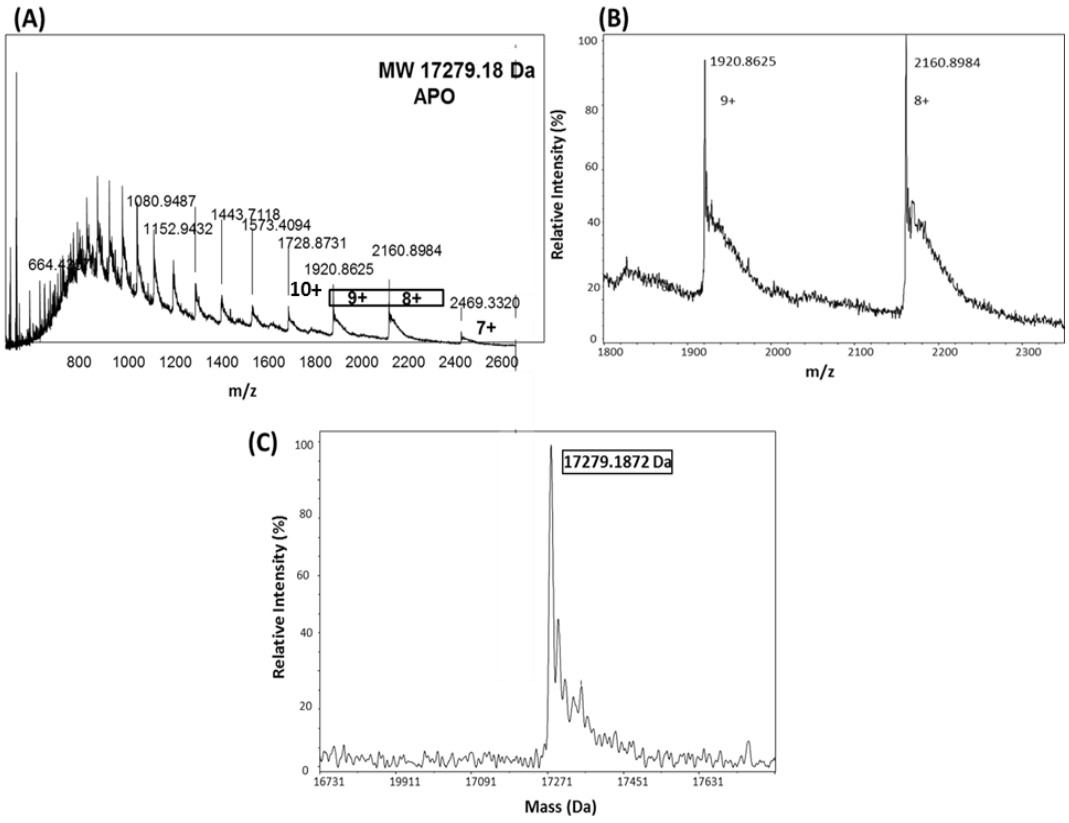


Figure 6-19: Identification of the S-tagged W08E12.3 protein in its apo state (pH 3.8) on ESI-MS. (A) Raw ESI-MS spectrum depicting the intensity of each peak (40 µM protein in 10 mM ammonium bicarbonate, 10% methanol and 2% formic acid) versus mass: charge ratio (m/z) depicting the multiple charge state of the S-tagged W08E12.3 demetallated protein. Each peak is representing a different charge state of the protein, some of which are labelled. (B) Zoomed region of the ESI-MS spectrum of the S-tagged W08E12.3 demetallated protein showing its 9+ and 8+ charge states. (C) The accurate mass spectrum of the tagged W08E12.3 demetallated protein obtained by ESI-MS depicting the deconvoluted mass of the neutral species.

The relative affinity of Zn/ Cd to the S-tagged W08E12.3 protein was checked by U.V absorption spectroscopy. The metal bound protein (5 µM) was titrated with 1N HCl to attain the metal devoid (apo) state of the protein. The demetallation of the protein commenced as hydrogen ions from the acid started to compete with the metals in the S-tagged W08E12.3 protein (**Figure 6-20**). When the Zn bound protein was titrated with the acid (1 M HCl) the A_{230} decreased from pH 7.98 to 6.76, which indicates a loss of Zn from the protein. However around pH 6.4 the absorbance across a wide range of wavelengths, including 230 nm, increased drastically, suggesting that the W08E12.3 tagged protein became insoluble and aggregated. On further reduction of the pH to 4.3 the protein re-solubilised and the A_{230} was lower than the A_{230} for pH 6.76. On further acidification until pH 2.8, there was hardly any change in the absorption spectra, suggesting the demetallation was complete by pH 4.3. Once the protein was demetallated, lowering the pH below 2.8 lead to insolubility and aggregation of the protein.

The addition of acid (1 M HCl) to the Cd bound protein resulted in a similar trend where a decrease in A_{254} (as LMCT band is at lower energy) indicated metal loss immediately after addition of the first aliquot of acid (i.e. when the pH was reduced from 7.92 to 7.54). The metal loss continued until pH 6.7, below which the protein became insoluble and precipitated. The protein re-solubilised at pH 3.3 and once the protein was in its apo form it precipitated at a pH lower than 3.

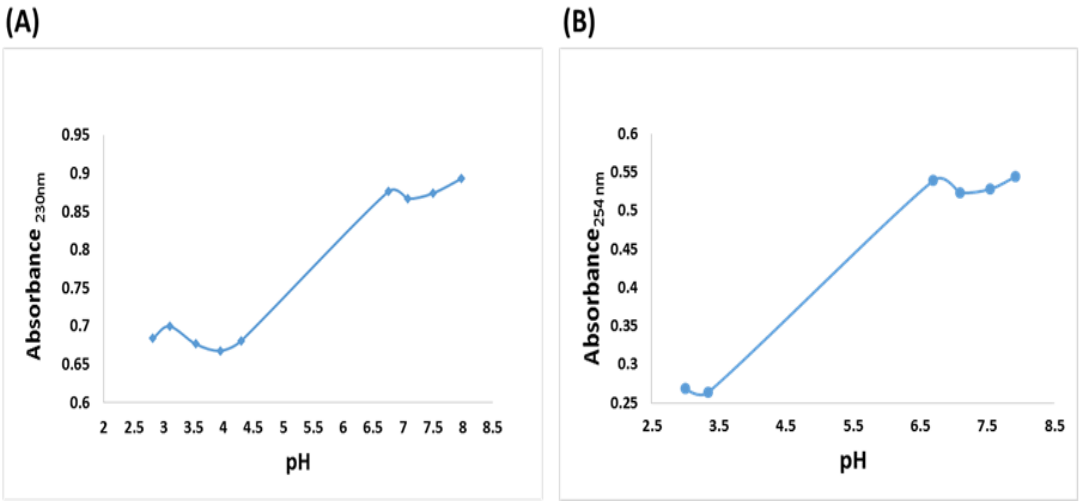


Figure 6-20: pH titration of S-tagged W08E12.3 monitored by the measurement of absorbance at 230 nm and 254 nm for Zn bound and Cd bound protein respectively. pH titration of (A) Zn bound W08E12.3 tagged protein and, (B) Cd bound W08E12.3 tagged protein. The W08E12.3 tagged protein was titrated with 1 M HCl to reduce the pH of the protein solution, with each point addition, the protein solution was left at room temperature for 10 minutes to permit equilibration of the solution and then the pH was recorded and the absorbance (200-600 nm) on the U.V spectrometer was measured. The absorbance measurements at 230 nm or 254 nm were plotted against the respective pH values excluding the outliers (where the protein aggregated).

It can be concluded from the pH titration assay that metals are lost from the Zn/Cd bound S-tagged W08E12.3 protein at a high pH (~7.5) and the apo protein is attained at low pH (~3). Thus, although it was not possible to determine a pH of half-displacement, as the S-tagged W08E12.3 protein was not soluble for a large section of the relevant pH range, it can be concluded from the immediate onset of demetallation even above pH 7, that neither Zn nor Cd were particularly firmly bound to W08E12.3 as isolated.

6.2.6 Metal exchange

To establish whether the bound Zn ions in the purified S-tagged W08E12.3 protein could be displaced by Cd (which generally has higher affinity for thiol-rich binding sites), a metal-displacemnt titration experiment was carried out. The aim was to understand the lability of bound metals and further investigate the binding stoichiometry (Nielson et al., 1985). Metal displacement would reveal information about the structural binding of Zn to the W08E12.3 protein, as the arrangement of the metal within the protein confirms its strength of binding to the protein (Sakulsak, 2012).

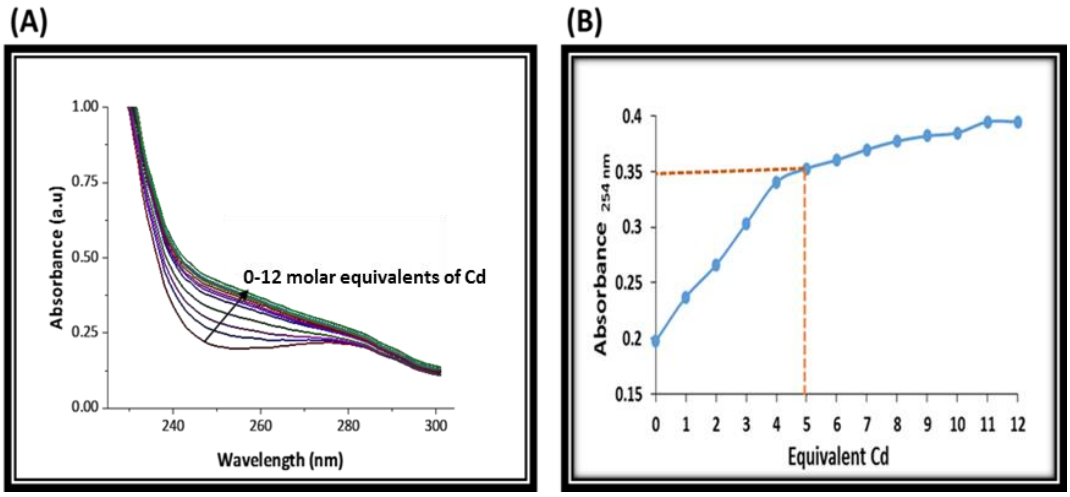


Figure 6-21: Spectrophotometric Cd (1-12 molar equivalents) titration of the 4.5 Zn(II) bound S-tagged W08E12.3 protein. (A) Full UV spectra recorded from 200-300 nm range after each addition of (0-12 molar equivalents) Cd(II), (B) The absorbance at 254 nm, which is associated with Cd-thiolate coordination, is plotted as a function of the molar ratio of Cd(II) added to the protein. After each addition of Cd, the protein was incubated in the quartz cuvette at room temperature for 10 minutes before the U.V spectrum was recorded. The red dotted line indicates the point of at which the slope of the line shifts less steeply where the majority of the Zn has been replaced by Cd.

The protein with 4.5 Zn(II) bound was incubated (10 minutes) at room temperature with various amounts of $CdCl_2$ (0-12 molar equivalents with respect to the total

protein used for the assay) and after each incubation the U.V absorption spectrum was recorded.

The metal stoichiometry was determined by the Cd titrations against the 4.5 Zn(II) loaded S-tagged W08E12.3 protein. It was noticed that upon Cd addition, the Zn atoms were being displaced from the protein until the protein was completely saturated with Cd ions (**Figure 6-21A**). The absorbance of the protein at 254 nm was plotted against the equivalent amount of Cd in order to define the point at which all the Zn ions from the protein were displaced by Cd (**Figure 6-21B**). It was observed that at ca. 5 molar equivalents Cd, the increase in absorbance became smaller. This value is similar to the amount of Zn(II) initially bound to the protein, and it may be speculated that the most favourable metal-binding sites that had been occupied by Zn(II) are occupied by Cd at this point. The further, less pronounced increase between 5 and ca. 11 molar equivalents may indicate the population of further binding sites with lower affinities (methionine or other sulfur containing groups followed by non-sulfur groups).

6.3 Discussion

Given that the previous chapters uncovered that the *W08E(12.2-12.5)* family is transcriptionally activated by metals, it was important to investigate whether the protein is also able to bind metals. The *W08E12.3* isoform was selected as it was the most characterised isoform at gene level. Expression of functional proteins in heterologous hosts forms the basis of modern biotechnology (Gustafsson et al., 2004) and cloning and expression in *E. coli* serves as a medium to obtain large amounts of protein. A popular expression system used to produce recombinant protein is the tag

fused protein approach which enables the selective purification of the protein of interest with minimal purification steps (Guan and Dixon, 1991).

The W08E12.3 protein was initially expressed as a fusion protein with an affinity tag (intein) based system. However, the intein fused W08E12.3 protein could not be purified as the tag cleaved off from the W08E12.3 protein during the induction process within the *E. coli* cells, likely due to the low intracellular pH. Therefore, an alternative affinity tag based system was implemented to express and purify the W08E12.3 protein, namely the S-tag system.

Attempts to purify the untagged form (i.e. in its native form) failed due to the inability to track and identify the protein, hence the S- tagged version of the protein was selected for further studies. The purification of the S-tagged W08E12.3 protein from the soluble fraction of the induced *E. coli* (DE3 pLysS) cells proved challenging and a Western blot with antibodies against the S-tag revealed that the protein was predominantly located in the insoluble fraction i.e. the inclusion bodies within the cells. The protein was attempted to be solubilised with the addition of Zn (0.4 mM) or Cd (0.1 mM) during induction, but the tagged W08E12.3 protein remained in the insoluble inclusion body fraction. As bacteria lack the capability of protein folding and quality control mechanisms found in eukaryotes, the possibility that recombinant proteins accumulate in inclusion bodies is high. Typically, the inclusion bodies isolated from over-expressing cultures mostly contain the recombinant protein expressed from the vector (Kurucz et al., 1995). Inclusion bodies consist of biologically inactive solid protein material and hence solubilisation of these aggregates is essential for refolding the protein in its native structure (Hlodan et al., 1991). Hence, a method developed to purify proteins from the inclusion bodies was implemented (Oneda and Inouye, 1999).

The tagged protein was isolated from the inclusion bodies under denaturing conditions in the presence of guanidine hydrochloride (6 M) to remove impurities such as nucleic acid, proteins and phospholipids (Maachupalli-Reddy et al., 1997) and refolded in the presence of L-arginine (0.25 M) and metals (Zn (0.1 mM), Cd (0.1 mM)). Cofactors such as Ca, Zn or Cu often accelerate proper folding and the retention of biological function of the proteins (Bushmarina et al., 2006). Furthermore, the physiological activity of cysteine rich proteins depend on the reactivity of the sulfhydryl group which is lost after oxidation (Dische and Zil, 1951). Hence refolding of the cysteine rich S-tagged W08E12.3 protein was performed by adding reducing agents (DTT) and also metals (Zn (0.1 mM), Cd (0.1 mM)). The re-folded, potentially metallated proteins were purified by gel filtration or anion exchange column chromatography, analysed by SDS-PAGE, and selected fractions were pooled. The identified pooled peak fractions (gel filtration fractions A_6-A_{11} 6.5 Zn(II) bound protein, anion exchange fractions B_{11}-C_1 4.5 Zn(II) bound and anion exchange fractions B_{11}-C_2 3 Cd(II) bound) were the full length S-tagged W08E12.3 protein, validated by Western blot analysis with an anti-S–tag antibody, ICP-OES and ESI-MS.

The ESI-MS analysis was performed to confirm the mass of the purified S-tagged W08E12.3 protein and also the metal binding stoichiometry based on the change in metallated mass from the metal-devoid (apo) protein. Unfortunately, ESI-MS spectra were only obtained for the metal-devoid (apo) form of the tagged protein, due to the inability of the metal bound form (4.5-6.5 Zn(II) at pH 7.8) to ionise well in the ESI-MS. It is possible that the protein was forming larger aggregates in neutral solution.

The metal-devoid form confirmed that the purified protein was intact S-tagged W08E12.3 protein displaying a deconvoluted mass (17279.18 Da) very close to its theoretical apo mass (17280.50 Da), and indicating a negligible amount of oxidation.

The peaks observed in the raw ESI-MS spectra of the S-tagged W08E12.3 (apo) protein signified the multiple charging of the protein, typical of ESI-MS spectra. This distribution of charge states observed in the ESI-MS is dependent on the molecular size as well as the electrospray conditions, and may be exploited to infer the tertiary state (folded/unfolded) of the protein (Konermann and Douglas, 1997). The intensities of the charge states of the demetallated protein at low pH indicated the presence of two conformations, one following a roughly Gaussian distribution ranging from +26 to +10, and a second with +9 and +8 as most intense peaks. This is most clearly seen by the higher intensity of the +9 charge state as compared to the +10 state. The charge state distribution at higher charges likely correspond to completely unfolded protein, whereas the second distribution at higher charge states is compatible with a more globular structure.

Metallothioneins are cysteine rich proteins and the thiol groups are instrumental in the binding of essential metals (e.g. Cu, Zn) and toxic metals (e.g. Cd, Pb) (Nordberg and Nordberg, 2000). Mammalian metallothioneins contain 20 cysteine residues capable of binding 7 Zn ions (Maret and Vallee, 1998). The S-tagged W08E12.3 protein comprises 19 cysteine residues, thus it is hypothetically able to bind 6-7 M(II) ions, based on typical metal:thiolate ratios observed for metallothioneins (Freisinger, 2011). Based on the sulfur content (3 methionines + 19 cysteines) of the protein, the ICP-OES data revealed that the S-tagged W08E12.3 protein binds 4.5-6.5 Zn(II) ions and 3.5 Cd(II) ions, depending on the extent of protein folding and method of isolation. It was observed that even the improperly folded protein within the

inclusion bodies could accommodate 5 Zn(II). This suggests a significant Zn binding affinity of the S-tagged W08E12.3 protein.

The interaction of metals with proteins plays an important role in determining their toxicity. There are at least two modes by which toxic metals may affect proteins i.e. either by binding to the free thiol groups or by displacing Zn or other essential metals bound to metal binding proteins (Sharma et al., 2008). To understand the metal-protein stoichiometry of the S-tagged W08E12.3 protein first the affinity of binding of metals to the protein was estimated (pH titration) and then a metal displacement assay was performed.

In order to estimate the Zn binding affinity of the tagged protein, metal loss was induced (pH of half metal displacement) by increasingly acidic pH conditions. However it was not possible to acquire a continuous titration curve, as the tagged W08E12.3 protein aggregated between pH 6.4 and 4, which coincides with its theoretical isoelectric point (pI=5.55). Solution pH defines the type (positive/negative) and total charge on the protein, thereby dictating the electrostatic interactions (Krebs et al., 2009). Many proteins aggregate around their pI as the net charge on the surface of the protein approaches zero and instead of interacting with the water molecules and ions in the solution, the protein molecules interact among themselves leading to aggregation. However, once below pH 4, the tagged W08E12.3 protein re-solubilized and displayed a significant reduction in 230 nm absorbance in comparison to the absorbance observed at pH 7.8. This confirmed that Zn loss was pH dependent and that Zn binding to the tagged W08E12.3 protein *in vitro* was in fact only moderate, when compared to the typical behaviour of true metallothioneins.

