TRANSCRIPTIONAL REGULATION OF *C. ELEGANS* EXCRETORY CELL-EXPRESSED GENES BY THE POU HOMEOBOX TRANSCRIPTION FACTOR CEH-6

by

Allan Mah
BSc, Simon Fraser University 2001

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

DOCTOR OF PHILOSOPHY

In the
Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

Spring 2009

APPROVAL

Name: **Allan Mah**

Degree: **Doctor of Philosophy**

Title of Thesis: **TRANSCRIPTIONAL REGULATION OF *C. ELEGANS* EXCRETORY CELL-EXPRESSED GENES BY THE POU HOMEOBOX TRANSCRIPTION FACTOR CEH-6**

Examining Committee:

Chair: **Dr. Timothy V. Beischlag**,
Associate Professor, Faculty of Health Sciences

_____

Dr. David L. Baillie,
Senior Supervisor
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Nicholas Harden,
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Barry M. Honda,
Supervisor
Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Jack N. Chen,
Supervisor
Associate Professor, Department of Molecular Biology and Biochemistry

_____

Dr. Bruce P. Brandhorst
Internal Examiner
Professor**,** Department of Molecular Biology and Biochemistry

_____

Dr. Bruce Bowerman,
External Examiner
Professor, Department of Biology
Institute of Molecular Biology, University of Oregon

Date Defended/Approved: _____

ABSTRACT

Caenorhabditis elegans' excretory cell is a large cell that extends the entire length of the nematode and is functionally analogous to a kidney. In order to develop a better understanding of genes involved in the development of the excretory cell, I have characterized a transcriptional regulatory mechanism involved in driving gene expression in this cell. Seven of the 13 vertebrate Aquaporins (AQPs; water channel proteins) express in the kidney where they assist in recapturing water lost because of renal filtration. There are also 12 *aqp*-encoding genes in *C. elegans* of which three express in the excretory cell. The expression pattern of *aqp-8* is limited to the excretory cell during post-embryonic developmental stages based on a GFP tagging assay. Analysis of 5' truncations of *aqp-8*'s promoter region, coupled with interspecies comparative analyses, revealed that an octamer DNA element (ATTTGCAT) is critical for driving excretory cell expression. The octamer element associates with POU homeobox proteins of which there are three in *C. elegans*. The class III POU transcription factor homolog, *ceh-6*, is an essential gene and is the only POU member with an expression pattern that overlaps with *aqp-8*'s. I have demonstrated, using both *in vitro* and *in vivo* approaches, that CEH-6 is the cognate transcription factor for the octamer motif. I have searched the genome for genes controlled by this transcriptional regulatory mechanism by locating interspecies conserved octamer motifs in gene-upstream regions. The candidate promoter regions were tested for their ability to drive expression in the excretory cell. I assessed the dependence on the *cis*-regulatory element for driving excretory cell expression by assessing the expression patterns of 5' promoter truncation constructs targeting the octamer motif. I have identified nine genes that are modulated by the CEH-6/octamer motif regulatory mechanism including the five: *M176.5* (a gene of unknown function), *sdr-2* (a ferric reductase), *F16F9.1* (a transcription factor), *twk-36* (a potassium channel), and *R02F2.8* (an amino acid transporter).

EXECUTIVE SUMMARY

Gene expression is modulated, both spatially and temporally via specific regulatory interactions. *Caenorhabditis elegans* is an ideal model organism for studying transcriptional regulation for several reasons including its sequenced genome, defined and invariant developmental cell lineage, and transparent body (for ease of observation). I characterized a transcriptional regulatory mechanism, which modulates the expression of genes, in the *C. elegans* excretory cell, the nematode functional equivalent of a kidney. The excretory cell is the largest cell in the nematode, consisting of two connected canals that run the entirety of the body forming an H-shape. The *C. elegans* excretory system is responsible for maintaining osmotic homeostasis and collecting metabolites for expulsion to the environment.

Aquaporins (AQPs) are transmembrane water channels, many of which are expressed in the kidney (7/13) where they aid in the recapture of water from the renal filtrate. I searched for an appropriate gene-upstream region for assessing excretory cell-modulating DNA element by determining the expression patterns of all 12 *aqps* encoded in the *C. elegans* genome. From this analysis, I have determined that *C. elegans aqp-8* is expressed exclusively in the excretory cell. I have discovered a DNA element (octamer motif) in the gene-upstream region of *aqp*-8 that is responsible for driving excretory cell expression. I identified the cognate *trans*-acting transcription factor that associates with the octamer element as CEH-6, a POU homeobox transcription factor that is associated with neuronal and renal tissue gene expression.

I searched the genomes of *C. elegans* along with the closely related nematode species, *C. remanei* and *C. briggsae*, to identify promoter regions that contain conserved octamer motifs. I assessed the expression patterns driven by these promoters and selected those which drove expression in the excretory cell in order to assess the effects of deleting the upstream octamer element. I found several genes that depend on the CEH-6/octamer motif regulatory mechanism for excretory cell expression.

ACKNOWLEDGEMENTS

I would like to thank my senior supervisor, Dr. David L. Baillie for his support and allowing me to have the opportunity to work in his laboratory. I would also like to thank members of my supervisory committee, Drs. Barry Honda, Nicholas Harden, and Jack Chen. I appreciate help provided by members of Dr. J. Chen's lab, especially Jeff Chu and Tao Luo for assistance in various methodologies towards the completion of this thesis. In the Baillie lab, I would like to thank Dr. Bob Johnsen for his encouragement during my early years of as a *C. elegans* geneticist-in-training. Much gratitude is given to Derek Chew, Keith Boroevich, Marco Gallo, and Victor Jensen for their exemplary volunteer work. I would also like to thank present members of the Baillie lab: Carrie Simms, ShuYi Chua, Heesun Shin, and Rylan Fernandes; and past members of the Baillie lab: Martin Jones and John Tyson.

Most of all, I am grateful for the help of Domena Tu for the construction of many strains integral to this project. In addition, without her encouragement, this thesis would have not been possible in the first place.

I would also like to acknowledge the funding of National Sciences and Engineering Research Council of Canada, David L. Baillie Graduate Fellowships in Molecular and Cellular Biology, Simon Fraser University's President's Ph.D. Research Stipend, and the Weyerhaeuser Molecular Biology Graduate Scholarship towards my graduate studies.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1: INTRODUCTION

1.1 Overview of *Caenorhabditis elegans* as a model organism for studies of metazoan transcriptional regulation

Transcriptional regulation is the coordinated and dynamic modulation of gene expression levels and patterns by regulatory factors that alter transcription rates. Transcription rates can be modulated indirectly in manners such as modifying chromatin structure and by histone modifications (for a review see (Li et al. 2004)). These types of changes influence accessibility to gene-promoters in a general way over a relatively large genomic region. Direct modulation of an individual gene's expression is regulated by the assembly of its transcription-initiation complex. This involves direct binding of sequence-specific *trans*-acting regulatory proteins to target DNA sequences. Transcriptional regulation of individual genes is a highly constrained process that ultimately controls a cell's ability to respond to external stimuli and controls the developmental programs of the cell.

Numerous studies have identified *trans*-acting factors and DNA sequences (*cis*-regulatory) involved in control of gene transcription in eukaryotes. Much of this work involved identifying factors governing general transcription, such as basal transcriptional machinery. One of these cognate *cis*-linked sites, the TATA box element, is usually located 25-30 bp upstream of most genes' transcriptional start site. The TATA box, together with other sites required for transcriptional initiation, are referred to as the core

promoter and is usually located close to the gene's translational start site (for a review see (Smale and Kadonaga 2003)).

In order to develop an in depth understanding of how a genome functions, it is essential to understand how gene transcription is regulated in specific tissues both temporally and spatially. These processes drive differential expression patterns that in turn lead to the differentiation of cell types and fates. Enhancer elements play a large role in this expression pattern specificity. Enhancer sequences are *cis*-acting elements that consisting typically short DNA fragments generally located close to a gene (although exceptions occur when the enhancer is located at a long distance). These sequences may increase transcriptional activity in a general or tissue-specific manner. Due to their short length and propensity to be located at variable distances from the translational start site; these elements can be difficult to locate within extensive genomic regions that are under low evolutionary constraints. Further adding to the difficulty of detection, the sequence of the *cis*-regulatory sequences may not be well conserved and may be orientation independent, therefore being more appropriately represented by a degenerate motif rather than a defined sequence. These situations make detection of these sequences difficult. Another problem is that the definition of tissue specific expression can be unclear as organs and tissues are composed of heterogeneous cell populations each with different developmental cues and expression patterns in themselves.

Caenorhabditis elegans, a multicellular eukaryote, is an ideal model organism for defining tissue-specific transcriptional regulatory mechanisms. The specific advantages, of *C. elegans* as opposed to other eukaryotes with more complex body plans and development include:

1. The entire developmental lineage of *C. elegans* from the fertilized embryo to gravid adult has been determined (Sulston et al. 1983). Their developmental lineage map provides the time of cell division, position, and identity of each cell, which ultimately lead to the adult set of 959 somatic cells. In addition, the nematode undergoes discrete stages during development, beginning with embryonic divisions leading to four larval stages (L1 to L4) and finally ending with the adult worm (edited by (Riddle et al. 1997)).

2. The variance of cell division timing during *C. elegans* development is very small as demonstrated using computer assisted developmental lineaging (Zhao et al. 2008).

3. *C. elegans* is the first multicellular organism to have its entire genome sequenced (Consortium 1998). The relatively compact 100,291,140bp genome contains sequences corresponding to 19,735 protein coding genes (Hillier et al. 2005). Of these genes, almost 5% encode transcription factors proteins (Reece-Hoyes et al. 2005). The relatively small size of the genome results in a high gene density. Consequently, the intergenic regions are short compared to many other organisms including humans, which has a thirty-fold larger genome with a similar number of genes (Genome sequence of the nematode C. elegans: a platform for investigating biology 1998; Lander et al. 2001; Venter et al. 2001; Consortium 1998). Due to *C. elegans*' small intergenic spacing, the density of functional DNA elements is presumably higher than in humans, or alternatively, the number of functional DNA elements in *C. elegans* is less than that of humans (which have more transcription factors encoded in their genome).

4. Analysis of gene expression in *C. elegans* has been greatly facilitated by the use of fluorescent reporter genes such as green fluorescent protein (GFP) (Chalfie et al. 1994) a protein originally isolated from the jellyfish *Aequorea victoria* (Prasher et al. 1992). To document expression patterns for individual genes, the genes' promoters were used to drive the expression of the reporter. Since GFP is not harmful to *C. elegans*, gene expression is observed in live worms during development in real time. GFP can also be fused in-frame to coding regions to allow the detection of the intracellular localization of the resulting chimeric proteins. These constructs are introduced into *C. elegans* using microinjection to introduce the exogenous DNA constructs into the distal arm of the gonad. This region of the gonad is comprised of a syncitium, which contains cytoplasm shared by many germ cell nuclei. The injected material freely mixes in the syncitium and is taken up by the naked nuclei. This method allows for transformation of many progeny from a single injection (Mello et al. 1991). The result of this method of transformation is the generation of large extrachromosomal DNA arrays in the worms.

5. RNA interference (RNAi) in *C. elegans* is a commonly employed technique to assay the effects of gene-expression knockdown. RNAi was pioneered in *C. elegans* by Andrew Fire and Craig Mello (Fire et al. 1998). It has become a pervasive technique in molecular biology involving the introduction of double-stranded (ds) RNA, corresponding to the sequence of a gene target, into organisms. The corresponding gene's mRNA is degraded thereby knocking down gene expression at a transcriptional level. Specifically, the dsRNA is cleaved, by a

type III endonuclease named Dicer, into small-interfering RNAs (siRNA) that are 21-23 nucleotides (nt) long (Bernstein et al. 2001). These siRNAs associate with the RNA-induced silencing complex (RISC) and act as a guide to the associated mRNA, which is then cleaved by RISC. Also, secondary siRNAs can be synthesized from the mRNA by RNA directed RNA polymerase (RdRP) in *C. elegans* and in turn can amplify the RNAi response ((for a review see, (Grishok 2005)). The first advantage is that the dsRNA can be introduced using a variety of methods including injection of dsRNA, soaking in dsRNA, or by feeding the worms dsRNA-expressing bacterial strains. The second advantage is that the RNAi effect is systemic (Fire et al. 1998). The third advantage is that the RNAi signal is amplified (Sijen et al. 2001).

6. The *C. elegans* genome has 934 predicted transcription factor coding sequences (Reece-Hoyes et al. 2005). This is a significantly smaller number of potential candidate transcription factors than the 1,962 transcription factor genes predicted in the Human genome (Messina et al. 2004).

7. To complement *C. elegans*' sequenced genome, several closely related nematodes have had their genomes sequenced including: *Caenorhabditis briggsae* (Stein et al. 2003), *Caenorhabditis remanei*, *Caenorhabditis japonica* and *Caenorhabditis brenneri*. The additional information provided by these complementary nematode genome sequences provide a valuable resource for identifying short conserved sequences buried within regions composed of less-conserved DNA sequences by using comparative methods.

The sequenced genome also facilitates the design of *promoter::reporter* constructs for predicted gene models. Gene-expression patterns derived from *promoter::reporter* transgenes in living nematodes are assayed by light microscopy coupled with fluorescence imaging. This is feasible due to the nematode's transparent body plan. The ease of determining expression patterns greatly aids the detection of genomic regions implicated in controlling gene expression.

Identification of *C. elegans cis*-regulatory elements has been performed for some gene-upstream regions. Examples of these studies include: the VPE1 and VPE2 elements upstream of *vit-2* (MacMorris et al. 1992), a short *cis*-element upstream of *dpy-7* (Gilleard, Barry, and Johnstone 1997), the PHA-4 binding site upstream of pharyngeally expressed genes (Gaudet and Mango 2002), the EX-1 motif upstream of *pgp-12* (Zhao et al. 2005), the X-box motif located upstream of genes expressed in the ciliated neurons (Swoboda, Adler, and Thomas 2000), the ASE motif found upstream of ASE (gustatory neuron) expressing genes (Etchberger et al. 2007), and the PGM1 motif upstream of pharyngeal gland expressing genes (Smit, Schnabel, and Gaudet 2008). In many of these prior cases, the cognate transcription factors have also been identified.

1.2 The *C. elegans* excretory cell as a model system for renal development

Nematode excretory systems are composed of a small number of cells. The compact size reflects the low complexity of the nematode body plan. Nematodes lack a circulatory system, but the fluid in the pseudocoelomic space carries out circulatory fluid function. The pseudocoelomic space separates the body wall from the inner tissues (alimentary system, gonad). The pseudocoelomic fluid provides hydrostatic pressure in the worm in

order to maintain body turgor, which is imperative for nematode mobility. The pseudocoelomic fluid is important for nutrient transport. This is due to its contact with all major cells and therefore is an important component for the maintenance of osmotic balance in each tissue system (Bird and Bird 1991).

The excretory system in *C. elegans* is comprised of only four cells. Each cell is descended from the AB cell lineage (Sulston and Horvitz 1977). The excretory canal cell is an H-shaped cell with four arms running anteriorly and posteriorly along the full length of the body with each arm tapering at the extremities. All four arms join at the cell body located ventral to the pharynx (Figure 1). Each of the fluid-filled arms is located within the pseudocoelomic space. Ultra structure analysis of the excretory cell shows that the canal cytoplasm contains high concentrations of mitochondria (Nelson, Albert, and Riddle 1983). This high metabolic demands of this tissue is consistent with the hypothesis that the canals are involved in active transport of substrates (Nelson and Riddle 1984). The arms of the excretory cell contain dead-end pores, known as canaliculi, which are continuous with the lumen of the excretory cell. The proposed function of the canaliculi is to increase the surface area of the excretory cell lumen. An excretory sinus is located on the excretory cell body. The excretory sinus is composed of a cluster of channels open to the excretory cell lumen. These channels provide a conduit for the removal of excretory cell contents. The excretory sinus connects to the excretory duct cell via tight junctions (Nelson, Albert, and Riddle 1983).

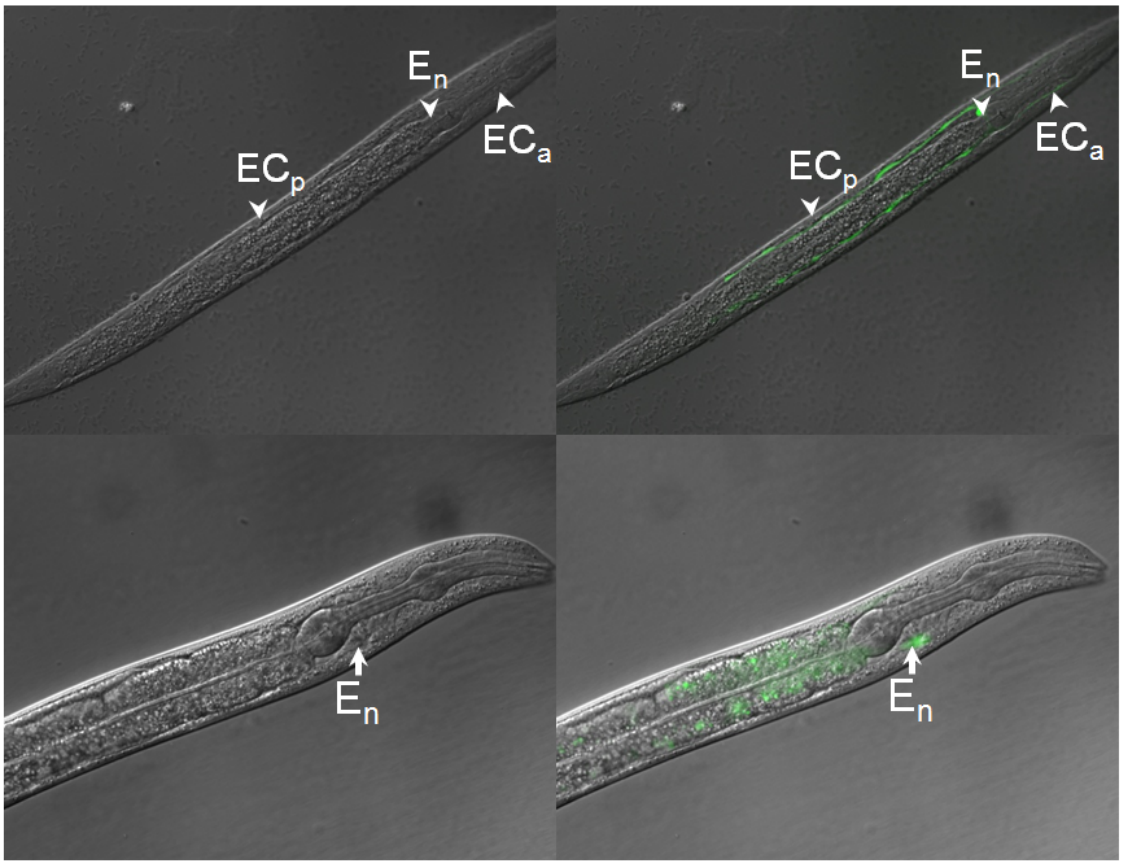

Figure 1. Location and structure of the excretory cell.

Top (*aqp-8*^{promoter} *::GFP*) – ventral/dorsal view of *C. elegans* showing the H-shaped structure of the excretory cell. The posterior canals (EC_p) and anterior canals (EC_a) meet at the cell body, which contains the nucleus (E_n).
Bottom (*Y19D10A.4*^{promoter} *::GFP*; contrast was increased in the GFP image for EC localization purposes) – side view of *C. elegans*. The excretory cell is typically ventral to the second pharyngeal bulb, but changes positions slightly during development as seen in the images on the top of a mid larval worm (posterior to second pharyngeal bulb) and the images on the bottom with the position moving slightly anterior of the second pharyngeal bulb.

The excretory duct cell associates with the excretory pore cell, a specialized hypodermal cell. The excretory pore cell encloses the terminal third of the excretory duct cell. The pore cell forms a base upon which the duct cell contacts the external environment. Measurement of excretory system function can be assayed indirectly via the observation of its pulse rate. By observation of dauer stage embryonic worms, an alternative post-L1 larval stage triggered by adverse environmental conditions such as over-crowding and/or low food levels, fluid expulsion is seen as swelling of the duct cell and releasing fluid through the pore upon relaxation (Nelson, Albert, and Riddle 1983). This pore cell pulsation activity has also been observed in other nematodes (P.P. 1952) .

The excretory gland cell is a binucleate cell, which associates with the excretory system by connecting to the excretory cell and the origin of the excretory duct. The high density of mitochondria in the excretory gland cell supports the hypothesis that it is required for the synthesis of a large amount of material. This hypothesis is also supported by the observation that the gland cell of the nematode *Phocanema decipiens* secretes enzymes with biological activity (Davey and Kan 1968).

Independent laser ablation studies of each of the four excretory system cells revealed that loss of the pore cell, duct cell, and canal cell lead to excess fluid accumulation in the form of sub-hypodermal vesicle-like particles (Nelson, Albert, and Riddle 1983). Ultimately, the effects of independent ablations of these excretory system cells, at various larval stages, are low fecundity and eventual premature death of the worm. This experiment supports the idea that the system is required for the removal of

excess waste fluids. Ablation of the excretory gland cell, however, does not affect survival or fecundity (Nelson and Riddle 1984).

The nematode excretory system is required for the maintenance of osmotic balance, removal of metabolic waste, secretion of fluid required for molting, and secretion of hormones (Nelson, Albert, and Riddle 1983). Indeed, this function has been demonstrated in *Trichinella spiralis* by showing that ammonia is a major component of the excreted fluid (Haskins and Weinstein 1957). The osmoregulatory function is particularly important in nematodes because the worms must counteract the effects of ever changing osmotic pressure in their native soil environments. The excretory system must be efficient to allow the worm to promptly adapt to changes in response to the environment's osmotic pressures. This adaptability is important for the worm to maintain internal turgor. The removal of environmentally acquired substances has been demonstrated in ascarid species by showing that nematodes rid themselves of environmentally acquired dyes via expulsion of the accumulated dye through the excretory duct. Although the excretory system is small and relatively simple, its morphology has variations among different nematode species. Even though the system is morphologically variable, the basic functions remain similar.

Excretory cell development begins at approximately 270 minutes after the first cellular division in *C. elegans* development. The cell at this stage is located approximately at the centre of the embryo. The developing excretory cell sends two processes dorsolaterally and subsequently each of the two processes branch canals which extend towards the anterior and posterior of the embryo. Upon hatching, the posterior canals are about half the length of the emerging L1 larva (Buechner 2002). By the end of

the first larval stage, the excretory cell is at its full length in relation to the worm (Buechner 2002). It has been suggested that developmental cues dictating the elongation and guidance of the excretory cell are shared with mechanisms involved in neuronal guidance and outgrowth (Buechner et al. 1999). In addition to outgrowth cues, excretory cell development is influenced also by its attachment to the hypodermis via gap junctions along the entire length of the cell and by anchoring of the canals to the basement membrane. It is via this physical association to the hypodermis that the excretory cell continues to extend its canals after the first larval stage (Buechner 2002).

Mutations of genes that affect various stages of excretory cell development have been identified. UNC-34, a protein required for early embryonic cell migrations,is required for initiation of the dorsolateral processes from the excretory cell body towards the left and right (Shakir, Gill, and Lundquist 2006). During extension of the four canals UNC-6 (netrin) and UNC-5 (netrin receptor in the immunoglobulin superfamily) are required for guidance of canal tips in addition to their roles in patterning of longitudinal nerves (Ren et al. 1999; Quinn et al. 2006). UNC-53 is required for anterior and posterior neuronal and excretory canal migrations. Loss of this protein leads to shortened canals; however, over-expression of this protein in muscle cells leads to elongated muscles in the anterior posterior dimension (Stringham et al. 2002). LIN-17 is required to stop elongation of the excretory cell canals posteriorly. In *lin-17* mutants, excretory canals over-elongate (Hedgecock et al. 1987). Maintenance of the tubular shape is also important in excretory cell morphology. 12 genes, labelled as the *exc* genes, have been identified that have mutant phenotypes that lead to excretory cell canals with various

degrees of canal defects ranging from distended canal lumens to engorgement of certain portions of the lumen (for a review see (Buechner 2002)).

Many genes affecting excretory cell morphology are not necessarily expressed in the excretory cell itself. A secreted mucin, *let-653* (Jones and Baillie 1995) is expressed in numerous tissues, but absent in the excretory cell (Dupuy et al. 2007; Hunt-Newbury et al. 2007). The phenotypes of *let-653* mutant alleles are lethal arrest between L1 and L2 larval stages coincident with a large vacuole corresponding to the position of the excretory cell signifying a defect in the duct or pore cell. The suggested role for the mucin is to protect tissues that have contact with the outside environment (Jones and Baillie 1995). The requirement of LET-653 for excretory cell development and function suggests that there may be other extracellular requirements that affect the development and morphology of the excretory cell.

Excretory cell development is highly specialized and complex sharing many mechanisms required for neural development. The dedicated physiological function of the excretory cell as an osmoregulatory organ, which removes aqueous waste, is consistent with the suggestion that the cell is the closest nematode functional equivalent to the kidney, a much more complex organ. Therefore, along with results from prior studies of the excretory cell, understanding genes controlled in the excretory cell will not only provide information regarding mechanisms involved in kidney function, but will also aid in the understanding genes involved in the function of requirements of neurons types that require similar growth and functional mechanisms.

In addition to the excretory system, the canal-associated neurons (CANs), the intestine, and the hypoderm are important in maintaining fluid homeostasis. The CANs

are closely associated with the excretory cell. Ablation of the CANs leads to a clear phenotype which is the result of the accumulation of fluid in the pseudocoelomic cavity, a phenotype similar to the one caused by ablating cells in the excretory system (Forrester and Garriga 1997).

1.3 Prior research on transcription regulation in the excretory cell

Previously, a study was carried out to determine one component of transcriptional regulation in the excretory cell. This study was based upon the ABC transporter gene, *pgp-12* (*P-GlycoProtein related*), which expresses in the excretory cell during all stages of development from embryo throughout to adulthood (Zhao et al. 2005). The study describes the identification of a novel DNA element, Ex-1, which is located upstream of *pgp-12*. Ex-1 was found to be responsible, at least in part, for *pgp-12*'s expression pattern. The cognate transcription factor for this *cis*-regulatory element is DCP-66/C26C6.5. The only prior attributed function for DCP-66 was as a transcriptional repressor, being the *C. elegans* homolog of a nucleosomal remodeling and deacetylation (NuRD) complex component. In *C. elegans*, the role of the NuRD complex is consistent with its role in other species as a gene expression repressor. For example, the NuRD complex represses the expression of *lag-2* in *C. elegans* (Poulin et al. 2005).

The Ex-1 motif was unable to drive expression ectopically when placed upstream of the Δ*pes-10* basal promoter (Kelly et al. 1997). To drive expression in an ectopic manner, additional downstream endogenous elements in the *pgp-12* promoter region must be required. Ex-1 is directional because it does not drive expression in its reverse complement direction (Zhao et al. 2005).

Other studies regarding excretory cell expression is the characterization of CEH-6, a transcription factor expressed in the excretory cell and also expressed in cells in the nervous system and epithelia (Burglin 2001, Reece-Hoyes 2005). CEH-6 is a member of the POU homeobox transcription factor family (derived from the founding members of the protein class: Pit-1, Oct-1, Oct-2, and UNC-86 (Herr et al. 1988)). In a study by Burglin and Ruvkun (2001), it was demonstrated that *ceh-6* mutants display phenotypes similar to *let-653* mutants.

1.4 Aquaporin function and expression patterns

CHIP28 (channel-like integral membrane protein 28) was originally identified as a integral membrane protein component of the lipid bilayer of red blood cells (Preston and Agre 1991). CHIP28 was classified as a member of the Major Intrinsic Protein (MIP) family of proteins. MIPs are transmembrane proteins that passively allow the passage water or small neutral solutes while excluding the passage of charged ions and non-selected solutes (Preston and Agre 1991). The insertion of CHIP28 into the lipid bilayer membrane of cells with low natural water permeability greatly enhanced those cells' water permeability (Preston et al. 1992). CHIP28 was subsequently renamed Aquaporin 1 (AQP1) to reflect its function. In general, AQPs (then nomenclature can be used interchangeably with MIPs) contain six transmembrane domains. The six trans-membrane structure of modern MIPs arose via duplication of the coding region of an ancestral protein form that had three transmembrane domains. Evidence of this duplication event can be seen upon alignment of the two halves of the proteins and observing the homology between the sequences (Pao et al. 1991; Reizer, Reizer, and Saier 1993). AQPs form membrane protein pores that facilitate the flux of water

molecules across cellular membranes. Although there are 12 known classes of MIPs/AQPs, animal AQPs can be divided into two general classes based upon their ability to transport different substrates (reviewed by (Park and Saier 1996)). The aquaporins only facilitate the passage of water whereas the aquaglyceroporins facilitate the passage of water and/or other uncharged solutes respectively (Froger et al. 2001).

AQPs are found in almost every organism and in almost every cell, likely due to their important function of facilitating water movement across the lipid bilayer membrane. For example, *Escherichia coli* has two MIP family proteins, AqpZ which permeates water, and GlpF which permeates glycerol but not water (Borgnia et al. 1999). *Saccharomyces cerevisiae* has four MIP proteins encoded in its genome. Plants have many MIP family proteins; for example, there are 23 MIP family proteins encoded in the *Arabidopsis thaliana* genome (Weig, Deswarte, and Chrispeels 1997). Thirteen AQPs (AQP0-AQP12) have been identified in mammals, and these are distributed in most tissues with higher concentrations of these proteins in water-transporting epithelia and endothelia of a variety of tissues ((for a review see, (Wang et al. 2006; Echevarria and Ilundain 1998; Yamamoto and Sasaki 1998)). Seven out of the 13 AQPs in humans are expressed in the kidney where they aid in optimizing water recapture from renal filtrate thereby maintaining osmotic homeostasis.

Physiological studies on the selectivity and permeability of individual AQP proteins have been facilitated by exogenously expressing the AQP proteins in *Xenopus laevis* oocytes and studying the resultant cell permeability. This *Xenopus laevis* oocyte expression system is the standard for addressing pore permeability properties due to the ease of generating such expressing oocytes and the low natural water permeability of the

oocyte (Preston et al. 1992). Another advantage of using *Xenopus* oocytes is due to their large size (diameter 1.0-1.3 mm) and robustness, providing an excellent platform for introducing mRNA or other substances. AQPs contain two well-conserved hydrophobic asparagine-proline-alanine regions referred to as NPA boxes. The NPA boxes are arranged on opposite sides of the lipid bilayer in the mature AQP protein. These NPA boxes invade the lipid bilayer to form the pore constriction, which permits selected substrate passage through the membrane; this structure is referred to as the hourglass model (Jung et al. 1994). Substrates pass through the pore in a single file manner because the pore in AQP is not large enough to accommodate more than a single water molecule in most regions along the channel (Jensen, Tajkhorshid, and Schulten 2003). The single file passage of substrates is important for the pore's inherent selectivity (Fu et al. 2000).