The metal displacement assay of the 4.5 Zn(II) bound S-tagged W08E12.3 protein, titrating with Cd(II) ions, investigated the lability of Zn(II) ions bound to the tagged protein. Given that the absorbance at 254 nm increases in proteins with Cd-cysteine coordination complexes (Nielson and Winge, 1983), the observed change in the 254 nm absorbance indicates Cd binding to the S-tagged W08E12.3 protein. The Cd ions first occupy the empty sites i.e. the free thiols present on the sub-metallated Zn4 or Zn5 bound protein leading to a slight change in the curve with the addition of 2.5 molar equivalent Cd(II) and then a further change in the curve is expected at 4-5 molar equivalents of Cd(II). This indicates that these sites have a good affinity for Cd. The titration curve did not completely linearize with the addition of 5 molar equivalent Cd(II), which suggests that the 4.5 Zn(II) bound protein contains further binding sites with lower affinities, resulting (if fully utilized) in binding of up to 7 M(II).

Overall, from the protein-metal studies of the S-tagged W08E12.3 it can be concluded that the protein accumulates within the inclusion bodies in the bacterial cells upon induction. The accumulation of the expressed protein within the inclusion body fraction of *E. coli* suggest the possibility of W08E12.3 to be a membrane protein as overexpression of membrane proteins within BL21 (DE3) cells either lead to cell death or accumulation of the protein within inclusion bodies (Miroux and Walker, 1996).

Within the inclusion bodies, the protein binds to at least five Zn(II) ions, suggesting that the S-tagged W08E12.3 protein, even when partially folded, is capable of accumulating Zn(II). After the purification from the inclusion bodies, upon refolding in presence of Zn(II), the protein was shown to bind up to 6.5 Zn(II), indicating that all 19 cysteine residues may take part in metal binding once the protein is properly

folded. Considering that different purifications yielded different stoichiometries (e.g. 4.5 Zn(II), 3 Cd(II) after anion exchange chromatography), it is likely that not all M(II)-binding sites have similar affinities, but that at least 2 or 3 have weaker sites. Alternatively, it is also possible that the lower metal:protein stoichiometry in some preparations was due to imperfect protein refolding. It is interesting to note that despite the high number of cysteine residues, and despite the generally higher affinity of Cd for cysteine-rich binding sites, the protein expressed or purified in the presence of Cd was not only less abundant, but also characterized lower metal:protein stoichiometry, which is in contrast to what may be expected.

6.4 Conclusion

The S-tagged W08E12.3 protein is capable of binding 4.5-6.5 Zn(II), with the higher value close to the theoretical maximum of 7 Zn per 19 cysteines. The Zn-cysteine coordination is redox active in proteins where oxidation leads to Zn mobilisation while reduction leads to Zn accumulation and binding. Some Zn proteins act as redox sensors indicating that Zn homeostasis is also essential for protection against cellular oxidative stress (Maret, 2006). The Zn in the S-tagged W08E12.3 protein is easily displaced by Cd upon titration. Both Cd and Zn do not bind very strongly to the tagged W08E12.3 protein as metal loss was observed even at slight increases in proton concentration of the protein solution. The relatively moderate metal affinity was also corroborated by the absence of metallated species in ESI-MS spectra. Overall, it can be stated that the *W08E(12.2-12.5)* family is, in addition to being metal responsive, also metal binding in nature.

The ability of the S-tagged W08E12.3 protein to bind metals and the observation that Zn accumulates in cells upon the knockdown of *W08E(12.3-12.5)* (as determined in

the previous chapter), supports the notion that *W08E(12.3-12.5)* may act as Zn sensors.

Chapter 7: General Discussion

7.1. Discussion

Global industrial development has led to an increased waste discharge in the environment especially in soil and water. The waste comprises of heavy metals which accumulate primarily in the urban areas. The industrial waste dumping poses a major health threat to the world, as the heavy metals accumulated cannot be broken down into non-toxic forms, ensuring long lasting effects on the eco system (Dixit et al., 2015). Many of the metals like Cd, Cr, Cu, Pb, Zn, Hg etc. are not just cytotoxic in low doses but also pertain mutagenic and carcinogenic properties (Hanaa et al., 2000).

The mechanisms by which heavy metals impose toxic effects have not been well defined, it is hypothesised that they are toxic due to their affinity towards essential ligands like sulfhydryl or nitrogen groups (Huang et al., 1980). As some of the heavy metals are not readily excreted from the system, detoxification mechanisms are essential (Durnam and Palmiter, 1981).

Metallothionein (MT), a ubiquitous, cysteine rich, small, metal-binding protein has been proposed to play a major role in heavy metal detoxification. The MT depicts a high affinity for d10 electron configuration metals like Zn, Cd, Cu and Hg (Isani and Carpene, 2014). The mammalian metallothionein comprises of 20 cysteines out of the 60 amino acid sequence and are capable of binding 7-12 metal ions per polypeptide (Robbins et al., 1991).

Regulation of the transcription of metallothionein is monitored by different stimuli, especially by heavy metals, glucocorticoids and oxidative stress. The metallothionein

promoter region in most of the organisms comprise of metal response elements (MREs) which regulate the cellular response to metal stress by binding to metal responsive transcription factor (MTF-1) (Zhang et al., 2003). MTF-1 not only deals with metal homeostasis and detoxification but also plays a role in dealing with hypoxia and oxidative stress (Green et al., 2001). MTF-1 is well characterised in human, fish, mouse, capybara and drosophila and also in invertebrates like hydra, sea urchin, sea anemone, typically present as a single copy gene (Günther et al., 2012). Although the presence of metallothionein has been confirmed in *C. elegans* and yeast, both lack MTF-1.

Among the nematoda family, Genbank comprises MT sequences of only four species namely *Trichinella spiralis*, *Caenorhabditis elegans*, *Caenorhabditis remanei* and *Caenorhabditis briggsae*, and in depth studies have only been carried on *C. elegans*. The *C. elegans* genome encodes for two multifunctional metal binding isoforms CeMT-1 and CeMT-2 (Hockner et al., 2011). CeMT-1 comprises of four histidines while the CeMT-2 has one histidine required for coordination performance of the two isoforms (Imagawa et al., 1990).

The MTs in the nematode differ from the mammalian MTs in the positioning of their cysteine residues and their *N*-terminal amino acids (alanine, valine in *C. elegans* while methionine in mammals). Both the metallothionein polypeptides are cysteine rich (25-30%) and display Cys-X-Cys and Cys-Cys motifs, but the metallothonein isoforms in *C. elegans* are not homologous to any other species (Freedman et al., 1993). The identified functions of MTs include essential trace metal homeostasis and metabolism (e.g. Zn, Cu) (Palmiter, 1998), prevention from oxidative stress (Colangelo et al., 2004), role in protein folding by acting as chaperones and protection against Cd and other toxic stressors (Beattie et al., 2005).

Reverse genetics describes the study from gene sequence to gene function rather than phenotype to gene sequence as in forward genetics. Although the entire genome of *C. elegans* has been sequenced it is still a challenge to determine the function of every gene in the genome. Reverse genetic investigations are typically performed by disrupting or modifying the gene or gene product and then assessing the phenotype. The range of approaches include gene silencing by RNAi, chemical mutagenesis, targeted gene disruption by homologous recombination and insertional/transposon mediated mutagenesis (Tierney, 2005). Advances in genomic technology allows the determination of gene and protein expression levels using microarray or mass spectrometry but they do not determine the gene function completely. Recently it was shown in yeast that functional inferences based on gene expression data alone can be misleading. Phenotype centric processes or forward genetics is not feasible for genome wide gene analysis due to the enormous efforts required to identify each gene function for an observable phenotype (Alonso and Ecker, 2006).

Researchers have realised that the availability of a sequenced genome alone cannot elucidate biological function of the genes as cell survival depends upon various metabolic and regulatory pathways. There is no strict linear relationship between genes and the proteins they encode however they are complementary to each other. Even after the advances in bioinformatics, the predicted genomic data is not always accurate. Hence, use of proteomic methods to verify the function of a gene product is an important first step in 'annotating the genome'. It is also necessary to determine the protein expression level as it may or may not be directly correlated with the mRNA expression levels (Pandey and Mann, 2000). Hence, both quantification of gene expression and protein expression level studies are essential for identifying the biological role of the *W08E(12.2-12.5)* isoforms.

The *C. elegans* genome is approximately 100 million bp with nearly 20,000 protein coding genes, indicating the presence of a compact genome architecture. Presence of four highly similar genes (≥90%) next to each other emphasises the importance of the *W08E(12.2-12.5)* family. The frequency of occurrence of four or more highly similar genes consecutively on the *C. elegans* genome was assessed by performing a genome wide screen using a bioinformatics program written in PERL, setting stringent criteria for the assessment of similarity (E=0.0001). The identified gene clusters were subjected to Gene Ontology to investigate their functions. The genome wide analysis revealed that the frequency of having four or more similar genes consecutively arranged on the genome is a rare event (0.6%). The Gene ontology analysis identified that the serpentine receptor family and the F-box protein family were most frequent. The existence of the large number of serpentine receptors throughout the worm genome reconfirms that chemoperception is an essential sense to the soil nematode and the F-box proteins are abundant to facilitate protein-protein interactions. Overall, the genome wide analysis highlighted that the occurrence of four highly similar genes as a rare event and that these clusters in the genome constitute highly essential genes. This suggests that the existence of the *W08E(12.2-12.5)* family within *C. elegans* is unique and important.

A protein-protein sequence alignment using BLAST unveiled a 100% identity of the *W08E(12.2-12.5)* with uncharacterised genes present in other *Caenorhabditis* species including *C. remanei* and *C. briggsae*. The sequence alignment also revealed a homology (~ 65-75 %) with keratin associated proteins in *Homo sapiens* and *M. musculus*. The hair keratin intermediate filaments in the hair cortex are embedded in a matrix comprising keratin associated proteins, which form disulphide bonds with abundant cysteine residues of hair keratins (Rogers et al., 2004). Cytokeratins have

proven to resist Cd induced apoptosis by representing an adaptive survival mechanism (Lau and Chiu, 2007). In the case of As poisoning, a temporary accumulation of the metal takes place in the soft tissue organs but the major long term accumulation takes place in keratin rich tissues such as nails, hair and skin (Hu, 2002). Cu, Hg, Ag, Cd, Pb, Cr and Al have been reported to have been taken up by wool keratin (Kar and Misra, 2004). Keratin protein fibres from hair, wool, horns and feather have been used to absorb heavy metals in industry for water purification and air cleaning (Aluigi et al., 2008). The homology of *W08E(12.2-12.5)* with keratin associated protein confirms the possible association of the gene family with heavy metals.

As nematodes lack MTF-1, transcription factors like ELT-2 and GATA regulate the expression of metallothionein (Moilanen et al., 1999). The *W08E(12.2-12.5)* isoforms code for cysteine rich (18-19 cysteines) proteins similar to the metallothionein isoforms (19 cysteines), suggesting a possible role in metal binding or metal response.

Manual screening of the upstream promoter region of the *W08E(12.2-12.5)* isoforms revealed transcription factor binding sites, namely metal binding factor (MBF-1_CS), metal response element (MRE_CS2), Zn response elements (EBV-ZRE5), Zn finger transcription factor (NF-E1_CS1), GATA elements, GR-MT-IIA (metallothionein specific transcription factor) as well as stress response elements such as hypoxia inducible factor (HIF) and hypoxia response element (HRE) indicating putative metal responsive or metal binding activity of the *W08E(12.2-12.5)* family.

A list of common Zn binding or coordinating transcription factors (BLMP-1, GATA, CHE-1 and ELT-3) were identified upon comparing the promoter regions of *W08E(12.2-12.5)* isoforms and the metallothionein isoforms on a pre-existing matrix (Zn binding/coordinating transcription factors) on JASPAR CORE database

suggesting a possible linkage between the two families. Overall the bioinformatics screening of the promoter region of the *W08E(12.2-12.5)* isoforms suggest that they are likely to be functionally similar to metallothionein and may be involved in metal homeostasis or detoxification.

The sequence information of the isoforms to perform the bioinformatics analysis was obtained from the WormBase (www.wormbase.org). The alignment of the amino acid sequence and nucleotide sequence of the coding region as well as the upstream promoter region of the four isoforms revealed them to be highly similar, conserved and cysteine rich. The *W08E12.4* and *W08E12.5* were found to be identical in their coding regions as well as non–coding intron regions but their promoters were different. In contrast, the promoter region (500-1000 bp sequence upstream to the genes) of *W08E12.3* and *W08E12.4* were identical to each other but not the coding region. The high level of similarity between the four genes suggested a recent duplication event. However *W08E12.2* was found to be comparatively different in certain parts of the coding region to the other three genes suggesting its evolution to be a result of an early divergence. Challenging nested PCRs, extra-long PCRs and cloning confirmed the existence and the order of all four isoforms within the worm genome. The sequencing of the cloned genomic region confirmed the sequence information provided by the WormBase database regarding the *W08E(12.2-12.5)* family.

In order to study global gene expression *in vivo*, *C. elegans* is a unique multicellular model. Precise, cell by cell analysis of the activity of the promoters throughout the development of the organism is possible via transgenic reporter lines (e.g. promoter::GFP) (McKay et al., 2003) in the nematode due to its transparent body and invariant cell-lineage (Sulston et al., 1983, Sulston and Horvitz, 1977). As the whole

genome is sequenced, a genome wide analysis of the promoters *in vivo* is possible (Dupuy et al., 2007). *C. elegans* has proved itself to be a potent model to study toxicity and toxicological mechanisms of various heavy metals such as As, Cd, Zn, Cu, Pb by either facilitating toxicity end point studies or via tracking reporter transgenic expression within the worm (Leung et al., 2008). As *W08E12.3* and *W08E12.4* share an identical promoter, the expression of *W08E12.3* and *W08E12.4* gene within the worm was tracked by creating a transgenic promoter reporter strain cop-136 (P*W08E12.3/4*::GFP) which is a single copy, genome integrated transgene, by Knudra Transgenics and following the reporter (GFP) gene expression. The transgene expression studies revealed that within the worm, *W08E12.3/4* expression is confined to the pharyngeal region, particularly defined in the posterior terminal bulb area, especially in the dorsal epithelial pm8 cells present above the pharyngeal-intestinal valve. The transgenic expression was also observed in the eggs indicating embryonic expression of *W08E12.3/4* within the nematode similar to that of the *mtl-2* isoform in the final stages of embryogenesis. Overall, the location of expression of *W08E12.3/4* within the nematode supports a possible linkage between metallothionein isoforms and *W08E12.3/4* expression.

Due to the linkage with metallothioneins, metal responsiveness of cop-136 (P*W08E12.3/4*::GFP) strain was tested by exposing the transgenic worms to defined concentrations of Cd, Zn and Cu. Due to fluctuations in the fluorescence signal and the low signal obtained due to single copy integration of transgene in cop-136(P*W08E12.3/4*::GFP), an extrachromosomal array of P*W08E12.3/4*::GFP construct was used to create the transgenic strain zsEx6(P*W08E12.3/4*::GFP). Based on the data obtained from both, the single copy genome integrated transgenic strain and the multiple copy, extrachromosomal array carrying transgenic strain, it can be

confirmed that upon exposure to low doses of Cd, *W08E12.3/4* might be upregulated but a dose higher than 10 µM Cd causes toxic effects to the nematode and hence, reduces the fluorescence intensity. Zn exposures enhances the expression of *W08E12.3/4* at doses up to 100 µM but again higher doses confer toxicity to the worms. Whereas in the presence of Cu (400 µM- 600 µM) *W08E12.3/4* was downregulated. It was observed that exposure to high concentrations of metal ions displayed drastic reduction in growth and led to starvation of the worms which was reflected in a reduced GFP signal. Overall it can be conceived that at low metal concentrations (Cu, Zn, Cd) the toxic effects are combated by the existing stress mechanisms within the worm, whereas beyond a threshold heavy metal toxicity is evident, signifying a direct relationship between metal concentration and toxicity. The pattern of expression of *W08E12.3/4* to heavy metals (Cd, Zn, Cu) was similar in the presence or absence of metallothionein although, the expression fold of *W08E12.3/4* was significantly higher in the metallothionein knockout in the wild type background. The toxicity observed in wild type transgenic worms and the metallothionein knock down worms was similar suggesting that the *W08E(12.2-12.5)* series might compensate for the absence of metallothioneins by playing a role in metal handling but, cannot protect against metal toxicity over a prolonged exposure period.

To quantify the level of expression of the entire *W08E(12.2-12.5)* family in response to heavy metals, qPCR assays were performed. As the *W08E12.3/12.4* displayed upregulation to Cd exposure in an metallothionein null background within the transgenic worm, the expression fold change of the *W08E(12.2-12.5)* family was tested in the presence of Cd. In order to characterise each isoform of the *W08E(12.2-12.5)* family, primers and probes for each isoform were initially designed. But, as the isoforms of *W08E(12.2-12.5)* family are highly similar to each other and reside next

to each other on the genome, it was challenging to design gene specific primer-probe combinations even with the use of Universal Probe Library assay design centre. The suggested probe-primer combinations were checked for gene specificity, however a certain degree of cross reactivity between the isoforms was observed. Hence, new primer-probe combinations were designed to quantify each of the isoform, namely *W08E12.2, W08E12.3* and *W08E12.4/12.5* individually. The new primer-probe combinations revealed a significant upregulation of all isoforms of the *W08E(12.2-12.5)* family in response to 10 µM Cd. A pattern of expression could be deduced for *W08E12.2* and *W08E12.3* i.e. a gradual increase in response to Cd (0-10 µM) was observed, although results were inconclusive for *W08E(12.4/12.5)*. The response of the *W08E(12.2-12.5)* isoforms to Cd in absence of metallothioneins was similar to the wild type, although the overall level of expression was amplified. A higher dose of Cd (30 µM) was not included in the tests as the worms suffered toxic effects leading to underdevelopment and larval arrest.

As *W08E(12.3-12.5)* depicted a possible functional relationship with metallothioneins, the phenotypic role of *W08E(12.2-12.5)* isoforms upon exposure to heavy metals (Cd, Zn) was tested by exploiting the most valuable reverse genetic tool applicable to *C. elegans* i.e. RNAi. The *W08E(12.2-12.5)* isoforms were deemed to be knocked down using RNAi by feeding, involving selection of the RNAi clones from the commercially sourced RNAi library which covers 19,762 protein coding genes. When the library was screened for the *W08E(12.2-12.5)* isoforms, only the clone for *W08E12.4* was available. *W08E12.4* RNAi clone would also enable the knock down of *W08E12.5* due to their 100 % sequence identity, therefore it was important for the *W08E12.2* and *W08E12.3* RNAi clones to be created in order to track the phenotypic importance of this gene family within the worm. A bespoke clone for the *W08E12.3* isoform was

created by amplifying the genomic region of the *W08E12.3* isoform and cloning it into the RNAi vector *pPD129.36.* The cloned vector was transformed into competent HT115 bacteria and fed to the worms to facilitate the knock down within the nematode.

Optimising a protocol for the RNAi by feeding to be effective for a double/triple gene knockdown posed a major challenge. After numerous trials, a protocol yielding optimal knock down of *W08E(12.3-12.5)* isoforms simultaneously, was developed. After the confirmation of the *W08E(12.3-12.5)* knockdown in the nematode, the strain was exposed to heavy metals (Zn and Cd) from L1 stage and a phenotypic growth and brood size assay were conducted.

The overall growth of the *W08E(12.3-12.5)* knockdown worms was significantly restricted in comparison to the wild type (N2) worms, especially in response to Cd, indicating a role of *W08E(12.3-12.5)* isoforms in the development and stress response pathway of the nematode. Whereas, the effect of knockdown of the *W08E(12.3-12.5)* isoforms on the brood size was negligible.