The NPA boxes have been used to find genes that encode AQPs and as such are considered a hallmark characteristic of these proteins. Peptide sequences flanking the two NPA boxes are also conserved, but to a lesser degree than the boxes themselves. An alignment of thirty-six AQPs, from different organisms, shows that the first NPA box can be represented by the motif SG(A/G)HXNPA and that the second NPA box can be represented by the motif NPAR(S/D/A) (Ishibashi 2006). Although the NPA boxes are particularly well conserved in AQPs, there are a few exceptions. AQPs which contain divergent NPA box regions are localized intracellularly and therefore designated as sub-cellular aquaporins (Ishibashi 2006). Passage of water through AQP11, a mammalian sub-cellular aquaporin, has been proven (Yakata et al. 2007).

In addition to their roles in transporting aqueous substrates, AQPs have been implicated in the transport of dissolved gases in plants and animals (Endeward et al.

2006) (Hanba et al. 2004), intercellular communication (Bok 1982), and the passage of anions by acting as chloride channels (Yasui et al. 1999; Hazama et al. 2002)

Research on gene knockouts in mice has shown that active fluid transport in kidney proximal tubules and salivary glands is seriously compromised by AQP5 and AQP1 deletions respectively (for a review see (van Os et al. 2000)). Impaired functions of AQP0 and AQP2 have been directly linked to cataracts and diabetes insipidus, respectively (Agre, Bonhivers, and Borgnia 1998; Borgnia et al. 1999; Deen and van Os 1998; Deen et al. 1994).

1.5 Prior research on aquaporins in *C. elegans*

The *C. elegans* genome contains 12 *aqp* genes (Consortium 1998) (Table 1). Studies have been performed to determine the physiological properties of AQPs 1-8 in *C. elegans* (Huang et al. 2007; Kuwahara et al. 2000; Kuwahara et al. 1998). Exogenous expression of the eight AQPs in *Xenopus* oocytes was performed to assess their substrate specifities. AQP-2 and AQP-4 is permeable to water and not urea or glycerol (Kuwahara et al. 1998; Kuwahara et al. 2000). Huang repeated the experiments for AQP-2 and AQP-4 and further determined the permeabilities of the AQPs (1, 3, 5, 6, 7, and 8).

Huang *et al.* (2007) have determined, using comparative analysis along with functional data, that AQP-4, AQP-5 and AQP-6 are members of the aquaporin family and that AQP-1, AQP-2, AQP-3, AQP-7, and AQP-8 are members of the aquaglyceroporin family. Moreover, they determined the expression patterns corresponding to *aqps (1-8)* using a combination of transcriptional and translational reporter constructs in *C. elegans*.

Gene locus	Common name	Protein / transcript size	Amino acids	Chromosome	Gene start	Gene end
F32A5.5a	AQP-1	915/2515 bp	304	II	7235774	7238289
C01G6.1a	AQP-2	873/2930 bp	290	II	9258835	9261916
Y69E1A.7	AQP-3	1266/2461 bp	421	IV	10961325	10963785
F40F9.9	AQP-4	822/6612 bp	273	V	9737563	9744202
C35A5.1	AQP-5	873/1098 bp	290	V	10492083	10493180
C32C4.2	AQP-6	735/1088 bp	244	V	10650708	10651795
M02F4.8	AQP-7	876/1351 bp	291	X	3025528	3026878
K02G10.7a	AQP-8	777/1930 bp	258	X	4702937	4704939
K07A1.16	AQP-9	735/2675 bp	244	I	9606253	9608928
ZK1231.3.a	AQP-10	843/1403 bp	280	II	9768387	9769774
ZK525.2	AQP-11	840/1400 bp	279	III	13673158	13674625
Y57A10A.35	AQP-12	732/5788 bp	243	II	12203224	12209011

Table 1. Aquaporins in *C. elegans*

Information retrieved from Wormbase WS190

1.6 Overview of POU homeobox structure and function

Homeobox genes encode transcription factors that are required for normal development. Homeobox is derived from the word "homeosis," the definition of which, is the transformation of a body structure into the homologous structure of another segment (for a review see (Gehring and Hiromi 1986)). Homeobox genes were originally identified in *Drosophila melanogaster* as mutants that transformed a body structure into a homologous body structure (Lewis 1978). An example of this effect, are the phenotypes arising from mutant alleles of the *antennapedia* homeobox gene in *D. melanogaster.* Antennapedia is a member of the HOX subfamily of homeobox protein which is Antennapedia is responsible for regulating the developmental decisions of the appendages (Munke et al. 1986). Loss of function mutations leads to the ectopic development of antennae at the position of the second leg pair. Conversely, a gain of function mutant will lead to the development of legs at the antennae positions ((for a review see, (Gehring 1987)).

Homeobox genes have been found in all animals where they are important as regulators of genes during development in processes such as patterning, differentiation and regional specification (for a review see (Duverger and Morasso 2008)). The proteins contain a conserved 60 amino acid motif referred to as the homeobox domain or homeodomain ((for a review see, (Ruvkun and Finney 1991)). Among the typical classes of homeodomain proteins, the POU homeobox genes are the most distant evolutionarily from the HOX class of proteins (Banerjee-Basu, Sink, and Baxevanis 2001). The POU homeobox transcription factor sub-family was originally identified as homeodomain containing genes that had an additional 150 to 160 amino acid long region known as the POU domain. The POU domain is named after the first identified members of these

proteins: Pit-1, Oct-1/2, and UNC-86 (Herr et al. 1988). The POU domain can be separated into two sub-domains. The POU-specific ($POU_{S)}$ sub-domain is located towards the N-terminal end of the POU domain while the POU-homeodomain (POU_{HB}) sub-domain is located near to the C-terminal end of the POU domain. While it is clear that the POU_{HB} sub-domain is derived from the homeodomain motif, the POU_S domain resembles the DNA-binding domains of bacterial transcription factors such as lambda repressor and thus might be virally or bacterially derived (Assa-Munt et al. 1993). Since these proteins can interact with TATA-box binding protein (TBP) directly, in a DNA-binding independent manner, they may play a role in directly stimulating transcription in cases where the POU binding site is within close proximity to the TATA-box associated transcriptional start (Zwilling, Annweiler, and Wirth 1994).

The two POU sub-domains are separated by a short flexible sequence linker region which is long enough to facilitate binding of the two modules in a bipartite manner to opposite sides of the DNA helix (Phillips and Luisi 2000). However each sub-domain, which interacts with four or five residues of the target sequence each, have low binding affinity alone (Herr and Cleary 1995). The consensus binding targets for POU_S and POU_{HB} are gAATAT(G/T)CA and RTAATNA respectively (Verrijzer et al. 1992).

The two halves of the POU motif co-operatively contact the target DNA consensus sequence conforming to 5'-ATGCAAAT-3' (octamer element). When the two sub-domains are bound to their targets in concert, the result is a co-operative high binding affinity. Site-directed mutagenesis studies which alter residues either in the POU_S or the POU_{HB} lead to dramatic decreases in DNA-binding affinity (Ingraham et al. 1990). For example, deletion of the POU_S sub-domain in Pit-1 (a pituitary specific POU TF) leads to

a 1000-fold decrease in Pit-1/binding motif affinity (Ruvkun and Finney 1991). Although both the POU_S and the POU_{HB} are both required for DNA contact (Verrijzer et al., 1990), the POU_S domain has a larger role in binding site affinity. This was demonstrated by replacing the POU_S domain in Pit-1 with the POU_S domain of Oct-1, another POU homeobox transcription factor that has a higher affinity to a derivative octameric binding motif (consensus of $(A/T)_4TNCA$). The sub-domain replacement led to an increase of the chimeric Pit-1's affinity to the octameric DNA sequence. However replacement of Pit-1's POU_{HB} with the POU_{HB} domain from Oct-1 did not lead to any changes in affinity to the octamer site (Ingraham et al. 1990). The incorporation of two DNA-binding sub-domains in a single protein enables POU transcription factors to have a large degree of target recognition sequences (Banerjee-Basu, Sink, and Baxevanis 2001). The fact that each sub-domain has an independent target sequence supports this target promiscuity. Even though each POU sub-domain has an independent binding site and the two domains do not interact, increasing the separation of the domains regions by changing the linker region length results in a decrease in the ability of the transcription factor to bind (van Leeuwen et al. 1997).

Although many of the POU proteins have similar binding sequences, they activate distinct sets of target genes. This ability to recognize different targets with same recognition sites is possible via differential tissue expression or association with different accessory proteins to further modulate their specificity or activity. The mammalian POU transcription factors Oct-1 and Oct-2 have high sequence identity. However, they activate genes in a differential manner. Their POU homeodomains are 87% identical, but Oct-1

activates histone and snRNA genes in a ubiquitous manner, whereas Oct-2 drives the expression of immunoglobulin genes in B-cells (Tanaka and Herr 1990).

The earliest studied POU transcription factor in *C. elegans* is UNC-86. This protein has been demonstrated to be required for expression of genes required for mechanosensation, chemosensation, thermosensation, mobility, and the ability to lay eggs (Chalfie and Au 1989; Finney, Ruvkun, and Horvitz 1988; Hodgkin, Horvitz, and Brenner 1979; Mori and Ohshima 1995; Ward et al. 1975). UNC-86 is active during the development of some neuroblast cell lineages where it is required for the differentiation of daughter cells from the mother cells (Finney and Ruvkun 1990). In the adult worm, UNC-86 is expressed in 57 neurons. This represents about a fifth of the cells in the worms 302-cell nervous system (Finney, Ruvkun, and Horvitz 1988).

Although UNC-86 is expressed in a large number of cells, it has been suggested that UNC-86 interacts with accessory transcription factor proteins in a combinatorial manner to modulate the expression of different target genes in different cells. For example, UNC-86 is responsible for driving expression of *mec-3*. MEC-3 is expressed in five of the 27 cell types that UNC-86 is expressed in, thus either, *mec-3* expression is repressed in the remaining cells or UNC-86 requires other factors in a combinatorial manner to drive expression in some cell types.

1.7 Thesis overview

In this thesis, I address the process of identifying a transcriptional regulatory mechanism involved in driving expression of genes in the nematode excretory cell and to other genes that are driven to express in the excretory cell by the same system.

In chapter two, I discuss the analysis of the expression patterns derived from promoter driven GFP constructs for all *C. elegans* aquaporins. From the expression pattern analyses, I selected suitable promoter regions for further analysis for determinants of excretory cell expression. I performed a preliminary analysis, determining the distribution of the AQPs into functional sub-classes to provide a link between aquaporin type and their range of tissue expression patterns.

In chapter three, I describe the identification of a putative *cis*-regulatory element upstream of the start codon of *aqp-8*. I generated and analyzed sequential 5' truncations of *aqp-8*'s promoter containing region and identified a restricted upstream section involved in excretory cell expression. The restricted section was then compared to orthologous regions in other nematodes to determine conserved regions within the functionally restricted section. The function of the only nested conserved element was characterized in transgenic worms containing transgene constructs with targeted mutations in conserved residues and determining the transgenes' ability to drive expression. By using this strategy, I show that there is a conserved region in *aqp-8*'s promoter that is responsible for driving expression in the excretory cell.

In chapter four, I first confirmed, by using electrophoretic mobility shift assays (EMSAs), that the putative *cis*-regulatory element binds proteins in a specific manner. I then determined that the *cis*-regulatory element is able to drive expression ectopically via

the generation and analysis of expression patterns resulting from chimeric promoter::GFP transgene constructs. The conserved element was then used as a database query to find suitable cognate transcription factors. I determine that CEH-6 is the cognate binding protein for the motif, now recognized as the octamer motif, via a supershift EMSA using CEH-6 specific antibodies.

In chapter five, I determine other genes that are regulated by this CEH-6/octamer-based mode of transcriptional regulation. There are almost two thousand genes with putative upstream octamer elements (within 1,200bp of the translational start site). From this set, many genes have had prior expression pattern analysis and are expressed in the excretory cell. Selecting a set of these genes for further analysis of their dependence on the octamer sequence for excretory cell expression turned out to be an inefficient means for determining genes that are regulated by CEH-6/octamer. A more direct approach was adopted. Genes with conserved upstream octamer sequences were identified and targeted for expression pattern analysis. Octamer elements, in promoters, which drove excretory cell expression, were tested to determine if they were necessary for gene-expression.

In the final chapter, I tie together information gathered from my research and discuss future applications that can be derived from this work.

2: ANALYSIS OF AQUAPORINS IN *C. ELEGANS*

2.1 Introduction

Aquaporins are membrane proteins which facilitate the passage of water and other small solutes across the cell lipid bilayer in a passive manner. The channel formed by the AQP may allow for passage of water molecules only or also permit the movement of other selected uncharged small solutes such as glycerol depending on the properties of residues adjacent to the pore. These proteins are typically composed of six transmembrane spanning regions with their characteristic NPA domains situated on opposite sides of the membrane (Jung et al. 1994).

Due to the universal requirement of water among organisms, AQPs are required for the flux of water in almost all cell and tissue types. Even with this universal requirement for water transport among cells, AQPs are expressed in tissue specific manners. Seven of the 13 known mammalian aquaporins are expressed in the various parts of the nephron in the kidney where they function to co-operatively recapture water lost from renal filtrate thereby preventing excessive water loss.

12 AQP-encoding genes have been identified in the *C. elegans* genome (WormBase WS190). Orthologs for each of the *C. elegans* AQP are found in the two related nematodes, *C. briggsae* and *C. remanei*. The genes are distributed across all six genetic linkage groups with *aqp-4, aqp-5* and *aqp-6* grouping closely together on *LGV*

(Figure 2) (Consortium 1998). The *C. elegans* AQPs range in size from 243aa for AQP-12 to 421aa in the case of AQP-3. Studies of the phenotypes resulting from RNAi against each of the 12 *aqp*s result in no obvious phenotypes in all cases (Kamath et al. 2003; Sonnichsen et al. 2005; Rual et al. 2004). In addition, knockout alleles of some *aqp* members have not produced visible phenotypes (Huang et al. 2006). The lack of phenotypic consequences from RNAi and gene-knock out mutants for most of the *aqp*s suggest that many of the nematode AQPs are either not required or are functionally redundant in nematodes raised in laboratory conditions.

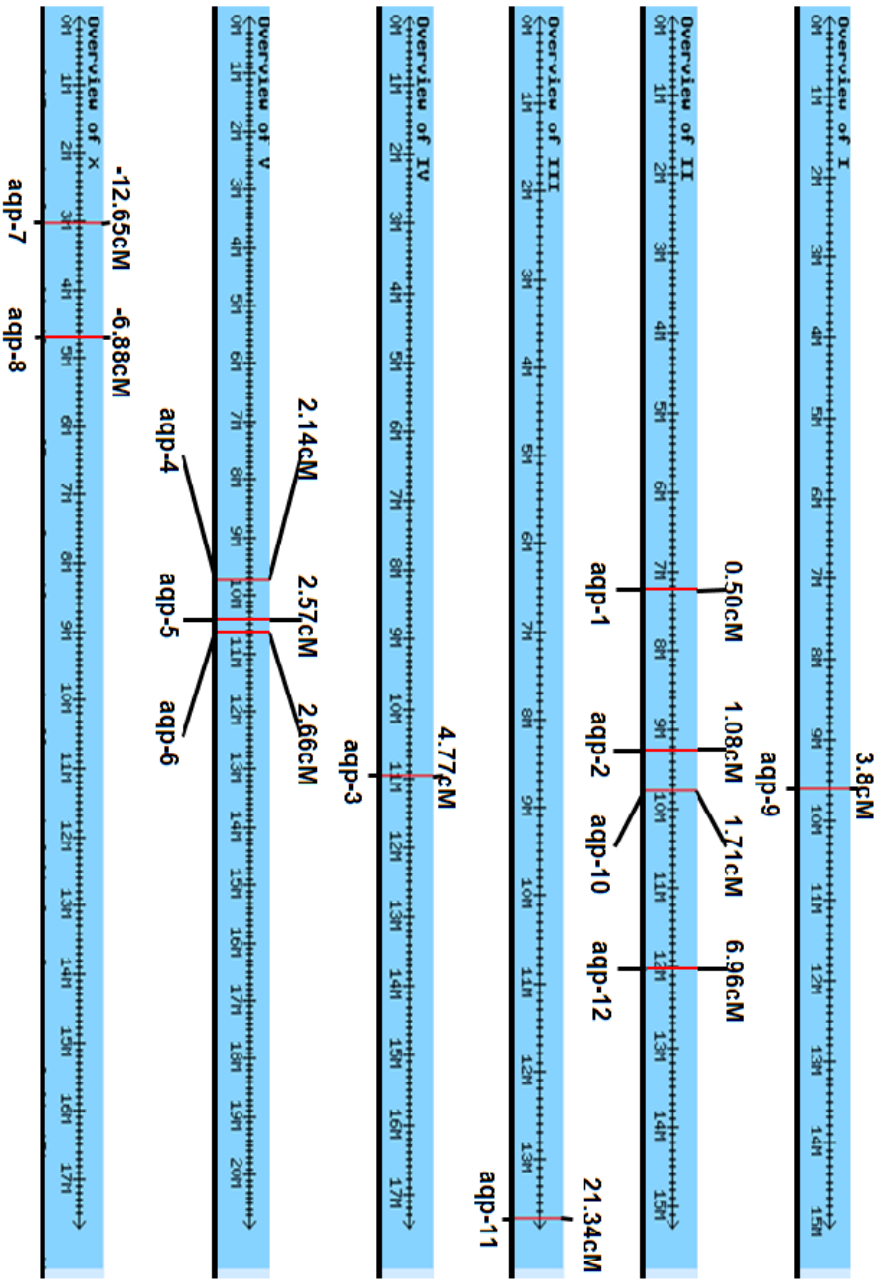

Figure 2. Genomic distribution of aquaporins and their approximate position on the genetic map.

Adapted from WormBase, WS190.

C. elegans AQPs-(1-8), have been designated as the canonical aquaporins (Huang et al. 2007). The AQPs were further classified into the aquaporin or aquaglyceroporin subfamilies of AQPs based upon evidence gathered by physiological studies and protein sequence comparisons (Huang et al. 2007) (Refer to Table 1). However, AQP-9, AQP-10, AQP-11, and AQP-12 were not included in this prior analysis. I provide preliminary evidence that these proteins form their own distinct aquaporin sub-class, by including these four "non-canonical" AQPs in a sequence comparison with the rest of the AQPs, based upon the structures of their NPA boxes.

Huang et al. (2007) have also identified expression patterns for *aqp*s-(*1-8*) in *C. elegans* using either transcriptional or translational GFP reporter fusions. I concurrently generated *promoter::GFP* constructs for all 12 *C. elegans* gene members and studied their spatial and temporal expression patterns independently. I found differences in expression patterns of *aqps* derived from the constructs made independently between our two groups. Moreover, I studied the expression patterns for the four AQP members not analyzed in the prior analysis (*aqp-9*, *aqp-10*, *aqp-11*, and *aqp-12*) to produce a complete *C. elegans* dataset of *aqp* expression patterns. The comprehensive analysis of expression patterns provides for a consistent assay for the assessment of expression patterns for the entire nematode *aqp* gene family. The study of their expression patterns also provides a basis for the understanding of the functions associated with the AQP members which have had minimal functional characterization.

The main objective of the *aqp* expression pattern analyses is to determine genes which express within the excretory cell, a tissue with an analogous function to the mammalian kidney. From the expression pattern analysis of all 12 *C. elegans* AQP genes,

I show that, although these *aqp*s are expressed in a wide range of tissues, a few are expressed within the nematode excretory cell as expected. A suitable candidate promoter region from the excretory cell-expressing *aqp*s will be used as a model to assess transcriptional mechanisms that modulate excretory cell gene-expression.

2.2 Materials and Methods

2.2.1 Phylogenetic analysis

Multiple sequence alignments were conducted using ClustalX (Thompson et al. 1997) using the longest isoforms of each AQP protein retrieved from WormBase (WS192). Default parameters were used for all sequence alignments. The PHYLIP format was selected as the output format option. For a list of sequences used for alignments, refer to Appendix 1.

2.2.2 Nematode strains and maintenance

Strains were maintained at 20°C on *E. coli OP50* inoculated nematode growth media (NGM) plates. All manipulations were conducted using standard procedures (Brenner 1974). For the list of *promoter::reporter* constructs used in this section, the locations of the primers relative to the translational start site and the size of the promoter region captured, refer to Appendix 2.

2.2.3 Transgene construction

DNA constructs were generated via fusion PCR as previously described by (Hobert 2002) using DNA template prepared from N2 genomic DNA (Bristol, Baillie Laboratory strain BC49). Phusion polymerase (Finnyzmes, New England Biolabs Cat: F530) was used for

all PCR reactions to ensure fidelity of the resultant construct. Promoter-containing sequences were fused upstream of the GFP coding region. The reverse promoter associated primer includes a segment complementary to the forward primer used for amplification of the GFP-reporter cassettes. GFP-coding cassettes used for expression pattern analysis are as follows: *pPD95.67* (*GFP*), *pPD95.75* (*GFP*), and *pAF207* (*GFP-PEST*, a PEST sequence inserted into the C-terminal portion of GFP from the *pPD95.81* vector). All GFP variants used were modified by the addition of a 5' NLS from SV40, 3' UTR derived from *unc-54*, S65C mutation, and additional synthetic introns. The primers used for amplification of the GFP-encoding region are as follows: GFP D*-GGA AAC AGT TAT GTT TGG TAT ATT GGG and GFP C- AGC TTG CAT GCC TGC AGG TCG ACT. Sequences used for *aqp-promoter* primer design are based upon sequences in WormBase (WS154). The forward/reverse primers and the distance upstream of the forward primer from the TSS are shown in Appendix 3. The reverse primer used for generating the translational AQP-8::GFP construct was AQP-8protB: TTT CTA CCG GTA CCC TCA AGG Gtc cac tac tgt cac tat act ctc tgt ca. The forward primer used for the translational construct corresponds to same left primer used to generate the *aqp-8/K02G10.7b* (Appendix 2). Additional primers were selected to encompass the entire *aqp-8 promoter* region, these primers are aqp-8B2(-2223): AGT CGA CCT GCA GGC ATG CAA GCT TTG AAA GAC ACC GAT ATC TAA AAA and aqp-8Afar'(+15): CCA TAG ATG GTT CTG CAA GGA.

2.2.4 Microinjection

1.0-mm, 6" filamented capillary tubes (World Precision Instruments) were pulled into needles using a Sutter P-97 horizontal needle puller. The needles were mounted into a

Leitz Wetzlar micromanipulator. All microinjections were conducted using either Olympus BH2-HLSH or Zeiss 47 3016 inverted microscopes. Worms were mounted in mineral oil (Sigma, M-3516) atop dry agarose pads laminated on 48 x 65mm microscope cover slips (Gold Seal Cover Glass, reorder number 3335). PCR constructs were injected into the syncitial portion of the nematode gonad at an average final concentration of 30ng/µl along with 100ng/µl of the marker construct, *pCeh361* (*dpy-5*(+) (a Dpy-5 rescuing construct) (Thacker, Sheps, and Rose 2006), into the somatic gonad of *dpy-5(e907)* worms. The injected worms were arranged five P_0s per plate. Dpy-5 rescued wild-type F_1s were individually plated. Wild-type F_2 lines were selected to establish the transgenic lines. Only one wild-type F_2 line was kept per original P_0 plate. If more than one P_0 plate produced viable wild-type F_2 lines, each were analyzed separately and designated as individual segregant*s*

2.2.5 Genome integration of transgene

A stable GFP-expressing line was generated by subjecting *aqp-8*$^{promoter(-711)}$*::GFP* (BC6835) P_0 worms to 1,500R X-irradiation (Torex 150D X-ray Inspection System, settings: 145kV @ 5mA on shelf 7 for 135 seconds). Spontaneous transgene integrants were isolated by selecting for F_3 lines which produced 100% rescued *dpy-5* progeny thus producing (BC7032) *dpy-5(e907);sIs1241 rCes[K02G10.7(-711)::GFP-PEST +pCeh361]*.

2.2.6 Microscopy

A Zeiss Axioscope equipped with a QImaging camera and the appropriate optical filter sets was used for GFP expression pattern analysis. Worms were immobilized on moist

agarose pads (2% in water) with a 5µl of 100mM sodium azide (in water) immediately prior to imaging. All images were taken with identical filter, lens, and camera settings for all image sets (exposure times are indicated in the figures). Images were captured using QCapture software and processed using Adobe Photoshop CS. Only worms that displayed GFP-expression were imaged.

2.3 Results

2.3.1 Comparative analysis of *C. elegans* AQPs

The *C.elegans* aquaporins, AQP-(1-8), are considered the canonical aquaporins due their conserved NPA box pairs. The exception to this rule is AQP-5, which has an NPV box in place of the second NPA box. The NPV substitution is a common variant that is found in other AQPs from organisms such as yeasts (Bill et al. 2001) and *Arabidopsis* (Wallace and Roberts 2004). The canonical AQPs each fall into one of two well-established sub-classes of AQPs, the aquaglyceroporins or the aquaporins (Huang et al. 2007). Including the bacterial aquaporins AqpZ and GlpF as examples of aquaporins and aquaglyceroporins (Fu et al. 2000) in a multiple alignment of all *C. elegans* AQPs, these proteins can be divided into these two classes based upon their groupings with the well-defined members of their respective sub-classes. The remaining four AQPs, AQP-(9-12) are non-canonical aquaporin members due to their lack of one or both NPA domains. These four AQPs are more similar to each other than to the canonical AQPs (Figure 3) and as such reside on a distinct branch of the phylogenetic tree that diverges from the canonical AQP branches (Figure 3).

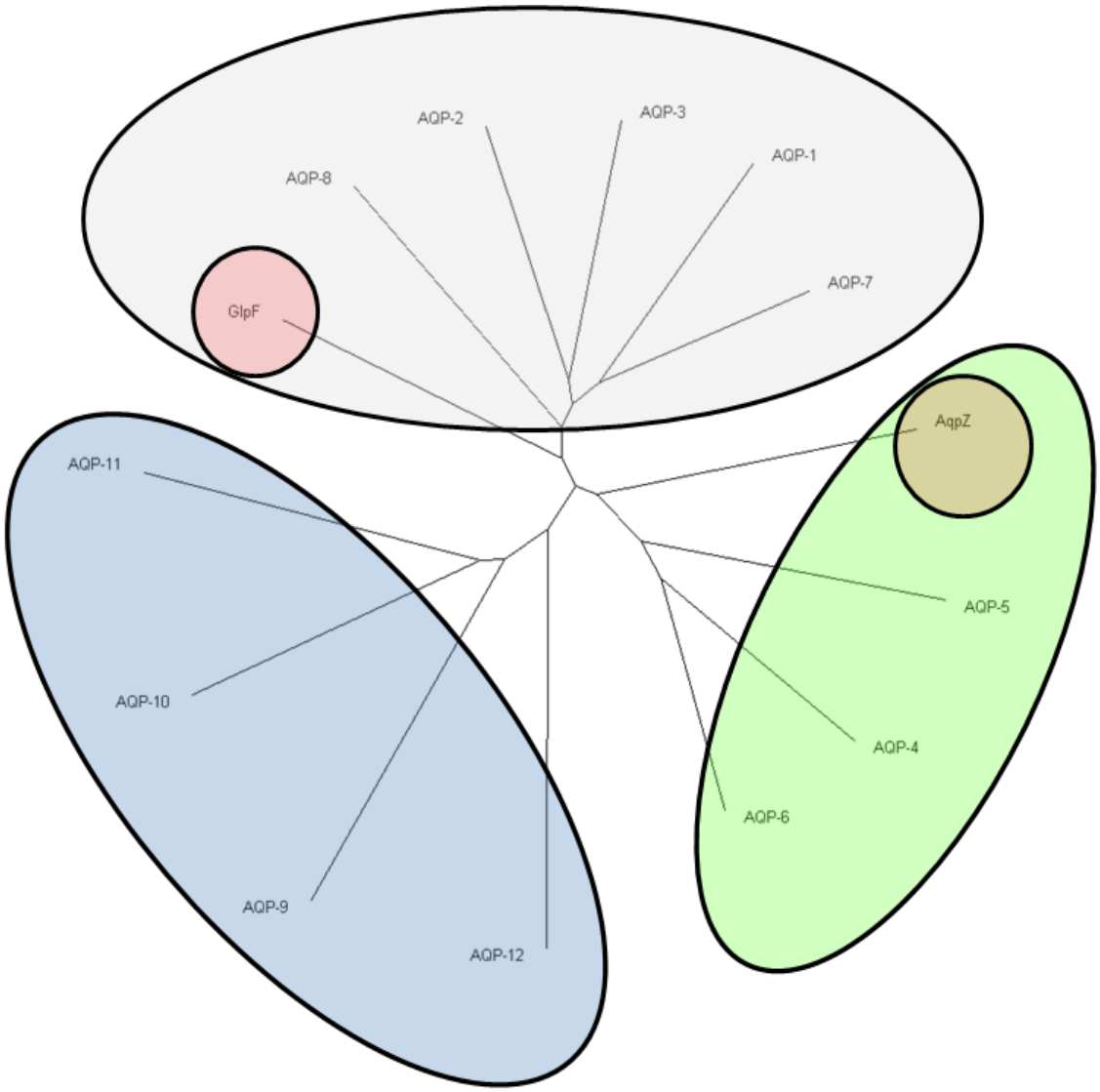

Figure 3. Phylogenetic tree representing all *C. elegans* AQP members.

AqpZ (orange) and GlpF (pink) are *E. coli* aquaporin and aquaglyceroporin sub-class members respectively and were included as representative members of their sub-families. As expected, AQP-8, AQP-2, AQP-3, AQP1 and AQP-7 (grey) branch along with the typical aquaglyceroporin and AQP-5, AQP-4, and AQP-6 (green) branch along with the typical aquaporin subclass member. AQP9, AQP-10, AQP-11, and AQP-12 (blue) group together along a branch distinct from the traditional sub-classes.
The longest isoforms of each respective AQP member was used to generate this tree.

Analysis of the *C. elegans* AQPs (9-12) sequences indicates that the proteins either completely lack conserved NPA sequence as in the cases of: AQP-9, AQP-10, and AQP-11, or are lacking conservation of one of the NPA sequences as in the case of AQP-12 (lacking an NPA motif in the second half of the protein) (Figure 4).