Changes in metal distribution within the worm in the absence of *W08E(12.3-12.5)* was tested with the aid of X-ray laser imaging of the *W08E(12.3-12.5)* knocked down strain after exposure to a mix of heavy metals (150 µM Zn, 30 µM Cd). No difference in Cd distribution or accumulation was observed between wild type and the *W08E(12.3-12.5)* knockdown worms. In contrast, significant amounts of Zn accumulated in the pharynx region of the *W08E(12.3-12.5)* knockdown worms, this was not observed in the wild type worms.

The *C. elegans* neuronal network is divided into the pharyngeal nervous system and the somatic nervous system. The pharyngeal neurons are connected to the extra pharyngeal nervous system via the gap junctions between I1 pharyngeal neurons and

the RIP extra pharyngeal neurons (Dent et al., 2000). In addition to the I1 neurons in the pharynx, there are various other neuronal networks involved in chemotaxis of the nematode. Of the numerous neurons in the pharyngeal nervous system ADL, ASE and ASH neuronal ablation leads to heavy metal ion avoidance in *C. elegans* (Sambongi et al., 1999). As *W08E(12.3-12.4)* is strongly expressed in the pharyngeal region of the nematode, the accumulation of Zn in absence of the *W08E(12.3-12.5)* isoforms is a strong indicator of disruption of a major Zn sensing neuronal or Zn homeostasis pathway.

In order to investigate if the *W08E(12.2-12.5)* family played any role in metal binding, characterization at protein level was conducted. *W08E12.3* was selected for the protein study as it was the most well characterised isoform at the gene level.

The first challenge to protein study was to clone the cDNA fragment of *W08E12.3* into a cloning vector. Due to the high similarity between the isoforms of *W08E(12.2-12.5)* family, it was difficult to design primers that amplify the isoform of interest from the genome. After several attempts of designing primers, amplifying and cloning the *W08E12.3* isoform, it was decided that only the genomic fragment (i.e. with the intron) could be amplified from the worm genome, therefore an artificial construct of *W08E12.3* was synthesised and sub-cloned into the expression vector of choice.

It was originally speculated that W08E12.3 could be expressed as a fusion protein with an affinity tag as it does not require steps involving nucleic acid removal or other cellular material, benefits from the mild elution conditions and moreover the possibility of purifying diversified proteins using generalised and simplified purification protocols (Lichty et al., 2005). Numerous affinity tag systems were deemed to be inappropriate as they require metal based columns for the purification

and, based on the cysteine content and metal responsiveness, W08E12.3 protein would likely bind to the metals embedded in the column and affect the purification process.

The IMPACT system was implemented to purify *W08E12.3* which was sub-cloned into the pTXB vector enabling the *C*-terminal fusion of the protein to an intein tag. The successful cloning was followed by expression profiling of the W08E12.3 protein. It was observed via Western blotting using intein specific antibodies that the tag cleaved off from the protein during induction with IPTG within the *E. coli* cells, likely due to the low intracellular pH. Therefore, an alternative affinity tag based system, well established for stable protein expression and purification was implemented, the S-tag system.

The synthesised *W08E12.3* was sub-cloned into the pET29a (+) vector both with and without the fusion S-tag. The sequencing of the successful clones revealed an addition of two extra amino acids (Methionine and Glutamate) in both the tagged and untagged constructs changing the overall mass of the S-tagged protein from 14.46 kDa to 17.28 kDa and the untagged from 14.20 kDa to 17.02 kDa. The W08E12.3 native protein without any tag was attempted to be purified but due to the inability to identify the protein over the course of purification and the inability to identify the untagged protein on ESI-MS on the basis of its mass, the S-tagged version of the protein was selected for further studies. The expression profile of the S-tagged W08E12.3 protein confirmed the stable induction of the protein within *E. coli* (DE3 pLysS) cells. The first attempt at purifying the S-tagged W08E12.3 protein was based on the assumption that the protein was in the soluble fraction of the cell lysate, which eventually was proven incorrect and it was identified that most of the expressed protein accumulated in the inclusion bodies of the *E. coli* (DE3 pLysS) cells. Attempts were made to get the protein in the soluble fraction by varying the *E. coli* strain, changing induction

conditions, adding metals to serve as stabilising factors, however due to the time constraints it was decided to isolate the S-tagged W08E12.3 protein from the inclusion bodies.

Inclusion bodies form upon overexpression or improper folding of recombinant protein within the cell cytoplasm, this is highly possible in *E. coli* cells as they lack the capability of protein folding and quality control mechanisms found in eukaryotes. Inclusion bodies are considered to be unimportant, dead end products of protein expression, although their formation holds a major advantage. The protein forming inclusion bodies are free from cell homogenates and are more pure (Rogl et al., 1998). As inclusion bodies are biologically inactive, it is essential to solubilise them and refold the protein (Hlodan et al., 1991). After several attempts involving modifications of established methods, the protocol designed to isolate and refold human matrix metalloproteinase from inclusion bodies formed in *E. coli* was implemented (Oneda and Inouye, 1999) to purify the S-tagged W08E12.3 protein.

Specifically, the protein was isolated from the inclusion bodies by solubilising them in guanidine hydrochloride (6 M) to remove nucleic acids, other cellular proteins and phospholipids. The isolated S-tagged W08E12.3 protein was then refolded in the presence of L-arginine (0.25 M) and metal cofactors (Zn or Cd). Cysteine molecules are reactive residues and their oxidation can lead to the formation of irreversible disulphide bonds (Lee and Helmann, 2006), these can alter the structure and function of the protein (Kiley and Storz, 2004) therefore (post isolation of the inclusion bodies) all the steps were carried out in the presence of reducing agents (DTT) and metal cofactors (Zn or Cd) in order to prevent the oxidation of the cysteine molecules present in the W08E12.3 protein.

The refolded S-tagged W08E12.3 protein was purified using classical chromatographic techniques involving gel filtration and anion exchange. In the first trial, the refolded protein was concentrated using several steps of centrifugation in order to accommodate the refolded protein in the gel filtration column, however during the subsequent attempt to generate additional protein, anion exchange chromatography was selected in order to avoid the multiple centrifugation steps and load a larger volume of refolded protein. The peaks present in the chromatogram of both gel filtration and anion exchange chromatography were examined by silver staining, ICP-OES, ESI-MS and Western blotting to identify the fractions containing the full length S-tagged W08E12.3 protein. Both the chromatographic techniques led to the isolation of pure S-tagged W08E12.3 protein, although the yield was much higher when anion exchange chromatography was used, probably due to prevention of loss of protein by the avoidance of multiple centrifugation steps prior to column purification.

The S-tagged W08E12.3 was subjected to ESI-MS and the deconvoluted spectra revealed a mass of 17,279.18 Da, which matched the theoretical apo mass of the protein. The metallated form of S-tagged W08E12.3 was also subjected to ESI-MS but did not yield the expected metal bound spectra, this could be due to a weak interaction of the protein with the metals that dissociate during the ESI process (Banci et al., 2006) or due to the heavy mass or folding of the metallated form which does not allow proper ionisation, therefore interfering with the ESI-MS.

The major challenge in purifying the S-tagged W08E12.3 protein was the accuracy of the refolding process. Maintaining the protein in the reduced state until refolding of the tagged W08E12.3 protein in the presence of metal cofactors (Zn or Cd) was critical to the amount of metals that bound to the protein. By ICP-OES, the amount of metal that bound to the protein was evaluated. The S-tagged W08E12.3 protein bound to 6.5

Zn(II) when purified using gel filtration column while 4.5 Zn(II) was revealed upon anion exchange chromatography. The inclusion body fraction before isolating the protein was also examined by ICP-OES. It was observed that the tagged W08E12.3 protein within the inclusion body binds to 5 Zn(II). The entire ICP-OES data signifies that the amount of metal bound to the protein is dictated by the refolding conditions. Upon proper refolding the tagged protein is capable of binding 6.5 Zn(II) while under partially folded condition it can bind to 4.5 Zn(II) and 3 Cd(II). Moreover, the presence of a Zn bound form of the S-tagged W08E12.3 protein within the inclusion bodies suggests that the protein is not completely unfolded inside the *E. coli* cells and is capable of sequestering Zn within the bacteria.

Protein-metal ion interactions play an essential role in many biological processes, metals act as cofactors for one-third of enzymes resulting in optimal catalytic activity (Waldron and Robinson, 2009), conformational stability (Berg and Shi, 1996) while conformational changes are introduced in others. Transport of metal ions within the cells and across membranes is based on protein-metal interactions (Anderson et al., 1987). There are two modes of action by which proteins are affected by metal ions: the free thiols of the protein and the other functional groups of native proteins bind metal ions or the metal ions displace other essential metals bound to metal-dependent proteins (Gurd and Wilcox, 1956, Vallee and Ulmer, 1972). The affinity with which a metal ion binds to a protein ligand is a key property of a metallo-protein and is defined by the dissociation constant (K_D). The determination of K_D is essential for the understanding of metal selection and speciation of the protein (Sharma et al., 2008).

Therefore in order to define the mechanisms by which metalloproteins conduct their functions, insights into the kinetic and thermodynamic parameters of the protein-metal interactions are essential (Deng et al., 2010). The metal-protein affinities are

determined *in vitro* and it is assumed that within biological cells under very different conditions, their determined relative values remain relevant (Spitzer and Poolman, 2005).

As protein ligands are generally Lewis bases and their pKa is altered upon the binding of metal ions at the binding site allows the metal-binding affinity of the protein ligand to be determined by potentiometric (acid-base) titration i.e. by competition with protons (Albert, 2012). As the pH electrode can accurately measure the proton concentration, it is a reliable method to measure the metal-binding constants (Xiao and Wedd, 2010). In order to investigate the strength of binding between the S-tagged W08E12.3 protein and the Zn(II), metal loss was attempted under increasing acidic pH. Unfortunately, it was not possible to derive the pH of half metal displacement or the K_D value for the S-tagged W08E12.3 protein by proton competition as it aggregated between pH 6.4 and pH 4, which coincides with its theoretical isoelectric point (pI = 5.55). It is usual for proteins to aggregate around their pI as the electrostatic interactions are dictated by the pH of the solution buffer and at pI the net charge on the protein becomes 0 which leads to precipitation (Krebs et al., 2009). However, the W08E12.3 protein re-solubilized and displayed reduced absorption at 230 nm in comparison to pH 7.8, which confirms that the Zn loss was pH dependent. It was also observed that the interaction between the Zn and the W08E12.3 protein was moderately weak as metal loss was observed with a minute increase in acidity of the buffer.

Another equilibrium competition method was employed to determine the metal-protein stoichiometry of the S-tagged W08E12.3 protein. The 4.5 Zn(II) loaded protein was titrated with increasing Cd(II) concentration and it was observed that the addition of a 4-5 molar equivalent of Cd led to the displacement of all the Zn bound to the

protein. The binding of Cd to the protein while displacing Zn was monitored by the absorption measurement at 250 nm (indicative of the Cd-cysteine coordination complex), which increased upon the addition of Cd. The Zn displacement by Cd followed a 1:1 ratio, suggesting the higher affinity of Cd towards W08E12.3 S-tagged protein in comparison to Zn (under *in vitro* conditions).

Overall the *in vitro* protein studies confirmed that the S-tagged W08E12.3 protein binds to metals and has a preference to sequester more Zn rather than Cd. The protein is capable of binding 5 Zn(II) even in its partially folded form and is not completely non- functional within the inclusion bodies (i.e. in its partially folded form) as it can bind 5 Zn(II). The S-tagged W08E12.3 protein, when properly folded, in its full potential binds to 6.5 Zn(II) indicating the participation of all cysteines in metal binding. Likewise, when the protein was partially folded it displayed a reduced metal binding capacity i.e. 4.5 Zn(II) and 3 Cd(II). The isoelectric point of the protein matches the theoretical pI i.e. 5.5 as the protein precipitates around pH 6.4 to 4.4. The metal displacement of Zn by Cd in the metallated S-tagged protein elucidated the metal-protein stoichiometry.

Zn-cysteine coordination is redox-active within the protein. When the sulphur ligands in the protein is oxidised, Zn becomes mobilised while reduced sulphur conditions promote Zn binding, indicating redox control over the available Zn ions. Many Zn proteins are considered redox sensors, where Zn release is related to conformational changes monitoring molecular chaperone activity, enzymatic activity and binding interactions, whereas the released Zn has no specific function and might bind to other proteins modulating signal transduction and gene expression. One such example is the Zn protein metallothionein which is a redox transducer, it plays an important role in cell signalling by redistributing cellular Zn, taking charge of the availability of Zn by

interconverting the redox and Zn signals together with its apo protein thionein (Maret, 2006). The W08E12.3 protein has an affinity towards Zn binding which is redox based, therefore there is a high possibility of it acting as a Zn sensor like metallothionein.

Most proteins are sulfhydryl proteins in the intracellular environment, and these proteins are maintained in their reduced state with the aid of glutathione (Barron and Singer, 1943), which gets reduced by NADPH via glutathione reductase (Fahey et al., 1977). In contrast to the intracellular proteins, the extracellular proteins can be classified as disulfide proteins. As these proteins function unprotected from the atmospheric oxygen, the thiols are observed more as cystines than cysteines in the extracellular environment. The occurrence of a protein intracellularly or extracellularly and the abundance of cysteines or cystines is defined by the thiol-disulfide redox state. For example the plasma membrane proteins when in intracellular compartment are exposed to a highly reduced state, while in the extracellular environment are exposed to a highly oxidized environment (Rosenberg and Guidotti, 1968).

Metallothionein, which plays an important role in cellular homeostasis, functions both in the intracellular and extracellular environment as its synthesis is associated with various cellular stressors present in both intra and extracellularly. The existence of extracellular MT is unusual as it lacks any signal peptide sequence or other signals involved in protein trafficking which would transport it to such an environment, suggesting the involvement of non-classical ways of secretion under cellular stress. MT when expressed intracellularly, sequesters the essential heavy metals, acts as an oxygen and nitrogen species scavenger and regulates the activity of transcription

factors whereas, extracellular existence of MT signifies as a danger signal, monitoring the handling of cellular damage acting as an antioxidant (Lynes et al., 2006).

The presence of many cysteine residues in the W08E12.3 protein dictated its role in metal binding, but the role of the cysteine is highly dependent on the cellular location of the protein. The cysteines are involved in the formation of disulphide bonds where pair of cysteines are oxidised to form covalent bonds, stabilising the structure of the protein in an extracellular environment. Whereas, within the intracellular environment, under the reducing conditions it is almost impossible for the cysteines to oxidise and form disulfide bonds (Betts and Russell, 2003). The metal binding properties of the S-tagged W08E12.3 protein is based on the assumption that the protein is maintained in a reduced state, in its native form (i.e. intracellularly expressed), therefore the reliability of the protein to participate in metal binding solely depends on where it is expressed.

In order to predict the location of the W08E12.3 protein, a combined transmembrane topology and signal peptide predictor software Phobius and SignalIP 4.1 were used (**Figure 7-1**).

From the prediction software, it can be assumed that W08E12.3 is a non-cytoplasmic protein which has a prominent signal peptide sequence attached to its *N*- terminus containing a signal peptide cleavage site between position 18 and 19 in the amino acid sequence. All the isoforms of the *W08E(12.2-12.5)* family, based on the Phobius and SignalIP software confirmed the presence of a signal peptide sequence with the same cleavage site between 18 and 19 amino acid (ASA-QC) in their sequence.

In eukaryotes, the insertion of the proteins into the membrane of endoplasmic reticulum for post translational modification is directed by signal sequences which are

16-30 a.a long. The signal peptides are presumably rapidly degraded after that but few have their own function. Therefore, signal peptides contribute to the function of membrane and secretory proteins (Kapp et al., 2000).

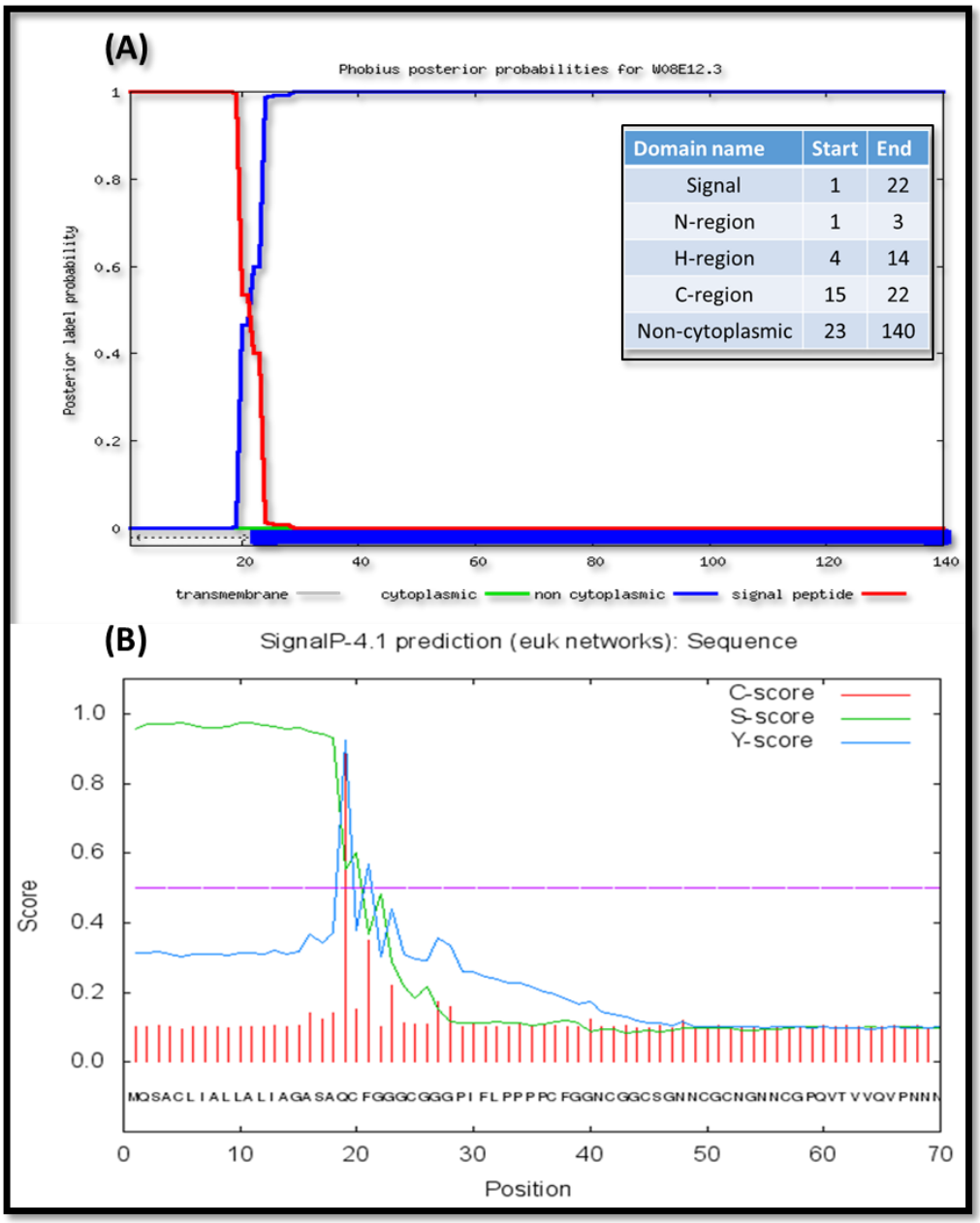


Figure 7-1: (A) Prediction of the probability of W08E12.3 protein being an intracellular protein with detailed annotation of the region spanning signal peptide, *N*, *C*, *H* region and the non-cytoplasmic region (http://phobius.sbc.su.se/cgi-bin/predict). (B) Prediction of the signal peptide region and the signal cleavage site in the W08E12.3 protein using SignalIP 4.0.