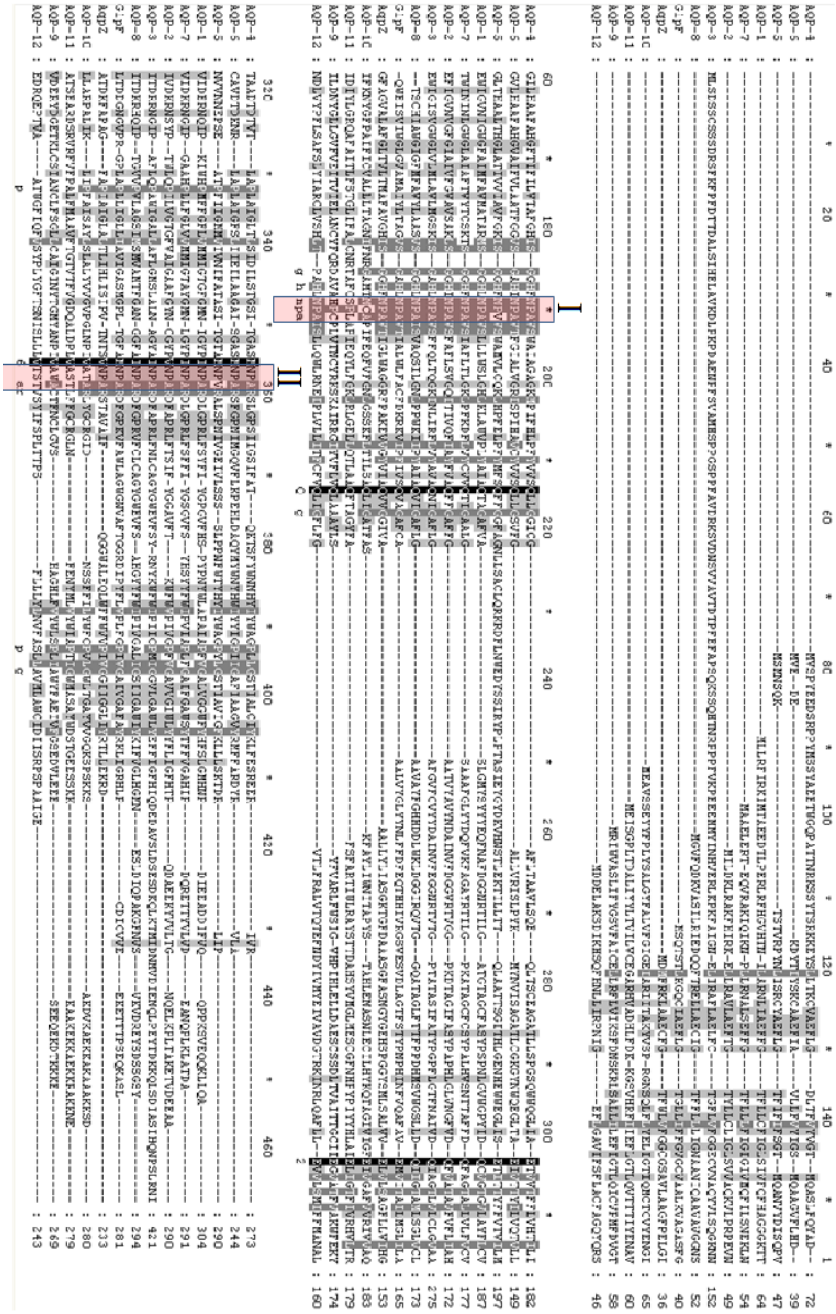

Figure 4. Locations of NPA domains in *C. elegans* AQPs.

The locations of the first and second NPA domain regions are indicated in the alignment by the pink highlights. AqpZ and GlpF are *E. coli* water and glycerol-specific transporters respectively. The first NPA box can be represented by the motif SG(A/G)HXNPA and that the second NPA box can be represented by the motif NPAR(S/D/A) (Ishibashi 2006).

The phylogenetic tree, as a result of a multiple alignment of all *C. elegans* AQPs against all human AQP proteins, showed that these non-canonical AQPs (9-12) cluster along with the mammalian aquaporins, hAQP11 and hAQP12. hAQP11 and hAQP12 are members of the recently characterized sub-cellular aquaporin class (Figure 5) (Ishibashi 2006; Yakata et al. 2007). Like AQPs (9-12), hAQP11 and hAQP12 both lack conserved second NPA boxes (Ishibashi 2006).

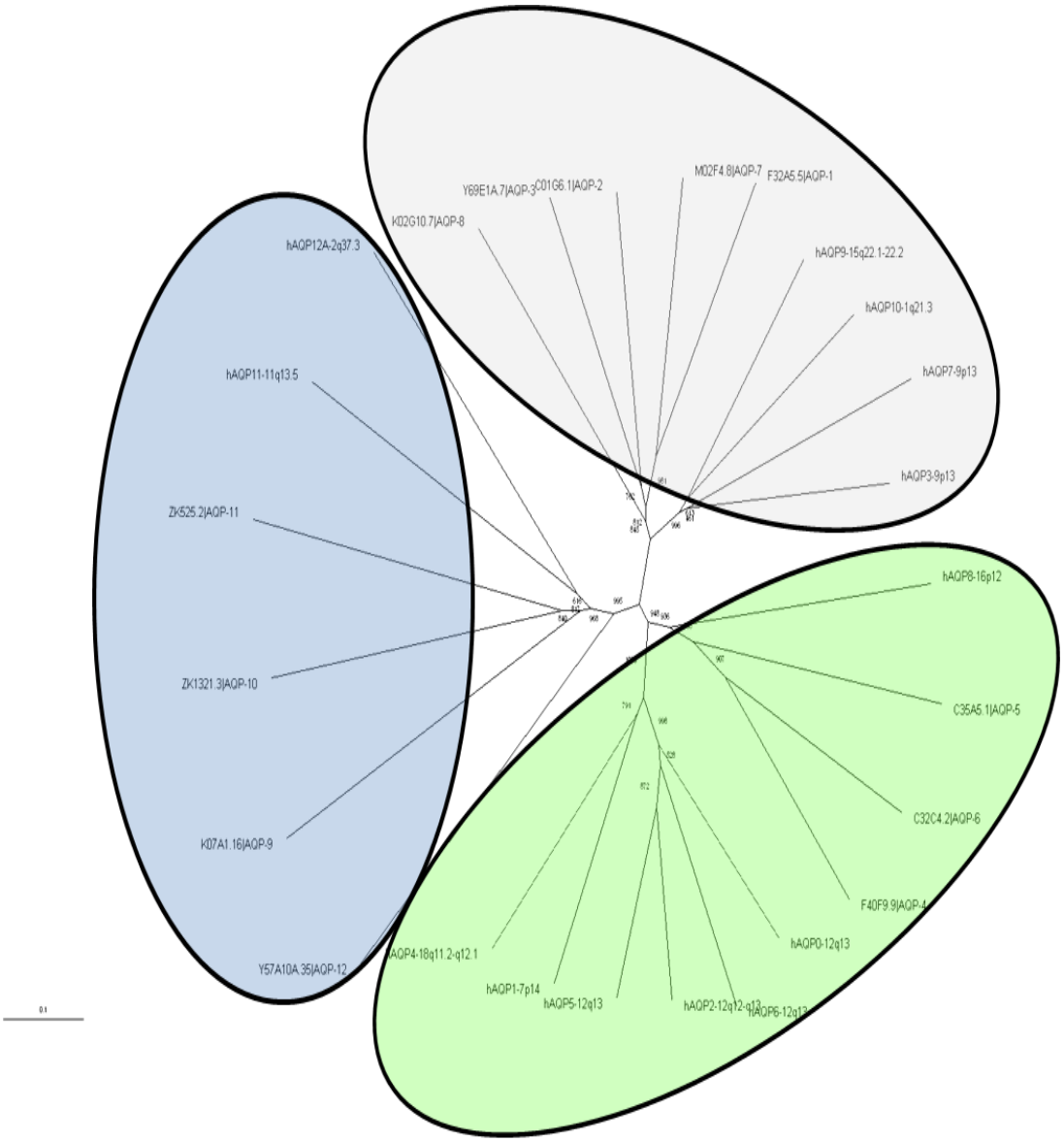

Figure 5. *C. elegans* AQPs vs. human AQPs.

The tree resulting from a multiple alignment of the *C. elegans* AQPs against mammalian (human) AQPs. AQP-8, AQP-2, AQP-3, AQP1 and AQP-7 (grey) branch along with the mammalian aquaglyceroporins and AQP-5, AQP-4, and AQP-6 (green) branch along with the mammalian aquaporin subclass group. AQP9, AQP-10, AQP-11, and AQP-12 (blue) group together along with hAqp11 and hAqp12, aquaporins that are localized into intracellular membranes.
The longest isoforms of each AQP was used to generate this tree.

2.3.2 Expression pattern analysis of each *aqp* member

The expression patterns of all 12 *C. elegans aqp* members have been determined via transgenic worms expressing *aqp-*^{promoter}*::reporter* constructs. The constructs' linear transgenes were generated via PCR fusion of the *aqp*-promoter regions to the reporter gene. The promoter region of each gene was defined as the smaller region of the two scenarios: up to 3kb upstream of the target gene's translational start site or the entire intergenic region of genes facing the same direction. A Green Fluorescent Protein (GFP) reporter optimized for use in *C. elegans* was used for the initial expression pattern analyses of the aquaporin family (construct from *pPd95.67*, courtesy of A. Fire, Stanford University). To confirm the expression patterns resulting from the *promoter::GFP* constructs and to obtain a more complete representation of the expression pattern, the same promoter regions were fused to the coding region for GFP-PEST protein, a short lived GFP variant based on a *C. elegans*-optimized GFP with a degradation signal inserted near the C-terminal end (Frand, Russel, and Ruvkun 2005) (Figure 6). The GFP-PEST reporter also provides ability to observe expression in tissues that underlies other tissue-types due to differences in temporal timing.

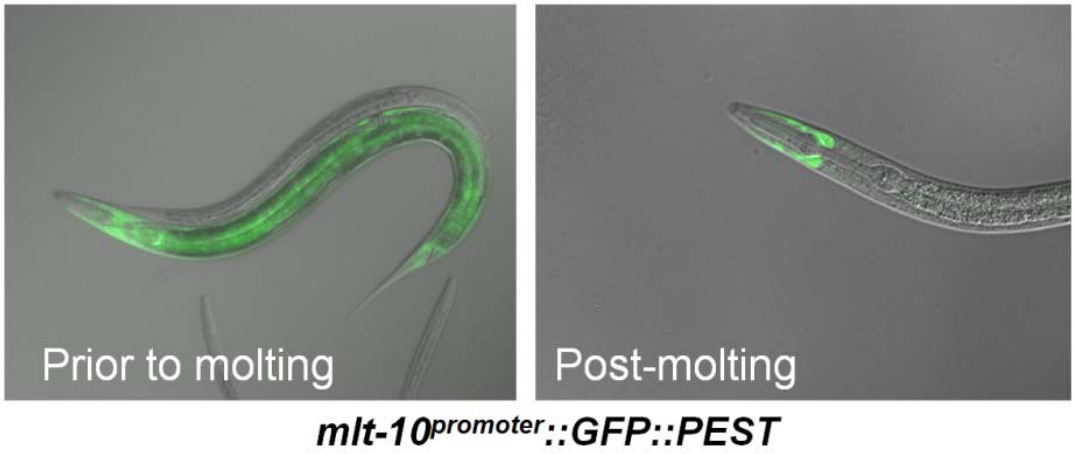

Figure 6. Example of GFP-PEST kinetics.

The rapid rate of GFP-PEST degradation is demonstrated using the *mlt-10* promoter region. The *mlt-10* promoter is active in hypodermal cells prior to molting at the end of each larval stage and is completely shut-down in hypodermal cell along the body post-molt (Frand, Russel, and Ruvkun 2005).

Combined with the *aqp* expression patterns reported by (Huang et al. 2007), all 12 of the *aqp* promoter regions drive visible expression of the GFP reporter *in vivo.* The *C. elegans aqp*s express in a variety of tissues. A table of expression patterns as a result of fusion of the promoter to GFP is shown in Table 2. A summary of expression patterns as a result of fusion of the promoter to GFP-PEST construct are shown in Table 3. A table of expression patterns as a result of fusion of the promoter to both reporter constructs and combined with previous expression patterns (Huang et al. 2007) are shown in Table 4. A sample of the expression patterns arising from both reporter constructs are shown in Figure 7.

When taking into account the expression patterns of specific AQP sub-classes, most of the aquaglyceroporin genes show multiple-tissue expression patterns. Of the five aquaglyceroporins, only *aqp-8*'s promoter drives expression in a single cell type. The expression patterns derived from promoters of the members of the aquaporin subclass display limited expression patterns. Two of the aquaporins each expressed in single tissues (*aqp-4* and *aqp*-5), while *aqp-6* expressed in two tissue types. The promoter regions of the sub-cellular aquaporins, in general, drove widely distributed multi-tissue expression patterns. *aqp-12* is the only member of the sub-cellular aquaporin subclass to be expressed in a single tissue type, although its pattern was not restricted to a single cell.

Table 2. GFP expression patterns of *C. elegans aqps*

Sequence name / gene name	Expression level	Expression location	Expression stage
F32A5.5/aqp-1	Low	1st/2nd bulb pharynx	Mid larval-adult
C01G6.1/aqp-2	High	Anterior neuronal, body muscle, possible excretory gland, hypoderm, excretory cell	All stage
Y69E1A.7/aqp-3	Low	Excretory cell	Mid larval-adult
F40F9.9/aqp-4	No expression	No expression	No expression
C35A5.1/aqp-5	No expression	No expression	No expression
C32C4.2/aqp-6	Low	Nerve ring, anterior neurons	Mid larval-adult
M02F4.8/aqp-7	High	Body muscle, anterior neurons, posterior neurons, intestine	All stage
K02G10.7/aqp-8	High	Excretory cell	Late embryo-adult
K07A1.16/aqp-9	No expression	No expression	no expression
ZK1321.3/aqp-10	High	Intestine, 1st/2nd bulb, muscle, spermatheca, (tail unknown), vulval muscle	All stage
ZK525.2/aqp-11	Medium	Intestine, body muscle, head muscle	All stage
Y57A10A.35/aqp-12	Medium	Anterior/posterior neurons	Early larval

Table 3. GFP-PEST expression patterns of *C. elegans aqps*

Sequence name / gene name	Expression level	Expression location	Expression stage
F32A5.5/aqp-1	No expression	No expression	No expression
C01G6.1/aqp-2	Medium	Anterior neuronal, body muscle, seam cell, canal associated neuron	larval-adult
Y69E1A.7/aqp-3	Low	No expression/male expression not observed	Mid larval-adult
F40F9.9/aqp-4	High	Intestine	Embryo-late larval
C35A5.1/aqp-5	No expression	no expression or v.low 1st/2nd bulb pharynx	No expression
C32C4.2/aqp-6	High	Developing vulva, anterior neuronal	Mid larval-adult
M02F4.8/aqp-7	High	Anterior neuronal, body muscle	All stage
K02G10.7/aqp-8	High	Excretory, no excretory in late adult	Early larval-adult
K07A1.16/aqp-9	High	Anterior neuronal??, body muscle, vulval muscle, anal depressor muscle	Early larval-adult
ZK1321.3/aqp-10	High	Intestine, 2nd bulb, developing vulva	Mid-adult
ZK525.2/aqp-11	Medium-low	Body muscle	Late adult
Y57A10A.35/aqp-12	no data	no data	no data

Table 4. Combined expression pattern analysis of *C. elegans aqps*

Sequence name / gene name	**Combined developmental stage of expression**	**Combined expression pattern**	Aquaporin class
F32A5.5/aqp-1	Mid larval - adult	*Pharynx, *intestine (basolateral membrane)	Aquaglyceroporin
C01G6.1/aqp-2	All stages	Anterior neuronal, *excretory cell, body muscle, hypoderm, canal associated neuron	Aquaglyceroporin
Y69E1A.7/aqp-3	Mid larval - adult	Excretory cell, Intestine, *seminal vesicle/vas deferens	Aquaglyceroporin
F40F9.9/aqp-4	Embryo - late larval	Intestine *(apical membrane)	Aquaporin
C35A5.1/aqp-5	No data	*I1 neurons	Aquaporin
C32C4.2/aqp-6	Mid larval - adult	Anterior neuronal, *IL1 neurons, developing vulva	Aquaporin
M02F4.8/aqp-7	All stages	Anterior neuronal, ventral nerve cord, body muscle, intestine, spermatheca, posterior neuronal	Aquaglyceroporin
K02G10.7/aqp-8	Late embryo - adult	Excretory cell	Aquaglyceroporin
K07A1.16/aqp-9	Early larval - adult	Anterior neuron, body muscle, vulval muscle, anal depressor muscle	Sub-cellular aquaporin
ZK1321.3/aqp-10	All stages	1st/2nd bulb pharynx, muscle, intestine, spermatheca, developing vulva	Sub-cellular aquaporin
ZK525.2/aqp-11	All stages	Gut, body muscle, vulval muscle, hypoderm	Sub-cellular aquaporin
Y57A10A.35/aqp-12	Early larval	Anterior neuronal, posterior neuronal, ventral nerve cord	Sub-cellular aquaporin

***(expression derived from Huang et al. 2007)**

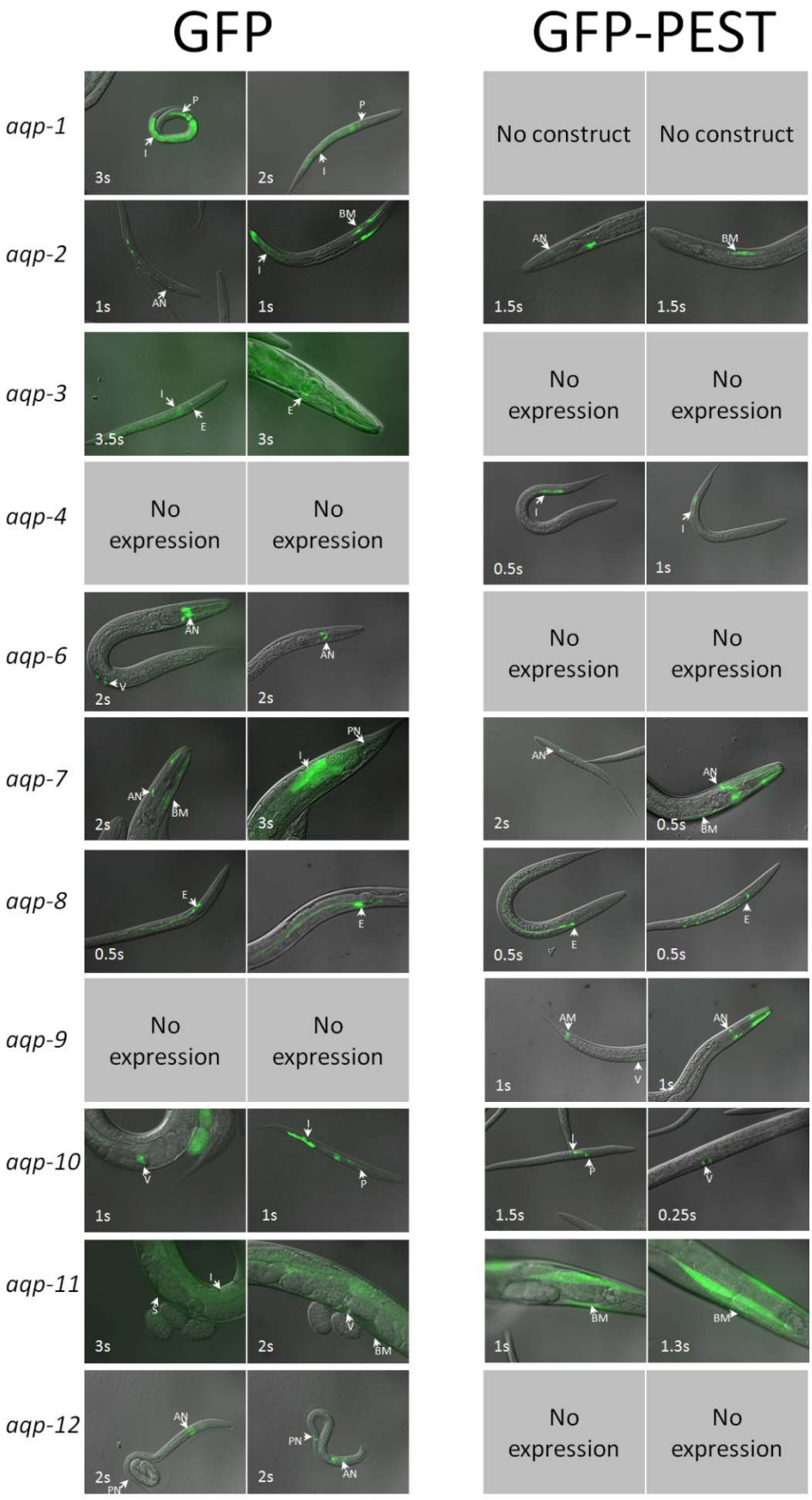
GFP
GFP-PEST
aqp-1
aqp-2
aqp-3
aqp-4
aqp-6
aqp-7
aqp-8
aqp-9
aqp-10
aqp-11
aqp-12
No construct
No construct
No expression
No expression
No expression
No expression
No expression
No expression
No expression
No expression
No expression
No expression
3s
2s
1s
1s
3.5s
3s
2s
2s
0.5s
1s
2s
2s
2s
3s
0.5s
0.5s
0.5s
0.5s
1s
1s
1s
1s
1.5s
0.25s
3s
2s
1s
1.3s
2s
2s
P
I
AN
BM
E
V
PN
S
AM

Figure 7. Sample of expression *aqp* expression patterns.

P: pharynx, I: intestine, AN: anterior neuron(s), PN: posterior neuron(s), BM: body muscle, E: excretory cell, V: vulva, AM: anal depressor muscle
aqp-5 was not included in the images as its promoter region did not drive expression of both GFP and GFP-PEST. A transgenic construct could not be generated for *aqp-1*^{promoter}*::GFP*.

2.3.3 *aqp-2*, *aqp-3* and *aqp-8* are excretory cell expressed *aqp*s

The promoter analysis indicated that three *aqp* genes are expressed within *C. elegans*' excretory cell. The only *aqp* that expressed exclusively in the excretory cell was *aqp-8*. The expression pattern driven by the *aqp-8* promoter region (a 1,556bp region spanning the positions -1,573bp → -20 bp upstream of the genes translational start site) was consistent when fused to both the GFP and GFP-PEST reporters. *aqp-2* and *aqp-3* are expressed in the excretory cell in addition to other tissues. The *aqp-2* promoter fragment (a 2,898bp fragment spanning the positions -2,987 bp →+11 bp of the TSS) drove non-identical expression patterns when fused to the two different reporters, GFP and GFP-PEST. The promoter fragment for *aqp-3* (a 2,987bp segment spanning the positions -2,986bp →+11bp relative to the TSS) also drove different expression patterns when fused to the two different transcriptional reporters (Figure 8 *A-C*).

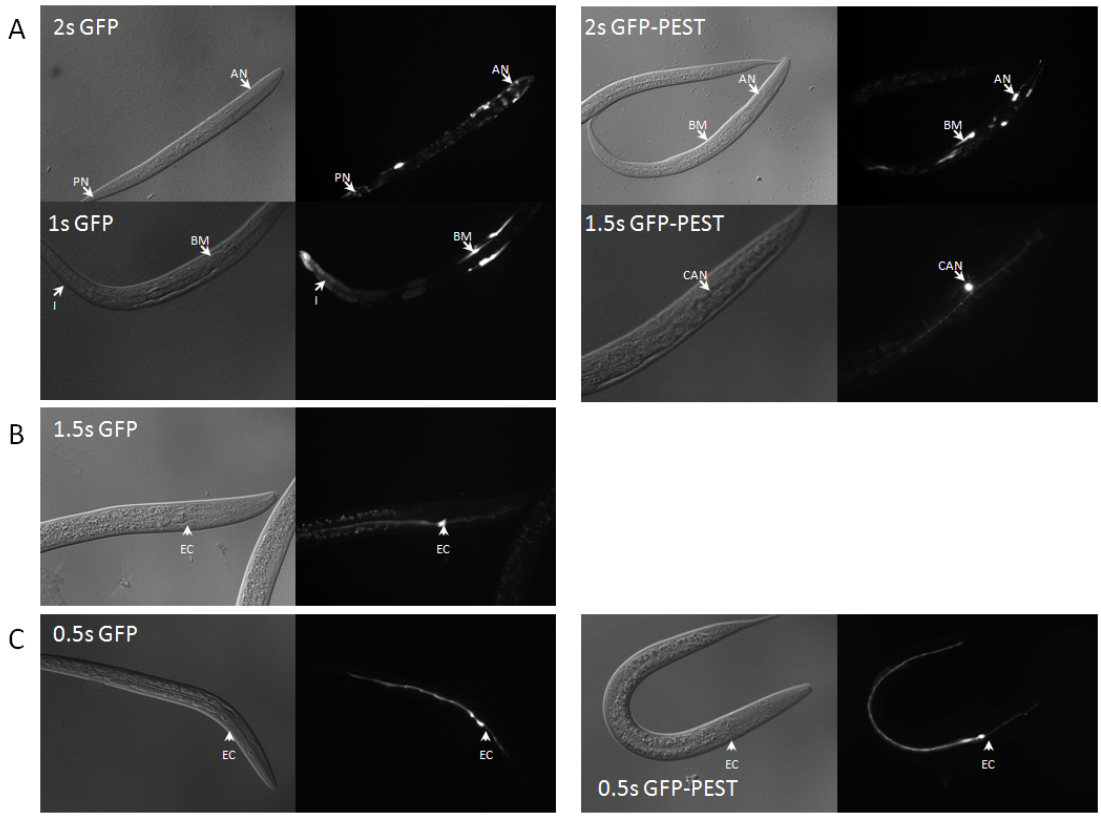

Figure 8. Sample of expression patterns from excretory cell expressed AQPs.

Animals at various developmental stages expressing *aqp-* ^{promoter}*::GFP* or *aqp-* ^{promoter}*::GFP-PEST* constructs

A. *aqp-2* ^{promoter}*::reporter constructs*

B. *aqp-3* ^{promoter}*::reporter constructs*

C. *aqp-8* ^{promoter}*::reporter constructs*

AN: anterior neuron(s), PN: posterior neuron(s), BM: body muscle, CAN: canal associated neuron(s), EC: excretory cell

2.3.4 Verification of *aqp-8* expression pattern via additional transgene constructs

The expression pattern driven by a -2,223bp→+15bp *aqp-8* promoter fragment drove an expression pattern identical to that of the previous constructs indicating that there are probably no required *cis*-regulatory elements located between -20bp and +15bp and between -1,573bp and 2,223bp that affect the expression pattern of *aqp-8* (image not shown). I selected a shorter 5'-truncated promoter fragment of *aqp-8*'s promoter region to fuse to the *GFP-PEST-coding* region to determine whether the fragment was still able to drive expression in the excretory cell. The fragment (-711bp →-20bp upstream of the TSS) was able to drive expression in the excretory cell much like the larger 1,556bp fragment (image not shown). This construct was then used as a basis for the generation of an X-ray irradiation-induced genome-integrated transgenic line. The integrated transgene was made to prevent somatic mosaic loss of transgenes, a feature common in extrachromosomal DNA arrays in *C. elegans* (Herman 1984). An AQP-8::GFP translational fusion transgenic line was also generated to determine whether protein localization corresponded to the patterns conferred by the *promoter::GFP* constructs and to also determine where the protein localized intracellularly. AQP-8 appears localized to the membrane of the entire excretory cell as opposed to the generalized cytoplasmic localization of GFP seen in the *aqp-8*^{promoter}*::GFP* (transcriptional) constructs, which show a diffuse expression along the entire excretory cell canals with a concentration of fluorescent signal at the cell body most likely due to the influence of the 5' nls on the *GFP*-coding sequence (Figure 9).

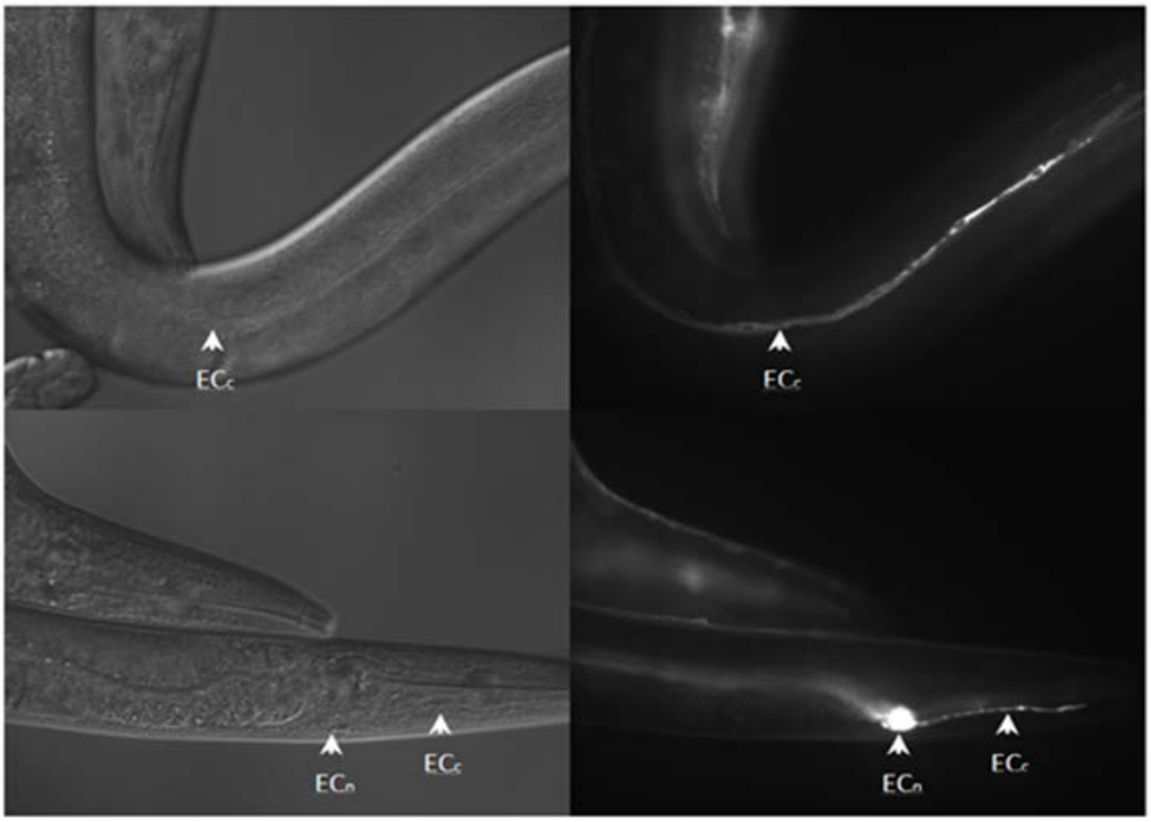

Figure 9. AQP-8 is localized to the periphery of the canal membranes.

The AQP-8::GFP fusion protein is located along the excretory canal membranes (EC_c) in the posterior canals (top) and the anterior canals (bottom).
The strains assayed were:
BC06925 *dpy-5(e907);sEx1318 rCes[K02G10.7 (translational)::GFP+pCeh361] (SegI)* and
BC06926 *dpy-5(e907);sEx1319 rCes[K02G10.7 (translational)::GFP+pCeh361] (SegII)*

2.4 Discussion

The *C. elegans* canonical aquaporins belong to one of two established subgroups. The two groups are aquaporins and aquaglyceroporins and, as their names imply, are defined by their ability to transport different substrates (Borgnia and Agre 2001). This AQP substrate selectivity is based upon amino acid residues surrounding the pore of the folded protein which either excludes substrates based upon size selection (Fu et al. 2000). The second NPA box of aquaporins typically follow the consensus NPA-(RVL)-(SA), whereas the second NPA box of aquaglyceroporins conforms to the consensus NPARD (Ishibashi, Kuwahara, and Sasaki 2000). Although these differences are also present in the two classes of *C. elegans* AQPs (Figure 4), it has been suggested that definitive sub-grouping should be based upon overall homology of the protein with representative members of either group (Ishibashi, Kuwahara, and Sasaki 2000).

The three closely linked *aqp*s located on *LGV* are the only members of the aquaporin subclass and show high sequence conservation. Their relative positions on that chromosome has been maintained in *C. elegans* and *C. briggsae*, and appears conserved in the genome assembly of *C. remanei* (crem_contig60 and crem_contig38), it is possible that this AQP subclass in nematodes is the result of an ancient aquaporin that underwent two replication events prior to the divergence of the sequenced nematodes.

I performed a comparative analysis with all *12 C. elegans AQPs*; taking into account those AQPs which have not been previously characterized phylogenetically (AQP-9, AQP-10, AQP11, and AQP-12; non-canonical AQPs). The previously unclassified AQPs group themselves phylogenetically and form a distinct sub-group. These non-canonical AQPs belong to another sub-class of aquaporins, which have yet to

be characterized in *C. elegans*. They contain at least one divergent NPA domain (usually the one closer to the N-terminus) and phylogenetically cluster along with hAQP11 and hAQP12, mammalian AQP proteins that also contain divergent NPA boxes.

It was previously believed that hAQP11 and hAQP12 do not conduct water (Gorelick et al. 2006; Itoh et al. 2005), but further analysis of members of this AQP class in other organisms has shown that these AQPs are localized to intracellular membranes (Nozaki, Ishii, and Ishibashi 2008; Yakata et al. 2007). The intracellular localization of the AQP makes expression of the AQP into *Xenopus* oocytes an ineffective strategy for determining the pore's substrate specificity. The names coined for these non-conventional AQPs are sub-cellular aquaporins (Yakata et al. 2007; Ishibashi 2006) and superaquaporins (Nozaki, Ishii, and Ishibashi 2008). Much like the *C. elegans* variants, the mammalian sub-cellular aquaporins are located on a distinct phylogenetic branch, distinguishing them from the canonical mammalian aquaporins.

Mammalian AQP11 has an altered first NPA box and expresses in both the endoplasmic reticulum and the plasma membrane in the testis, kidney, liver, and brain (Gorelick et al. 2006). In the brain, AQP11 is expressed in the: Purkinje cell dendrites, hippocampal neurons of CA1 and CA2, and cerebral cortical neurons (Gorelick et al. 2006). Mice with defective AQP11 appear normal when born, but die before weaning due to polycystic kidneys (Morishita et al. 2005). Mammalian AQP11's substrate specificity has been determined using an alternative method. The intact AQP-11pore proteins were isolated from cellular membranes and reconstituted into liposomes. These liposomes were subjected to hydrostatic and osmotic pressures. The *C. elegans* sub-cellular aquaporin, AQP-11, has been demonstrated to be a water channel with a capacity of

water transport similar to the non-sub-cellular mammalian aquaporin, AQP1 (Yakata et al. 2007). Studies of mammalian AQP12 have been more limited. The protein was found to be localized intracellularly in acinar cells of the pancreas (Itoh et al. 2005) .

The challenges in studying the substrate specifities of the *C. elegans* sub-cellular AQP members are similar to those of the mammalian sub-cellular aquaporins. The technique of reconstituting the AQPs in liposome membranes developed by Yakata *et al. 2007* appears to be the appropriate technology for characterization of these channels. Nevertheless, preliminary studies for determining nematode AQP-10 and AQP-11 solute specificity have been attempted by analyzing their ability to transport water and glycerol using the traditional system by expressing the proteins in *Xenopus* oocytes. The results of these studies on nematode sub-cellular aquaporins did not indicate any change in water permeability of the cells possibly as a result of their intracellular localizations (Sakube et al. 2003).

RNAi treatment any of the *C. elegans* sub-cellular aquaporins does not elicit a detectable phenotype in RNAi screens (Sonnichsen et al. 2005; Rual et al. 2004; Kamath et al. 2003). This is the same as the effects of RNAi-mediated knockdown of the rest of the canonical AQPs, which do not lead to any overt phenotypes arising from the gene knockdown.

Single-gene knockout mutants of *aqp-2*, *aqp-3*, *aqp-4*, and *aqp-8* do not produce any assayable phenotypes (Huang et al. 2007). Analysis of a triple mutant *aqp-2*, *aqp-3*, and *aqp-8* to determine the effects of loss of the excretory cell AQP function also leads to no assayable phenotype. However, making a quadruple mutant consisting of the knockout alleles of the genes, *aqp-2*, *aqp-3*, *aqp-4*, and *aqp-8*, leads to the subtle phenotype of a

decreased ability of the worm to recover from hypotonic stress (Huang et al. 2007). The RNAi studies, taken together with the studies with the gene knockout mutants, indicate that there might be a high level of functional redundancy among these proteins and that most of these proteins are not critically important for worms raised in standard laboratory conditions.

To provide a better understanding of their functions, I analyzed GFP expression patterns driven by the promoters of each of the 12 *aqps*. For my study, I defined the promoter region as the region directly upstream of the gene's translational start site within 3kbp (or to the next upstream gene). I ignored the possibility of expression modulating elements within the *aqp* introns since it has been suggested that the *cis*-regulatory elements located within introns are largely responsible for regulation of alternative splicing (Kabat et al. 2006).

Two types of GFP reporter variants were used in the *aqp* expression pattern analyses. The first reporter used is a standard GFP-coding cassette used for *C. elegans* expression pattern analysis, *pPD96.67*. *pPD96.67* contains a 5' *nuclear localization sequence* derived from the virus *SV40* (5' *nls*), five artificial introns, and *a 3' untranslated sequence* (3' UTR) isolated from the *unc-54* gene in *C. elegans* (Kelly et al. 1997). The second reporter used was *pAF207* (Frand, Russel, and Ruvkun 2005). This reporter consists of all of the features of the first construct; in addition, it contains a modified mouse ornithine decarboxylase (MODC)-derived PEST-sequence (sequence rich in Proline, Glutamic acid, Serine, and Threonine). MODC has a half-life of approximately a half-hour (Li et al. 1998). The PEST sequence is a proteolytic signal originally identified in a comparative analysis of proteins with short half-lives (Rogers,

Wells, and Rechsteiner 1986). The PEST sequence was isolated from mouse ornithine carboxylase containing vector, *pd1EGFP-N1* (nucleotides 1,399–1,521, Clontech, Palo Alto, California, United States), and inserted between the last coding codon and the stop codon of GFP (Frand, Russel, and Ruvkun 2005). The half-life of the reporter construct was determined previously by treating CHO-K1 cells transiently expressing the d1EGFP-PEST protein with cyclohexamide, a protein synthesis inhibitor (Sibler et al. 2005).

The rapid turnover rate of the GFP-PEST construct allows for the evaluation of expression kinetics *in vivo*. A positive side effect of the rapid turnover conferred by the PEST sequence is the ability to assess expression of genes in tissues that would not usually be visible using a time-stable reporter due to high-level expression in overlapping or neighbouring tissues. The use of the GFP-PEST reporter, although it did not uncover any additional excretory cell-expressing genes, did provide useful information which is additive to the expression pattern derived from the GFP reporter. An example of this is seen when comparing the expression patterns as a result of the *aqp-2*^{promoter}*::GFP* and *aqp-2*^{promoter}*::GFP-PEST* constructs. The pattern derived from the *aqp-2*^{promoter}*::GFP-PEST* construct includes signal seen in the CANs, a tissue that might have been missed when assessing the expression pattern of the *aqp-2*^{promoter}*::GFP* construct due to the expression in the overlapping intestinal cells.

Expression pattern analysis for all *C. elegans aqp*s has revealed a wide range of expression patterns among the gene members. This reflects the requirement for water transporters in virtually all tissue types. The expression patterns derived from the constructs generated for this thesis and of those found in a previous study by Huang *et al.* (2007) show pattern overlaps, but there are a few significant differences between our two

studies. For example, the promoter region used in this study for *aqp-1* 2,952bp fragment consisting of the region spanning -2,968bp → -20bp upstream of the *aqp-1*'s TSS) drives expression of GFP in the first and second pharyngeal bulbs. In contrast, the fragment by Huang *et al.* 2007 (approximately 1.5kbp upstream of *aqp-1*'s the TSS) drove expression of GFP in the pharynx and an additional tissue, the basolateral membrane of the intestine. Conversely, the expression pattern of *aqp-2p::GFP-PEST* (2,897bp fragment consisting of the region spanning -2887bp →+10bp relative to the *aqp-1*'s TSS) drove expression in a group of neurons in the anterior portion of the worm whereas an approximately 5kbp promoter fragment used by Huang *et al.* failed to show expression in this tissue. For a summary of differences of promoter regions used, see Table 6. These instances of differences in expression patterns between genes from our two groups demonstrate the variability of expression patterns that can be conferred even by subtle differences in promoter sequences used to generate the *promoter::reporter* strains. In both cases (*aqp-1* and *aqp-2* promoter constructs), the shorter upstream region led to expression in a greater number of tissues. This might be due to the loss of *cis*-regulatory elements in the shorter promoter constructs that act to repress downstream gene expression in the short constructs. A summary of the promoter regions used in this study and by Huang *et al* 2007 are summarized in Table 5.

Table 5. Promoter regions used for transcription pattern analysis of *aqp*'s in *C. elegans*

Sequence name / gene name	Promoter region captured Huang et al. 2007	Construct size (this study) size / primer location
F32A5.5/aqp-1	1.5	2952bp / (-)2968 - (-)16
C01G6.1/aqp-2	5	2897bp / (-)2887 - (+)10
Y69E1A.7/aqp-3	2.8	2987bp / (-)2986 - (+)11
F40F9.9/aqp-4	2.2	2111bp / (-)2157 - (-)47
C35A5.1/aqp-5	3	2925bp / (-)2951 - (-)27
C32C4.2/aqp-6	3.7	2925bp / (-)2951 - (+)15
M02F4.8/aqp-7	4.5	2968bp / (-)2982 - (+)5
K02G10.7/aqp-8	2.5	1556bp / (-)1575 - (-20)
K07A1.16/aqp-9	n/a	2072bp / (-)2061 - (+)10
ZK1321.3/aqp-10	n/a	921bp / (-)933 - (+)13
ZK525.2/aqp-11	n/a	2952bp / (-)2937 (+)14
Y57A10A.35/aqp-12	n/a	2507bp / (-)2563 - (-)56

The expression patterns derived from the three water-specific aquaporin genes each express only one or two tissue types. The lack of a wide distribution of expression of these genes and their lack of observed *RNAi*-induced phenotypes suggest that these genes play highly specific tasks in the nematode that may not be required in worms raised in laboratory conditions.

The expression patterns derived from aquaglyceroporin promoters tend to display expression in more than one tissue for each gene. The fact that aquaglyceroporins tend to express in a generalized pattern might indicate that these genes act as general purpose transporters of both water and other small solutes and therefore are useful in most tissues. The expression pattern of *aqp-8* is unique among the five aquaglyceroporin members because it is the only member of that group which expresses in a single tissue - the excretory cell. AQP-8's substrate is also unknown, suggesting that this channel might play a very specific role in the regulation of the osmotic environment within the worm. Furthermore, subjecting *C. elegans* to hypertonic stress induces *aqp-8* mRNA expression eight-fold over unstressed worms (*pers. comm.* Lamitina T. and Strange K).

In addition to the canonical AQPs, I have determined the expression patterns of the sub-cellular AQPs. Each of these genes are expressed in more than one tissue. The multi-tissue expression patterns are probably due to the need for intracellular transport of water and uncharged solutes and having this duty spread among only four genes. This result indicates that these types of AQPs are required in almost all cells much like the canonical AQPs. For this study, I am interested, in particular, in genes expressed within the excretory cell. The expression pattern analyses indicate that *aqp-2*, *aqp-3*, and *aqp-8* are expressed in this cell.

The excretory cell-expressing aquaporins represent one-quarter of the members in *C. elegans.* Of 1,885 *promoter::GFP* expressing constructs, 193 (10.2%) express in the excretory cell (Hunt-Newbury et al. 2007). This shows that that excretory cell expression might be over-represented in the *aqp* family. Due to the small sample size (12 *aqp* members), the significance of this number can be debated. Only one in five genes in *C. elegans* are expressed in tissue-specific patterns and gene promoters that drove expression only in the excretory cell represent only 0.3% (6/1,885 gene promoters) of the expression pattern dataset; reflecting the rarity of exclusive excretory cell gene-expression (Table 6) (Hunt-Newbury et al. 2007).

Table 6. Genes that are expressed exclusively in the excretory cell.

Gene locus	Gene common name	Strain name	Physical location
C18C4.2	*cft-1*	BC10556	V:5540195..5541388
F22E10.1	*pgp-12*	BC10089	X:12728916..12731782
F44B9.10	n/a	BC15256	III:8033441..8036404
F48E8.3	n/a	BC11730	III:5458970..5461386
F56E10.2	*fhod-2*	BC14072	V:70819..73630
K12G11.2	*sulp-5*	BC13952	V:11881677..11882663
Y8G1A.2	*inx-13*	BC11504	I:3753911..3755565

cft = Cystic Fibrosis Transmembrane conductance regulator homolog
pgp = P-GlycoProtein related
fhod= Formin HOmology Domain
sulp = SULfate Permease family
inx = INneXin

The expression pattern of *aqp-8* was confirmed by assaying the expression patterns conferred by transgene constructs and publicly available expression profiling data as follows:

1. A smaller promoter fragment from *aqp-8* (-711 → -20bp upstream of *aqp-8*'s TSS) was able to drive GFP-PEST expression in the excretory cell much like the larger 1,556bp fragment used in the previous analysis. In addition to confirming the function of *aqp-8*'s promoter, this shorter construct effectively restricted *aqp-8*'s excretory cell-modulating *cis*-regulatory element to roughly a 700bp window.

2. Putative genome-integrated transgenes (stable lines) were analysed to determine whether the expression pattern derived from the *aqp-8*^{promoter}*::GFP*(*-PEST*) constructs were incomplete due to the possibility of mosaic expression as a result of transgene loss during somatic cell divisions. This is a potential problem in the analysis of extrachromosomal transgene arrays. The stable transgenic lines display expression patterns that are consistent with all prior *aqp-8*^{promoter}*::reporter* constructs and therefore mosaic loss of the extrachromosomal transgene is not an issue for analyses of expression patterns based on the *aqp-8* promoter region.

3. A translational fusion consisting of *aqp-8*'s 1,573bp promoter region fragment and coding region of AQP-8 fused to GFP (lacking a 5' NLS) at the AQP-8's C-terminus. The expression pattern of this construct is similar to those of the *aqp-8*^{promoter}*::reporter* constructs. The differences between the localization of signal from AQP-8::GFP and *aqp-8*^{promoter}*::reporter* constructs appear to be non-existent

except that the AQP-8::GFP fusion protein appears to be localized at the cell surface.

4. A larger promoter fragment (-2,223 → +15bp relative to *aqp-8*'s TSS) was used to determine whether there are expression pattern modulating elements upstream and downstream of the previous construct. Expression was only observed in the excretory cell.

5. The expression profile from *aqp-8* Serial Analysis of Gene Expression (SAGE) (Velculescu et al. 1995) data indicate that *aqp-8* message expression is low in embryos-L1 (Figure 10 *A*) with expression gradually increasing in the L2-L3 larval stages, with the expression level peaking at L4 stage worms (Figure 10 *B*) (McKay et al. 2003). It is important to note that *aqp-8* mRNA is probably localized entirely to the excretory tissue as determined from the assay of prior *aqp-8*^{promoter} *::reporter* transgene constructs. The mRNA used for SAGE expression profiling was extracted from whole animals and therefore the level of *aqp-8* expression is possibly underestimated using this method. Looking at SAGE tags derived from embryonic dissected tissues (purified oocytes, purified oocytes, FACS sorted muscle cells, FACS sorted pan-neural cells, FACS sorted pharynx cells, N2 Embryos (longSAGE), FACS sorted hypodermal cells, FACS sorted ciliated neurons, FACS sorted pharyngeal marginal cells, FACS sorted punc-4::GFP cells, FACS sorted AFD neurons, FACS sorted ASER neurons, and FACS sorted pharyngeal gland cells) single SAGE tags are only observed in each of the libraries: purified oocytes, N2 Embryos (longSAGE), and FACS sorted AFD neurons. The rarity of tags in these libraries indicates that the mRNA is not

found in assayable quantities in the FACS sorted cell types (except for the AFD neurons, in which the tag could represent background levels of contaminant in the message preparation).

Results from the *aqp-8*^{promoter}*::GFP-PEST* constructs, which allow accurate analysis of kinetics of transient gene expression, along with the SAGE expression profile data, reveal that *aqp-8*'s promoter is active from the earliest observed GFP-PEST expression with strong levels of excretory cell expression being maintained throughout to early adulthood in *C. elegans*. Nuclear proteins that regulate the expression of *aqp-8* should express earlier than the earliest observed incidence of *aqp-8* expression indicated above. Furthermore, the expression pattern of these nuclear proteins should also overlap with the pattern driven by *aqp-8*'s promoter region.

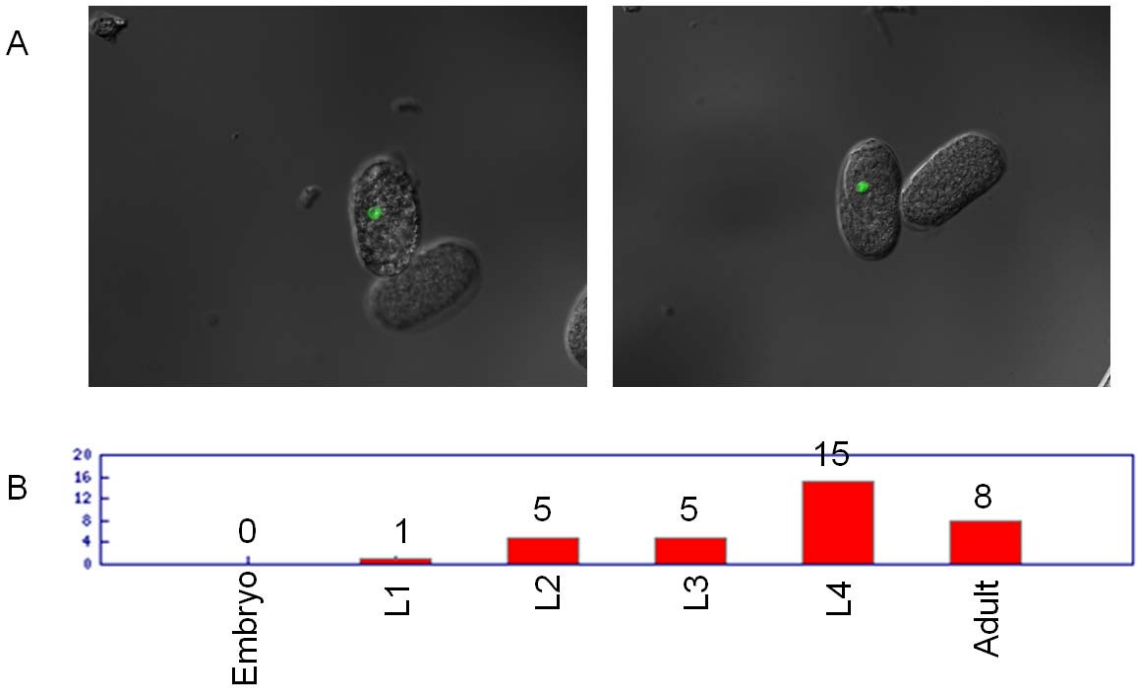

Figure 10. Expression of *aqp-8* during development.

A. Both images show the earliest observed expression for *aqp-8* is approximately the plum stage; 7.0-7.25 hours post-fertilization (BC6925, *aqp-8::GFP* stabilized transgene expressing strain).

B. Developmental SAGE profile of *aqp-8/K02G10.7* showing the gradual increase in message as the worm develops, eventually reaching a plateau at L4-stage (http://tock.bcgsc.ca/cgi-bin/sage140)

3: MAPPING OF CIS-REGULATORY ELEMENT(S) REQUIRED FOR EXCRETORY CELL EXPRESSION OF *AQP-8*

3.1 Introduction

Transcriptional regulation can be described by different steps in control: control at the chromatin level, control at the nucleosome level, control by DNA modification (*e.g.* methylation of nucleotides), and control at the level of the individual gene.

Studying DNA elements required to drive individual gene expression patterns (in both the spatial temporal dimensions) provides important initial understanding of the function of the gene. These elements, commonly referred to as activators and repressors (generically referred to as *cis*-regulatory elements), bind nuclear proteins which affect the levels of RNA polymerase interaction at the gene's promoter. Analysis of *cis*-regulatory elements that modulate expression patterns in one gene provides a basis for determining the function of related nucleotide sequences in the vicinity of other genes. The study of these functional non-coding DNA elements has been aided by the increasing number of genomic sequences available for study which provide material for comparative analyses. Many DNA elements, which control downstream genes, have been isolated within intergenic regions of *C. elegans*. Fortunately, the intergenic regions in *C. elegans* are usually fairly short - a consequence of the high gene density of the genome. Some *cis*-regulatory elements in *C. elegans* have, however, been found within the intron sequences

(Li et al. 1999), *3' untranslated region* (3' UTR) (Sarin et al. 2007), and the intergenic spaces between genes within an operon (Huang et al. 2007). As mentioned in the previous chapter, many of the intron-associated *cis*-regulatory elements may play a role in regulating alternative splicing decisions (Kabat et al. 2006). The *cis*-regulatory elements in the 3' UTR, however, might have roles in microRNA (miRNA) binding, a post-transcriptional control mechanism; for examples see (John et al. 2004; Krek et al. 2005; Xie et al. 2005). Binding of factors to sites within the 3' UTR can affect stability of the mRNA, modulate translation and influence localization.

The focus of this study is to determine elements required for the transcriptional regulation of genes expressed within the excretory cell, the largest mononucleate cell in *C. elegans.* Previous studies involving mechanisms driving excretory cell expression in the nematode excretory cell have been carried out. Zhao Z. *et al.* 2005 discovered a group of *cis*-regulatory elements upstream of *pgp-12*'s (*P-GlycoProtein related-12*) TSS. PGP-12 belongs to one of the largest protein families in *C. elegans*, the ATP-binding cassette family transporter family of proteins (ABC transporters). PGP-12 is homologous to the PGP proteins found in mammals which play a role in the development of multi-drug resistance during cancer chemotherapy (Sheps et al. 2004).

In the previous chapter, I identified three *aqp*s which express in the excretory cell. Although, both *aqp-2* and *aqp-3* express in the excretory cell, their expression patterns indicate that these genes are also expressed in other cells and tissues. The complexity of their expression patterns translates into difficulty for finding *cis*-regulatory elements which specifically modulate expression in the excretory tissue. One of the difficulties arising from working with promoters, which drive expression in multiple tissues, is the

possibility of isolating *cis*-regulatory elements which modulate expression in more than one tissue. Comparison of the promoter regions of each of the excretory cell-expressing *aqp*s with their orthologous region in *C.briggsae* shows that each gene has many conserved gene-upstream sequences (data not shown). The conservation of upstream non-coding sequences between the two species might indicate that at least some of these regions are functionally significant.

To simplify the analysis, and to find *cis*-regulatory sequences that modulate excretory cell expression specifically, I focused my analysis on *aqp-8*'s promoter for determining *cis*-regulatory sequence(s) due the promoter region's ability to drive expression exclusively in the excretory tissue much like the *pgp-12*-promoter based system used in the previous study by Zhao *et al.* 2007, which led to the discovery of the EX-1 *cis*-regulatory element. To find these DNA-regulatory elements, I used a combination of genetic and computational approaches to arrive at a region suspected to have an excretory cell modulating *cis*-regulatory element.

3.2 Materials and methods

3.2.1 Sequences

The -500bp → +50bp regions relative to the translational start site of *C. elegans aqp-8* and its orthologous regions in *C. briggsae*, *C. brenneri*, *C. japonica* and *C. remanei* were downloaded from WormBase (WS190), for the sequences, refer to Appendix 4.

3.2.2 Multiple sequence alignments

Default parameters were used for ClustalX (Thompson et al. 1997) except for the gap opening parameter was changed to 5.00/100 and gap extension parameter changed to

4.00/100 to relax the gap introduction and elongation stringencies. The alignments were visualized using GeneDoc (http://www.psc.edu/biomed/genedoc). Triple alignment of the non-coding region upstream of *aqp-8*and its orthologs from *C. briggsae* and *C. remanei* was done using the FamilyRelations II program (Brown et al. 2005) (http://family.caltech.edu/), using the same regions as above. The sequence identity stringency was set to 90% and 96% with the triple filter on and using a 10bp search window.

3.2.3 Transgene construction and strains

See section 2.2.3 Materials and methods for a detailed description of transgene construction. The right primer remained consistent (aqp-8R: agt cga cct gca ggc atg caa gct tag aaa cgg atc gca gaa aa) for the *aqp-8* 5' promoter truncation constructs and the site-directed mutagenized constructs. For a list of primers used for generating *aqp-8* 5' promoter truncation constructs and regulatory site mutagenized constructs and the resulting strains refer to Appendix 5 and Appendix 6, respectively.

3.3 Results

3.3.1 *aqp-8* promoter region analysis via sequential deletion constructs

The original promoter region used for expression pattern analysis of *aqp-8* was defined as a fragment spanning -1,575bp to -20bp upstream of *aqp-8*'s TSS. For further studies of this region, a -711bp to -20bp fragment upstream of the TSS construct was used for reporter fusion studies in the previous section. These shorter constructs, including a stabilized transgenic line, drove expression patterns identical to the original promoter constructs. The -711bp to -20bp lines narrowed the search window for excretory cell

modulating regions in *aqp-8*'s promoter region essentially by about one half. From the -711bp starting point, I generated a series of 5' *aqp-8* promoter truncation primers in an unbiased manner to generate sequential deletions to attempt to narrow the region(s) containing DNA elements necessary for modulating excretory cell expression (Figure 11). The 3' end of the promoter region (-20bp) was consistent for all promoter truncation constructs.

The shortest *aqp-8* 5' truncated promoter element which drove expression in the excretory cell, in this initial survey, was one that consisted of the fragment -279bp to -20bp (*aqp-8*$^{promoter(-279)}$*::GFP*). The next candidate promoter region downstream was a construct consisting of the fragment -261bp to -20bp (*aqp-8*$^{promoter(-261)}$*::GFP*). (*aqp-8*^{promoter}*GFP(-PEST)*). Transgenic lines were generated for the remaining smaller 5' promoter truncation constructs to confirm that excretory cell expression did not return upon further 5' deletions of the gene-upstream region. None of the shorter 5' truncated constructs could restore excretory cell expression or drive expression in any other tissues indicating that other excretory cell modulating enhancers and repressors are likely not present downstream of the -261bp promoter truncation fragment. The results of the analysis of the promoter fragments indicates the presence of a *cis*-regulatory element located downstream of the -279bp site and either upstream of or including the -261bp site in *aqp-8*'s promoter region.

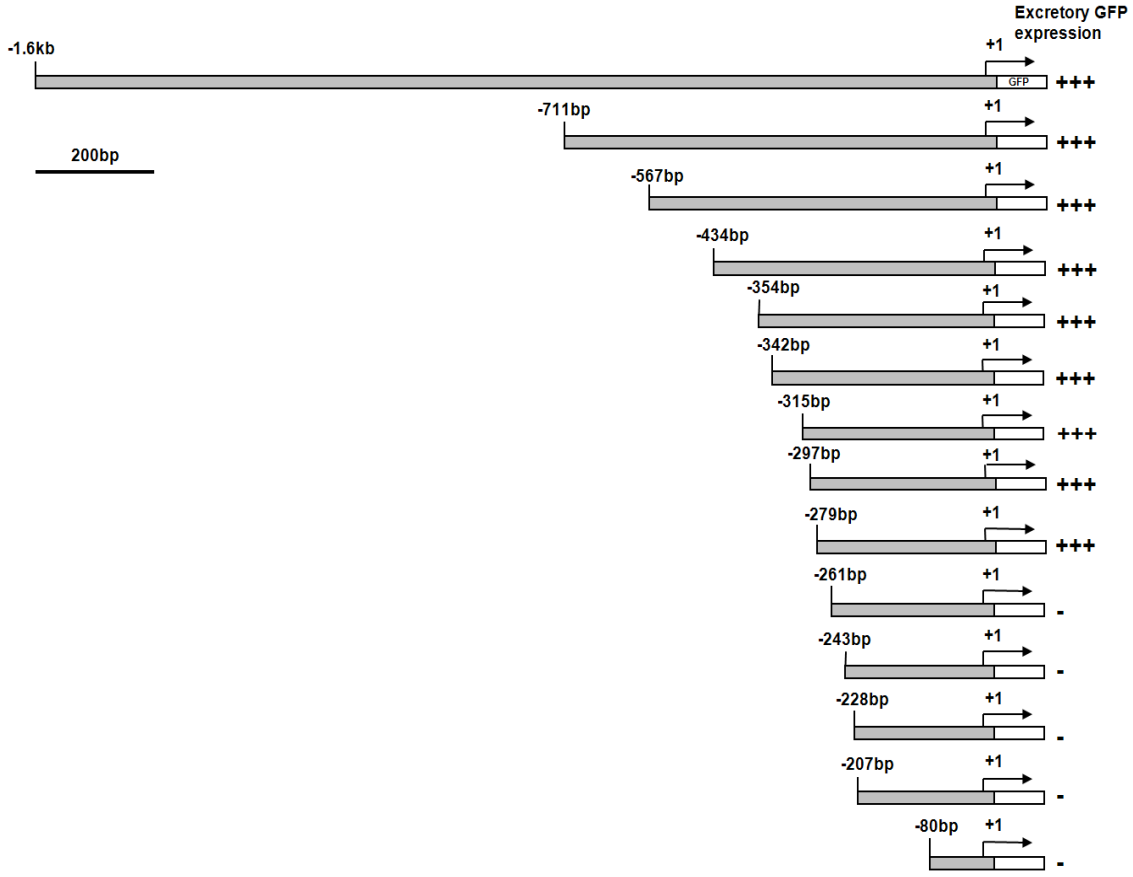

Figure 11. Deletional analysis of the *aqp-8* promoter region.

The promoter region of *aqp-8* was truncated in an unbiased manner. When available, more than one segregate line was assayed for transcriptional activity. Expression remained consistent until GFP expression was lost in the -261 lines. Since the -279 and -261 constructs were critical in this analysis, second independently isolated lines for these transgene were assayed to confirm their transcriptional activities.

3.3.2 Phylogenetic footprinting of the excretory cell modulating DNA region in *aqp-8*'s promoter

I identified evolutionarily conserved non-coding elements upstream of *aqp-8* in *C. elegans* by comparing a portion of the *C. elegans aqp-8* promoter to its orthologous regions in the other *Caenorhabditis* nematodes species: *C. briggsae*, *C. remanei*, *C. brenneri*, and *C. japonica*. The regions chosen for the promoter alignments were defined as the region 550bp upstream of the TSS to 50bp downstream of the TSS for each of the *aqp-8* orthologs (sequences were obtained from WormBase release WS190). The multiple sequence alignments were generated using ClustalX (Thompson et al. 1997). Parameters in ClustalX were changed from the default set to accommodate comparison of regions further upstream of the gene's translational start site. The gap opening parameter was changed to 5.00 and the gap extension parameter was changed to 4.00. These settings relaxed the conditions of gap opening and gap extension to provide for expansion or contraction of DNA regions where sequence conservation starts breaking down.