It can be hypothesised that the signal peptide directs the W08E12.3 protein to the endoplasmic reticulum and then it is processed into a secretory pathway directing the protein to either an intracellular or extracellular location. Based on the outcome provided by Phobius, W08E12.3 mature protein generated after the cleavage of the signal peptide region is categorized to be non-cytoplasmic. In order to identify the fate of the W08E12.3 protein within the cell, based on the amino acid sequence of the protein, PSORTII database was chosen to assess the eukaryotic protein localisation *in silico*. Based on the Reinhardt's method for cytoplasmic or nuclear discrimination of proteins, it was predicted that W08E12.3 is a nuclear protein.

Results of the k- Nearest Neighbors (*k*-NN) prediction (non-parametric method used for regression and classification) using PSORTII (Horton and Nakai, 1997):

Table 7-1: Probability (%) of the expression of W08E12.3, MTL-1 and MTL-2 being intracellular of extracellular using prediction software PSORTII.

location / Protein	Nuclear	Plasma membrane	Cytoplasmic	Cytoskeletal	Vesicles of secretary system	Golgi	Mitochondrial	Vacuolar
W08E12.3	65%	13%	9%		4%	4%	4%	
MTL-1	56%		9%	4%	4%		28%	4%
MTL-2	48%		30%	9%			13%	

Overall, from the PSORTII prediction (k = 9/23 for all the three genes), W08E12.3 is similar in location to the metallothionein isoforms of *C. elegans* and they are both secreted intracellularly (**Table 7-1)**. On the basis that a signal sequence is present on the *N*-terminus of the W08E12.3 protein suggests that, like metallothioneins, W08E12.3 protein may also have dual roles i.e. intracellular metal handling and extracellular anti-oxidant to handle cellular stress. This again reiterates that the

W08E(12.2-12.5) and the metallothionein isoforms might have functional similarity and might work in concert in metal homeostasis and/or metal detoxification.

7.2. Future work

The *W08E(12.3-12.5)* knock down worms can be exploited to investigate other phenotypic assays like life span, pharyngeal pumping etc. and their response to stress conditions including temperature or hypertonic stress. This may reveal further information regarding their biological role within the worm. Quantification of the isoforms of *W08E(12.2-12.5)* in response to heavy metals in the absence of metallothioneins and phytochelatin will provide insight into the relationship between these metalloprotein families. The W08E12.3 protein can be further studied to gain information about the structure and folding of the protein by NMR, this will unveil more about the protein folding, the domains and the metal binding ability of the protein.

Cleaving off the region coding for the highly hydrophobic signal peptide region of the W08E12.3 protein and then attempting to purify might enable overcoming the issues of heterologous expression and achieving a better characterization of the metal-protein complex of W08E12.3.

In order to stabilise the transgenic expression of P*W08E12.3/4*::GFP, a control metal responsive gene tagged with mCherry could be introduced. Also, another transgenic line with the fusion of promoters of all the three genes P*WO8E12.2,12.3/4,12.5*::GFP in the worm genome, will enable overall transcriptional profiling of all four genes.

In order to trace the location of the protein expression within the nematode, a single copy, genome integrated transgenic strain carrying the coding region of the *W08E12.3*

(*W08E12.3*:GFP) can be created. Also, further study suggested the protein to be secretory while PSORTII predicted it to be a nuclear protein, therefore to trace the exact cellular location of the protein within the nematode a sub cellular characterisation by immuno-detection could be carried out.

The homology of the *W08E(12.2-12.5)* gene family with keratin- associated proteins can be further explored to identify any correlation with As binding and detoxification. The homology of W08E12.3 protein with other characterised proteins especially membrane-bound proteins can be explored to estimate their other possible functions.

7.3. Conclusion

The data obtained by investigating the bioinformatics based genome screening, *in vivo* gene location and expression analysis, quantification and deciphering expression profile of the isoforms and protein level studies has led to the successful characterization of the *W08E(12.2-12.5)* family. *W08E(12.2-12.5)* are metal responsive and metal binding in nature, especially towards Zn. The *W08E(12.2-12.5)* family are functionally similar to metallothionein isoforms indicating their potential role in metal sensing, metal homeostasis and/or metal detoxification. In conclusion, the toxicogenomic studies have uncovered a novel metal responsive and metal binding cysteine rich family which may have the potential to act as a Zn sensor or a biomarker of metal toxicity.

Chapter 8: References

ALBERT, A. 2012. *The determination of ionization constants: a laboratory manual*, Springer Science & Business Media.

ALLEY, R., AOYAGI, H., CLENDENIN, J., FRISCH, J., GARDEN, C., HOYT, E., KIRBY, R., KLAISNER, L., KULIKOV, A., MILLER, R., MULHOLLAN, G., PRESCOTT, C., SÁEZ, P., SCHULTZ, D., TANG, H., TURNER, J., WITTE, K., WOODS, M., YEREMIAN, A. D. & ZOLOTOREV, M. 1995. The Stanford linear accelerator polarized electron source. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment,* 365**,** 1-27.

ALONSO, J. M. & ECKER, J. R. 2006. Moving forward in reverse: genetic technologies to enable genome-wide phenomic screens in Arabidopsis. *Nat Rev Genet,* 7**,** 524-536.

ALTUN, Z. F. A. H., D.H. 2015. *Handbook of C. elegans Anatomy.* .

ALUIGI, A., VINEIS, C., VARESANO, A., MAZZUCHETTI, G., FERRERO, F. & TONIN, C. 2008. Structure and properties of keratin/PEO blend nanofibres. *European Polymer Journal,* 44**,** 2465-2475.

ANDERSON, B. F., BAKER, H. M., DODSON, E. J., NORRIS, G. E., RUMBALL, S. V., WATERS, J. M. & BAKER, E. N. 1987. Structure of human lactoferrin at 3.2-A resolution. *Proceedings of the National Academy of Sciences,* 84**,** 1769-1773.

APOSTOLI, P., TELIŠMAN, S. & SAGER, P. R. 2007. Chapter12 - Reproductive and Developmental Toxicity of Metals. *In:* NORDBERG, G. F., FOWLER, B. A., NORDBERG, M. & FRIBERG, L. T. (eds.) *Handbook on the Toxicology of Metals (Third Edition).* Burlington: Academic Press.

BANCI, L., BERTINI, I., CALDERONE, V., CIOFI-BAFFONI, S., MANGANI, S., MARTINELLI, M., PALUMAA, P. & WANG, S. 2006. A hint for the function of human Sco1 from different structures. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 8595-8600.

BANERJEE, S. & MAZUMDAR, S. 2012. Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte. *International journal of analytical chemistry,* 2012.

BANEYX, F. 1999. Recombinant protein expression in Escherichia coli. *Current Opinion in Biotechnology,* 10**,** 411-421.

BARGMANN, C. I. & HORVITZ, H. R. 1991. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in C. elegans. *Neuron,* 7**,** 729-742.

BARKER, K., HUSSEY, R., KRUSBERG, L., BIRD, G., DUNN, R., FERRIS, H., FERRIS, V., FRECKMAN, D., GABRIEL, C. & GREWAL, P. 1994. Plant and soil nematodes: societal impact and focus for the future. *Journal of Nematology,* 26**,** 127.

BARNES, T., KOHARA, Y., COULSON, A. & HEKIMI, S. 1995. Meiotic recombination, noncoding DNA and genomic organization in Caenorhabditis elegans. *Genetics,* 141**,** 159-179.

BARRETT, P. L., FLEMING, J. T. & GÖBEL, V. 2004. Targeted gene alteration in Caenorhabditis elegans by gene conversion. *Nature genetics,* 36**,** 1231-1237.

BARRON, E. G. & SINGER, T. 1943. Enzyme systems containing active sulfhydryl groups. The role of glutathione. *Science,* 97**,** 356-358.

BEATTIE, J. H., OWEN, H. L., WALLACE, S. M., ARTHUR, J. R., KWUN, I. S., HAWKSWORTH, G. M. & WALLACE, H. M. 2005. Metallothionein overexpression and resistance to toxic stress. *Toxicol Lett,* 157**,** 69-78.

BEIJER, K. & JERNELOV, A. 1986. Sources, transport and transformation of metals in the environment. *Handbook on the Toxicology of Metals,* 1**,** 68-74.

BEREZIKOV, E., BARGMANN, C. I. & PLASTERK, R. H. 2004. Homologous gene targeting in Caenorhabditis elegans by biolistic transformation. *Nucleic acids research,* 32**,** e40-e40.

BERG, J. M. & SHI, Y. 1996. The galvanization of biology: a growing appreciation for the roles of zinc. *Science,* 271**,** 1081.

BERKOWITZ, L. A., KNIGHT, A. L., CALDWELL, G. A. & CALDWELL, K. A. 2008. Generation of stable transgenic C. elegans using microinjection. *Journal of visualized experiments: JoVE*.

BESSEREAU, J.-L., WRIGHT, A., WILLIAMS, D. C., SCHUSKE, K., DAVIS, M. W. & JORGENSEN, E. M. 2001. Mobilization of a Drosophila transposon in the Caenorhabditis elegans germ line. *Nature,* 413**,** 70-74.

BETTS, M. J. & RUSSELL, R. B. 2003. Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists,* 317**,** 289.

BEYERSMANN, D. & HARTWIG, A. 2008. Carcinogenic metal compounds: recent insight into molecular and cellular mechanisms. *Archives of toxicology,* 82**,** 493-512.

BLINDAUER, C. A. & LESZCZYSZYN, O. I. 2010. Metallothioneins: unparalleled diversity in structures and functions for metal ion homeostasis and more. *Natural product reports,* 27**,** 720-741.

BOFILL, R., ORIHUELA, R., ROMAGOSA, M., DOMENECH, J., ATRIAN, S. & CAPDEVILA, M. 2009. Caenorhabditis elegans metallothionein isoform specificity–metal binding abilities and the role of histidine in CeMT1 and CeMT2. *FEBS journal,* 276**,** 7040-7056.

BONGERS, T. 1990. The maturity index: an ecological measure of environmental disturbance based on nematode species composition. *Oecologia,* 83**,** 14-19.

BOYD, W. A., SMITH, M. V., KISSLING, G. E. & FREEDMAN, J. H. 2010. Medium-and high-throughput screening of neurotoxicants using C. elegans. *Neurotoxicology and teratology,* 32**,** 68-73.

BOYD, W. A. & WILLIAMS, P. L. 2003a. Availability of metals to the nematode Caenorhabditis elegans: toxicity based on total concentrations in soil and extracted fractions. *Environ Toxicol Chem,* 22**,** 1100-6.

BOYD, W. A. & WILLIAMS, P. L. 2003b. Comparison of the sensitivity of three nematode species to copper and their utility in aquatic and soil toxicity tests. *Environmental toxicology and chemistry,* 22**,** 2768-2774.

BRENNER, S. 1974. THE GENETICS OF CAENORHABDITIS ELEGANS. *Genetics,* 77**,** 71-94.

BROEKS, A., GERRARD, B., ALLIKMETS, R., DEAN, M. & PLASTERK, R. 1996. Homologues of the human multidrug resistance genes MRP and MDR contribute to heavy metal resistance in the soil nematode Caenorhabditis elegans. *The EMBO Journal,* 15**,** 6132.

BURCELIN, R., LI, J. & CHARRON, M. J. 1995. Cloning and sequence analysis of the murine glucagon receptor-encoding gene. *Gene,* 164**,** 305-310.

BUSHMARINA, N. A., BLANCHET, C. E., VERNIER, G. & FORGE, V. 2006. Cofactor effects on the protein folding reaction: Acceleration of α-lactalbumin refolding by metal ions. *Protein Science : A Publication of the Protein Society,* 15**,** 659-671.

BUSTIN, S. A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of molecular endocrinology,* 25**,** 169-193.

CASTRO-GONZÁLEZ, M. I. & MÉNDEZ-ARMENTA, M. 2008. Heavy metals: Implications associated to fish consumption. *Environmental Toxicology and Pharmacology,* 26**,** 263-271.

CHAN, J., HUANG, Z., MERRIFIELD, M. E., SALGADO, M. T. & STILLMAN, M. J. 2002. Studies of metal binding reactions in metallothioneins by spectroscopic, molecular biology, and

molecular modeling techniques. *Coordination Chemistry Reviews,* 233–234**,** 319-339.

CHANG, C. C., LIAO, W. F. & HUANG, P. C. 1998. Cysteine contributions to metal binding preference for Zn/Cd in the beta-domain of metallothionein. *Protein Engineering,* 11**,** 41-46.

CHEN, N., PAI, S., ZHAO, Z., MAH, A., NEWBURY, R., JOHNSEN, R. C., ALTUN, Z., MOERMAN, D. G., BAILLIE, D. L. & STEIN, L. D. 2005. Identification of a nematode chemosensory gene family. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 146-151.

CIOCI, L., QIU, L. & FREEDMAN, J. H. 2000. Transgenic strains of the nematode Caenorhabditis elegans as biomonitors of metal contamination. *Environmental toxicology and chemistry,* 19**,** 2122-2129.

COBBETT, C. & GOLDSBROUGH, P. 2002a. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annual review of plant biology,* 53**,** 159-182.

COBBETT, C. & GOLDSBROUGH, P. 2002b. Phytochelatins and metallothioneins: roles in heavy metal detoxification and homeostasis. *Annu Rev Plant Biol,* 53**,** 159-82.

COLANGELO, D., MAHBOOBI, H., VIARENGO, A. & OSELLA, D. 2004. Protective effect of metallothioneins against oxidative stress evaluated on wild type and MT-null cell lines by means of flow cytometry. *Biometals,* 17**,** 365-70.

CONSORTIUM, T. C. E. S. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science,* 282**,** 2012-8.

COYLE, P., PHILCOX, J., CAREY, L. & ROFE, A. 2002. Metallothionein: the multipurpose protein. *Cellular and Molecular Life Sciences CMLS,* 59**,** 627-647.

CUI, C., ZHAO, W., CHEN, J., WANG, J. & LI, Q. 2006. Elimination of in vivo cleavage between target protein and intein in the intein-mediated protein purification systems. *Protein expression and purification,* 50**,** 74-81.

CULOTTI, J. G. & RUSSELL, R. L. 1978. Osmotic avoidance defective mutants of the nematode Caenorhabditis elegans. *Genetics,* 90**,** 243-256.

DALLINGER, R. 1996. Metallothionein research in terrestrial invertebrates: Synopsis and perspectives. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology,* 113**,** 125-133.

DAULBY, J., ISAAC, R. E. & COATES, D. Glutathione-s-transferases of C.elegans. 1996.

DE CASTILHOS GHISI, N. 2012. *Relationship between biomarkers and pesticide exposure in fishes: A review*, INTECH Open Access Publisher.

DENG, L., SUN, N., KITOVA, E. N. & KLASSEN, J. S. 2010. Direct Quantification of Protein–Metal Ion Affinities by Electrospray Ionization Mass Spectrometry. *Analytical chemistry,* 82**,** 2170-2174.

DENT, J. A., SMITH, M. M., VASSILATIS, D. K. & AVERY, L. 2000. The genetics of ivermectin resistance in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences,* 97**,** 2674-2679.

DEPLEDGE, M. H., AAGAARD, A. & GYÖRKÖS, P. 1995. Assessment of trace metal toxicity using molecular, physiological and behavioural biomarkers. *Marine Pollution Bulletin,* 31**,** 19-27.

DI MARCO, V. B. & BOMBI, G. G. 2006. Electrospray mass spectrometry (ESI-MS) in the study of metal–ligand solution equilibria. *Mass Spectrometry Reviews,* 25**,** 347-379.

DIAMENT, A., PINTER, R. Y. & TULLER, T. 2014. Three-dimensional eukaryotic genomic organization is strongly correlated with codon usage expression and function. *Nature communications,* 5.

DISCHE, Z. & ZIL, H. 1951. Studies on the Oxidation of Cysteine to Cystine in Lens Proteins During Cataract Formation*. *American Journal of Ophthalmology,* 34**,** 104-113.

DIXIT, R., MALAVIYA, D., PANDIYAN, K., SINGH, U. B., SAHU, A., SHUKLA, R., SINGH, B. P., RAI, J. P., SHARMA, P. K. & LADE, H. 2015. Bioremediation of Heavy Metals from Soil and Aquatic Environment: An Overview of Principles and Criteria of Fundamental Processes. *Sustainability,* 7**,** 2189-2212.

DONKIN, S. G. & DUSENBERY, D. B. 1993. A soil toxicity test using the nematode Caenorhabditis elegans and an effective method of recovery. *Archives of Environmental Contamination and Toxicology,* 25**,** 145-151.

DUDEV, T. & LIM, C. 2001. Metal Selectivity in Metalloproteins: Zn2+ vs Mg2+. *The Journal of Physical Chemistry B,* 105**,** 4446-4452.

DUDEV, T. & LIM, C. 2008. Metal Binding Affinity and Selectivity in Metalloproteins: Insights from Computational Studies. *Annual Review of Biophysics,* 37**,** 97-116.

DUPUY, D., BERTIN, N., HIDALGO, C. A., VENKATESAN, K., TU, D., LEE, D., ROSENBERG, J., SVRZIKAPA, N., BLANC, A., CARNEC, A., CARVUNIS, A.-R., PULAK, R., SHINGLES, J., REECE-HOYES, J., HUNT-NEWBURY, R., VIVEIROS, R., MOHLER, W. A., TASAN, M., ROTH, F. P., LE PEUCH, C., HOPE, I. A., JOHNSEN, R., MOERMAN, D. G., BARABASI, A.-L., BAILLIE, D. & VIDAL, M. 2007. Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. *Nat Biotech,* 25**,** 663-668.

DURNAM, D. M. & PALMITER, R. D. 1981. Transcriptional regulation of the mouse metallothionein-I gene by heavy metals. *J Biol Chem,* 256**,** 5712-6.

DUVERGER, Y., BELOUGNE, J., SCAGLIONE, S., BRANDLI, D., BECLIN, C. & EWBANK, J. J. 2007. A semi-automated high-throughput approach to the generation of transposon insertion mutants in the nematode Caenorhabditis elegans. *Nucleic acids research,* 35**,** e11-e11.

EATON, D. L. & BAMMLER, T. K. 1999. Concise review of the glutathione S-transferases and their significance to toxicology. *Toxicological sciences,* 49**,** 156-164.

ELLMAN, G. L. 1959. Tissue sulfhydryl groups. *Archives of Biochemistry and Biophysics,* 82**,** 70-77.

ERCAL, N., GURER-ORHAN, H. & AYKIN-BURNS, N. 2001. Toxic metals and oxidative stress part I: mechanisms involved in metal-induced oxidative damage. *Current topics in medicinal chemistry,* 1**,** 529-539.

EVANS, T. C. 2006. Transformation and microinjection. Worm book.

FAHEY, R. C., HUNT, J. S. & WINDHAM, G. C. 1977. On the cysteine and cystine content of proteins. *Journal of molecular evolution,* 10**,** 155-160.

FÉLIX, M.-A. & BRAENDLE, C. 2010. The natural history of Caenorhabditis elegans. *Current Biology,* 20**,** R965-R969.

FIGUEROA, E. 2008. Are more restrictive food cadmium standards justifiable health safety measures or opportunistic barriers to trade? An answer from economics and public health. *Science of the total environment,* 389**,** 1-9.

FIRE, A., ALBERTSON, D., HARRISON, S. W. & MOERMAN, D. 1991. Production of antisense RNA leads to effective and specific inhibition of gene expression in C. elegans muscle. *Development,* 113**,** 503-514.

FREEDMAN, J., SLICE, L., DIXON, D., FIRE, A. & RUBIN, C. 1993. The novel metallothionein genes of Caenorhabditis elegans. Structural organization and inducible, cell-specific expression. *Journal of Biological Chemistry,* 268**,** 2554-2564.

FREISINGER, E. 2011. Structural features specific to plant metallothioneins. *JBIC Journal of Biological Inorganic Chemistry,* 16**,** 1035-1045.

FRØKJÆR-JENSEN, C., DAVIS, M. W., HOPKINS, C. E., NEWMAN, B. J., THUMMEL, J. M., OLESEN, S.-P., GRUNNET, M. & JORGENSEN, E. M. 2008. Single-copy insertion of transgenes in Caenorhabditis elegans. *Nature genetics,* 40**,** 1375-1383.

FROKJAER-JENSEN, C., WAYNE DAVIS, M., HOPKINS, C. E., NEWMAN, B. J., THUMMEL, J. M., OLESEN, S.-P., GRUNNET, M. & JORGENSEN, E. M. 2008. Single-copy insertion of transgenes in Caenorhabditis elegans. *Nat Genet,* 40**,** 1375-1383.

GIESY, J. P., VERSTEEG, D. J. & GRANEY, R. L. 1988. A review of selected clinical indicators of stress-induced changes in aquatic organisms. *Toxic Contaminants and Ecosystem Health: A Great Lakes Focus, John Wiley & Sons, New York***,** 169-200.

GILES, N. M., GILES, G. I. & JACOB, C. 2003. Multiple roles of cysteine in biocatalysis. *Biochemical and Biophysical Research Communications,* 300**,** 1-4.

GOLDEN, J. W. & RIDDLE, D. L. 1984. The Caenorhabditis elegans dauer larva: developmental effects of pheromone, food, and temperature. *Developmental biology,* 102**,** 368-378.

GOYER, R. A. 1997. Toxic and essential metal interactions. *Annual review of nutrition,* 17**,** 37-50.

GOYER, R. A. & CLARKSON, T. W. 1996. Toxic effects of metals. *Casarett & Doull's Toxicology. The Basic Science of Poisons, Fifth Edition, Klaassen, CD [Ed]. McGraw-Hill Health Professions Division, ISBN,* 71054766.

GREEN, C. J., LICHTLEN, P., HUYNH, N. T., YANOVSKY, M., LADEROUTE, K. R., SCHAFFNER, W. & MURPHY, B. J. 2001. Placenta Growth Factor Gene Expression Is Induced by Hypoxia in Fibroblasts A Central Role for Metal Transcription Factor-1. *Cancer research,* 61**,** 2696-2703.

GUAN, K. & DIXON, J. E. 1991. Eukaryotic proteins expressed in Escherichia coli: an improved thrombin cleavage and purification procedure of fusion proteins with glutathione S-transferase. *Analytical biochemistry,* 192**,** 262-267.

GÜNTHER, V., LINDERT, U. & SCHAFFNER, W. 2012. The taste of heavy metals: Gene regulation by MTF-1. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* 1823**,** 1416-1425.

GUO, S. & KEMPHUES, K. J. 1995. par-1, a gene required for establishing polarity in C. elegans embryos, encodes a putative Ser/Thr kinase that is asymmetrically distributed. *Cell,* 81**,** 611-620.

GUPTA, S. C., SHARMA, A., MISHRA, M., MISHRA, R. K. & CHOWDHURI, D. K. 2010. Heat shock proteins in toxicology: How close and how far? *Life Sciences,* 86**,** 377-384.

GURD, F. & WILCOX, P. E. 1956. Complex formation between metallic cations and proteins, peptides and amino acids. *Advances in protein chemistry,* 11**,** 311.

GUSTAFSSON, C., GOVINDARAJAN, S. & MINSHULL, J. 2004. Codon bias and heterologous protein expression. *Trends in biotechnology,* 22**,** 346-353.

HA, S.-B., SMITH, A. P., HOWDEN, R., DIETRICH, W. M., BUGG, S., O'CONNELL, M. J., GOLDSBROUGH, P. B. & COBBETT, C. S. 1999. Phytochelatin synthase genes from Arabidopsis and the yeast Schizosaccharomyces pombe. *The Plant Cell,* 11**,** 1153-1163.

HAMMARLUND, M., PALFREYMAN, M. T., WATANABE, S., OLSEN, S. & JORGENSEN, E. M. 2007. Open syntaxin docks synaptic vesicles. *PLoS Biol,* 5**,** e198.

HANAA, M., EWEIDA, A. & FARAG, A. Heavy metals in drinking water and their environmental impact on human health. International conference on environmental hazards mitigation, Cairo University, Egypt, 2000. 542-556.

HE, Z. L., YANG, X. E. & STOFFELLA, P. J. 2005. Trace elements in agroecosystems and impacts on the environment. *Journal of Trace Elements in Medicine and Biology,* 19**,** 125-140.

HLODAN, R., CRAIG, S. & PAIN, R. H. 1991. 2 Protein Folding and its Implications for the Production of Recombinant Proteins. *Biotechnology and Genetic Engineering Reviews,* 9**,** 47-88.

HO, C. S., LAM, C. W. K., CHAN, M. H. M., CHEUNG, R. C. K., LAW, L. K., LIT, L. C. W., NG, K. F., SUEN, M. W. M. & TAI, H. L. 2003. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *The Clinical Biochemist Reviews,* 24**,** 3-12.

HOBERT, O. 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques,* 32**,** 728-30.

HOCKNER, M., DALLINGER, R. & STURZENBAUM, S. R. 2011. Nematode and snail metallothioneins. *J Biol Inorg Chem,* 16**,** 1057-65.

HÖCKNER, M., STEFANON, K., DE VAUFLEURY, A., MONTEIRO, F., PÉREZ-RAFAEL, S., PALACIOS, Ò., CAPDEVILA, M., ATRIAN, S. & DALLINGER, R. 2011. Physiological relevance and contribution to metal balance of specific and non-specific Metallothionein isoforms in the garden snail, Cantareus aspersus. *Biometals,* 24**,** 1079-1092.

HORTON, P. & NAKAI, K. Better Prediction of Protein Cellular Localization Sites with the it k Nearest Neighbors Classifier. Ismb, 1997. 147-152.

HU, H. 2002. Human health and heavy metals. *Life support: the environment and human health***,** 65.

HUANG, P. C., SMITH, B., BOHDAN, P. & CORRIGAN, A. 1980. Effect of zinc on cadmium influx and toxicity in cultured CHO cells. *Biological Trace Element Research,* 2**,** 211-220.

HYNEK, D., KREJCOVA, L., SOCHOR, J., CERNEI, N., KYNICKY, J., ADAM, V., TRNKOVA, L., HUBALEK, J., VRBA, R. & KIZEK, R. 2012. Study of interactions between cysteine and cadmium (II) ions using automatic pipetting system off-line coupled with electrochemical analyser. *Int J Electrochem Sci,* 7**,** 1802-1819.

IMAGAWA, M., ONOZAWA, T., OKUMURA, K., OSADA, S., NISHIHARA, T. & KONDO, M. 1990. Characterization of metallothionein cDNAs induced by cadmium in the nematode Caenorhabditis elegans. *Biochem J,* 268**,** 237-40.

ISANI, G. & CARPENE, E. 2014. Metallothioneins, unconventional proteins from unconventional animals: a long journey from nematodes to mammals. *Biomolecules,* 4**,** 435-57.

JAGER, T., ÁLVAREZ, O. A., KAMMENGA, J. E. & KOOIJMAN, S. A. L. M. 2005. Modelling nematode life cycles using dynamic energy budgets. *Functional Ecology,* 19**,** 136-144.

JAKUBOWSKI, M. & PALCZYNSKI, C. 2007. Chapter 21 - Beryllium. *In:* NORDBERG, G. F., FOWLER, B. A., NORDBERG, M. & FRIBERG, L. T. (eds.) *Handbook on the Toxicology of Metals (Third Edition).* Burlington: Academic Press.

JIANG, G. C., HUGHES, S., STÜRZENBAUM, S. R., EVJE, L., SYVERSEN, T. & ASCHNER, M. 2009. Caenorhabditis elegans metallothioneins protect against toxicity induced by depleted uranium. *Toxicological Sciences***,** kfp161.

KAEGI, J. H. R. & SCHAEFFER, A. 1988. Biochemistry of metallothionein. *Biochemistry,* 27**,** 8509-8515.

KALETTA, T. & HENGARTNER, M. O. 2006. Finding function in novel targets: C. elegans as a model organism. *Nat Rev Drug Discov,* 5**,** 387-399.

KAMATH, R. S. & AHRINGER, J. 2003. Genome-wide RNAi screening in Caenorhabditis elegans. *Methods,* 30**,** 313-321.

KAPP, K., SCHREMPF, S., LEMBERG, M. K. & DOBBERSTEIN, B. 2000. Post-targeting functions of signal peptides.

KAR, P. & MISRA, M. 2004. Use of keratin fiber for separation of heavy metals from water. *Journal of Chemical Technology & Biotechnology,* 79**,** 1313-1319.

KARIN, M. 1985. Metallothioneins: proteins in search of function. *Cell,* 41**,** 9-10.

KILEY, P. J. & STORZ, G. 2004. Exploiting thiol modifications.

KIPREOS, E. T. & PAGANO, M. 2000. The F-box protein family. *Genome Biol,* 1**,** R3002.

KLAASSEN, C. D., LIU, J. & CHOUDHURI, S. 1999. Metallothionein: an intracellular protein to protect against cadmium toxicity. *Annu Rev Pharmacol Toxicol,* 39**,** 267-94.

KLAASSEN, C. D., LIU, J. & DIWAN, B. A. 2009. Metallothionein Protection of Cadmium Toxicity. *Toxicology and applied pharmacology,* 238**,** 215-220.

KLASS, M. R. 1977. Aging in the nematode Caenorhabditis elegans: major biological and environmental factors influencing life span. *Mechanisms of ageing and development,* 6**,** 413-429.

KONERMANN, L. & DOUGLAS, D. 1997. Acid-induced unfolding of cytochrome c at different methanol concentrations: electrospray ionization mass spectrometry specifically monitors changes in the tertiary structure. *Biochemistry,* 36**,** 12296-12302.

KOSAK, S. T. & GROUDINE, M. 2004. Gene order and dynamic domains. *Science,* 306**,** 644-647.

KREBS, M. H., DOMIKE, K. & DONALD, A. 2009. Protein aggregation: more than just fibrils. *Biochemical Society Transactions,* 37**,** 682.

KULON, K., WOŹNIAK, D., WEGNER, K., GRZONKA, Z. & KOZŁOWSKI, H. 2007. Specific interactions of metal ions with Cys-Xaa-Cys unit inserted into the peptide sequence. *Journal of Inorganic Biochemistry,* 101**,** 1699-1706.

KURUCZ, I., TITUS, J. A., JOST, C. R. & SEGAL, D. M. 1995. Correct disulfide pairing and efficient refolding of detergent-solubilized single-chain Fv proteins from bacterial inclusion bodies. *Molecular Immunology,* 32**,** 1443-1452.

LAU, A. T. & CHIU, J.-F. 2007. The possible role of cytokeratin 8 in cadmium-induced adaptation and carcinogenesis. *Cancer research,* 67**,** 2107-2113.

LEE, J.-W. & HELMANN, J. D. 2006. The PerR transcription factor senses H2O2 by metal-catalysed histidine oxidation. *Nature,* 440**,** 363-367.

LEIERS, B., KAMPKÖTTER, A., GREVELDING, C. G., LINK, C. D., JOHNSON, T. E. & HENKLE-DÜHRSEN, K. 2003. A stress-responsive glutathione S-transferase confers resistance to oxidative stress in Caenorhabditis elegans. *Free Radical Biology and Medicine,* 34**,** 1405-1415.

LESLIE, E. M., DEELEY, R. G. & COLE, S. P. C. 2001. Toxicological relevance of the multidrug resistance protein 1, MRP1 (ABCC1) and related transporters. *Toxicology,* 167**,** 3-23.

LESZCZYSZYN, O. I. & BLINDAUER, C. A. 2010. Zinc transfer from the embryo-specific metallothionein EC from wheat: a case study. *Physical Chemistry Chemical Physics,* 12**,** 13408-13418.

LEUNG, M. C., WILLIAMS, P. L., BENEDETTO, A., AU, C., HELMCKE, K. J., ASCHNER, M. & MEYER, J. N. 2008. Caenorhabditis elegans: an emerging model in biomedical and environmental toxicology. *Toxicological Sciences,* 106**,** 5-28.

LICHTY, J. J., MALECKI, J. L., AGNEW, H. D., MICHELSON-HOROWITZ, D. J. & TAN, S. 2005. Comparison of affinity tags for protein purification. *Protein Expression and Purification,* 41**,** 98-105.

LIEBERMAN-AIDEN, E., VAN BERKUM, N. L., WILLIAMS, L., IMAKAEV, M., RAGOCZY, T., TELLING, A., AMIT, I., LAJOIE, B. R., SABO, P. J. & DORSCHNER, M. O. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science,* 326**,** 289-293.

LITHGOW, G. J. & WALKER, G. A. 2002. Stress resistance as a determinate of C. elegans lifespan. *Mechanisms of Ageing and Development,* 123**,** 765-771.

LIU, K. S. & STERNBERG, P. W. 1995. Sensory regulation of male mating behavior in Caenorhabditis elegans. *Neuron,* 14**,** 79-89.

LUND, J., TEDESCO, P., DUKE, K., WANG, J., KIM, S. K. & JOHNSON, T. E. 2002. Transcriptional profile of aging in C. elegans. *Current Biology,* 12**,** 1566-1573.

LYNES, M. A., ZAFFUTO, K., UNFRICHT, D. W., MARUSOV, G., SAMSON, J. S. & YIN, X. 2006. The physiological roles of extracellular metallothionein. *Experimental Biology and Medicine,* 231**,** 1548-1554.

MA, H., GLENN, T. C., JAGOE, C. H., JONES, K. L. & WILLIAMS, P. L. 2009. A transgenic strain of the nematode Caenorhabditis elegans as a biomonitor for heavy metal contamination. *Environmental Toxicology and Chemistry,* 28**,** 1311-1318.

MAACHUPALLI-REDDY, J., KELLEY, B. D. & CLARK, E. D. B. 1997. Effect of inclusion body contaminants on the oxidative renaturation of hen egg white lysozyme. *Biotechnology progress,* 13**,** 144-150.

MADDEN, T. 2013. The BLAST sequence analysis tool.

MARET, W. 2006. Zinc coordination environments in proteins as redox sensors and signal transducers. *Antioxid Redox Signal,* 8**,** 1419-41.

MARET, W. & VALLEE, B. L. 1998. Thiolate ligands in metallothionein confer redox activity on zinc clusters. *Proceedings of the National Academy of Sciences,* 95**,** 3478-3482.

MARGOSHES, M. & VALLEE, B. L. 1957. A cadmium protein from equine kidney cortex. *Journal of the American Chemical Society,* 79**,** 4813-4814.

MARIOL, M.-C., WALTER, L., BELLEMIN, S. & GIESELER, K. 2013. A rapid protocol for integrating extrachromosomal arrays with high transmission rate into the C. elegans genome. *JoVE (Journal of Visualized Experiments)***,** e50773-e50773.

MARTINEZ-FINLEY, E. J. & ASCHNER, M. 2011. Revelations from the Nematode Caenorhabditis elegans on the Complex Interplay of Metal Toxicological Mechanisms. *Journal of Toxicology*.

MCCARTHY, J. F. & SHUGART, L. R. 1990. *Biomarkers of environmental contamination*.

MCKAY, S., JOHNSEN, R., KHATTRA, J., ASANO, J., BAILLIE, D., CHAN, S., DUBE, N., FANG, L., GOSZCZYNSKI, B. & HA, E. Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans. Cold Spring Harbor symposia on quantitative biology, 2003. Cold Spring Harbor Laboratory Press, 159-170.

MEJÁRE, M. & BÜLOW, L. 2001. Metal-binding proteins and peptides in bioremediation and phytoremediation of heavy metals. *Trends in Biotechnology,* 19**,** 67-73.

MELLO, C. & FIRE, A. 1995. DNA transformation. *Methods Cell Biol,* 48**,** 451-482.

MELLO, C. C., KRAMER, J. M., STINCHCOMB, D. & AMBROS, V. 1991. Efficient gene transfer in C. elegans: extrachromosomal maintenance and integration of transforming sequences. *The EMBO journal,* 10**,** 3959.

MENZEL, R., RÖDEL, M., KULAS, J. & STEINBERG, C. E. 2005. CYP35: xenobiotically induced gene expression in the nematode Caenorhabditis elegans. *Archives of biochemistry and biophysics,* 438**,** 93-102.

MING-HO, Y. 2005. Environmental Toxicology: biological and health effects of pollutants. *CRC Press. Boca Ratón, Fia. USA,* 1**,** 1-4.

MIROUX, B. & WALKER, J. E. 1996. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology,* 260**,** 289-298.

MOILANEN, L. H., FUKUSHIGE, T. & FREEDMAN, J. H. 1999. Regulation of metallothionein gene transcription identification of upstream regulatory elements and transcription factors responsible for cell-specific expression of the metallothionein genes from Caenorhabditis elegans. *Journal of Biological Chemistry,* 274**,** 29655-29665.

MORAIS, S., E COSTA, F. G. & DE LOURDES PEREIRA, M. 2012. *Heavy metals and human health*, INTECH Open Access Publisher.

NEMERY, B. 1990. Metal toxicity and the respiratory tract. *European Respiratory Journal,* 3**,** 202-219.

NIELSON, K. B., ATKIN, C. & WINGE, D. 1985. Distinct metal-binding configurations in metallothionein. *Journal of Biological Chemistry,* 260**,** 5342-5350.

NIELSON, K. B. & WINGE, D. 1983. Order of metal binding in metallothionein. *Journal of Biological Chemistry,* 258**,** 13063-13069.

NORDBERG, M. & NORDBERG, G. 2000. Toxicological aspects of metallothionein. *Cellular and molecular biology (Noisy-le-Grand, France),* 46**,** 451-463.

ONEDA, H. & INOUYE, K. 1999. Refolding and recovery of recombinant human matrix metalloproteinase 7 (matrilysin) from inclusion bodies expressed by Escherichia coli. *Journal of biochemistry,* 126**,** 905-911.

PALMITER, R. D. 1998. The elusive function of metallothioneins. *Proc Natl Acad Sci U S A,* 95**,** 8428-30.

PANDEY, A. & MANN, M. 2000. Proteomics to study genes and genomes. *Nature,* 405**,** 837-846.