The result of this alignment shows that there are five regions within the 550bp upstream flanking region of the TSS that consist of greater than six perfectly conserved residues (Figure 12). Removal of the evolutionarily most distant of the five strains, *C. japonica* sequence (Kiontke and Fitch 2005; Cutter 2008), from the comparison expands some of the perfectly aligned regions and also identifies two additional regions of greater than six perfectly conserved residues (data not shown). In both multiple alignments, the conserved region most distant from the TSS is identical and consists of the sequence, AATTTGCATA. This sequence corresponds to a 10bp sequence starting at position -269bp in the *C. elegans aqp-8* gene-upstream region (hereafter referred to as the putative *cis*-regulatory element).

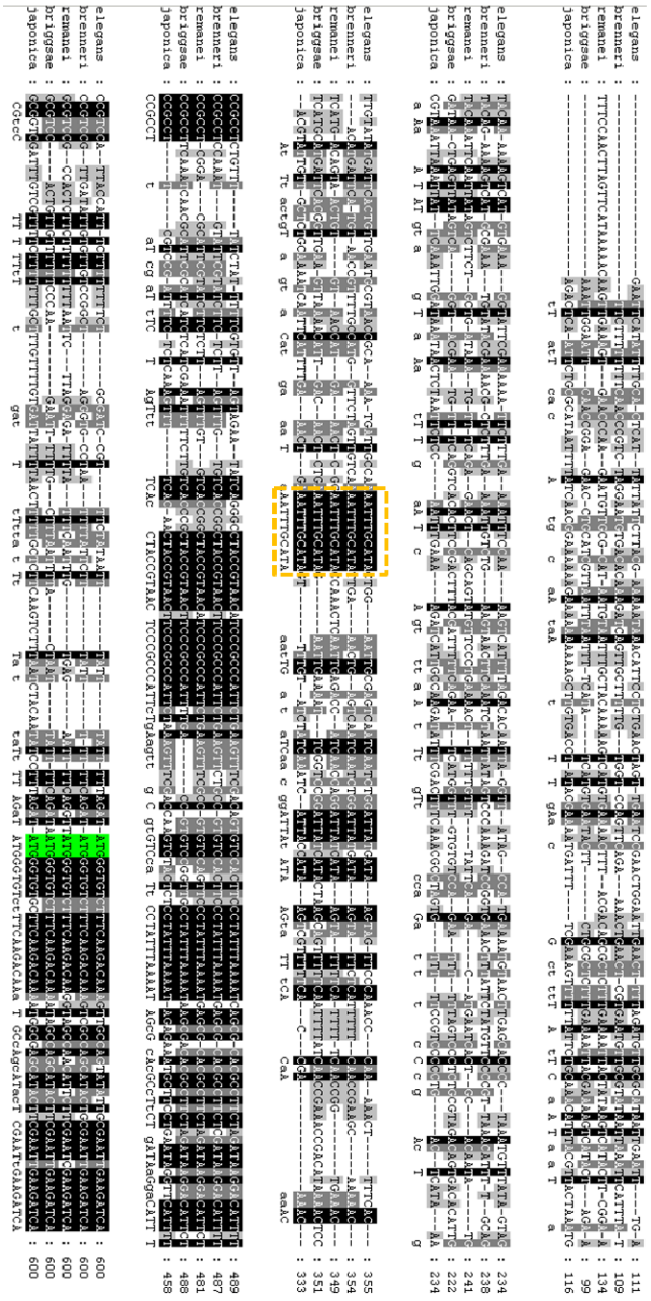

Figure 12. Multiple sequence alignment of orthologous upstream regions of *aqp-8*.

The regions corresponding to -550→+50bp (relative to the TSS; green) were selected from the five nematode species: *C. elegans, C. briggsae*, *C. remanei*, *C. brenneri*, and *C. japonica*. The region within the orange box corresponds to position -269bp in *C. elegans*, -270bp in *C. brenneri*, -275bp in *C. japonica*, -276bp in *C. remanei*, and -284bp in *C. briggsae.*

The putative *cis*-regulatory element appears to also have a positional constraint in relation to the translational start site. The elements are located between 284bp to 269bp upstream of the translational start site in all five species with an average of about 275bp upstream.

Analysis of the same regions (-550bp → +50bp) in *C. elegans*, *C. briggsae* and *C. remanei* using the FamilyRelations II program (Brown et al. 2002) shows that the region does not contain any inverted conserved sequences in the region that affects excretory cell expression even with a low stringency (Figure 13 *A*). Increasing the stringency of the comparison reveals only sites that have complete conservation (Figure 13 *B*). Looking closer at the alignment shows that the second most distal site from the translational start site corresponds to the site identified in the ClustalX analysis (Figure 13 *C*).

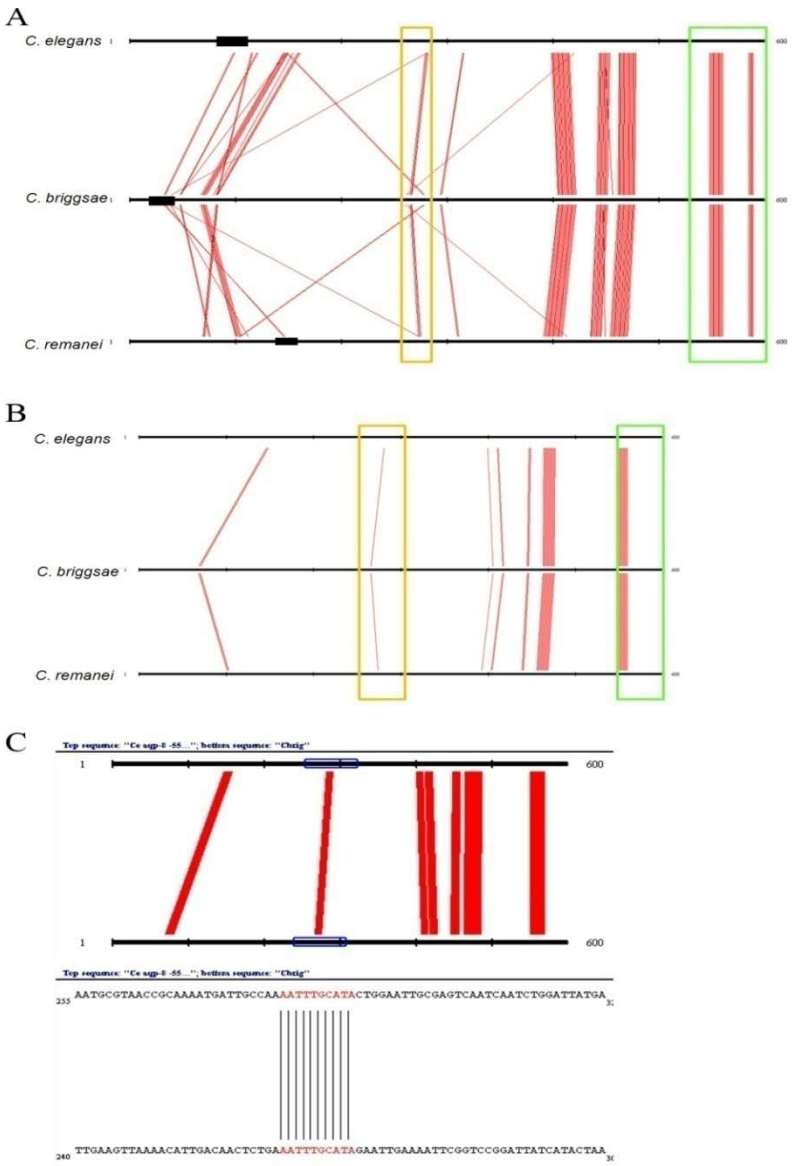

Figure 13. Interspecific comparison of *C. elegans*, *C. briggsae*, and *C. remanei* - 550→+50bp regions.

Comparison of the -550→+50bp regions relative to the TSS (green) of *C. elegans*, *C. briggsae*, and *C. remanei* was compared using FamilyRelations II (Brown et al. 2002) to detect conserved sequences that might not be represented when using ClustalX (inverted sequences and sequences that are not in the same order). *A*, using a 90% stringency and 10bp window, conserved inverted sequences were identified (black bars) upstream of the *C. elegans* -267bp site (orange). *B*, increasing the stringency of the comparison to 100% removes most background while the -267bp site remains as the second most distal conserved element. *C*, close-up of -267bp site in *C. elegans* vs. -284bp site in *C. briggsae*.

3.3.3 Determination of *cis*-regulatory element function

The position of the -269bp evolutionarily conserved site falls within the region delimited by the *aqp-8* promoter truncation constructs (-279bp → -261bp window). Two additional *aqp-8*^{promoter}*::GFP* constructs were generated to confirm the site's function. The first construct consisted of a promoter element with the 5'-end terminating at -272bp (*aqp-8*$^{promoter(-272)}$*::GFP*). This construct contained the entire ten base-pair putative *cis*-regulatory element. The second construct consisted of a promoter element with the 5'-end terminating at -267bp upstream of *aqp-8*'s TSS (*aqp-8*$^{promoter(-272)}$*::GFP*) which excludes the last two nucleotides of the conserved element. The *aqp-8*$^{promoter(-272)}$*::GFP* construct drove GFP expression within the excretory cell whereas the shorter, *aqp-8*$^{promoter(-267)}$*::GFP* construct, failed to drive expression of the GFP-coding cassette. The resolution provided by this targeted 5' promoter-truncation analysis provided strong evidence that the putative *cis*-regulatory element is required for directing *aqp-8*'s expression in the excretory cell.

3.3.4 Analysis of the putative motif by mutagenesis studies

Mutagenesis of two residues within putative *cis*-regulatory element was performed to determine whether its sequence must be completely conserved to drive gene-expression in the excretory cell. The underlined residues in the sequence AATTTGCATA are located at -264bp and -263bp respectively and were the residues targeted for mutagenesis. All possible single base substitutions of these residues were tested. The constructs were fused to the *GFP*-coding region of *pPD95.67* to assay for any variation in the expression patterns resulting from the residue change. The 5'-end of each construct in the mutagenesis study was defined as -276bp upstream of the *aqp-8*'s TSS since a

shorter construct starting at -272bp (*aqp-8*$^{promoter(-272)}$*::GFP*) was sufficient to drive expression in the excretory cell.

The expression patterns driven by the mutagenized promoter constructs were compared to a reference transgenic carrying an approximately 1.6kb promoter region fused to GFP (Figure 14*A*) (GFP cassette - *pPD95.67*; *C. elegans strain* – BC20052). The -264 G→A residue change in the construct *aqp-8*$^{promoter(-264G→A)}$*::GFP* did not alter the expression level or pattern to the expression pattern derived from the *aqp-8*$^{promoter(-272)}$*::GFP* construct (Figure 14*B*). A G→T change at the same site *aqp-8*$^{promoter(-264G→T)}$*::GFP* led to a significant decrease in the expression level (Figure 14*C*). Even with the decrease in expression, the GFP expression pattern remained localized to the excretory cell. The G→C substitution (*aqp-8*$^{promoter(-264G→C)}$*::GFP*) led to a loss of the ability of the promoter to drive expression of the reporter.

The C→A residue substitution at position -263bp upstream of *aqp-8*'s TSS, in the construct *aqp-8*$^{promoter(-263C→A)}$*::GFP*, led to a lower level of GFP expression, which remained localized to the excretory cell (Figure 14*D*). Both the C→G and C→T substitutions at position -263bp, using the transgene constructs *aqp-8*$^{promoter(-263C→G)}$*::GFP* and *aqp-8*$^{promoter(-263C→T)}$*::GFP*, led to complete loss of visible GFP expression.

I also analyzed a transgene construct containing a two-residue change in the conserved region. The changes consisted of substitution of the GC pair at -264 to an AG pair resulting in the *aqp-8*$^{promoter(-264GC→AG)}$*::GFP* construct (Figure 14*E*). The two-residue substitution led to low-level expression localized only in the excretory cell. For a summary of the effects of all mutagenized motif constructs, see Table 7.

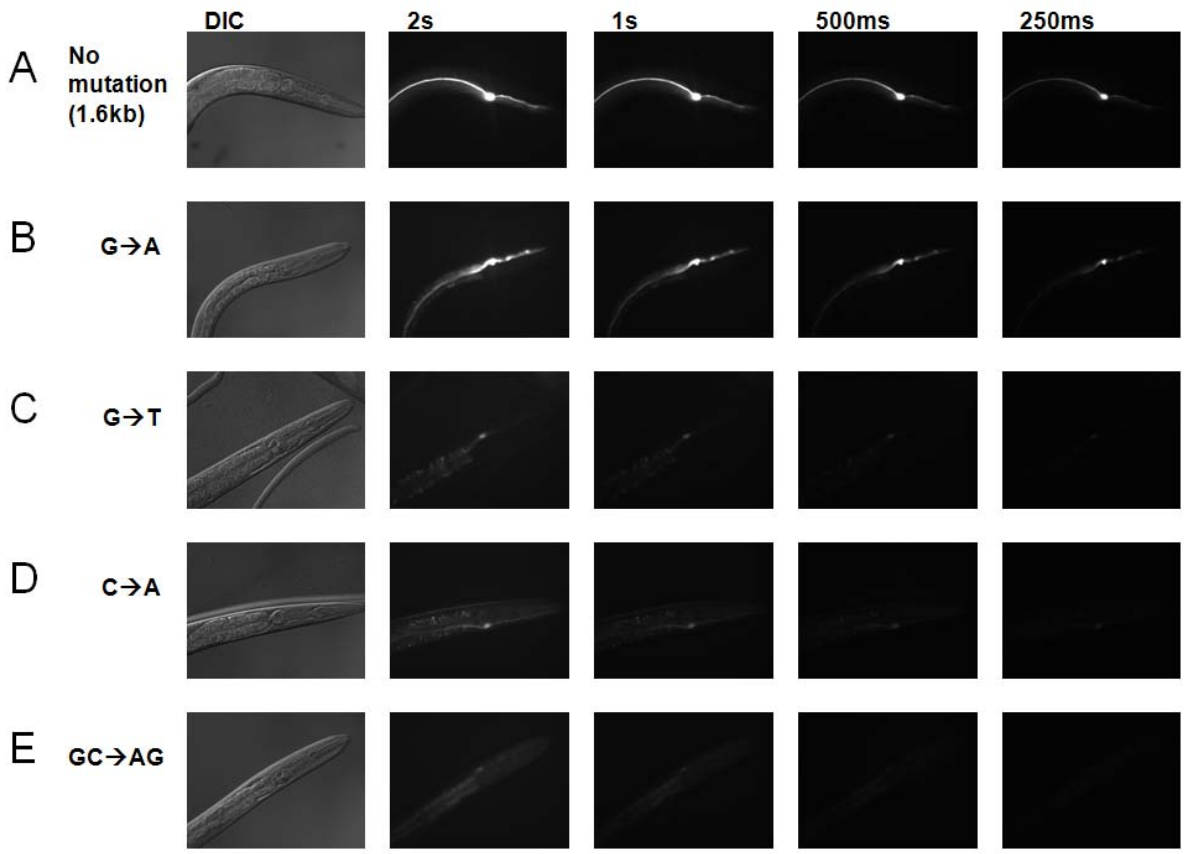

Figure 14. Effects of conserved element mutagenesis towards expression levels.

The results of site-directed mutagenesis of the -264G and -263C sites upstream of *aqp-8*. Expression levels were either completely unaffected, diminished, or completely abolished. *A*, *aqp-8*^{promoter}*::GFP* reference strain. *B*, a -264G→A change led to no change in expression level. *C*, *D*, *E*, Mutations in the form of -264G→T, -263C→A, and -264GC→AG all lead to appreciable loss of resultant GFP expression.

Table 7. Summary of effects of conserved motif mutagenesis on excretory cell expression levels.

Position relative to TSS	Residue(s) change	Effect on expression
-264	G>A	expression
-264	G>T	weak expression
-264	G>C	no expression
-263	C>G	no expression
-263	C>A	weak expression
-263	C>T	no expression
-263	GC>AG	weak expression

3.4 Discussion

A previous study has described the binding of DCP-66 (*Deacetylase Complex Protein – 66*) to the *cis*-regulatory element Ex-1 is required for excretory cell expression of *pgp-12* (*P-GlycoProtein related – 12*) in adult nematodes. Endogenous PGP-12 is expressedat a high level in the excretory cell during all worm stages (Zhao et al. 2005). The transcription factor, DCP-66, is an ortholog of the p66 member of the nucleosomal remodelling and deacetylase (NuRD) complex which plays a role histone deacetylation. Genes in the NuRD complex are broadly expressed in human tissues (Denslow and Wade 2007). Histone deacetylation tends to lead to silencing of adjacent genes via condensation of the nearby regions making the DCP-66 gene-activating mechanism described in the previous study paradoxical. Three additional elements in *pgp-12*'s promoter region function to modulate the levels of expression but not its spatial pattern**.**

The loss of the Ex-1 element resulted in an overall decrease in the expression level and loss of reporter-gene expression in embryos (Zhao et al. 2005). A scan of promoter regions of excretory cell-expressing genes shows that four of the eight sampled promoter regions contain the Ex-1 motif. Additionally, not including *pgp-12*'s promoter, the other positive hits only managed 80% identity in the Ex-1 region.

I set out to find additional mechanisms governing transcriptional regulation in the excretory cell using a different promoter region to complement this previous study. The promoter region of *aqp-8* is an ideal system for determination of these elements because it drives strong expression exclusively in the excretory cell like *pgp-12* used in the previous study. Prior analyses, from the survey of all *C. elegans* aquaporin expression patterns (previous chapter), indicated the presence of an excretory cell-expression

modulating *cis*-regulatory element located between p the positions -711bp to -20bp upstream of *aqp-8*'s translational start site. 5' deletion constructs of *aqp-8*'s promoter region revealed an approximately twenty base-pair window upstream of *aqp-8*'s translational start site, defined by the constructs *aqp-8*$^{promoter(-279)}$*::GFP* and *aqp-8*$^{promoter(-261)}$*::GFP,* that contains element(s) which are involved in driving expression of AQP-8 in the excretory cell. Further 5' truncations of *aqp-8*'s promoter did not lead to appearance of any expression.

Since two other nematode species (*C. briggsae* and *C. remanei*) have genes corresponding to orthologs of all 12 *C. elegans* aquaporins in their genomes, and also given that developmental programs are similar between these three nematode species (Zhao et al. 2008; Bao et al. 2008), each ortholog group should perform similar functions *in vivo.* Mechanisms regulating these genes are probably similar or the same mechanisms driving *aqp* expression patterns of the ortholog between the different *Caenorhabditis* species. I performed a multiple alignment of *aqp-8*'s orthologous promoters using the five *Caenorhabditis* species with sequenced genomes to find conserved non-coding regions. The alignment revealed a ten base pair conserved sequence starting at position -269bp in the *C. elegans aqp-8* promoter. The distance of this putative element agreed with the window containing the functional DNA element identified in the deletion analysis. The perfect conservation of an element within this region provides additional evidence that there is indeed a *cis*-regulatory element present and that this regulatory mechanism has been preserved between the *Caenorhabditis* species with sequenced genomes.

SOLiD core alignments (http://genome.ucsc.edu/cgi-bin/hgGateway), display representations of nucleosome positioning density at genomic loci (Valouev et al. 2008). The region roughly 300bp upstream of *aqp-8*'s translational start has a conserved non-coding region in addition to a low nucleosome position stringency (NSome) score (Figure 15 *A,B*). The chromatin used to determine the Nsome scores was isolated from mixed stage worms and therefore is representative of overall preference of nucleosome occupancy. The NSome stringency score represents the degree of nucleosome positioning at the corresponding genomic position. Nucleosomes are composed of 147bp of DNA wound around the histone octamer with an average periodicity of 175bp (Luger et al. 1997; Valouev et al. 2008). The positioning of nucleosomes along the genome plays a large part in the spatial accessibility of the sequence element to DNA-binding proteins. Since active *cis*-regulatory elements are usually depleted of nucleosomes, non-nucleosome occupied DNA sites has been used as a basis to predict transcription factor binding sites in yeast (Narlikar, Gordan, and Hartemink 2007). The same logic can be extended to looking at potential *cis*-regulatory sites in *C. elegans.*

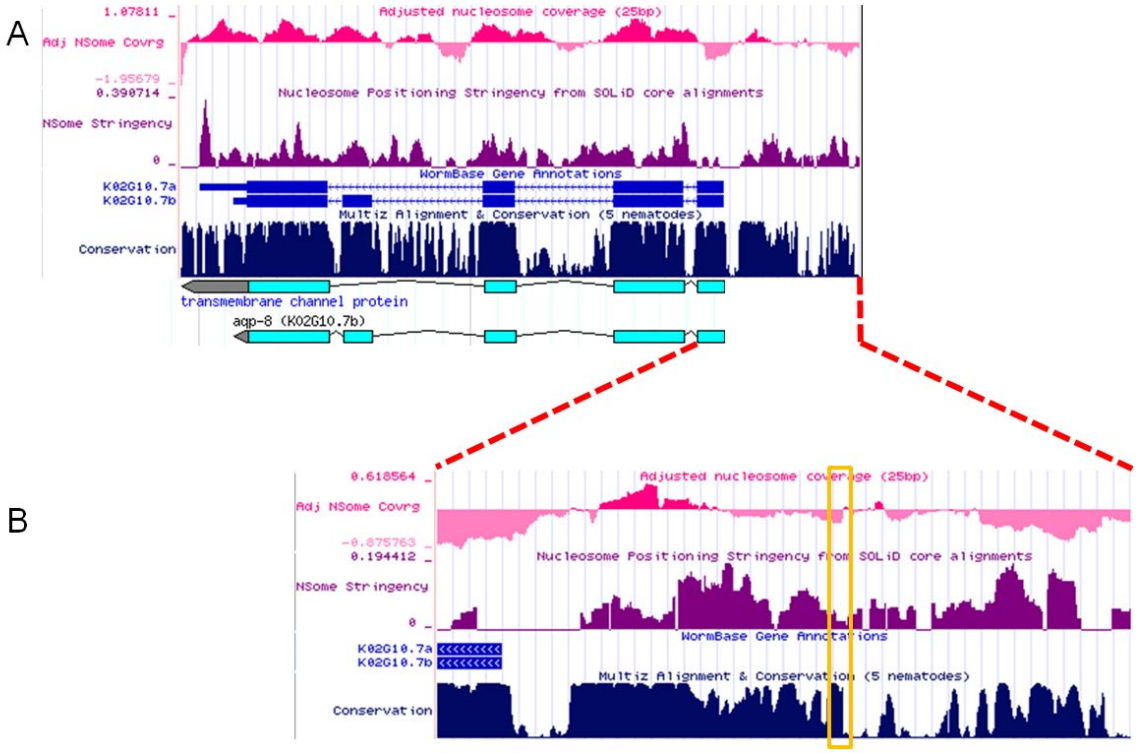

Figure 15. Nucleosome positioning stringency and sequence conservation at -269bp of *C. elegans aqp-8*.

A, Coding region → -550bp region of *aqp-8*. The coding sequences , along with an upstream block have high degrees of sequence conservation. Nucleosome coverage downstream of the translational start site has an expected nucleosome coverage periodicity (top; pink track) for genomic regions that are not required to be accessible to *trans*-acting factors.
B, First intron → -550bp of *aqp-8.* The -269bp element (orange box) corresponds to a region of high sequence conservation (blue track) and low NSome stringency (purple track) indicating that this region is generally "open" to *trans*-acting factors.
Images are modified from (http://genome.ucsc.edu/cgi-bin/hgGateway)

The distances of the putative *cis*-regulatory elements from the translational start site in the promoters of the *aqp-8* orthologs had a small variance. The distances ranged from extremes of -269bp in C. *elegans* to -284bp in *C. briggsae* with an average distance of -275bp among the five *Caenorhabditis* species. The tight grouping of the regulatory site's distance from the translational start among the orthologs might indicate that its location and possibly its arrangement relative in relation to other *cis*-regulatory elements might be subject to functional constraints. Distance and topological restriction of *cis*-elements are observed in other promoter regions. One case of distance conservation is the X-box *cis*-regulatory elements which cluster roughly 100bp upstream of gene translational start sites (Efimenko et al. 2005). Distance dependent enhancer element effects gene has been shown in adenovirus, in which moving a GC-box *cis*-regulatory element (cognate transcription factor Sp1) further from the TATA box led to a decrease in expression levels even though Sp1 bound to the DNA with approximately the same affinity (Wu and Berk 1988). The β-actin promoter is an example of transcription factor binding site clustering topography affecting gene expression. The β-actin promoter contains two *cis*-acting sequences, which bind different transcription factors independently. These two *cis*-regulatory sequences must be constrained spatially in order to elicit a transcriptional response. When the inter-sequence spacing between the *cis*-regulatory elements were contracted or expanded, detrimental effects on the levels of transcription were observed (Danilition et al. 1991).

Although the putative *cis*-regulatory region appeared to require absolute sequence conservation, as shown by the multiple alignments of the promoter regions, I addressed the effects of perturbing specific sites within the putative element via targeted

mutagenesis and fusing of these fragments to a *GFP*-coding cassette. Changes to the putative element indicate that the guanine residue at position -264bp upstream of the TSS might only be a mildly important site for driving excretory cell expression. Two different residue changes at this position did not affect the anatomical expression pattern although one change did lead to a reduced level of expression as reflected by the -246G→T change.

I performed the single-residue mutation analysis at the downstream adjacent residue (residue -263bp upstream of the TSS). The only change at this position that did not result in a loss of ability to drive expression in the excretory cell was a -263C→A change. The mutation caused the construct to drive expression in the excretory cell at a reduced level compared to the unmutated putative *cis*-regulatory element.

A final mutated motif construct was generated consisting of a dual residue replacement in the form of aatttGCata→atttAGata (-264GC→AG). The result of this change was a construct that could still drive expression in the excretory cell, but at a lower level than the reference construct. Recall that the single site mutation, in the form of 263C→A, eliminated the promoter fragment's ability to drive visible expression whereas the single site mutation consisting of -264G→A did not affect the expression levels or pattern when compared to *aqp-8*^{promoter}*::GFP* constructs containing whole and unmutated putative *cis*-regulatory motifs.

Since the dual residue mutation in the putative element drove an identical spatial expression pattern to the unmutated *aqp-8*^{promoter}*::GFP* constructs, I concluded that the double mutant still has sufficient ability to recruit the DNA-binding proteins, but in a reduced capacity than the unmutated element, for the purpose of driving excretory cell

expression. Although the putative *cis*-regulatory site is perfectly conserved among the five genome-sequenced *Caenorhabditis* species in the *aqp-8* promoter region, the alternate versions of the *cis*-regulatory sequences might represent a mechanism for an additional level of transcriptional regulation. Identification of the complementary binding proteins to this sequence will provide more insight as to how these residue changes affect the *cis*-regulatory element's ability to bind the *trans*-acting factor.

In this section, I have used *aqp-8*'s gene-upstream promoter as a representative model for genes, which express within the excretory cell. Using this region, I identified a site containing a *cis*-regulatory element required for excretory cell expression. The characterisation of this DNA element will provide a basis for the determination of *trans*-acting factors, which can bind to this sequence to regulate transcription of *aqp-8* and possibly other genes in the excretory cell.

4: DETERMINATION OF COGNATE TRANSCRIPTION FACTOR FOR THE EXCRETORY CELL MODULATING *CIS*-REGULATORY ELEMENT IN THE *AQP-8* PROMOTER REGION

4.1 Introduction

The binding of DCP-66 to the EX-1 motif upstream of *pgp-12* is required for expression of the gene in the *C. elegans* excretory cell. DCP-66, an ortholog of the mammalian p66 protein is a component of the NuRD complex. This complex plays a role in histone deacetylation and chromatin remodeling. Histone deacetylation typically leads condensation of the nucleosome complexes due to the increase of positive charge to the histone tails. This encourages high-affinity binding between the histones and DNA backbone (Zupkovitz et al. 2006). The condensation, in turn, leads to silencing of genes within the affected region. Therefore, it appears that the mechanism that relies on positive regulation of genes in a DCP-66 dependent manner may be more of an exception rather than a common mode of regulation. I use *aqp-8*'s promoter region as a tool to determine additional mechanisms for gene modulation in the excretory cell. My analysis provides further information on how the excretory cell, a morphologically and functionally unique body structure, functions. Additionally, the mechanism(s) inferred from this analysis in nematodes should provide insight into proteins involved in kidney function.

Many transcription factors have been found to influence renal development and function. A sample of these transcription factors include: HNF1β, a homeodomain-

containing transcription factor, loss of which leads to several kidney abnormalities, including the formation of cysts, oligomeganephronia, renal agenesis, renal hypoplasia, and familial juvenile hyperuricemic nephropathy (Dudziak et al. 2008), Kid-1, a C2H2 class of zinc finger genes, which is transcriptionally regulated upon kidney trauma (Witzgall et al. 1993), and Ets-1, which transcriptionally regulates FREAC-4 a winged helix transcription in renal tissue (Cederberg et al. 1999).

In the previous chapter, I discovered a putative ten base-pair *cis*-regulatory element (AATTTGCATA) that is responsible, at least in part, for excretory cell expression of *aqp-8* in *C. elegans.* To date, there have been no previous studies which have characterized the function this DNA element in nematodes. The *cis*-regulatory element resides 269bp upstream of *aqp-8*'s translational start site in *C. elegans* and is highly conserved among five *Caenorhabditis* species with sequenced genomes (WormBase, WS190). I tested the function of this site using constructs which affect the conserved element and also by mutating certain residues within the element. These constructs were then tested for their ability to ability to drive expression in the excretory cell. With the certainty that at least a portion of the conserved DNA region is a *bona fide* *cis*-regulatory element; the next challenge was to determine *trans*-acting factors.

In this chapter, I describe the identification of a cognate transcription factor which binds to the putative *cis*-regulatory element located in *aqp-8*'s promoter region. I tested the affinity, *in vitro*, of the putative *cis*-regulatory element for *C. elegans* proteins by performing electrophoretic mobility shift assays (EMSA). This method allows for the visualization of nucleic acid-protein interactions by resolving specific DNA/protein interactions as slower moving bands relative to unbound DNA. The use of different

cellular fractions as input in the EMSA reaction also provides additional insight regarding mechanisms of DNA-binding protein regulation.

I also tested whether the putative *cis*-regulatory element can drive expression outside of the other factors associated with *aqp-8′*s promoter region. This provides information as to whether the element can work alone or if it requires the cooperation of other factors in *aqp-8′*s promoter region to drive expression in the excretory cell. For this purpose, I used a well established system for assaying *cis*-regulatory function, assessing the ability of the putative *cis*-regulatory element to drive expression aided only by a minimal promoter (*Δpes-10*, clone *pPD97.78*). A second type of chimeric promoter construct based upon the *vit-2* promoter region was used to determine whether positional effects are important for the *cis*-regulatory element.