PASSERINI, A., PUNTA, M., CERONI, A., ROST, B. & FRASCONI, P. 2006. Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks. *Proteins: Structure, Function, and Bioinformatics,* 65**,** 305-316.

PEREDNEY, C. L. & WILLIAMS, P. L. 2000. Utility of Caenorhabditis elegans for assessing heavy metal contamination in artificial soil. *Arch Environ Contam Toxicol,* 39**,** 113-8.

PETERSON, P. J. 2007. Assessment of Exposure to Chemical Pollutants in Food and Water. *Mineral Components in Foods***,** 413-431.

POTOCKI, S., ROWINSKA-ZYREK, M., VALENSIN, D., KRZYWOSZYNSKA, K., WITKOWSKA, D., LUCZKOWSKI, M. & KOZLOWSKI, H. 2011. Metal Binding Ability of Cysteine-Rich Peptide Domain of ZIP13 Zn2+ Ions Transporter. *Inorganic Chemistry,* 50**,** 6135-6145.

RAINES, R. T., MCCORMICK, M., VAN OOSBREE, T. R. & MIERENDORF, R. C. 2000. [23] The S·tag fusion system for protein purification. *Methods in Enzymology.* Academic Press.

RASMUSSEN, J. P., ENGLISH, K., TENLEN, J. R. & PRIESS, J. R. 2008. Notch signaling and morphogenesis of single-cell tubes in the C. elegans digestive tract. *Dev Cell,* 14**,** 559-69.

REICHERT, K. & MENZEL, R. 2005. Expression profiling of five different xenobiotics using a Caenorhabditis elegans whole genome microarray. *Chemosphere,* 61**,** 229-237.

ROBBINS, A. H., MCREE, D. E., WILLIAMSON, M., COLLETT, S. A., XUONG, N. H., FUREY, W. F., WANG, B. C. & STOUT, C. D. 1991. Refined crystal structure of Cd, Zn metallothionein at 2.0Åresolution. *Journal of Molecular Biology,* 221**,** 1269-1293.

ROBERT, V. & BESSEREAU, J. L. 2007. Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. *The EMBO journal,* 26**,** 170-183.

ROBERTSON, H. M. & THOMAS, J. H. 2006. The putative chemoreceptor families of C. elegans.

ROGERS, M. A., LANGBEIN, L., WINTER, H., BECKMANN, I., PRAETZEL, S. & SCHWEIZER, J. 2004. Hair Keratin Associated Proteins: Characterization of a Second High Sulfur KAP Gene Domain on Human Chromosome 211. *Journal of investigative dermatology,* 122**,** 147-158.

ROGL, H., KOSEMUND, K., KÜHLBRANDT, W. & COLLINSON, I. 1998. Refolding of Escherichia coli produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Letters,* 432**,** 21-26.

ROH, J.-Y., LEE, J. & CHOI, J. 2006. Assessment of stress-related gene expression in the heavy metal-exposed nematode Caenorhabditis elegans: A potential biomarker for metal-induced toxicity monitoring and environmental risk assessment. *Environmental Toxicology and Chemistry,* 25**,** 2946-2956.

ROH, J.-Y., SIM, S. J., YI, J., PARK, K., CHUNG, K. H., RYU, D.-Y. & CHOI, J. 2009. Ecotoxicity of silver nanoparticles on the soil nematode Caenorhabditis elegans using functional ecotoxicogenomics. *Environmental science & technology,* 43**,** 3933-3940.

ROSENBERG, S. A. & GUIDOTTI, G. 1968. The protein of human erythrocyte membranes I. Preparation, solubilization, and partial characterization. *Journal of Biological Chemistry,* 243**,** 1985-1992.

SAHU, S. N., LEWIS, J., PATEL, I., BOZDAG, S., LEE, J. H., SPRANDO, R. & CINAR, H. N. 2013. Genomic analysis of stress response against arsenic in Caenorhabditis elegans. *PloS one,* 8**,** e66431.

SAKULSAK, N. 2012. Metallothionein: An overview on its metal homeostatic regulation in mammals. *Int. J. Morphol,* 30**,** 1007-1012.

SALT, D. E. & RAUSER, W. E. 1995. MgATP-dependent transport of phytochelatins across the tonoplast of oat roots. *Plant Physiology,* 107**,** 1293-1301.

SAMBONGI, Y., NAGAE, T., LIU, Y., YOSHIMIZU, T., TAKEDA, K., WADA, Y. & FUTAI, M. 1999. Sensing of cadmium and copper ions by externally exposed ADL, ASE, and ASH neurons elicits avoidance response in Caenorhabditis elegans. *Neuroreport,* 10**,** 753-757.

SAPER, R. B., PHILLIPS, R. S., SEHGAL, A. & ET AL. 2008. LEad, mercury, and arsenic in us- and indian-manufactured ayurvedic medicines sold via the internet. *JAMA,* 300**,** 915-923.

SHA, K. & FIRE, A. 2005. Imprinting capacity of gamete lineages in Caenorhabditis elegans. *Genetics,* 170**,** 1633-1652.

SHARMA, S. K., GOLOUBINOFF, P. & CHRISTEN, P. 2008. Heavy metal ions are potent inhibitors of protein folding. *Biochemical and biophysical research communications,* 372**,** 341-345.

SIEGEL, F. 2002. Geochemistry in Ecosystem Analysis of Heavy Metal Pollution. *Environmental Geochemistry of Potentially Toxic Metals.* Springer Berlin Heidelberg.

SINGH, R., GAUTAM, N., MISHRA, A. & GUPTA, R. 2011. Heavy metals and living systems: An overview. *Indian Journal of Pharmacology,* 43**,** 246-253.

SKOOG, D. A. & WEST, D. M. 1980. *Principles of instrumental analysis*, Saunders College Philadelphia.

SPITZER, J. J. & POOLMAN, B. 2005. Electrochemical structure of the crowded cytoplasm. *Trends in Biochemical Sciences,* 30**,** 536-541.

STOHS, S. & BAGCHI, D. 1995. Oxidative mechanisms in the toxicity of metal ions. *Free Radical Biology and Medicine,* 18**,** 321-336.

STÜRZENBAUM, S. R., WINTERS, C., GALAY, M., MORGAN, A. J. & KILLE, P. 2001. Metal ion trafficking in earthworms Identification of a cadmium-specific metallothionein. *Journal of Biological Chemistry,* 276**,** 34013-34018.

SUGIMOTO, M. 1995. [Glutathione S-transferases (GSTs)]. *Nihon rinsho. Japanese journal of clinical medicine,* 53**,** 1253-1259.

SULSTON, J. E. & HORVITZ, H. R. 1977. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Developmental biology,* 56**,** 110-156.

SULSTON, J. E., SCHIERENBERG, E., WHITE, J. G. & THOMSON, J. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Developmental biology,* 100**,** 64-119.

SUTHERLAND, D. E. & STILLMAN, M. J. 2014. Challenging conventional wisdom: single domain metallothioneins. *Metallomics,* 6**,** 702-28.

SWAIN, S., KEUSEKOTTEN, K., BAUMEISTER, R. & STÜRZENBAUM, S. 2004. C. elegans metallothioneins: new insights into the phenotypic effects of cadmium toxicosis. *Journal of molecular biology,* 341**,** 951-959.

SZUNYOGH, D. M., GYURCSIK, B., LARSEN, F. H., STACHURA, M. K., THULSTRUP, P. W., HEMMINGSEN, L. & JANCSÓ, A. 2015. Zn II and Hg II binding to a designed peptide that accommodates different coordination geometries. *Dalton Transactions*.

TAMAS, M. J. & MARTINOIA, E. 2006. *Molecular Biology of Metal Homeostasis and Detoxification: From Microbes to Man,* Heidelberg, Springer.

TANIZAWA, H., IWASAKI, O., TANAKA, A., CAPIZZI, J. R., WICKRAMASINGHE, P., LEE, M., FU, Z. & NOMA, K.-I. 2010. Mapping of long-range associations throughout the fission yeast genome reveals global genome organization linked to transcriptional regulation. *Nucleic acids research,* 38**,** 8164-8177.

TCHOUNWOU, P. B., YEDJOU, C. G., PATLOLLA, A. K. & SUTTON, D. J. 2012. Heavy metal toxicity and the environment. *Molecular, Clinical and Environmental Toxicology.* Springer.

TERPE, K. 2003. Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology,* 60**,** 523-533.

THELLMANN, M., HATZOLD, J. & CONRADT, B. 2003. The Snail-like CES-1 protein of C. elegans can block the expression of theBH3-only cell-death activator gene egl-1 by antagonizing the function of bHLH proteins. *Development,* 130**,** 4057-4071.

TIERNEY, M. B. A. L., K.H. 2005. An Introduction to Reverse Genetic Tool for Investigating Gene Function. *The Plant Health Instructor*.

TROEMEL, E. R. 1999. Chemosensory signaling in C. elegans. *BioEssays,* 21**,** 1011-1020.

TROEMEL, E. R., CHOU, J. H., DWYER, N. D., COLBERT, H. A. & BARGMANN, C. I. 1995. Divergent seven transmembrane receptors are candidate chemosensory receptors in C. elegans. *Cell,* 83**,** 207-218.

VALLEE, B. L. & ULMER, D. D. 1972. Biochemical effects of mercury, cadmium, and lead. *Annual review of biochemistry,* 41**,** 91-128.

WALDRON, K. J. & ROBINSON, N. J. 2009. How do bacterial cells ensure that metalloproteins get the correct metal? *Nature Reviews Microbiology,* 7**,** 25-35.

WILHELM, J., PINGOUD, A. & HAHN, M. 2003. Real-time PCR-based method for the estimation of genome sizes. *Nucleic Acids Research,* 31**,** e56.

WILLIAMS, P. L., ANDERSON, G. L., JOHNSTONE, J. L., NUNN, A. D., TWEEDLE, M. F. & WEDEKING, P. 2000. Caenorhabditis elegans as an alternative animal species. *J Toxicol Environ Health A,* 61**,** 641-7.

WILLIAMS, P. L. & DUSENBERY, D. B. 1988. Using the Nematode Caenorhabditis Elegans To Predict Mammalian Acute Lethality To Metallic Salts. *Toxicology and Industrial Health,* 4**,** 469-478.

XIAO, Z., LOUGHLIN, F., GEORGE, G. N., HOWLETT, G. J. & WEDD, A. G. 2004. C-Terminal Domain of the Membrane Copper Transporter Ctr1 from Saccharomyces cerevisiae Binds Four Cu(I) Ions as a Cuprous-Thiolate Polynuclear Cluster: Sub-femtomolar Cu(I) Affinity of Three Proteins Involved in Copper Trafficking. *Journal of the American Chemical Society,* 126**,** 3081-3090.

XIAO, Z. & WEDD, A. G. 2010. The challenges of determining metal–protein affinities. *Natural product reports,* 27**,** 768-789.

YADAV, S. K. 2010. Heavy metals toxicity in plants: An overview on the role of glutathione and phytochelatins in heavy metal stress tolerance of plants. *South African Journal of Botany,* 76**,** 167-179.

YAO, Y., NELLÅKER, C. & KARLSSON, H. 2006. Evaluation of minor groove binding probe and Taqman probe PCR assays: Influence of mismatches and template complexity on quantification. *Molecular and Cellular Probes,* 20**,** 311-316.

YOU, C., MACKAY, E. A., GEHRIG, P. M., HUNZIKER, P. E. & KÄGI, J. H. 1999. Purification and characterization of recombinant Caenorhabditis elegans metallothionein. *Archives of biochemistry and biophysics,* 372**,** 44-52.

YU, M.-H. 2004. Environmental Toxicology. *Biological and Health Effects of Pollutants.* 2nd Edition ed. BocaRaton, USA: CRC Press LLC.

ZEITOUN-GHANDOUR, S., CHARNOCK, J. M., HODSON, M. E., LESZCZYSZYN, O. I., BLINDAUER, C. A. & STÜRZENBAUM, S. R. 2010. The two Caenorhabditis elegans metallothioneins (CeMT-1 and CeMT-2) discriminate between essential zinc and toxic cadmium. *FEBS Journal,* 277**,** 2531-2542.

ZEITOUN-GHANDOUR, S., LESZCZYSZYN, O. I., BLINDAUER, C. A., GEIER, F. M., BUNDY, J. G. & STÜRZENBAUM, S. R. 2011. C. elegans metallothioneins: response to and defence against ROS toxicity. *Molecular BioSystems,* 7**,** 2397-2406.

ZHANG, B., GEORGIEV, O., HAGMANN, M., GÜNES, Ç., CRAMER, M., FALLER, P., VASÁK, M. & SCHAFFNER, W. 2003. Activity of metal-responsive transcription factor 1 by toxic heavy metals and H2O2 in vitro is modulated by metallothionein. *Molecular and cellular biology,* 23**,** 8471-8485.

Chapter 9:Appendices

9.1 Appendix: List of the gene chains occurring across the entire worm genome and their functions.

Chromosome	Chain	Family	Isoforms	Function
1	1	Unidentified	WBGene00014698 WBGene00007989 WBGene00007990 WBGene00014699 WBGene00014700 WBGene00007992	Unidentified
1	2	sri Serpentine Receptor, class I	sri-13 sri-15 sri-16 sri-17	7TM GPCR, serpentine chemoreceptor class i.
1	3	sra Serpentine Receptor, class A (alpha)	sra-20 sra-21 sra-22 sra-24 T06G6.11 (Pseudogene)	G-protein coupled receptor activity, based on protein domain information.
1	4	clec C-type Lectin	clec-95 clec-94 clec-96 clec-97 clec-98 clec-99 ZK39.9	Carbohydrate binding activity.
2	1	Unidentified	WBGene00015108 WBGene00022870 WBGene00022869 WBGene00015114 WBGene00022873	Unidentified
2	2	Unidentified	WBGene00015111 WBGene00015112 WBGene00015112 WBGene00015113	Ortholog of human TNFAIP1 (tumor necrosis factor, alpha-induced protein 1 (endothelial)), KCTD13 (potassium channel tetramerization domain.
2	3	Sri Serpentine Receptor, class I	sri-28 sri-29 sri-30 B0454.21 (non coding RNA)	7TM GPCR, serpentine chemoreceptor class i.
2	4	F-box A protein	fbxa-164 fbxa-166 fbxa-158 fbxa-159	motif predicted to mediate protein-protein interactions either with homologs of yeast Skp-1p or with other proteins;

				this gene's encoded protein also contains an FTH/DUF38 motif, which may also mediate protein-protein interaction.
2	5	No similarity	fbxa-13 C08F1.6 CO8F1.10 C08F1.11	No common function.
2	6	math Meprin-associated Traf homology domain	math-3 math-4 math-27 math-28 F36H5.10 (no data)	May be involved in apoptosis.
2	7	chil Chitinase like	chil-1 chil-3 chil-5 chil-6	Predicted to have chitinase activity, based on protein domain information.
2	8	math Meprin-associated Traf homology domain	math-16 math-17 math-5 math-6 math-7	May be involved in apoptosis.
2	9	math Meprin-associated Traf homology domain	math-14 math-15 math-10	May be involved in apoptosis.
2	10	Unidentified	WBGene00015849 WBGene00015850 WBGene00015851 WBGene00015852 WBGene00195070 WBGene00195071 WBGene00015853	Unidentified
2	11	Unidentified	WBGene00015853 WBGene00015842 WBGene00195071 WBGene00195070	Unidentified
2	12	F-box C protein	fbxc-16 fbxc-15 fbxc-14 fbxc-17	
2	13	Unidentified	C40A11.2 C40A11.6 C40A11.7	Encodes an ortholog of human potassium channel tetramerization domain
2	14	Unidentified	WBGene00016573 WBGene00016574 WBGene00016575 WBGene00019511 WBGene00016575	Unidentified

			WBGene00019511 WBGene00016576	
2	15	math Meprin-associated Traf homology domain	math-22 math-23 math-24 math-30	May be involved in apoptosis.
2	16	Unidentified	C52E2.4 C52E2.5 T08E11.8 Linc-16 (long non-coding RNA)	Unidentified
2	17	Unidentified	WBGene00008494 WBGene00008493 WBGene00008495 WBGene00012727	Unidentified
2	18	F-box B protein	fbxb-105 sdz-10 sdz-9 fbxb-111	Encodes a protein containing an F-box, a motif predicted to mediate protein-protein interactions either with homologs of yeast Skp-1p or with other proteins; its encoded protein also contains a Pfam-B45 motif of presently unknown function.
2	19	F-box B protein	sdz-11 fbxb-91 fbxb-92 fbxb-90	Same as mentioned above.
2	20	sre Serpentine receptor class epsilon	sre-33 sre-32 sre-34 tag-280	*tag-280* encodes an ortholog of human plasminogen receptor, C-terminal lysine transmembrane protein
2	21	nep Neprilysin metallopeptidase family	nep-7 nep-8 nep-9 nep-10 nep-20	neprilysins are thermolysin-like zinc metallopeptidases, found on the outer surface of animal cells, that negatively regulate small signalling peptides (e.g., enkephalin, tachykinin, insulin, and natriuretic peptides) by cleaving them; F18A12.3 has no clear orthologs in other organisms.
2	22	clec C-type Lectin	clec-62 clec-63 clec-64	
2	23	No similarity	F36H5.8 F36H5.9 fbxb-12 F36H5.13	No common function
2	24	No similarity	F36H5.4 fbxb-53 F36H5.10	No common function

2	25	gst Glutathione s transferase	gst-12 gst-14 gst-15 gst-20 gst-24	Encodes an ortholog of human hematopoietic prostaglandin D synthase (HGNC:HPGDS).
2	26	gst Glutathione s transferase	gst-16 gst-17 gst-18 gst-19	Same as mentioned above.
2	27	Unidentified	WBGene00044317 WBGene00044316 WBGene00044316 WBGene00044315 WBGene00018303	Unidentified
2	28	No similarity	WBGene00018384 WBGene00018385 WBGene00206413 WBGene00022872	No common function.
2	29	Unidentified	WBGene00018747 WBGene00018748 WBGene00018750 WBGene00018750 WBGene00018749	Unidentified
2	30	F-box B protein	fbxb-25 fbxb-26 fbxb-18 fbxb-19	Encodes a protein containing an F-box, a motif predicted to mediate protein-protein interactions either with homologs of yeast Skp-1p or with other proteins.
2	31	BTB and MATH domain containing	bath-21 bath-1 bath-3 bath-5	Encodes an ortholog of human kelch repeat and BTB (POZ) domain.
2	32	No similarity	ift-74 H43E16.1 T19D12.2 T19D12.3 T19D12.10	No common function
2	33	Unidentified	WBGene00086561 WBGene00019278 WBGene00019283 WBGene00019279	Unidentified
2	34	Chil Chitinase like	chil-13 chil-16 chil-17 21ur-12211 (21 u RNA gene)	Predicted to have chitinase activity, based on protein domain information.
2	35		cyp-13A7 cyp-13A5 cyp-13A6 cyp-13A8 T10B9.13 (ncRNA)	cyp-13A7 encodes a homolog of cytochrome P450 proteins; these proteins are membrane proteins with a heme prosthetic group that catalyze the synthesis of steroid hormones (and bile salts), and

				also detoxify foreign substances (xenobiotic compounds); loss of cyp-13A7 activity in large-scale RNAi screens results in uncoordinated locomotion and decreased and/or slow growth; a cyp-13A7::gfp promoter fusion is induced in the intestine in response to treatment with the antibiotic rifampicin.
2	36	irld Insulin/EGF-Receptor L Domain protein	irld-49 irld-50 irld-65 irld-66 irld-67	Insulin/EGF-Receptor L Domain protein.
2	37	srh Serpentine Receptor, class H	srh-61 srh-69 srh-70 srh-73	7TM GPCR, serpentine receptor class h (Srh).
2	38	clec C-type lectin	clec-131 clec-133 clec-132 clec-130 clec-129	
2	39	No similarity	W10G11.3 W10G11.4 clec-132 dnc-3	No common function.
2	40	Nematode Specific Peptide family, group E	nspe-4 nspe-3 nspe-1 nspe-6 nspe-2 nspe-8	
2	41	Glutathione s transferase	gst-26 gst-27 gst-28 gst-29 gst-39	
2	42	Unidentified	WBGene00013263 WBGene00198206 WBGene00013256 WBGene00044051 WBGene00013250	Unidentified
2	43	F-box B protein	fbxb-15 fbxb-21 fbxb-20 fbxb-16	
2	44	Unidentified	ZC239.13 ZC239.14 ZC239.15 ZC239.16	Unidentified

			ZC239.17	
2	45	sri Serpentine Receptor, class I	sri-48 sri-50 sri-51 sri-53	
2	46	Unidentified	ZK1240.2 ZK1240.3 ZK1240.6 ZK1240.8 ZK1240.9	ZK1240.2 encodes an ortholog of human tripartite motif containing 32(HGNC:TRIM32); ZK1240.2 is predicted to have zinc ion binding activity, based on protein domain information.
3	1	No similarity	WBGene00015639 WBGene00015643 WBGene00015640 WBGene00015641	No common function.
3	2	Unidentified	WBGene00015642 WBGene00017847 WBGene00015642 WBGene00017849	Unidentified
3	3	srg Serpentine Receptor, class G (gamma)	srg-1 srg-2 srg-4 srg-5 srg-6 srg-8 srg-9	Predicted to have transmembrane signaling receptor activity, based on protein domain information.
3	4	F-box protein A	fbxa-10 fbxa-11 fbxa-67 fbxa-23	
3	5	F-box protein A	fbxa-22 fbxa-59 fbxa-60 fbxa-23	
3	6	Unidentified	WBGene00012109 WBGene00012110 WBGene00012109 WBGene00012111 WBGene00194852	Unidentified
3	7	Unidentified	WBGene00022327 WBGene00022319 WBGene00022341 WBGene00022340 WBGene00022342	Unidentified
3	8	F-box A protein	fbxa-29 fbxa-30 fbxa-31 fbxa-32	
4	1	srv Serpentine Receptor, class V	srv-28 srv-30 srv-31	Encodes an ortholog of human coagulation factor II (thrombin) receptor-like 1 (HGNC:F2RL1).