Computational database searches, using the putative *cis*-regulatory element as a query, were performed. The results of this search provided a set of candidate transcription factor types which provided a preliminary understanding of the types of proteins which bind to the putative *cis*-regulatory sequence. The computational screen identified POU homeobox transcription factors as the most likely cognate transcription factors. The homeobox transcription factor family represents a diverse range of proteins as reports of between five to seven classes of homeobox genes have been published (Banerjee-Basu, Sink, and Baxevanis 2001; Holland, Booth, and Bruford 2007).

The POU homeobox transcription factors are bipartite transcription factor proteins that usually have roles in development. The bipartite region is composed of two distinct domains: the C-terminal-most homeodomain region (POU_{HB}) and the N-terminal-most POU-specific motif (POU_S), which was first identified in the genes: Pit-1, Oct-1/2, and

UNC-86 (Herr et al. 1988). The POU_S and the POU_{HB} regions are separated by a short flexible linker region.

The computational search also identified residues within the ten-base pair putative *cis*-regulatory element (AATTTGCATA) that are important for the docking of these transcription factors. The nested sequence (ATTTGCAT), often referred to as the octamer motif or element, has been extensively characterized in other organisms. For the remainder of this thesis, I refer to ATTTGCAT as the forward octamer configuration and its reverse complementary sequence, ATGCAAAT, as the reverse configuration, although both are referred to as the octamer element in literature.

There are three well-characterized POU homeobox transcription factors encoded in the *C. elegans* genome (Table 8). Their structures are typical of POU homeobox transcription factor structure with a single POU_S sub-domain followed by a single POU_{HD} sub-domain. Three predicted POU members are also present in the genome (Table 9). Two of these genes have had their expression-pattern analyzed and one has no prior analysis. Transcriptional reporter-gene expressing transgenics for each POU homeobox transcription factor member in *C. elegans* which did not have prior expression pattern analysis were generated and analyzed for its expression pattern. The POU promoter regions which drove excretory cell expression were considered the best candidates for being the corresponding transcription factor for the octamer element in *aqp-8*'s promoter region.

Table 8. POU transcription factors in *C. elegans.*

Gene Public Name	Sequence Name (Gene)	Location	Peptide Length (AAs)
ceh-18	*ZC64.3*	X:3843354..3857778	542
ceh-6	*K02B12.1*	I:8498107..8502405	380
unc-86	*C30A5.7*	III:8212678..8216546	429

Table 9. Additional predicted POU TFs in *C. elegans.*

Sequence Name (Gene)	Location	Peptide Length (AAs)
F45C12.15	X:14141868..14144189	175
F45C12.15	II:17337191736536	543
Y38E10A.6	II:1258209412603594	1345

I identified CEH-6, a well characterized POU transcription factor candidate, as the cognate transcription factor for the octamer motif found in *aqp-8's* promoter region. This was proven by RNAi-mediated gene knock-down of *ceh-6* in an AQP-8::GFP background. I confirmed the results of the *in vivo* analysis with an *in vitro* assay of the binding specificity of CEH-6 to the octamer motif. Finally, I assessed the ability of the reverse orientation of the octamer motif to drive expression via the analysis of a chimeric promoter construct.

4.2 Materials and Methods

4.2.1 Electrophoretic Mobility Shift Assay (EMSA)

Plate-grown (NGM) *N2* worms were harvested in M9 buffer and centrifuged to acquire approximately 100µl of wet packed nematodes. The worms were washed, lightly centrifuged and aspirated an additional five times in 1/10 v/v ratio of phosphate buffered solution (PBS) to remove excess bacteria from the packed worm pellet. The worms were re-suspended in a final volume of 200µl PBS and Dounce homogenized using five strokes. Nuclear and cytoplasmic extracts were isolated from this preparation using the NE-PER Nuclear and Cytoplasmic Extraction Reagents Kit (Pierce, Cat No: 78833). The synthetic biotinylated oligonucleotides used in this study contain the consensus octamer oligonucleotide along with short flanking regions on either end: OctFBio - 5'-ATTGCCAAAATTTGCATACTGGAAT-3' and its complement OctRBio 5'-ATTCCAGTATGCAAATTTTGGCAAT-3'. Unlabeled versions of these oligos were also used and labeled as OctFnonBio and OctRnonBio respectively. EMSA reactions were carried out using the LightShift Chemiluminescent EMSA Kit (Pierce, Cat No: 20148). Samples were loaded into an 8% non-denaturing polyacrylamide gel and

electrophoresed in pre-chilled 0.5X Tris/Borate buffer at 100 V for 1 h. The entire gel electrophoresis apparatus was kept cool in an ice-bath during operation. The samples were electro-blotted onto nylon filters (Bio-Rad Transblot Cell) in 0.5X Tris/Borate buffer at 100 V for 30 minutes and cross-linked to the membranes using a UV Stratalinker 2400 (Stratagene) using the autocrosslink function. Visualization of the gel was carried out using the LightShift Chemiluminescent EMSA Kit (Pierce, Cat No: 20148). For supershift EMSA, cellular extracts were pre-incubated for 30 minutes at room temperature with CEH-6 specific antibodies (kindly provided by (Burglin and Ruvkun 2001)). For the cross-competition (unrelated competition), the following unlabeled oligos were used let-721RepF - 5'-TTT TGT CCC TCG TGG GAG ACA CAT-3' and let-721RepR - 5'-ATG TGT CTC CCA CGA GGG ACA AAA-3'.

All oligonucleotide-pairs used for EMSAs were annealed by mixing equal molar amounts of each fragment in 1XTE buffer. The mixture was heated to 65^{o}C and allowed to cool at a rate of 0.1^{o}C/s until it reached to 4^{o}C.

4.2.2 Transgene construction

See section 2.2.3 Materials and methods for a description on transgene construction.

4.2.3 Ectopic expression constructs

Primers used for tandem forward motif fusion to Δ*pes-10::GFP* are as follows: 4XOCTR: agt cga cct gca ggc atg caa gct tat gca aat tta tgc aaa ttt a and 4XOCTL: aat ttg cat aaa ttt gca taa att tgc ata aat ttg cat a, representing the reverse complement and forward configurations respectively. Primers used for tandem reverse motif fusion to Δ*pes-10::GFP* are as follows: 4XOCTR-AGT CGA CCT GCA GGC ATG CAA GCT

TAA ATT TGC ATA AAT TTG CAT A and 4XROCTL – TAT GCA AAT TTA TGC AAA TTT ATG CAA ATT TAT GCA AAT T.

The primers were added to a template-less PCR reaction using a standard cycling program, but with a short extension time (15 second extension). Fusion of this fragment to, Δ*pes-10::GFP* (*pPD97.78*, Δ*pes-10::GFP*-coding cassette) was carried out using the standard fusion PCR protocol as described in section 2.2.3. The resulting strains were: BC06979 *dpy-5(e907); sEx1345 rCes[4xAATTTGCATA::*Δ*pes-10::GFP+ pCeh361](SegI)* and BC07030 *dpy-5(e907); sEx1387rCes[4xAATTTGCATA::*Δ*pes-10::GFP]+ pCeh361](SegII)*.

The same method was used for the reverse octamer motif construct fused to Δ*pes-10::GFP*. The primers used for this were 4XrevOCTR- AGT CGA CCT GCA GGC ATG CAA GCT TAT GCA AAT TTA TGC AAA TTT A and 4XrevOCTL- AAT TTG CAT AAA TTT GCA TAA ATT TGC ATA AAT TTG CAT A. The resulting strains were: BC07816 *dpy-5(e907);sEx1710 rCes[4x(rev oct)TATGAAATT::* Δ*pes-10::GFP + pCeh361]*, BC07817 *dpy-5(e907);sEx1711 rCes[4x(rev oct)TATGAAATT: :*Δ*pes-10::GFP + pCeh361]*, and BC07818 *dpy-5(e907);sEx1712 rCes[4x(rev oct)TATGAAATT::* Δ*pes-10::GFP + pCeh361]*.

Three tandem repeats of AATTTGCATA were fused upstream of various 5' truncations of the *vit-2* promoter region (see Appendix 7 for left primer sequences and resulting strains). All downstream (right) primers were the same (vit2reverse- AGT CGA CCT GCA GGC ATG CAA GCT CGA CCT GAT GGC TGA ACC G).

4.2.4 Transcriptional element prediction

Prediction of putative transcriptional binding sites was done using the Transcriptional Element Search Software (TESS) (http://www.cbil.upenn.edu/cgi-bin/tess/tess) (Schug J 1996). This program searches predicts transcription factor binding sites based on weighted matrices from training set of data from the TRANSFAC, IMD and CBIL-GibbsMat databases. The string-based search query option was selected for the prediction using default parameters. Genomic regions containing putative regulatory sites identified by deletion analysis and phylogenetic footprinting were used as input.

4.2.5 Local BlastP of homeodomain regions

The homeodomain portions of all homeoboxes listed in the Homeodomain Resource were downloaded from (http://genome.nhgri.nix.gov/homeodomain/) (Assa-Munt et al. 1993) and used as the local BlastP database. The predicted POU_S sub-domain and POU_{HD} sub-domains of UNC-86, CEH-6, CEH-18, F28H6.2, Y38E10A.6, and F45C12.15 were independently used as queries against this database.

4.2.6 RNAi

K. Armstrong (Chamberlin Laboratory, Ohio State University) carried out this entire analysis. The RNAi injections were preformed according to method by (Fire et al. 1998). Portions of the coding regions for *ceh-6* (1.2kb) and *eri-1* (0.5kb) were *in vitro* transcribed from pBlueScript II vectors to generated double stranded RNA. 200 ng/ul of dsRNA corresponding to either *eri-1* (control) or a mixture of both *eri-1* and *ceh-6* dsRNA (experimental) were injected into the syncitial gonad of Adult BC6925 (*aqp-8::GFP* stabilized transgene) worms. The injected worms were transferred to plates to

recover for 24 hours before being transferred to fresh plates to lay offspring. Thirty progeny of the dsRNA injected worms were scored 48 hours post-injection (L2 stage) for the presence or absence of GFP fluorescence in the excretory cell using a standard image exposure time of one second with identical camera settings for all images. 15 lines were scored to ensure reproducibility of the result. F_1 progeny, from identically treated plates, were scored for developmental delay and lethality 72 hours post injection to confirm the efficiency of the *ceh-6* knock down (the scoring protocol is summarized in Figure 16).

Fluorescent and DIC microscopy were performed using a Zeiss Axioplan 2 microscope and images were captured using a Hammamatsu digital camera. Nematodes were immobilized using M9-agarose pads containing 0.03% sodium azide. Images were edited with Adobe Photoshop.

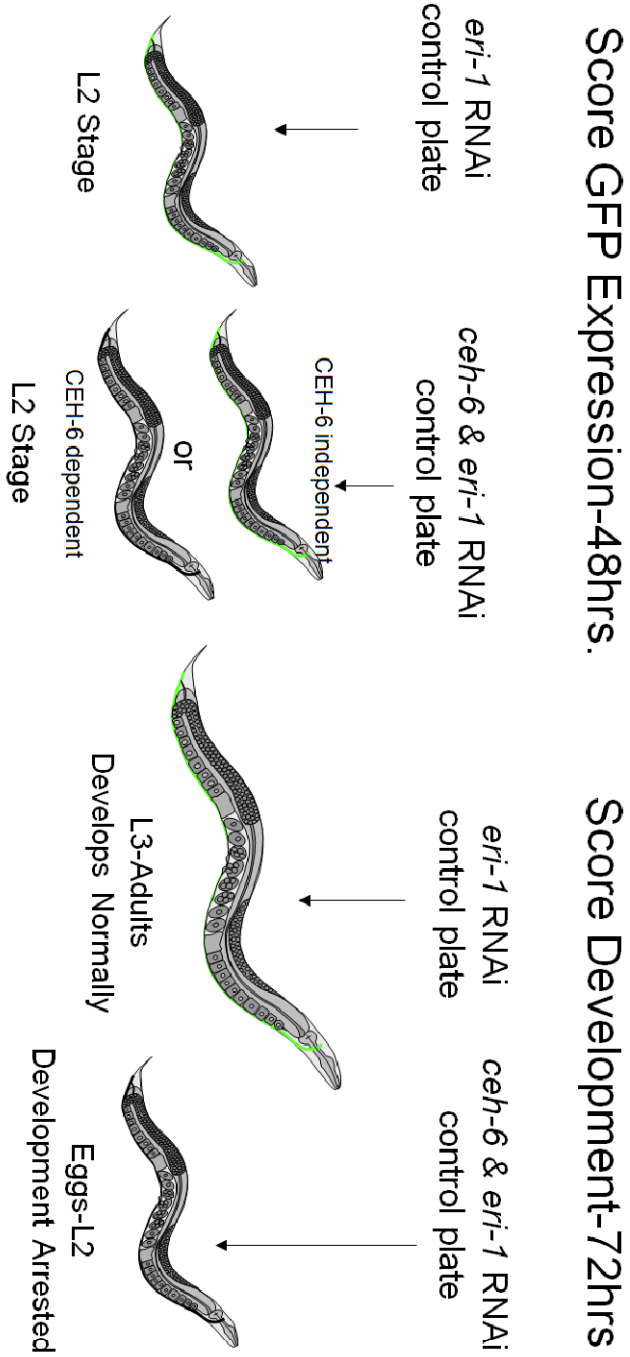

Figure 16. *ceh-6* RNAi screening protocol.

(*left*) worms were scored at after 48 hours at the L2 larval stage for presence or absence of GFP (green).
(*right*) after 72 hours the plates were screened again to determine the confirm the effects of *ceh-6* knockdown (L2 arrest and/or unhatched eggs).

4.3 Results

4.3.1 The sequence AATTTGCATA binds proteins *in vitro*

To determine if the putative *cis*-regulatory element has affinity for any *C. elegans* proteins *in vitro*, I performed EMSA using bait double-stranded (ds)DNA consisting of the putative motif and short 5' and 3' native flanking sequences from *aqp-8*'s promoter region. The bait probe was tested for binding interactions against both nuclear and cytoplasmic *C. elegans* mixed-stage cellular protein extract fractions. Proteins from both of the cellular extracts bound the dsDNA bait fragment as represented by slower migrating bands in a non-denaturing polyacrylamide gel (Figure 17 *A*, *B*). The same EMSA reactions were performed with the addition of a 1000-fold excess of unlabeled dsDNA (same sequence as probe) to determine whether general DNA binding proteins are binding the labelled probe or if the interactions between the labelled probe and protein(s) in the shifted bands are with a specific protein or protein complex. The loss of a shifted band when compared to the lane without competing probe reveals that, although there is some general DNA-binding protein interaction (as represented by the large amount of material remaining in the wells), there is a loss of a specific shifted band with either the nuclear or cytosolic protein extracts (Figure 17 *A*, *B*). The loss was a result of titration of the reaction by binding to the excess unlabelled competitor DNA fragment in incubations. In another test of DNA/protein interaction specificity, I performed an EMSA containing an excess molar amount of an unrelated unlabelled dsDNA probe. This sequence has been demonstrated to have weak specific nuclear protein affinity in addition to a stronger affinity to general DNA-binding protein(s) via a prior EMSA analysis (Figure 17 *C*). Despite the presence of this unrelated DNA, an interaction was still

observed suggesting that the binding interaction is between the DNA element to a specific protein or complex that is present in both cellular fractions tested (Figure 17 *D*).

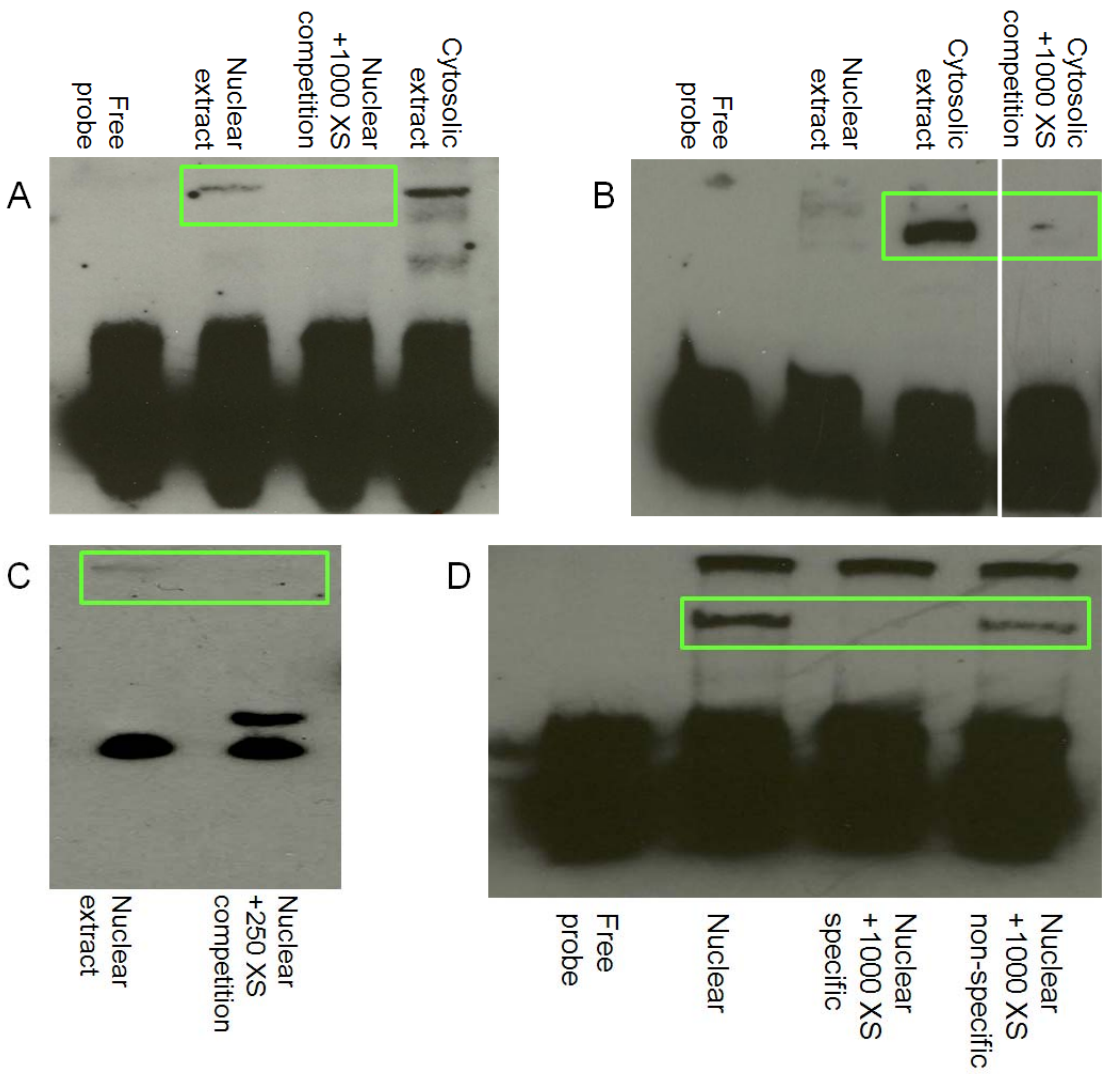

Figure 17. EMSA using conserved element.

A, The conserved element can bind nuclear proteins. The addition of excess unlabeled probe leads a competitive loss of binding of the labelled probe.
B, The conserved element can bind cytosolic proteins. The addition of excess unlabeled probe leads a competitive loss of binding of the labelled probe. The white line indicates that the last lane was cut from another part of the same gel and spliced alongside the lanes for presentation purposes.
C, EMSA of unrelated competitor shows that it has a mild affinity for DNA (top faint band) and that the reaction can be titrated with excess competition probe.
D, The binding interaction between the conserved element and binding protein cannot be titrated out with the dsDNA construct used in Figure 17 *C*. This indicates that the interacting protein is sequence specific.

4.3.2 The conserved element can drive expression ectopically

The conserved motif was tested for its ability to drive expression in the absence of additional DNA sequences found within *aqp-8*'s promoter region by fusing the conserved element to a generic promoter to generate a chimeric promoter construct. Four tandem repeats of the sequence (AATTTGCATA) were fused upstream of the *pes-10* minimal promoter in the *Δpes-10::GFP* construct and the resulting expression pattern was assayed. The construct drove expression of GFP in the excretory cells in addition to the AUA and AVH cells (Figure 18). This is in contrast to the expression pattern of the unmodified *Δpes-10::GFP* construct which only drives weak expression in intestinal cells and the pharynx (not shown) (Harfe and Fire 1998).

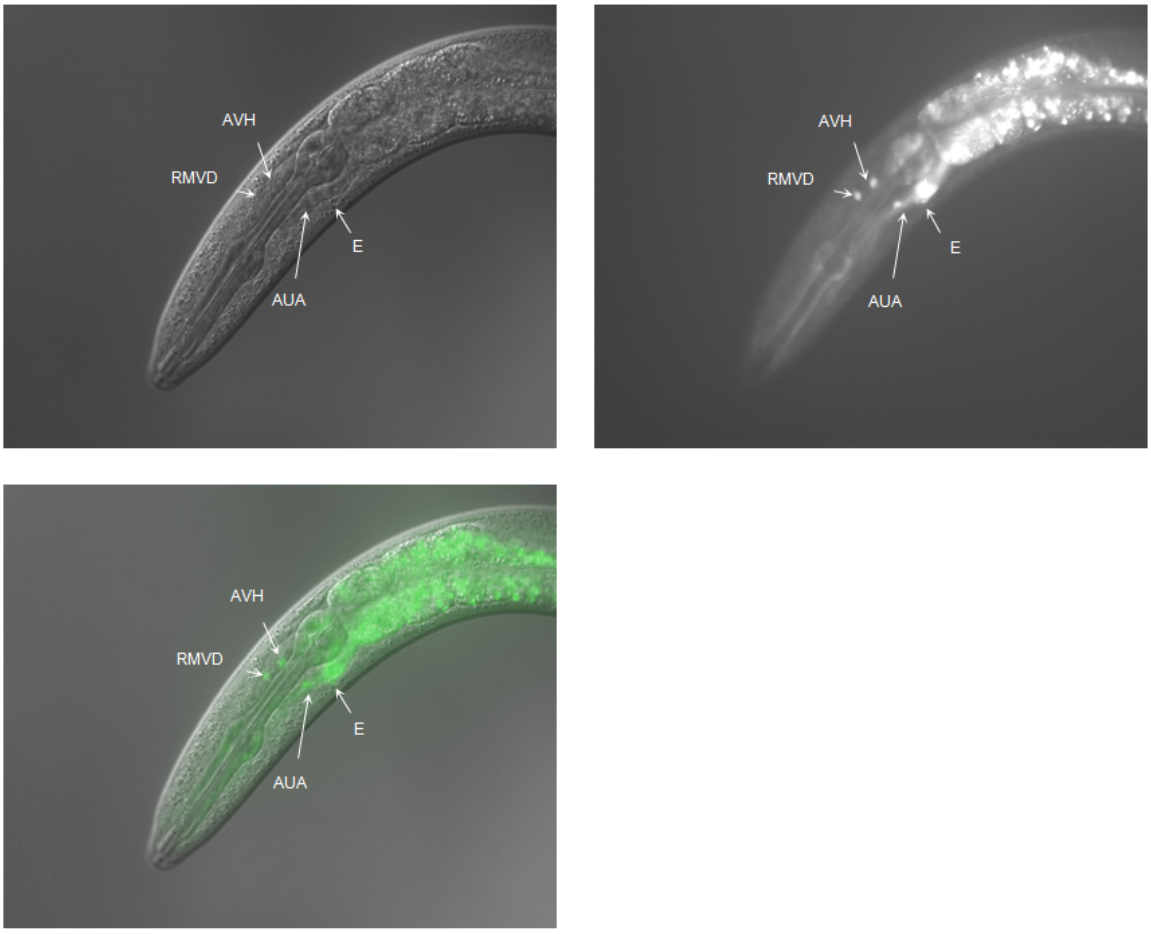

Figure 18. Conserved element fused to minimal promoter element.

Adult *C. elegans* expressing GFP under control of four tandem (AATTTGCATA) elements fused to a minimal promoter. Expression is mainly localized to the excretory cell (E). Expression is also driven in the AUA cell pair (L/R), the AVH cell (L/R), and the RMDV cell pair (L/R).

The images were captured at 400X magnification with a 3 second exposure time for the fluorescence channel. The focal plane only allows the visualization of one of the two neuron pairs.

4.3.3 Reverse complement of the octamer cannot drive expression ectopically

The reverse complement of the putative *cis*-regulatory element (TATGCAAATT) did not drive expression when fused in tandem at the 5' end of the *Δpes-10::GFP* construct. No GFP expression was detected in either of the two strains carrying this transgene construct (Images not shown).

4.3.4 The octamer sequence can drive promoter constructs at different upstream distances

For independent confirmation of the previous chimeric promoters, I used the *vit-2* promoter as a basis for generating chimeric promoter constructs consisting of a single copy of the putative regulatory element fused at various distances upstream of *vit-2*'s translational start site. VIT-2 is a vitellogenin structural protein (yolk protein). The unmodified *vit-2* promoter drives expression of VIT-2 in the intestine of the worm. A short 247bp upstream region of *vit-2* is sufficient for driving high level expression in the intestine (MacMorris et al. 1992). Three tandem repeats of the octamer-containing sequence (AATTTGCATA) were appended onto the 5' end of the *vit-2* promoter elements. The constructs placed the tandem octamer repeats at positions ranging from 258bp upstream of the translational start site to 700bp upstream of the translational start site (Figure 19 *A*).

The octamer element could not drive expression when fused to any *vit-2* promoter less than 450bp upstream of the gene's translational start site despite the ability of the promoter construct to drive expression in the intestine (Figure 19 *B*). When the octamer was placed at the 5' end of any *vit-2* promoter construct longer than 450bp, low levels of excretory cell expression were observed in addition to the *vit-2* intestinal pattern (Figure

23 *C*) except for a construct containing -652bp *vit-2* promoter region, which, although it drove reporter expression in the intestine, failed to drive expression in the excretory cell. This could indicate the presence of an excretory cell-specific transcriptional repressor located in the *vit-2* promoter region between the -600bp and -652bp upstream of the translational start site.

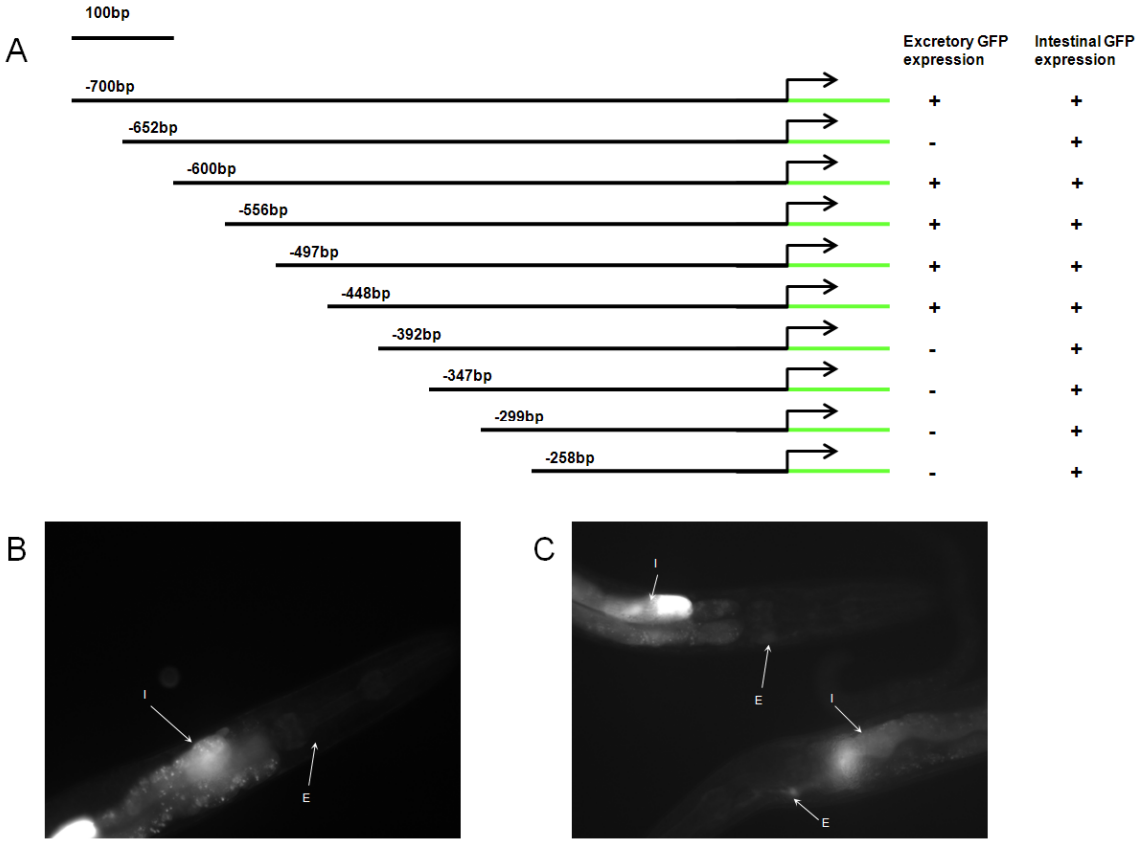

Figure 19. The effects of placing the octamer at various distances upstream of a gene's translational start site.

A, The constructs were based upon octamer sequences appended onto the 5' end of the *vit-2* promoter region (distance upstream indicated).
B, All *vit-2* promoter constructs shorter than the -392bp construct failed to drive expression in the excretory cell (E, expression absent in this construct) although each of these constructs still expressed GFP in the intestine (I).
C, Constructs with *vit-2* promoter regions larger than and including the -448bp construct drove weak expression in the excretory cell in addition to the expected high level of expression in the intestine except for the -652bp construct which failed to drive excretory cell expression.

All fluorescent images were captured with 2 second exposure times.

4.3.5 POU homeobox transcription factors are the most likely cognate binding proteins

With the element able to bind cellular proteins *in vitro* and also able to drive ectopic expression using the basal transcriptional machinery in other genes' promoters, I used the *cis*-regulatory element sequence as a query in the transcription factor binding site database search tool, Transcription Element Search System (TESS), which searches for experimentally validated DNA binding using site or consensus strings and positional weight matrices in the TRANSFAC, JASPAR, IMD, CBIL-GibbsMat databases (Schug J 1996). The results of this search reveal that the best candidates for cognate binding proteins for the binding to the motif were the POU homeobox transcription factors (POU TFs). The binding sites for POU TFs appeared to conform to the motif ATTTGCAT (commonly known as the octamer motif); the core sequences within in the 10bp element identified earlier or the reverse complement sequence (ATGCAAAT).

There are three functionally verified POU TFs encoded in the *C. elegans* genome (*unc-86, ceh-6*, and *ceh-18*) and another three other computationally predicted members each with limited amount of analyses (*F28H6.2*, *Y38E10A.6*, and *F45C12.15*) (Reece-Hoyes et al. 2005).

4.3.6 Analysis of POU homeobox proteins and their expression patterns

The three confirmed POU transcription factors in *C. elegans* have had extensive expression pattern analyses. Moreover, the spatial arrangement of their two POU sub-domains conforms with the established structure of POU TFs with POU_S sub-domain

followed by a single POU_{HD} sub-domain (Figure 20 *A*). Of the three additional putative POU TF members, only *F28H2.6* has had no prior expression pattern analysis. Since information was limited for *F28H6.2*, *Y38E10A.6*, and *F45C12.15*, I generated and analyzed *promoter::GFP* constructs representing all three predicted POU TFs.

The *Y38E10A.6*^{promoter}*::GFP* construct (BC15197) did not drive GFP expression, however, an independent study shows that this promoter drives expression ubiquitously and during all developmental stages (Reece-Hoyes et al. 2007). In addition, using simple modular architecture research tool (SMART) (Schultz et al. 1998) to search for protein domains reveals ten homeodomains, with two of these regions corresponding to POU_{HD} domains, but no POU_S domains (Figure 20*B*) (Table 10). The alignment of Y38E10A.6 against the POU_S domain of mammalian Oct1 confirms the lack of this domain.

The *F45C12.15*^{promoter}*::GFP* construct (BC15191) drove expression only in the intestines of embryonic worms. The earliest observed stage during which expression was observed was the plum stage of the embryo, which corresponds to an age of about 7.0-7.25 hours post-fertilization. (Figure 20*C*). SMART analysis indicates that this gene only contains two homeobox domains but lacks any POU_S-like domains (Figure 19 *B*). This expression pattern overlaps with the previously reported expression pattern which identifies expression in all cells but with stronger signal in internal posterior regions (Reece-Hoyes et al. 2007). Once again, aligning the protein against a POU_S domain reveals little sequence identity even though this protein has two homeodomains. The first homeodomain corresponds to the POU_{HB} domains of POU homeobox proteins, but the second homeodomain did not appear to be related to POU homeobox proteins (Table 10).

The *F28H6.2*^{promoter}*::GFP* construct (BC15200) did not drive expression of GFP, which agrees with the analysis of a previous transgenic strain (Reece-Hoyes et al. 2007). The gene encodes a protein containing a single homeodomain motif, which corresponds to the extended HOX class (Table 10). None of the three newly predicted POU homeobox genes have regions that are similar to POU_S sub-domains.

Table 10. Search of DNA-binding sub-domains against a homeodomain DB.

The DNA-binding sub-domains of each gene were searched against a local Blast database consisting of homeodomains (HDs).

Homeodomain portions of all homeoboxes were isolated from the Homeodomain Resource (http://genome.nhgri.nih.gov/homeodomain/) (Assa-Munt et al. 1993).

The POU domains were included in the search for the sake of comparisons.

	Motif	Position start	Position end	Protein length	Best BLASTP match	HD Class	Organism
F45C12.15	HD1	85	149	543	OTF11/SKN1	POU	Rattus norvegicus
	HD2	467	529		HAT1	NK	Mus musculus
F28H6.2	HD1	115	179	283	TIX1	Extended HOX	Mus musculus
Y38E10A.6	HD1	109	171	1345	HOX-1.9	HOX	Mus musculus
	HD2	172	229		SIX1	Atypical	Mus musculus
	HD3	286	349		Brn5	POU	Rattus norvegicus
	HD4	350	406		THG1	Paired	Homo sapiens
	HD5	487	549		POUc	POU	Danio rerio
	HD6	556	603		DTH-1	Extended HOX	Dugesia tigrina
	HD7	604	661		SMOX-5	Atypical	Schistosome mansoni
	HD8	710	769		CEH-11	Extended HOX	Caenorhabditis elegans
	HD9	771	827		Rough Sheath 1	Atypical	Zea mays
	HD10	1078	1183		OPTX/SIX9	Atypical	Homo sapiens
CEH-6	HD1	268	348	380	POU-M1	POU	Bombyx mori
	POU1	192	266		no significant hits	x	x
UNC-86	HD1	368	430	467	Brn-3	POU	Homo sapiens
	POU1	274	347		no hits found	x	x
CEH-18	HD1	424	486	542	Brn3c	POU	Mus musculus
	POU1	293	367		no significant hits	x	x

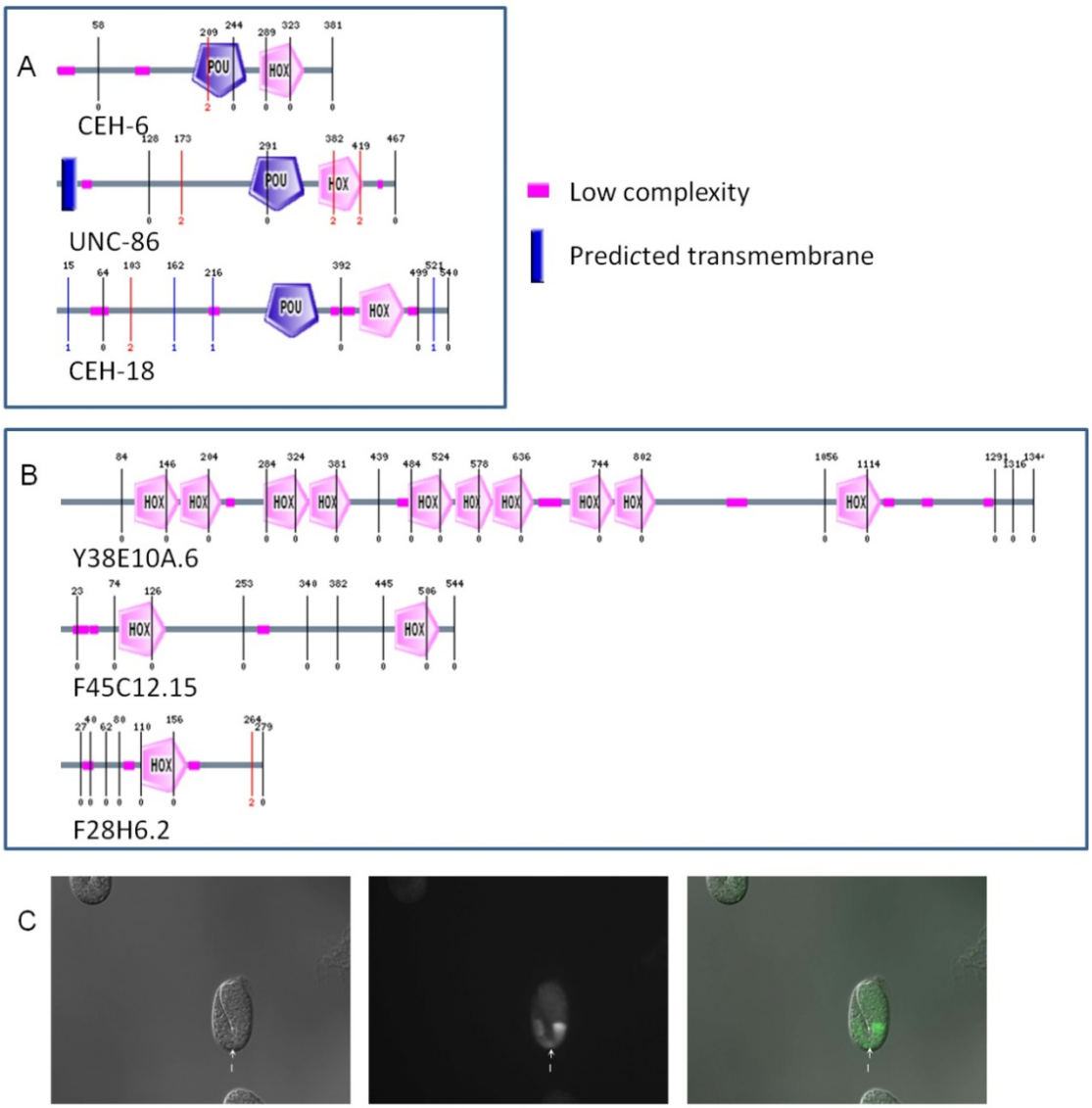

Figure 20. Characterized POU TF vs. predicted POU TF.

A. SMART analysis of well characterized POU TFs.
B. SMART analysis of newly predicted POU TFs from (Reece-Hoyes et al. 2007).
C. Expression pattern of *F45C12.5*^{promoter}*::GFP*. Plum stage embryo with GFP expression observed in the intestine (I). Fluorescent image was captured with a 3 second exposure time.

POU (label) = POU_S, HOX (label) = POU_{HD}

The three confirmed POU homeobox transcription factors (UNC-86, CEH-18, and CEH-6) all contained both a POU_{HD} domain and a POU_S domain. Their POU_S domains did not have any significant matches to peptide fragments represented in the local BLASTP database representing only homeodomain fragments (Table 10) (Banerjee-Basu, Sink, and Baxevanis 2001). This is consistent with the notion that the POU_S domain is evolutionarily related to bacterial transcription factors such as lambda repressor (Assa-Munt et al. 1993).

Of the three well characterized POU TFs, only CEH-6's expression pattern showed a non-ubiquitous expression pattern that overlapped with the excretory cell (Burglin and Ruvkun 2001; Reece-Hoyes et al. 2007). The expression patterns were derived from the transcriptional (Figure 21 *A*) and translational constructs (Figure 21 *B*). GFP fusions were consistent with between the constructs (*ceh-6*^{promoter}*::GFP*; BC16839; *sEx16839* and CEH-6::GFP; CM1250, *guEX929*; kindly provided by K. Armstrong and H. Chamberlin of Ohio State University) except for the strict nuclear localization from the CEH-6::GFP translational construct.

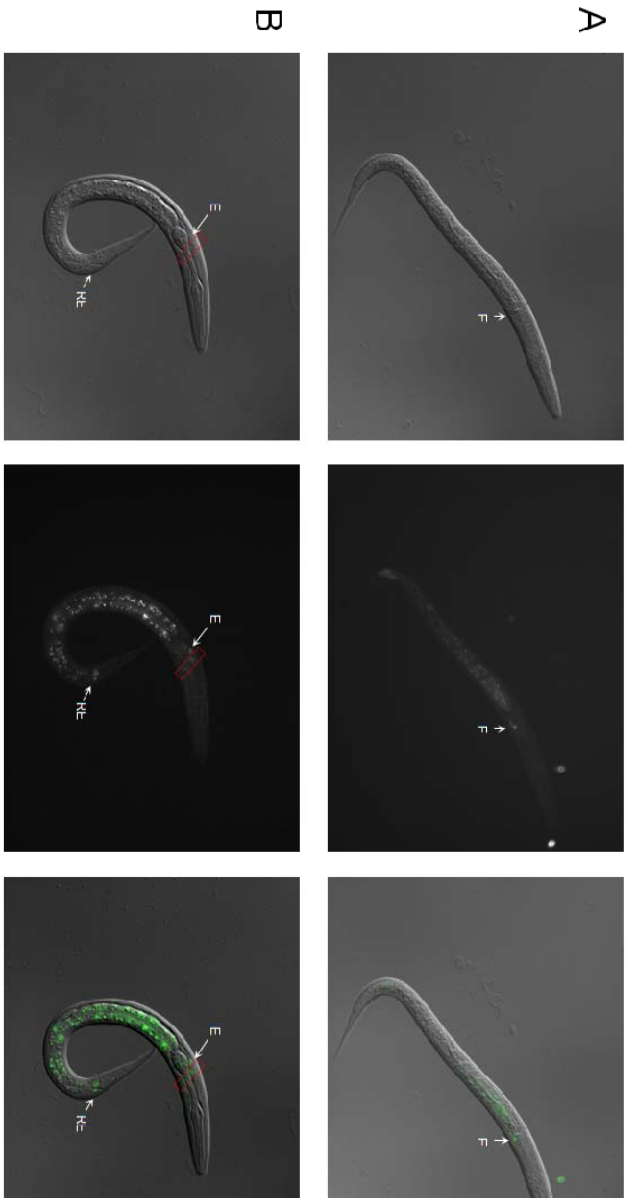

Figure 21. Expression pattern of *ceh-6.*

A, *ceh-6*^{promoter}*::GFP* construct showing expression in the excretory cell (E). Fluorescent image was captured at 3 seconds.
B, CEH-6::GFP is observed in the excretory cell (E), rectal epithelia (RE) along with a cluster of head neurons (red box).

The GFP signal in the merged pictures were altered to enhance the weak GFP signal.

4.3.7 CEH-6 is a POU homeobox transcription factor that is the cognate *trans*-acting factor for the octamer DNA element in *aqp-8*'s promoter region

CEH-6 is a POU homeobox transcription factor that is expressed in the excretory system and also in a limited number of other cells mostly in the hypoderm and nervous system. Mutants of *ceh-6* display phenotypes associated with defects in the excretory cell among other problems. To determine whether CEH-6 functions as a potential cognate transcription factor for the octamer element in *aqp-8*'s promoter region, RNAi against *ceh-6* in an AQP-8::GFP was performed (experiment performed by K. Armstrong and H. Chamberlin, Ohio State University).

Treatment of nematodes by injection of dsRNA corresponding to *ceh-6* alone was not sufficient to knockdown excretory cell GFP-expression. To increase the sensitivity of the RNAi effect, double RNAi injections consisting of a *ceh-6* and *eri-1* dsRNA co-injection was performed. The knockdown of *ceh-6* expression in the *eri-1* sensitized background led to the loss of AQP-8::GFP expression. This indicates that the presence of CEH-6 is required to drive expression of *aqp-8* via elements located in its promoter region (Figure 22).

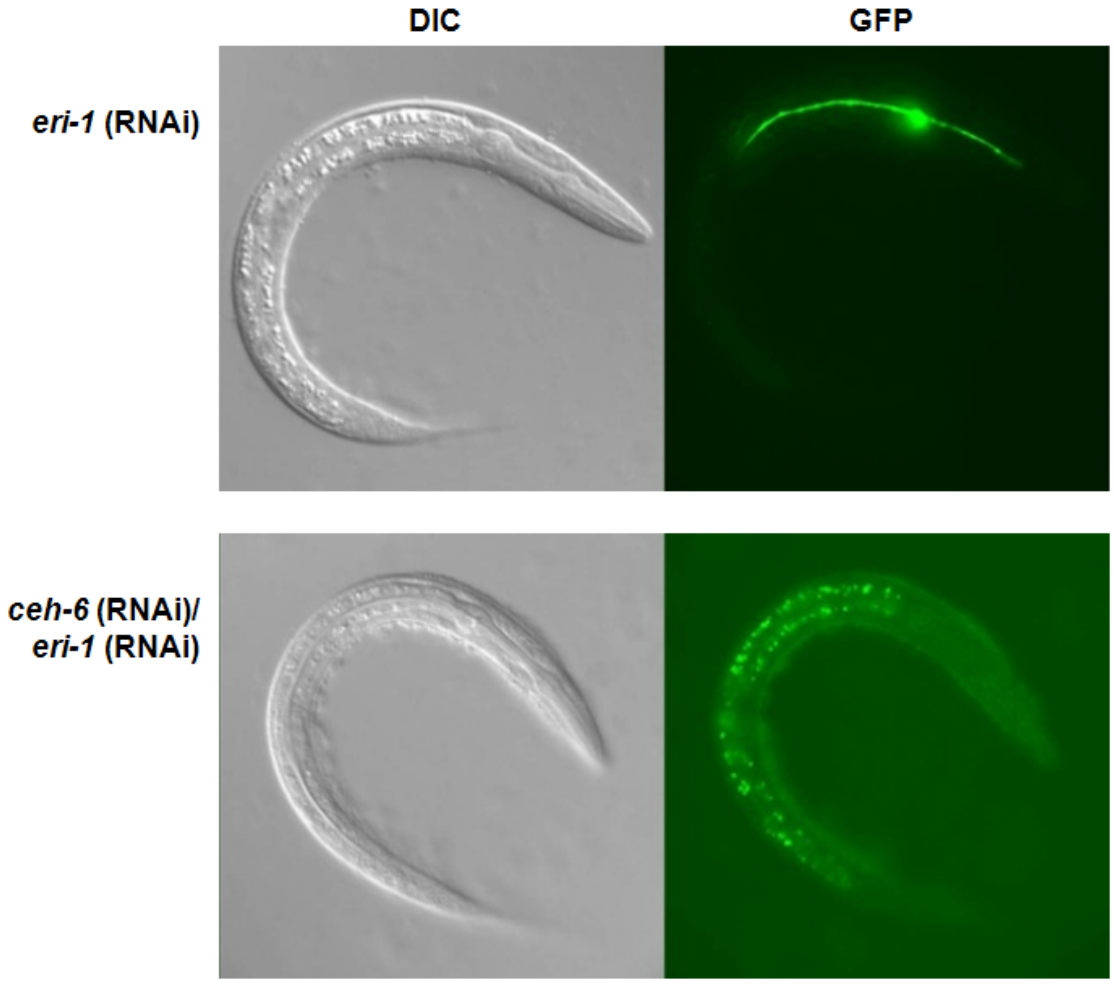

Figure 22. CEH-6 is required for *aqp-8::GFP* expression.

Injection of dsRNA, corresponding to the *eri-1* gene, into AQP-8::GFP-expressing worms alone did not lead to a loss of GFP expression pattern in the excretory cell (top). Injection of two types of dsRNAs corresponding to *eri-1* and *ceh-6* leads to a consistent complete loss of GFP expression in the excretory cell of the AQP::GFP-expressing line (bottom).

*Images are of L2 larval *C. elegans* 48-hours post dsRNA injection
**Camera conditions and exposure times were the same for all images.

4.3.8 *In vitro* validation of the octamer motif /CEH-6 interaction by supershift EMSA

Although I have demonstrated via RNAi that CEH-6 is required in conjunction with *aqp-8*'s promoter region to drive expression in the excretory cell, the physical interaction between the octamer DNA element and CEH-6 has not yet been proven. To determine whether these two elements are required to physically interact to drive expression in the excretory cell, I used supershift EMSAs. The cellular extracts were incubated with CEH-6 specific antibodies prior to incubations with the probes to provide sufficient formation of the CEH-6::antibody complexes.

Using identical labeled probes as the ones used in the previous EMSA, even slower migrating bands were observed in lanes loaded with antibody-incubated *C. elegans* cellular extracts. This indicates that the larger antibody-CEH-6 complex is binding the labeled probe (Figure 23). I show that the antibody-protein-DNA interaction can be competed out using unlabeled probes with the same sequence in the same manner demonstrated in the earlier EMSA. This shows that it is a specific interaction between the octamer motif containing probe and the antibody-CEH-6 complex.

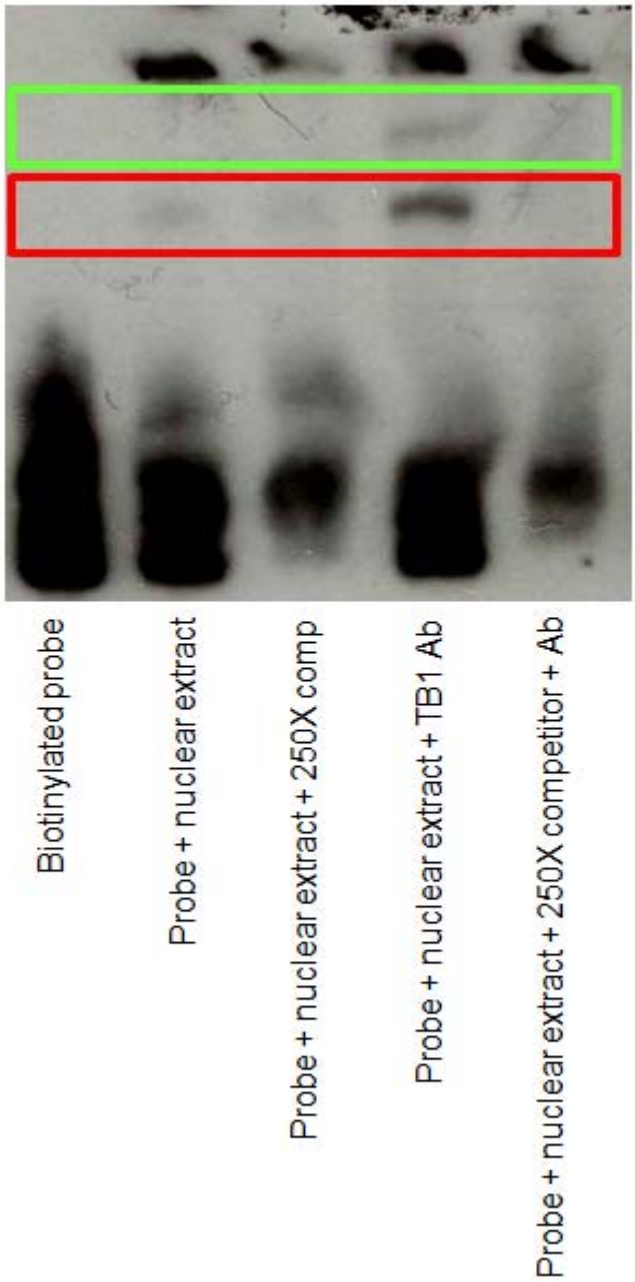

Figure 23. Supershift EMSA with CEH-6 specific antibodies.

Incubation of the biotinylated probe with nuclear extracts provides a faint shifted band (red) which can be readily outcompeted using unlabeled probe. The addition of CEH-6 specific antibody (TB1, kindly provided by T. Burglin), leads to further retardation of the antibody-CEH-6-probe complex (green) that can also be readily outcompeted using unlabeled probe.

4.4 Discussion

The motif identified in the previous section was tested for its ability to bind proteins in different cellular fractions by EMSA. The DNA fragment, containing the conserved motif, has affinity to proteins found in both nuclear and cytoplasmic mixed stage *C. elegans* cellular extracts. The shifted bands from both the nuclear and cytoplasmic fractions appear to be the same size and thus indicate the binding of an identical protein(s). The result of DNA-binding affinity from cytosolic proteins could be due to contamination of the cytoplasmic fraction with some nuclear proteins, but more likely, due to the high level of binding observed, there might to be a population/pool of cognate binding protein(s) localised within the cytoplasm. The shifted band disappears when titrated out by the addition of unlabelled dsDNA with the identical sequence as the probe. This indicates that the unlabelled probe can successfully compete with the labelled probe for the limited amount of DNA-binding protein specific to the sequence. Some of the bound proteins can be attributed to general DNA-binding factors. This includes the bound biotinylated product that remained in the well of the gel and a band which mobilized in the gel that was not outcompeted by unlabeled probe.

I also used a dsDNA probe specific to another region of *C. elegans*' genome for competition studies against the conserved element. The non-specific competitor probe corresponds to a putative transcriptional repressor site located upstream of the *let-721* gene's translational start site (unpublished, D. Chew, Baillie Laboratory, Simon Fraser University). This fragment has a low affinity for a specific nuclear protein(s). The

addition of excess amounts of the unrelated dsDNA competitor probe could not out-compete the protein causing the distinct shifted band in the EMSA.

Since the conserved element was able to recruit nuclear proteins, I tested the element's ability to drive expression independently of other possible regulatory sequences located within *aqp-8*'s promoter region. I generated a chimeric promoter constructs. The chimeric promoter construct is based upon the widely used *Δpes-10::GFP* cassette (Kelly and Fire 1998). The *Δpes-10::GFP* cassette contains a minimal (or basal) promoter, which is composed of a region containing sequences required for basal transcription without the ability to drive any specific expression itself, fused to the GFP reporter. The *Δpes-10::GFP* construct alone drives faint expression of GFP in the intestine and pharynx. This low-level background expression does not interfere with the assay due to lack of overlap with the excretory cell. The *Δpes-10::GFP* chimeric promoter system has been used successfully in previous studies to determine the function of *cis*-regulatory elements. These studies include the regulation of *lin-3*, a gene required in the anchor cell for vulval development (Hwang and Sternberg 2004) and *pha-4*, a gene expressed in the pharynx (Gaudet et al. 2004).

Multiple copies of the conserved ten base-pair element were placed upstream of the *Δpes-10* sequence to ensure adequate room for *trans*-acting factor docking. This construct was sufficient to drive expression of GFP at a low level in the excretory cell in addition to a several neurons located in the anterior portion of the worm. This result shows that the conserved element works as a *cis*-regulatory element that can drive expression in the excretory cell in a generic manner. The reverse complement sequence could not drive expression at an assayable level when tested using the same system.

I tested the effect of placing the octamer element at different positions of in relation to the translational start site with respect to expression patterning. To achieve this, I generated chimeric promoter constructs by placing the *cis*-regulatory element at various distances upstream of an already functional promoter element. The purpose of this experiment was to determine whether the biological function of the *cis*-regulatory element changed relative to its position, as there have been reports of positional effects affecting the final expression pattern. For instance, the analysis of a set of human putative promoters showed that 1,226 putative *cis*-regulatory elements had a significant positional preference relative to the translational start site of the gene (Tharakaraman et al. 2008). The authors also found that some *cis*-regulatory clustered at multiple places upstream of the gene with each clustering position associated with a different transcriptional modulatory function (activation or repression) or expression pattern. The motif CCGCCGCC, specifically, was found to have differential effects upon regulation of genes based upon its relative position from the gene (Tharakaraman et al. 2008) (consensus 5'-(C/g/a)(G/l)(C/t/a)CATN(T/a)T/g/c)-3', preferred residues capitalized). The CCGCCGCC associates with Yin-Yang 1 (YY1), a zinc finger transcription factor that expresses in many tissues (Hyde-DeRuyscher, Jennings, and Shenk 1995). The CCGCCGCC clustered to two different locations at +13bp and +63bp relative to the translational site. When this *cis*-regulatory site is situated at +13bp, gene-transcripts were underrepresented in the medulla oblongata whereas when the *cis*-regulatory site was situated at postion +63bp, genes were found to be up-regulated in T-cells (Tharakaraman et al. 2008).

To generate the chimeric transgene constructs, I used the *vit-2* promoter, which has already been extensively analyzed for conserved regulatory elements located directly upstream of the gene, as a basis for fusing the *cis*-regulatory element. The upstream region of *vit-2* contains multiple copies of the VPE1 (TGTCAAT) and VPE2 (CTGATAA) elements (MacMorris et al. 1992), which are conserved between *C. elegans* and *C. briggsae* (Zucker-Aprison and Blumenthal 1989). The VPE sites work in cooperation to drive expression of genes in the intestine of the worm. A 247bp upstream fragment of *vit-2* containing these VPE elements is sufficient to drive expression (MacMorris et al. 1992).

A -258bp→+9bp fragment of *vit-2'*s promoter was used as a starting point to determine whether the octamer motif behaviour changes as a function of its distance from the translational start site. The tandem repeats of the *cis*-regulatory motif, fused to various lengths of promoter fragments 258bp upstream to 700bp upstream, were unable to drive expression in the excretory cell until the element was place 448bp upstream of the translational start site. This is in contrast with the position of the *cis*-regulatory element in *aqp-8*'s promoter region, which drives expression at a distance of -268bp upstream of the translational start site. Although these *vit-2* promoter-based constructs did not provide much information about how close the *cis*-regulatory can get to the translational start site before being ineffective in driving excretory cell expression, I did show that the octamer element still had the capacity to drive excretory cell expression at distances of over 600bp upstream of the translational start site. However, I already showed that the octamer could drive expression at smaller distances relative to the translational start via the prior of *pes-10* chimeric construct, in which the tandem repeats

of the octamer were able to drive expression of GFP when placed 183bp upstream (Figure 18).

POU homeobox transcription factors were the most likely candidates to be the cognate *trans*-acting factor to the *cis*-regulatory element. This was determined by using the conserved element as a query in databases containing documented DNA-protein interactions. The database searches also revealed the identity of a binding sequence as an octamer motif, which is nested within the conserved element. The octamer motif is comprised of the sequence: ATTTGCAT or the reverse complement of the sequence ATGCAAAT. These motifs were originally identified in the promoter region of mouse immunoglobulin light chain genes and heavy chain genes (Parslow et al. 1984). The forward conformation was always found in the promoter regions of light chain genes whereas the reverse conformation was found in the promoter regions of heavy chain genes (Parslow et al. 1984). Octamer motifs are also found in the introns and 3' enhancer regions of immunoglobulin heavy chain genes where they also have expression modulating activity (Prabhu et al. 1996). The proximity of these sequences to the transcriptional initiation site indicated that the sequences were most likely involved in transcriptional regulation. The octamer was later realized to be an active *cis*-regulatory motif for the simian virus 40 (SV40) early genes and for the U2 snRNA gene of *Xenopus laevis* (Bohmann et al. 1987).

There are three well-established POU homeobox transcription factors in *C. elegans*: UNC-86/C30A5.7, CEH-18/ ZC64.3, and CEH-6/K02B12.1. Each of these POU transcription factors has had extensive phenotypic analysis and known expression

patterns. Recently, bioinformatic analyses have predicted three other potential POU transcription factors: F28H6.2, Y38E10A.6, and F45C12.15 (Reece-Hoyes et al. 2005).

My expression pattern analyses of these genes provided different expression patterns for Y38E10A.6 and F45C12.15 than previously published. The differences in the expression patterns might be attributable to the use of different promoter regions or the fact that the previous transgenic lines were plasmid-based transgene constructs as opposed to PCR-based transgene constructs, which were analyzed in this study. Previous studies have shown that transgenics derived from the introduction of linear DNA constructs more accurately represent the expression pattern of genes (Etchberger and Hobert 2008).

An alignment and the resulting phylogenetic tree of all six *C. elegans* POU proteins along with the *Rattus norvegicus* Oct1 (octamer binding POU transcription factor protein) indicate that the three newly predicted POU homeoboxes are either distant members of the POU homeobox family, or that they may be representatives of different homeobox transcription factor classes altogether. Typical POU homeobox proteins contain a POU_{HD} sub-domain and a separate POU_S sub-domain. The number of homeodomains in the newly predicted POU proteins range from one to ten (Figure 19 *B*). The homeodomains of these newly predicted POU TFs show similarity to different classes of homeodomains. In addition, F28H6.2, Y38E10A.6, and F45C12.15 all lack POU_S domains.

POU homeobox transcription factor proteins contain nuclear localization sequences. The nuclear localization sequence is a basic signal (consensus GRKRKKRT) which precedes the first α-helix of the POU_{HD} domain and its function has been

confirmed by mutagenesis studies in Oct6/Tst1 (Sock et al. 1996). The NLS is notably missing from the POU_{HD} domains of the putative members F28H6.2, Y38E10A.6, and F45C12.15 (Figure 24 *A*). Comparison of the well-studied *C. elegans* POU TFs shows that the NLS is highly conserved among nematode and mammalian POU homeobox proteins (Figure 24 *B*).