			T13A10.63: (ncRNA)	
4	2		F55G11.2 dod-22 (Downstream Of DAF-16 (regulated by DAF-16) F55G11.6:Unknown F55G11.7 F55G11.8	Encodes a protein containing a CUB (complement C1r/C1s, Uegf, Bmp1)-like domain; F55G11.2 expression is upregulated in response to infection with the human bacterial pathogen Pseudomonas aeruginosa; F55G11.2 expression is regulated by the ELT-2 GATA transcription factor.
4	3	F-box C protein	fbxc-4 fbxc-2 fbxc-1 fbxc-8 fbxc-3 fbxc-12 fbxc-11 fbxc-10 fbxc-9	
4	4	F-box C protein	fbxc-4 fbxc-8 fbxc-3 fbxc-12 fbxc-11 fbxc-9	
4	5	scl SCP-Like extracellular protein	scl-1 scl-2 scl-3	Encodes a predicted secretory protein that is a member of the cysteine-rich secretory protein (CRISP) family; scl-1 activity positively regulates longevity and stress resistance; in wild-type animals, scl-1 mRNA is detected solely in eggs, while in daf-2 mutants it is detected in eggs and late adult stages; scl-1 expression appears to be under the control of the DAF-2/insulin-like signaling pathway as, in mixed-stage cultures,scl-1 mRNA levels are increased in daf-2 andage-1 mutants and undetectable in daf-16mutants; scl-1 contains a DAF-16 consensus binding element within its predicted regulatory regions.
4	6	Unidentified	WBGene00170935 WBGene00174799 WBGene00016758 WBGene00016763 WBGene00016764 WBGene00016759 WBGene00016760	Unidentified

			WBGene00016763	
4	7	Unidentified	WBGene00016759 WBGene00016760 WBGene00016763 WBGene00016763	Unidentified
4	8	sru Serpentine Receptor, class U	sru-17 sru-18 sru-19 sru-20	7TM GPCR, serpentine receptor class u (Sru).
4	9	clec C-type lectin	clec-167 clec-166 F38A1.6: no data clec-168 clec-165 F38A1.13: no data clec-169	
4	10	clec C-type lectin	clec-74 clec-75 clec-76	
4	11	No similarity	WBGene00017484 WBGene00044479 WBGene00017485 WBGene00017488	No common function.
4	12	Scramblase (phospholipid scramblase)	K08D10.5: K08D10.5 encodes an ortholog of human DNA replication helicase/nuclease 2 (HGNC:DNA2). scrm-8 scrm-5 K08D10.18	Encodes a putative phospholipid scramblase.
4	13	No similarity	WBGene00018650 WBGene00235368 WBGene00018649 WBGene00018651 WBGene00235367	No common function.
4	14	Unidentified	WBGene00219497 WBGene00018082 WBGene00018083 WBGene00170053 WBGene00018084 WBGene00018084	Unidentified
4	15	No similarity	sru-9 sru-16 21ur-6542 Y45F10B.60	No common function.
4	16	Unidentified	WBGene00019931 WBGene00219216 WBGene00019533	Unidentified

			WBGene00219216	
4	17	Unidentified	WBGene00044308 WBGene00044307 WBGene00044309 WBGene00044306 WBGene00019059	Unidentified
4	18	W08E12.x	WBGene00021084 W08E12.3 WBGene00021085 W08E12.4 WBGene00021083 W08E12.2 WBGene00021086 W08E12.5	Encodes an ortholog of human keratin associated protein 5-1
4	19	Nuclear Hormone Receptor family	nhr-87 nhr-92 nhr-122 nhr-287 nhr-146	Encodes an ortholog of human hepatocyte nuclear factor 4, alpha (HGNC:HNF4A); nhr-87 is predicted to have sequence-specific DNA binding transcription factor activity, steroid hormone receptor activity, zinc ion binding activity, and sequence-specific DNA binding activity, based on protein domain information.
4	20	No similarity	str-156 Y9C9A.17 21ur-13648 21ur-13487	No common function.
4	21	No similarity	str-174 21ur-5006 21ur-8485 21ur-10569	No common function.
5	1	cyp Cytochrome P450 family	cyp-34A5 cyp-34A6 cyp-34A7 cyp-34A8	
5	2	cyp Cytochrome P450 family	cyp-34A8 cyp-34A9 cyp-34A10	
5	3	Unidentified	C01G10.4 C01G10.5 C01G10.6 C01G10.8 C01G10.15 C01G10.17 C01G10.18	Unidentified
5	4	srh Serpentine receptor class h	srh-20 srh-21 srh-22 srh-23	
5	5	No similarity	str-231 str-233 T10C6.2	No common function.

			C06C6.12	
5	6	nhr Nuclear Hormone Receptor family	nhr-136 nhr-154 nhr-207 nhr-209 C13C4.11: ncRNA	Encodes an ortholog of human nuclear receptor 1H3 (HGNC:NR1H3); nhr-136 is predicted to have sequence-specific DNA binding transcription factor activity, steroid hormone receptor activity, zinc ion binding activity, and sequence-specific DNA binding activity, based on protein domain information.
5	7	serpentine receptor class bc (Srbc)	str-259: Seven TM Receptor srbc-10 srbc-9 srbc-7 srbc-8	7TM GPCR, serpentine receptor class bc
5	8	Serpentine Receptor, class G	srg-56 srg-57 srg-61 srg-62	Predicted to have transmembrane signaling receptor activity, based on protein domain information.
5	9	Nuclear Hormone Receptor family	nhr-161 nhr-139 nhr-140 nhr-162 nhr-163	Predicted to have sequence-specific DNA binding transcription factor activity, steroid hormone receptor activity, zinc ion binding activity, and sequence-specific DNA binding activity, based on protein domain information.
5	10	srbc Serpentine receptor class bc	srbc-21: srbc-22 srbc-23 srbc-24	7TM GPCR, serpentine receptor class bc
5	11	Unidentified	WBGene00016788 WBGene00016783 WBGene00077593 WBGene00016785 WBGene00016784	Unidentified
5	12	cdr Cadmium responsive	cdr-4 cdr-6 cdr-7 C54D10.18: ncRNA anr-30: *C. elegans* anti-sense RNA gene	Encodes a predicted transmembrane protein that is a member of the CDR (cadmium responsive) family of proteins; cdr-4 expression is upregulated in response to cadmium treatment.
5	13	Unidentified	C55A6.3 sdz-8 C55A6.6 C55A6.7	Encodes orthologs of human carbonyl reductase/retinol dehydrogenase/hydroxysteroid ; predicted to have oxidoreductase activity based on protein domain information.
5	14	Unidentified	E02C12.6	Predicted to have transferase activity, transferring phosphorus-containing groups,

			E02C12.8: non-coding Transcript Isoform E02C12.9 E02C12.10 E02C12.11 E02C12.12: no data	based on protein domain information.
5	15	srd Serpentine receptor class D	srd-25 srd-29 srd-30 srd-31	
5	16	clec C-type lectin	clec-227 clec-57 clec-56 clec-54 clec-55	
5	17	Serpentine receptor class x	srx-134 srx-135 srx-136 srx-137	7TM GPCR, serpentine receptor class x (Srx).
5	18	str Seven TM Receptor	str-108 str-109 str-111 F10A3.12: no data K05D4.9: no data	G-protein coupled receptor.
5	19	F-box protein B	fbxb-62 fbxb-61 F55C9.11 (fbxb-63) F55C9.14: no data F55C9.15: no data	
5	20	UDP-Glucuronosyl Transferase	ugt-8 H23N18.4 ugt-9 ugt-11 ugt-10	Encodes an ortholog of human UDP glycosyltransferase 3 family, polypeptide A2 (HGNC:UGT3A2); ugt-8 is predicted to have transferase activity, transferring hexosyl groups, based on protein domain information
5	21	cyp Cytochrome P450 family	cyp-35B3 cyp-35B2 cyp-35B1 cyp-35A5	Encodes an ortholog of human cytochrome P450 2C8 (HGNC:CYP2C8); cyp-35B3 is predicted to have iron ion binding activity, oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, and heme binding activity, based on protein domain information.
5	22	Nuclear Hormone Receptor family	nhr-206 nhr-207 nhr-208 nhr-209	
5	23	Keratin associated protein	T05B4.8 T05B4.9 T05B4.10	Encodes an ortholog of human keratin associated protein.

			phat-5: Pharyngeal gland Toxin-related T05B4.12	
5	24	No similarity	WBGene00020243 WBGene00020244 WBGene00020238 WBGene00020237	No common function.
5	25	Serpentine Receptor, class G (gamma)	srg-26 srg-27 srg-28 srg-29	Predicted to have transmembrane signaling receptor activity, based on protein domain information
5	26	hsp Heat shock protein	hsp-16.1 hsp-16.48 hsp-16.49 T27E4.22: small nuclear-RNA gene	Encodes a 16-kD heat shock protein (HSP) that is a member of thehsp16/hsp20/alphaB-crystallin (HSP16) family of heat shock proteins, and that is identical to the protein encoded by hsp-16.48; HSP-16.49 is likely to function as a passive ligand temporarily preventing unfolded proteins from aggregating.
5	27	srh Serpentine Receptor, class H	srh-287 srh-291 srh-296 Y94A7B.12	7TM GPCR, serpentine receptor class h.
X	1	nspc Nematode Specific Peptide family, group C	nspc-1 nspc-2 nspc-3 nspc-4 nspc-5	Nematode Specific Peptide family, group C.
X	2	nspc Nematode Specific Peptide family, group C	nspc-20 nspc-18 nspc-17 nspc-19	Nematode Specific Peptide family, group C.
X	3	Nuclear Hormone Receptor family	nhr-27 nhr-272 nhr-21 nhr-215 nhr-26	Encodes an ortholog of human nuclear receptor 1H4 (HGNC:NR1H4); nhr is predicted to have sequence-specific DNA binding transcription factor activity, steroid hormone receptor activity, zinc ion binding activity, and sequence-specific DNA binding activity, based on protein domain information.
X	4	srd Serpentine receptor class delta-41	srd-41 srd-44 srd-45 srd-46 srd-47 srd-48	Serpentine receptor class delta-41
X	5	Unidentified	WBGene00003998 WBGene00003997	Unidentified

			WBGene00003997 WBGene00003998	
X	6	nspc Nematode Specific Peptide family, group C	nspc-13 nspc-14 nspc-15 nspc-12	Nematode Specific Peptide family, group C
X	7	Unidentified	WBGene00018534 WBGene00219411 WBGene00018535 WBGene00018536 WBGene00020515	Unidentified
X	8	spp Saposin-like Protein family	spp-2 spp-3 spp-4 spp-5 spp-6	Encodes an antimicrobial peptide that belongs to the SPP-protein family, called caenopores, with structural similarity to saposin-like proteins (SAPLIPs), the amoebapores and the mammalian NK-lysin and granulysin; caenopores are part of the worm's innate immune system and interact with and eliminate bacteria via membrane pore-forming activity; C. elegans has at least 33 putative SPP proteins
X	9	col Collagen	col-167 col-169 col-170 col-171 T10E10.11: ncrna	Structural constituent of cuticle, based on protein domain information.
X	10	Unidentified	WBGene00022728 WBGene00022729 WBGene00022730 WBGene00022731	Unidentified

9.2 Appendix: List of common transcription factor binding sites in the promoter region upstream of *W08E(12.2-12.5)* and their frequency of occurrence in between the isoforms.

Common factors between the *W08E(12.2-12.5)* promoters	**Frequency of occurrence**
T00104 C/EBPalpha T00105 T00107 T00108 T01388 C/EBP	48
T01484 Cdx-1	44
_00000 TFIID	36
T00395 Hb	28
T00306 GATA-1	26
T00305 GATA-1	25
T00997 T00996 SRY	24
T00036 AP-4	22
T00923 Zta	20
T00029 T00030 T00031 T00032 T01115 T01140 T01156 T00027 AP-1	19
T00627 NIT2	19
T00114 c-Ets-1_54	18
T00535 NF-1	18
_00000 DBP	17
T00267 GATA-1	17
_00000 GHF-1	16
T00321 GCN4	16
T00385 HSF1	16
T00642 POU2F1a T00643 T00644 T00959 T01031 T01157 T01466 T00641 POU2F1	16
_00000 Hb	15
_00000 SEF4	15
_00000 Sp1	15
_00000 Zta	15
T00138 T00139 T00137 c-Myb	15

T00304 T00305 GATA-1A T00306 T00267 GATA-1	15
T00304 T00305 GATA-1A T00306 T00267 GATA-1	15
T00798 TBP	15
T01599 LCR-F1	15
T01806 T00671 p53	15
T02072 Msx-1	15
T02691 Dof3	15
_00000 NF-S	14
T01476 Abd-B	14
T02016 En-1	14
_00000 CAP/CRP	13
_00000 c-Jun	13
_00000 CP2	13
_00000 erg	13
_00000 PEA3	13
T00625 AREB6	13
T02983 Pax-4a	13
_00000 DEF	12
_00000 EFII	12
_00000 LBP-1	12
_00000 PHO2	12
T00304 GATA-3 T00305 GATA-1A T00306 GATA-2 T00307 NF-E1b T00308 T00309 T00310 T00311 T00312 T00313 T00314 T00567 T01302 T00267 GATA-1	12
T00487 MATalpha2	12
T01467 deltaEF1	12
_00000 Ftz.2	11
_00000 GCN4	11
_00000 Oct-1	11
_00000 PEB1	11
_00000 V$CAP_01	11
T00011 ADR1	11
T00029 AP-1	11
T00107 C/EBPalpha	11

T00120 CF2-II	11
T00311 GATA-3	11
T00333 GR	11
T00691 Pit-1a	11
T00759 Sp1	11
T00895 v-Myb	11
_00000 AP-1	10
_00000 AP-3	10
_00000 Elk-1	10
_00000 GATA-1	10
_00000 GR/PR	10
_00000 IL-6.RE-BP	10
_00000 PUT3	10
_00000 Tll	10
_00000 Zen	10
T00031 AP-1	10
T00137 c-Myb	10
T00788 T-Ag	10
T00872 USF-1 T00874 T00875 T00876 USF1 T00877 T00870 USF	10
T01059 MNB1a	10
T01429 Sox-5	10
T01481 Pbx-1a	10
T02637 RAV1	10
T02841 FACB	10
_00000 C/EBP	9
_00000 CAC-binding	9
_00000 DEP2	9
_00000 E12	9
_00000 Eve	9
_00000 Ftz.1	9
_00000 TCF-1	9
_00000 TEF	9
_00000 TTF-1	9
T00108 C/EBPalpha	9

T00123 c-Fos	9
T00386 HSTF	9
T00479 LyF-1	9
T01042 T01525 HSF1_(long) T01044 HSF1	9
T02251 T02250 Lmo2	9
_00000 AP-2	8
_00000 f_alp-f_eps	8
_00000 GAGA	8
_00000 myogenin	8
_00000 OBF	8
_00000 Pap1	8
_00000 PR	8
_00000 SEF1	8
T00028 T00029 YAP1 T00030 T00031 T00032 T01115 T01132 T01140 T01156 T00027 AP-1	8
T00307 GATA-2	8
T00702 PU.1	8
_00000 Gt	7
_00000 Oct-4	7
_00000 PU.1	7
_00000 RAR-gamma	7
_00000 USF	7
T00111 c-Ets-1	7
T00114 c-Ets-1_54 T00111 c-Ets-1	7
T00253 En	7
T00529 MZF-1	7
T01043 T01045 T00972 HSF2	7
T02692 PBF	7
_00000 ABF-2	6
_00000 c-Fos	6
_00000 EcR	6
_00000 Eryf1	6
_00000 GATA-3	6
_00000 GT-1	6

_00000 MAF	6
_00000 MCBF	6
_00000 NBF	6
_00000 Oct-2	6
_00000 twi	6
_00000 YY1	6
T00032 AP-1	6
T00032 c-Fos T00122 T00123 T00124 T00125 T00131 T00132 T00133 T00134 T01156 c-Jun T01976 T02205 T02830 T00029 AP-1	6
T00167 CRE-BP1	6
T00794 TBP	6
T00844 Ttk_88K	6
T00915 delta_factor T00278 YY1	6
_00000 B-factor	5
_00000 c-Myb	5
_00000 Elf-1/NTF-1	5
_00000 HNF-5	5
_00000 HOXA5	5
_00000 LVc	5
_00000 LyF-1	5
_00000 MAT-alpha-2	5
_00000 MNB1b	5
_00000 MNF1	5
_00000 MyoD	5
_00000 Net	5
_00000 Prd	5
_00000 R2	5
_00000 RAP1/SBF-E/TUF	5
_00000 SGF-1	5
_00000 STE12	5
_00000 TBP	5
_00000 TFII-I	5
T00033 AP-2alpha T00035 AP-2alphaA	5
T00133 c-Jun	5

T00193 Dfd	5
T00250 Elk-1	5
T00322 GCR1	5
T00459 T00581 T00017 C/EBPbeta	5
T00525 T00526 T01128 T00524 MyoD	5
T00528 myogenin	5
T00539 NF-1	5
T00630 POU3F2 T01524 POU3F2 T01873 POU3F2	5
T00772 STE12	5
T00794 TBP T00796 TBP T00798 TBP T00820 TFIID	5
T00796 T00797 T00794 TBP	5
T00796 TFIID T00820 T00821 T01159 T01175 T00794 TBP	5
T01427 p300	5
T01479 BR-C_Z3	5
T02068 PU.1	5
_00000 ATF-1	4
_00000 CBF1	4
_00000 CBP/CRF	4
_00000 CP1	4
_00000 E2F-1	4
_00000 E74A	4
_00000 Elf-1	4
_00000 HAP	4
_00000 HEB	4
_00000 NFAT-1	4
_00000 NF-IL6	4
_00000 R1	4
_00000 SBF-1	4
_00000 SGF-2/3/4	4
_00000 TAF-1	4
_00000 T-Ag	4
_00000 TEF1/GT-IIC	4
_00000 TFE3-S	4
_00000 TREB-1	4

_00000 Ttk	4
_00000 zeste	4
T00076 CAC-binding_protein	4
T00111 c-Ets-1 T00112 c-Ets-1 T00114 c-Ets-1 54 T00115 c-Ets-1 68 T00684 PEA3 T00685 PEA3 T00686 PEA3	4
T00115 c-Ets-1_68	4
T00132 c-Jun T00133 T00134 T00167 T00131 CRE-BP1	4
T00164 T00989 T01307 T01380 T01385 T00163 CREB	4
T00267 GATA-1 T00305 GATA-1 T00306 GATA-1	4
T00305 GATA-1 T00306 GATA-1 T00308 GATA-2 T01690 elt-2	4
T00534 T00535 T00536 T00537 T00538 T00539 T01036 T01298 T00533 NF-1	4
T00599 NF-1/L	4
T00751 Sn	4
T00829 TGA1a	4
T00968 ATF-1	4
T01101 SEF4	4
T01470 Ik-2	4
T01542 E2F-1	4
T01675 Nkx2-5	4
_00000 ASF-1 _00000 MSN4 _00000 deltaCREB	3
_00000 ATF-CREB	3
_00000 CYP1.HAP1	3
_00000 E2A	3
_00000 Ets-1	3
_00000 fl-fII	3
_00000 H-2RIIBP	3
_00000 HSTF	3
_00000 IHF	3
_00000 MBF-1	3
_00000 MIG1	3
_00000 NFIII	3
_00000 NF-InsE3	3
_00000 NF-uE2	3
_00000 NF-uE5	3

_00000 NF-X	3
_00000 PEBP2	3
_00000 SRF	3
_00000 Tal-1	3
_00000 v-Jun	3
T00027 AP-1 T00028 YAP1 T00029 AP-1 T00030 AP-1 T00031 AP-1 T00032 AP-1 T01115 AP-1 T01132 AP-1 T01140 AP-1	3
T00036 AP-4 T00204 E12 T00526 MyoD T00906 XPF-1 T01447 HEN1 T01654 HEN1 T01655 HEN1	3
T00051 ATF	3
T00133 c-Jun T00167 CRE-BP1	3
T00184 DBP	3
T00267 GATA-1 T00305 GATA-1 T00794 TBP T00821 TFIID	3
T00295 Ftz	3
T00369 HNF-1 T00889 HNF-1B	3
T00371 HNF-3	3
T00378 HOXA3	3
T00456 Kr	3
T00489 Max1	3
T01094 DSC1 T01120 MCBF	3
T01517 Twi	3
T02857 TCF-3	3
T02878 TCF-4E	3
T03388 Meis-1a T03389 Meis-1b T03390 Meis-1-1 T03391 Meis-1-2 T03411 Meis-1-3 T03419 Meis-1	3
T03461 Crx	3
_00000 CF1	2
_00000 HBP-1	2
_00000 IUF-1	2
_00000 Ker-1	2
_00000 Kruppel	2
_00000 LRF-1	2
_00000 MalT	2
_00000 P$GBP_Q6	2

_00000 PHO4	2
_00000 RC2	2
_00000 RFX2	2
_00000 T3R-beta	2
_00000 V$GC_01	2
_00000 VDR	2
_00000 vHNF-1	2
_00000 Vmw65	2
_00000 Zmhoxla	2
T00029 AP-1 T00036 AP-4 T00097 CCK-1a T00126 CG-1 T00140 c-Myc T00141 c-Myc T00142 c-Myc T00143 c-Myc T00163 CREB T00489 Max1 T00871 USF T00874 USF1 T00877 USF-1 T01125 TDEF T01517 Twi T02330 G/HBF-1 T02378 USF-1 T04353 CAN T04363 DPBF-1 T04364 DPBF-2	2
T00033 AP-2alpha T00035 AP-2alphaA T02466 AP-2alphaB T02468 AP-2gamma T02469 AP-2beta T02470 T00952 AP-2	2
T00035 AP-2	2
T00049 ATF T00050 atf1 T00051 ATF T00132 c-Jun T00163 CREB T00164 CREB T00166 deltaCREB T00167 CRE-BP1 T00354 HBP-1 T00846 TREB-1 T00937 HBP-1a T00938 HBP-1b T00942 EivF T01095 ATF3	2
T00051 ATF T00052 ATF-a T00053 ATF-adelta T00054 ATF-like T00442 47-kDa CRE bind. prot. T00968 ATF-1	2
T00051 ATF T00167 CRE-BP1 T00223 E4F1	2
T00051 ATF T00354 HBP-1 T00937 HBP-1a T00938 HBP-1b T01078 GBF1 T01393 HBP-1b(c1) T01394 HBP-1a(1) T01395 HBP-1a(c14) T02659 OBF4 T02661 OBF5 T02662 OBF3.1 T02663 OBF3.2 T02789 bZIP910 T02794 TGA1 T02796 TGA3 T02797 TGA6	2
T00152 CP2	2
T00163 CREB	2
T00163 CREB T00164 CREB T00165 deltaCREB T00166 deltaCREB T00989 CREB T01311 deltaCREB T02361 CREBbeta	2
T00164 deltaCREB T00165 T00166 T00989 T01311 T00163 CREB	2
T00216 C/EBPgamma	2
T00321 GCN4 T00829 TGA1a	2
T00557 NF-E2	2
T00573 NF-GMb T01218 NF-GMb	2

T00668 Opaque-2 T00829 TGA1a T00830 TGA1b T00937 HBP-1a T01090 TAF-1 T01091 CPRF-1 T01092 CPRF-2 T01093 CPRF-3 T02669 EmBP-1a T02999 OCSBF-1	2
T00668 Opaque-2 T00829 TGA1a T00830 TGA1b T00937 HBP-1a T01090 TAF-1 T01092 CPRF-2 T01093 CPRF-3	2
T00690 PHO4	2
T00752 Sp1	2
T00754 T00759 T00752 Sp1	2
T00803 TCF-2alpha	2
T00803 TCF-2alpha	2
T00997 SRY	2
T01071 Hlf	2
T01199 T01198 NRF-2	2
T01275 mat1-Mc	2
T01445 N-Myc	2
T01445 N-Myc	2
T02256 AML1a	2
T02450 GKLF	2
T03461 Crx T03489 Crx	2
T04362 IPF1	2

9.3 Appendix: Quantitation of *W08E(12.2 -12.5)* and unravelling their role in metal handling

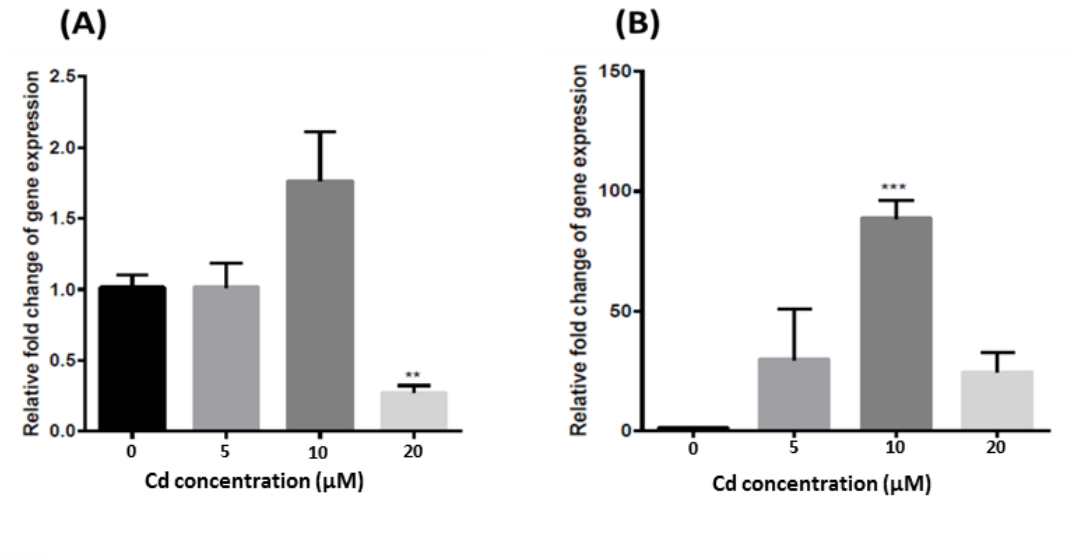


Figure 9-1: Differential *W08E12.2* expression in cDNA samples from metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) worms exposed to defined cadmium concentrations for 48 hr. Nematodes were exposed to Cd (0-20 µM) from L1-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA (1000 ng/µl) was synthesised. (A), (B) and (C) are biological repeats of the cDNA samples. Average fold change expression difference of *W08E12.2* when exposed to defined doses of Cd has been represented. Fold changes of the respective isoforms were calculated upon stable small ribosomal RNA control (*rla-1*) normalization. (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test; *** $P < 0.001$, ** $P < 0.005$, error bars: SEM The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

The response of *W08E(12.2-12.5)* expression upon Cd exposure was tested in an metallothionein null background in order to assess the response in comparison to the wild type (N2). The *W08E12.2* in the metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*)) followed a dose dependent increase in expression up to 10 µM Cd. The biological repeats followed a similar pattern, displaying peak response at

10 µM Cd, although the fold change in the expression of *W08E12.2* between the two repeats varied by a magnitude of 50 (**Figure 9-1**).

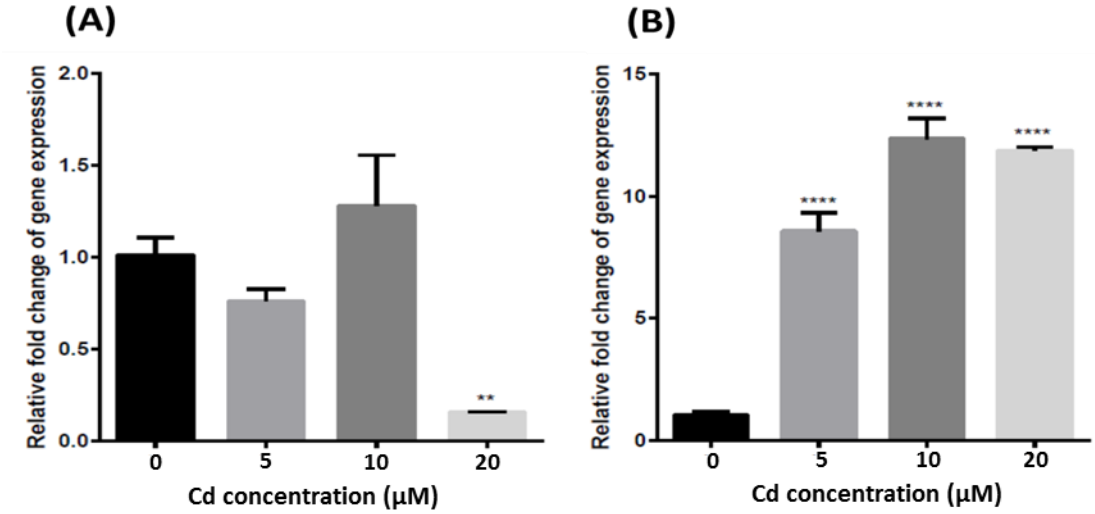


Figure 9-2: Differential *W08E(12.3-12.5)* expression in cDNA samples metallothionein double knockout (*mtl-1(tm1770); mtl-2(gk125)*) worms exposed to defined cadmium concentrations for 48 hr. Nematodes were exposed to Cd (0-20 µM) from L1-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA (1000 ng/µl) was synthesised. (A) and (B) are biological repeats of the cDNA samples. Average fold change expression difference of *W08E12.3, W08E12.4* and *W08E12.5* when exposed to defined doses of Cd has been represented. Fold changes of the respective isoforms were calculated upon stable small ribosomal RNA control (*rla-1*) normalization. (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$, ** $P < 0.005$, error bars: SEM. The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

The *W08E(12.3-12.5)* expression in the metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*)) nematodes displayed an increase upon exposure to 10 µM Cd (**Figure 9-2**). Based on the biological repeats it is difficult to deduce a definitive pattern of *W08E(12.3-12.5)* expression to Cd. But, both the repeats signify that there is an upregulation in the cumulative response of *W08E(12.3-12.5)* at this dose.

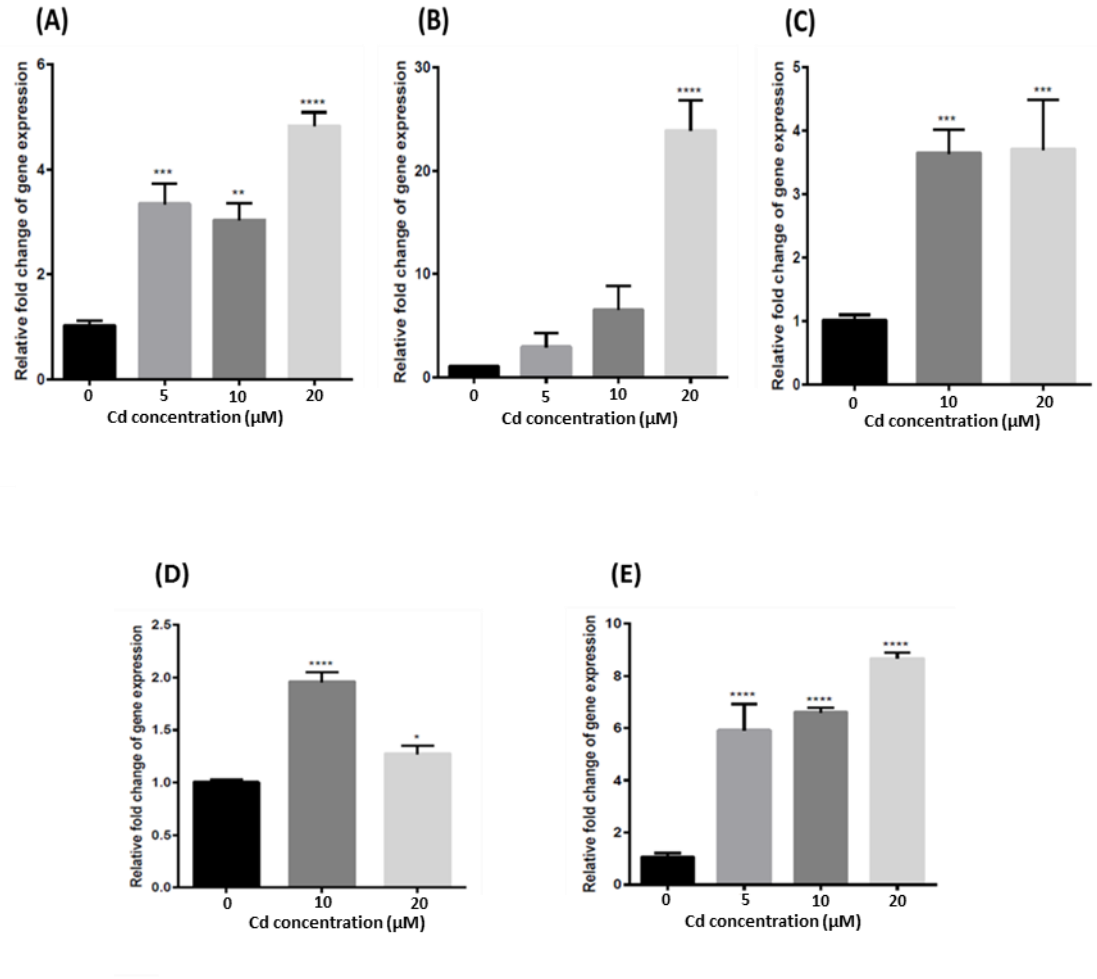


Figure 9-3: Differential *W08E(12.2-12.5)* expression in cDNA samples from metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) worms exposed to Cd for 24 hr amplified with new primer-probe combination. Nematodes were exposed to defined concentrations of Cd (0-20 µM) from L2-L4 stage. Total RNA was extracted from the exposed nematodes and cDNA (1000 ng/µl) was synthesised. Average fold change in expression of *W08E12.2* (biological repeats A, B, C) and *W08E(12.3-12.5)* (biological repeats D,E) when exposed to defined doses of Cd has been represented. Fold changes of the respective isoforms were calculated following the normalization to the invariant control (*rla-1*) normalization (The statistical analysis was performed using a one-way ANOVA and Dunnett's multiple comparison test; **** $P < 0.0001$, *** $P < 0.001$; error bars: SEM. The statistical test defines the significance compared to the unexposed worms (0 µM Cd)).

The response of *W08E(12.2-12.5)* expression upon acute Cd exposure was tested in an metallothionein null background in order to assess if there was a deviation in their response when compared to the chronic exposure (Cd 48 hr) and the wild type (N2).

The *W08E12.2* expression was shown to increase in nematodes exposed to Cd (5-20 µM). All three biological repeats signify that there was a dose dependent upregulation of *W08E12.2* upon Cd (10 µM, 20 µM) exposure. The fold change in *W08E12.2* expression was approximately the same for two of the biological repeats (**Figure 9-3A and C**), whereas the magnitude of increase was fivefold higher in one of the repeats (**Figure 9-3B**). Nevertheless, the overall pattern followed by the *W08E12.2* expression was the same for all the three biological repeats.

The *W08E(12.3-12.5)* expression in the metallothionein double knockout (*mtl-1(tm1770);mtl-2(gk125)*) increased upon acute Cd (24 hr) exposure (**Figure 9-3D**). There was a difference in the pattern of expression of *W08E(12.3-12.5)* between the biological repeats even though both displayed an upregulation of the cumulative expression of *W08E(12.3-12.5)* in response to Cd. Due to the differences between the biological repeats it was difficult to draw a conclusion about the expression of *W08E(12.3-12.5)*.

Chapter 10: Scientific output

1. **Poster presentation:** PhD Symposium 2012, Analytical and Environmental Science Division, King's College London

2. **Poster presentation:** PhD Symposium 2013, Analytical and Environmental Science Division, King's College London

3. **Poster presentation:** 19th *C. elegans* International meeting at UCLA, Los Angeles, 2013.

4. **Oral presentation:** PhD Symposium 2014, Analytical and Environmental Science Division, King's College London

5. **Poster presentation:** European Worm Meeting, Berlin, 2014.

6. **Paper in progress**, 2015.