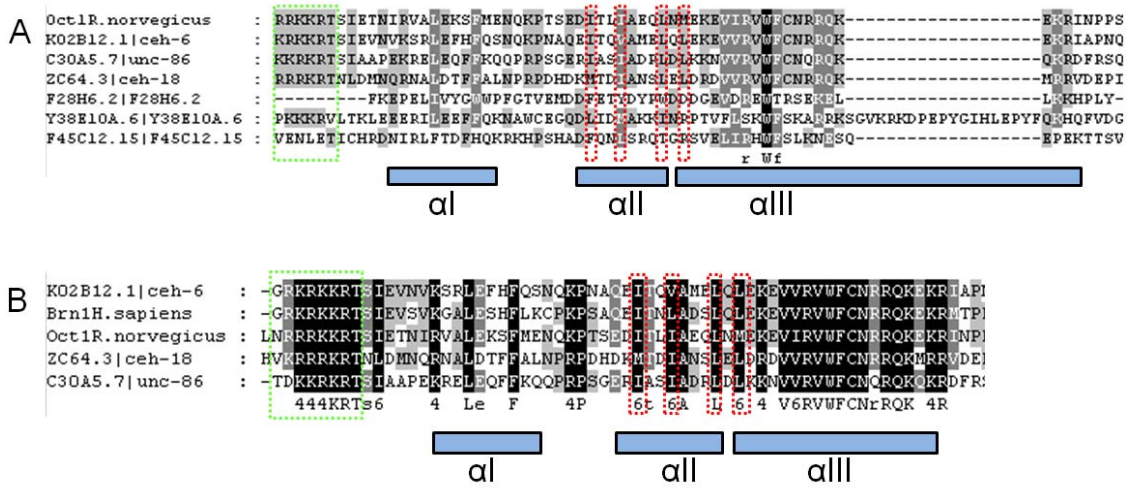

Figure 24. Comparison of the POU_{HD} sub-domains of the POU TFs in *C. elegans*.

A, Comparison of the POU_{HD} sub-domains of the six POU TFs in *C. elegans* indicates that there is little similarity among the three newly predicted members.
B, The nuclear localization signal (green) is highly conserved and precedes the first α-helix in the POU_{HD} region. The nuclear export sequence is also strongly conserved (red).

POU TFs also contain nuclear export sequences (NES). This sequence was originally found in the Oct6/Tst1 protein, a protein that is predominantly nuclear, but is observed in the cytosol in some cell sub-populations (Ilia, Bazigou, and Price 2003). Residues affecting nuclear export are located in the second and third α-helices of the POU_{HD} domain and are characterized by being leucine rich (Baranek, Sock, and Wegner 2005). These leucine rich NES sequences were originally identified in HIV type-1 (Fischer et al. 1995) and protein kinase inhibitor (Wen et al. 1995). Once again, the NES sequence is highly conserved among POU homeobox proteins. The NES is an attribute of both mammalian and nematode POU TFs (Figure 24 *A, B*). The presence of an NES and an NLS in their respective positions appears to be a conserved theme in POU proteins throughout the animal kingdom.

With all these factors taken together, it is unlikely that F28H6.2, Y38E10A.6, and F45C12.15 are *bona fide* POU homeobox genes and therefore I did not consider them as candidate cognate *trans*-acting factors for the octamer motif in *aqp-8*'s promoter region.

Turning to the established members of the POU homeobox genes in *C. elegans,* CEH-18 is required for directing proper gonadal sheath differentiation (Rose et al. 1997). Loss of CEH-18 function leads to defective oocyte maturation. It is expressed in the: gonadal sheath cells, distal tip cell, anterior touch neurons, and many epidermal cells and is also implicated in directing gonadal migration and epidermal differentiation (Greenstein et al. 1994).

UNC-86 is the most intensively studied of the three POU transcription factors in *C. elegans*. Homozygous loss of function mutants have nervous system defects resulting

from losses and gains of neurons because of altered development. Because of the loss of the HSN neuron, the worms have defects in egg laying

(Finney, Ruvkun, and Horvitz 1988; Sulston and Horvitz 1981). Most of the defects in *unc-86* mutants are attributed to cell lineage defects (Finney, Ruvkun, and Horvitz 1988). UNC-86 is observed in the cells: RIH, RIR, PVR, IL2L/R, URYVL/R, RIPL/R, AIZL/R, FLPL/R, ADAL/R, RMGL/R, UL/R, PLML/R, ALML/R, ALNL/R, HSNL/R, URBL/R, NSML/R, URADL/R, URADL/R, IL2DL/R, I2L/R, IL2VL/R, URAVL/R, URXL/R, AIML/R, URYDL/R, PQR, PVM, SDQL/R, PVDL/R, PHCL/R, and PLNL/R (Finney, Ruvkun, and Horvitz 1988). The DNA binding site for UNC-86 has been determined as CATnnnT/AAAT, which is the same as the binding sequence for its vertebrate ortholog Brn-3 (Wang and Way 1996). Both CEH-18 and CEH-6 have unknown DNA binding sites

The length of the linker region between the POU_{HD} and the POU_S sub-domains has been shown to determine whether the protein binds an octamer sequence (ATGCAAT) or an inversely oriented sequence ($GCATN_xTAAT$) produced by shuffling of the POU_{HD} and the POU_S consensus binding sequences (van Leeuwen et al. 1997) by defining the extent of the proteins conformational flexibility. The linker region is accessible to proteases when in either in its DNA-bound (Aurora and Herr 1992) or unbound (Botfield, Jancso, and Weiss 1992) states suggesting that the linker region provides slack between the two DNA-binding sub-domains. The linker regions of POU TFs tend to not be well conserved except for being enriched in non-polar amino acids. This agrees with its generic role as a flexible region in which conservation of sequence is not crucial to its function. Although it is possible that UNC-86 and CEH-18 might have

the ability to bind to promiscuously to different DNA sequences based upon the flexibility of the linker or have degenerate binding motif preferences, their spatial distribution patterns make them unsuitable candidates for being the cognate transcription factors for the octamer motif upstream of *aqp-8*.

CEH-6 is the strongest candidate as a cognate transcription factor for the octamer site in *aqp-8*'s promoter region. In addition to expression in the excretory cell, the protein is expressed in the bilaterally symmetric neurons: RMDDLR, RMDVLR, AUALR and AVHLR neurons, theP.na cells in the ventral nerve cord, the five rectal cells: B, Y, U, F and K (Burglin and Ruvkun 2001) 2 tail nerve cells, ventral hypoderm, anterior body wall muscle, body wall muscle cells and the intestine (Reece-Hoyes et al. 2007). The *ceh-6* expression pattern also overlapped with the expression patterns driven by the octamer motif in the chimeric *octamer::promoter::GFP* (both *vit-2* and *Δpes-10* promoter-based) transgenic lines. The *Δpes-10::GFP*-based construct also drove expression in the AUA, AVH, and RMDV anterior neuronal cell pairs. Although the reverse complement of the octamer motif failed to drive expression when fused to a generic promoter in *C. elegans*, the reverse conformation acts as an enhancer element for the mammalian immunoglobulin heavy chain genes and therefore might play a role in *C. elegans* under other circumstances.

The function of CEH-6 is involved in the development of the excretory cell as loss of the protein leads to morphological defects in worms, which are consistent with the phenotypes of mutations that specifically affect the excretory cell. Homozygous *ceh-6*(*mg60*) mutant embryos die with defects affecting a variety of tissues. Homozygous embryos lack excretory cells, and fluid filled vacuoles occupied the area that the

excretory cell normally occupies (Burglin and Ruvkun 2001). The vacuole formation phenotype is consistent with the excretory cell defects observed in homozygous *let-653*(*s1733*) mutants.

Knocking down *ceh-6* by RNAi in the AQP-8::GFP background did not change the expression level of AQP-8::GFP. A double gene RNAi via the co-injection of dsRNA corresponding to *ceh-6* and *eri-1* was used to elicit gene knockdown in the excretory cell. The double RNAi method was required to increase the effectiveness of the technique via the knockdown of *eri-1*, a gene encoding a protein that contains DEDDh-like exonuclease and SAP/SAF-box nucleic acid binding domains which confer an siRNase activity (Kennedy, Wang, and Ruvkun 2004). ERI-1 is predominantly localized in the gonads and neurons, thus eradication of ERI-1 leads to persistence of dsRNA in neuronal cell lineages.

The affinity between the octamer motif and CEH-6 was confirmed *in vitro* by analysis of supershift EMSAs using antibodies that were previously used for antibody staining for CEH-6 expression (Burglin and Ruvkun 2001). The immunostaining pattern derived from the antibodies matched the expression patterns derived from *ceh-6::lacZ* and CEH-6::GFP expressing worms which indicated that the antibodies specifically bind to CEH-6 and most likely do not cross-react with the other POU TFs.

CEH-6 is required for the transcriptional control of *aqp-8* via the binding of an octamer motif. The conservation of the residues flanking the octamer motif in the promoters of *aqp-8* orthologs in five species of nematodes, however, implies a functional significance for these nucleotides. It is possible that the two flanking bases, although not required for driving expression for all CEH-6 modulated genes, might contribute

stability, specificity, and/or, recognition of accessory nuclear factors to *aqp-8*'s promoter region. Oct-1, a POU homeobox transcription factor expressed ubiquitously in mammals, preferentially binds the octamer motif-centred consensus sequence, a(a/t)TATGC(A/T)AAT(t/a)t (Verrijzer et al. 1992), indicating that residues adjacent to the canonical motif are often required for transcription factor recruitment. To obtain a better understanding of whether the flanking bases surrounding the octamer element are important for transcription factor binding, additional genes that are transcriptionally regulated by CEH-6 must identified.

An alternative approach for finding transcription factors, which bind the octamer element, is the sampling of target transcription factors via the yeast one-hybrid method. The yeast one-hybrid method is an *in vivo* sampling of DNA-binding proteins for a specific target DNA sequences. An advantage of using this assay is the potential to return a greater number of candidate binding proteins. However, along with greater number positive hits, there is also the potential for this method to return large number of false positive hits. These false positives could include general DNA-binding proteins, which might test positive in supershift EMSAs. As seen in the regular EMSA and the supershift EMSA, a large proportion of the protein bound to the dsDNA probes represented general DNA-binding factors. Another possible cause for false positives in yeast one-hybrid may be due to proteins being able to bind their own target and a subset of weaker affinity targets. This could be the case for the members of the POU TFs, which could bind similar octameric sequences.

CEH-6 was found in both the nuclear and cytosolic cellular protein fractions as indicated in the EMSA experiments. In fact, CEH-6 appears to localize in a greater

quantity in the cytosol. This result shows that CEH-6, like the mammalian class III POU Oct6/Tst-1, is a nucleocytoplasmic shuttling protein (Baranek, Sock, and Wegner 2005). The cytosolic populations of CEH-6 protein have been observed, *in vivo*, in the actively dividing ventral neuroblasts, P11.a and P12.a cells (Burglin and Ruvkun 2001).

SOX POU homeobox transcription factors have a similar leucine rich NES. This NES facilitates its transport into the cytosol in a CRM1-dependent manner (Rehberg et al. 2002). The export factor CRM1/Exp1 recognizes the NES which is comprised of hydrophobic and leucine-rich amino acid sequences (Kutay and Guttinger 2005). CEH-6 contains a similar leucine-rich NES sequence. The *C. elegans* genome contains an ortholog of CRM1/Exp1, IMB-4 (importin-beta-like protein-4, ZK742.1). The GFP signal resulting from the transgene *ZK742.1*^{promoter}*::GFP* (BC10706) was too weak to be detected, and therefore I could not determine if CEH-6 and IMB-4 are co-localized. The NES and NLS in CEH-6 is likely used as a modulatory system that facilitates the rapid activation of deactivation of the transcription factor via nucleocytoplasmic shuttling of the protein with the cytosol acting as a repository. Such a means of transcription factor activity modulation has been observed for other classes of transcription factors. Not only does transport of the transcription factor to the cytosol act as an efficient means to inactivate the transcription factor, but also may act as a mechanism for facilitating post-translational modifications to the transcription factor, which can occur in the cytoplasm. These modifications may be necessary for the function of the transcription factor in the nucleus (Rehberg et al. 2002).

With the interaction between the octamer motif and CEH-6 established, and with prior mutagenesis studies on the octamer motif carried out (see last chapter) with their

resulting expression patterns, inferences can be made about how these motif residue changes affect the binding of the transcription factor. I generated mutations in fifth and sixth residues of the octamer motif ($A^1T^2T^3T^4G^5C^6A^7T^8$). These sites are associated with the POU_S sub-domain binding. Fortunately the POU_S sub-domain provides a greater influence on binding specificity than the POU_{HD} since it is amino acid differences in the A-box region of the POU_S sub-domain that specifies nucleotide sequence specification (see fig below for A-box and B-box locations) (Aurora and Herr 1992). Base-pair five associates via hydrogen bonding with arginine 48 (number relative to the start of the POU_S sub-domain). Base-pair six associates via hydrogen bonding with threonine 44. Arginine 48 also associates with the nucleotide on the complementary strand of base-pair six, also via hydrogen-bonding (Phillips and Luisi 2000). A summary of the interactions are shown in (Figure 25).

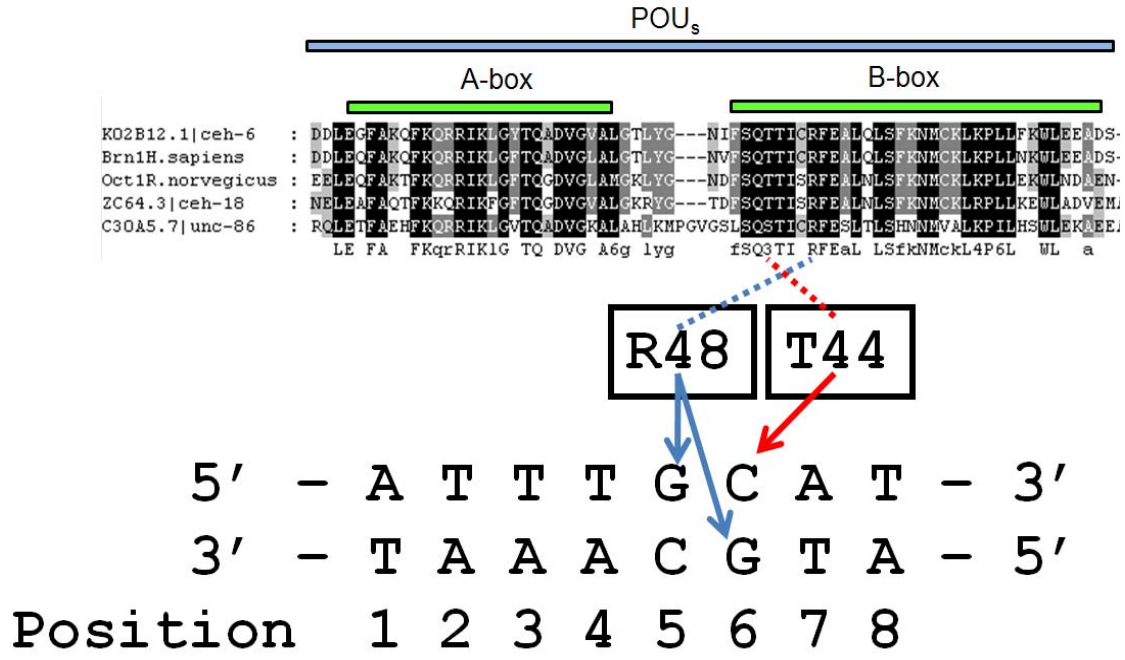

Figure 25. Summary of POU TF interactions at positions 5 and 6 of the octamer element.

An alignment of POU_S sub-domains show that the residues that contact positions 5 and 6 of the octamer element are both in the B-box region of the sub-domain and that the amino acid residues are highly conserved.

I did not observe a change in expression level or pattern when position five of the octamer was mutated from one purine to another (ATTTGCAT→ATTTACAT). The catfish class III POU TF, Oct2, has an only a slightly lower affinity for this variant (ATTTACAT) octamer motif than the consensus octamer (Hikima et al. 2006). This alternate octamer sequence was also determined to be a functional POU binding site in the upstream regions of a surface antigen of an unknown protein in hepatitis B virus and a mouse mammary tumor virus protein respectively (Fogel et al. 2004). Therefore it appears that this residue change creates a octamer element variant that still has the ability to interact with the POU_S sub-domain binding consensus (TG(C/A)ATattc) (Verrijzer et al. 1992).

The mutation involving the fifth residue changing to thymidine (ATTTGCAT → ATTTTCAT) resulted in weak expression still localized to the excretory cell. This site was able to bind to Oct1 in EMSA experiments (Givens et al. 2004), however in other reports this variant of the octamer was not able to drive reporter expression in human cell lines (Brabletz et al. 1993). These results, taken together, might indicate that this octamer variant is not an optimal POU TF binding site.

A mutation of the sixth site (ATTTGCAT→ATTTGAAT) also resulted in weak expression localized to the excretory cell. This site has been verified to be functional in the promoter of the *D. melanogaster* gene, *choline acetyltransferase* (*ChAT*), where it recruits the POU homeobox transcription factor, dPOU-19 (Kitamoto and Salvaterra 1995).

A double residue replacement at positions five and six of the octamer motif (ATTTGCAT→ATTTAGAT) leads to an ability to drive low-level gene expression.

There have been no previous reports of this dual nucleotide substitution variant of the octamer motif associating with POU TFs and therefore may represent a novel POU TF binding site.

There are six POU TF classes, which are sorted based upon homology of the entire POU domain (POU_S, variable linker, and POU_{HD}) (Wegner, Drolet, and Rosenfeld 1993). CEH-6 is orthologous to mammalian class III POU TFs. Class III POUs express predominantly in the central nervous system where they have important functions in development. There are four members of this class in mammals: POU3F1 (aka, Tst-1, Oct6, SCIP, or Otf6), POU3F2 (aka. Brn2, N-Oct3, N-Oct5, or Otf7), POU3F3 (aka. Brn1 or Otf9), and POU3F4 (aka. Brn4, RHS2, N-Oct4, or Otf9). These transcription factors are intronless in mice (Alvarez-Bolado, Rosenfeld, and Swanson 1995), an aspect not reflected in the *C. elegans ceh-6* locus since it's coding region contains five introns. The function of some class III POU TFs can be used interchangeably as demonstrated by rescuing of mutants POU3F1/Oct6 with an exogenously expressed alternate POU class III member, POU3F3/Brn1. This rescue can occur despite the obvious differences in their respective N and C terminal regions (Friedrich et al. 2005; Schreiber et al. 1997).

POU3F3/Brn1 expresses in the distal tubule and the loop of Henle of the mouse kidney where it is required for proper formation of these structures. The phenotype associated with *brn-1* loss of function mutants is mild effect on branching morphogenesis of the tubes and death occurs within 24 hours as a result. Closer inspection of these tissues revealed a shortened loop of Henle, and suppressed differentiation in all three tissues (Nakai et al. 2003). The Brn1 ortholog in quail, qBrn1, localizes to neuronal tissue and in the mesodermal sections of the developing kidney in day-5 old embryonic quail. In

addition, qBrn1 has been detected as early as day-2 old in embryonic tissue sections (Lan et al. 2007). In Zebrafish, Brn1 has been localized via whole-mount *in situ* expression patterning to neuronal tissue, the pronephric primordium and the pharyngeal arches (Hauptmann and Gerster 2000). Moving away from vertebrate systems, in echinoderms, Brn1 expression has also been detected in the gastrointestinal tract of embryonic sea urchins (Yuh, Dorman, and Davidson 2005).

The *Drosophila* POU III TF ortholog, Drifter/Ventral Veinless/Cf1a, expresses in neurons and the trachea where the TF is involved in development. The trachea is a tubular system that develops by migration and fusion of clusters of ectodermal cells. Loss of *ventral veinless* leads to defects in the cellular migration (Llimargas and Casanova 1997). The only class III POU homeobox in the crustacean *Artemia franciscana* is APH-1 that expresses during larval stages in the salt gland (Chavez et al. 1999). The salt gland's function, like the *C. elegans* excretory cell and mammalian kidney, is an osmoregulatory organ.

The POU III ortholog in *Neochildia fusca,* a member of the Aceola taxon (flatworms), *NceoBrn-1*, is expressed from mid-embryonic stages throughout to adulthood. The expression pattern is confined to neuronal and ventrally located intra-epidermal glandular cells of the flatworm (Ramachandra et al. 2002).

The *Bombyx mori* (silkworm) class III POU, *SGF-3/POU-M1*, is localized to the brain, prothoracic glands, silk glands, oenocytes, parts of the hindgut, parts of the anus, some cells of the central nervous system, parts of lateral ectoderm, adductor plates, abductor plates, subbuccal gland, corpora allata, and salivary glands (Xu et al. 1994). Looking specifically at the silk gland, *SGF-3/POU-M1* is localized to the entire

developing organ during embryonic stages and eventually becomes restricted to areas within the silk gland. One of its target genes, *sericin-1*, which encodes a serine rich protein, is expressed in the same tissue but at a later developmental stage (Matsunami et al. 1998). SGF-3/POU-M1 binds to the SC region (defined as -204bp → 183bp) of the *sericin-1* promoter. The SC region in *Bombyx mori* appears to contain an alternate octamer motif (ATTTACAT) (Fukuta et al. 1993). The dependence of *sericin-1* upon transcriptional activation by possible binding of a class III POU TF to an octamer-like site is evidence that this system (TF and DNA element relationship) may be conserved between arthropods and nematodes. The consistency of expression patterns for class III POU TF orthologs across phyla points toward their express localizing to neuronal, nephric, and glandular tissues being a common theme across phyla. It is likely class III POU TFs are involved in driving equivalent sets of genes (orthologs) required for the function of these tissues.

5: DETERMINATION OF GENES TRANSCRIPTIONALLY REGULATED BY CEH-6/OCTAMER ELEMENT

5.1 Introduction

The physical interaction between the class III POU homeobox transcription factor, CEH-6, and the octamer motif drives *aqp-8* expression in *C. elegans*' excretory cell. I demonstrated that the octamer motif is able to drive expression independently of other DNA factors associated with *aqp-8*'s promoter region by generating chimeric promoters with the octamer artificially inserted into the 5' end of various *promoter::reporter* constructs. These artificial promoters could drive expression in a pattern similar to the pattern driven by *aqp-8*^{promoter}*::GFP* constructs. The octamer motif is absolutely conserved in the upstream region of *aqp-8*'s orthologs in four other *Caenorhabditis* species. The sequence conservation indicates that the orthologous genes are likely regulated in the same CEH-6 dependent manner in other nematode species. With this transcription factor and binding site relationship established, I set out to determine whether this mechanism of transcriptional regulation is required for the expression of other genes in the excretory cell.

Two other genes have been demonstrated to rely on CEH-6 function in order to express in the excretory cell. *nac-2* and *clh-4* both express in the excretory cell. Using the same double RNAi method outlined in the previous chapter, expression was lost when *ceh-6* was knocked down in the *eri-1* RNAi sensitized background. The *nac-2* gene

upstream sequence contains a conserved octamer sequence between *C. elegans* and *C. briggsae* located at -259bp and -268bp respectively. A sequence alignment of the *clh-4* upstream region of this gene in *C. elegans* and *C. briggsae* does not reveal a conserved octamer element (data not shown). When the *clh-4* gene-upstream sequences are individually queried into the Transcriptional Element Search System (Schug J 1996), potential POU homeobox binding sites are identified at -131bp upstream of the *C. elegans clh-4* (CTTTGCAT) and at position -178bp upstream of *clh-4* in *C. briggsae* (TTTTACAT). The site -131bp in the *C. elegans clh-4* resembles an octamer motif and has been referred to occasionally as a "divergent octamer motif" or the "imperfect octamer motif". The CTTTGCAT divergent octamer motif is necessary for the expression of the immunoglobulin Vκ19 in mammals (Schwarzenbach, Newell, and Matthias 1995). The POU homeobox transcription factors OCT4 and SOX2 can bind to the divergent octamer motif (TTTTGCAT) *in vitro*. This variant of the octamer motif is responsible for driving *Nanog* expression in *Drosophila* (Kuroda et al. 2005). In the previous chapter, I have also shown that some variants of the octamer motif, with mutations in the POU_S A-box-interacting residues, could drive excretory cell expression in *C. elegans*.

Not all excretory cell expressed genes are CEH-6 modulated. The Vacuolar H ATPase, *vha-5*, expresses in the excretory cell. *vha-5*^{promoter}*::GFP* expression is not affected by knockdown of *ceh-6* in an *eri-1* sensitized background indicating that CEH-6 is not involved in the expression of all excretory cell-expressing genes (K. Armstrong, H. M. Chamberlin, *pers. comm*).

Due to the complexity when taking into account all possible permutations of alternate functional octamer *cis*-regulatory elements, this study focuses on the octamer

element conforming to the consensus octamer motif (ATTTGCAT) or its reverse complementary sequence. Roughly ten percent of nematode promoters from a sample set drive expression in the excretory cell (193/1,886-genes with visible expression) From this set, only seven promoters regions drove expression in an excretory cell-exclusive pattern (Hunt-Newbury et al. 2007). Two of these seven genes have upstream octamer elements. The *cft-1/C18C4.2* promoter region contains an octamer in the reverse orientation at position -967bp. The octamer is also found upstream of the *cft-1* orthologs in *C.briggsae* (-266bp) and *C. remanei* (-83bp) in the same direction. *C. elegans sulp-5/K11G11.2* has an octamer element located at position -979bp upstream of the translational start site. Although the *C. briggsae sulp-5* ortholog does not have an octamer element in the 1,200bp upstream flanking region, the *C. remanei sulp-5* promoter contains two octamer elements in the same orientations at positions -783bp and -442bp respectively.

There are 1,855 genes in *C. elegans* that contain an octamer motif (either direction) within the 1,200bp upstream region flanking the translational start site (WormBase, WS155). I chose several gene-promoter candidates from this list to ascertain the function of the octamer sequence via targeted deletions of the element. Due to the limited positive results from this strategy, a combination of comparative genomics and large-scale transgenic assay was employed to determine functional octamer motifs. The comparative genomics component consisted of interspecies comparisons of gene-upstream regions to find conserved octamer sequences (performed by J.S. Chu, Chen Laboratory, Simon Fraser University). These promoter regions were targeted for transcriptional reporter constructs. Finally, targeted 5' truncations of the promoter regions

that tested positive for expression in the excretory cell were carried out to assess the function of the octamer motif.

5.2 Materials and methods

5.2.1 Search of all genes with upstream octamer elements

All genes, which contained an octamer element within 1,200bp upstream of a protein-coding gene in C. elegans, were identified from WormBase, WS155.

5.2.2 Transgene construction and strains

See section 2.2.3 Materials and methods for a description on transgene construction. For a list of strains and oligos for: 5' truncations of promoter regions of genes with non-conserved octamer elements, promoter region s of genes with conserved octamer elements, and 5' truncations of promoter regions of genes with conserved octamer elements, for this section, see Appendices 8, 9, and 10 respectively.

5.2.3 Search of all genes with interspecies conserved upstream octamer motif

J. S. Chu (Chen Lab, Simon Fraser University) carried out this entire analysis. The putative gene-upstream regions (in this analysis defined as 1,000 bp upstream of the translational start site (ATG)) of all *C. elegans* protein-coding genes (WormBase, WS170), as well as the gene-upstream regions of genes in the related nematodes *C. briggsae* (WormBase, CB25) and *C. remanei* (WormBase, supercontig 2006 assembly), for the motif ATTTGCAT or the reverse complementary sequence ATGCAAAT. A *C. elegans* gene was considered if its *C. briggsae* and *C. remanei* ortholog both contain one or more octamer motifs as well. To achieve this, genome sequences of these three *Caenorhabditis* species and the predicated motifs were loaded into a MySQL database

using the GFF3 format. Comparative analysis is done by Perl using Bio::DB::GFF module (Stein et al. 2002). Candidate *C. elegans* CEH-6-regulated genes were examined for their expression patterns by searching a *C. elegans* GFP expression database (Hunt-Newbury et al. 2007).

5.2.4 Determination of significance of interspecies conserved octamers

J. S. Chu (Chen Lab, Simon Fraser University) carried out this entire analysis. Statistical significance was determined by 10,000 random selections of the number of candidate genes with expression pattern and calculating the probability of the observed number of genes in the excretory cell. The probability of the expression pattern occurring at random was calculated by counting the number of times, out of 10,000 that a value is greater than or equal to the observed value over the total number of trials. Mathematically, this can be represented by letting v be the observed value and letting N be the set of excretory cell observations from random selections. S is the largest subset of N such that $\forall\ \sigma \in S, \sigma \geq v$. The resulting probability is $| S | / | N |$.

5.3 Results

5.3.1 Initial test candidates

There are 1,855 genes with octamer motifs, taking into account both forward and reverse orientations, within 1,200bp upstream of the gene translational start site (WormBase, WS155). When taking into account the two nucleotides flanking the octamer element in *aqp-8*'s promoter (AATTTGCATA), which are conserved among five *Caenorhabditis* species, there are only 261 genes containing this sequence within 1,200bp upstream of their translational start site in *C. elegans*. I selected seventeen of these genes, which have

had their expression pattern analyzed previously (Hunt-Newbury et al. 2007), to assess whether the excretory cell expression was dependent upon the upstream octamer element. Transgene constructs which either contains the octamer or lack the octamer were generated and compared to the original expression pattern. I considered that the octamer motif was required if excretory cell expression was diminished or completely lost (Table 11).

Table 11. 5' deletion of promoter regions containing upstream octamer elements.

Promoter regions that drove excretory cell expression were cut 5' either truncated in an octamer-element targeted manner or using a naïve approach. The 5' end of the PCR primer and octamer element location are relative to the genes' translational start sites. The stages of expression are designated as E: embryonic, L: larval, and A: adult. The expression level are designated as L: low, M: medium, or H: high and are subjective.

Gene	5, end PCR primer	Octamer location	Octamer necessary for excretory cell expression	Stage	Level	Expression Pattern
C01B12.3	2445	1055	Yes	ELA	H	Excretory cell, hypoderm, spermatheca, anal depressor muscle
	879			ELA	L	Excretory cell
	505			ELA	L	Excretory cell
	286			ELA	L	Excretory cell
	247			ELA	L	Excretory cell
ZC395.10	3380	120	Yes	LA	M	Anterior neuronal, intestine, vulva
	135			LA	M	Anterior neuronal, excretory cell, intestine, rectal epithelia
	105			LA	L	Anterior neuronal, intestine
C02B8.4	2918	582	Yes	A	L	Vulva, anterior neuron
	589			A	L	Excretory cell, 2nd bulb pharynx
	577					no expression
R10H1.2	3403	99	No	ELA	M	Anterior neuronal, excretory cell, nerve cord
	118			ELA	M	Excretory cell, anterior neuronal, 2ndb bulb pharynx, all stage, medium
	95					No expression
Y69E1A.7	2303	1197	No	LA	L	Excretory cell
	1534					No expression
	1252					No expression
	1201					No expression
	1183					No expression
F58B4.1	3265	237		ELA	H	Phasmids, excretory cell, posterior neuronal
	1227		No	LA	H	Excretory cell
	729			LA	H	Excretory cell, posterior neuronal
	390			LA	H	Excretory cell, posterior neuronal
	189			LA	H	Excretory cell, posterior neuronal
C45G9.5	1118	752	No	ELA	H	Excretory cell, intestinal, spermatheca, vulva, anal depressor muscle
	762			A	L	Excretory cell, vulva
	742			LA	L	Excretory cell, hypoderm, vulva
ZK470.5	2449	1051, 775	No	LA	H	Excretory cell
	1634			LA	L	Anterior neuronal, excretory cell, ventral nerve cord, hypoderm, posterior neuronal
	856			ELA	H	Anterior neuronal, excretory cell
	483			LA	L	Pharyngeal gland, excretory cell, hypoderm, intestine
	216			LA	L	Excretory cell
	91			LA	M	anterior neuronal, excretory cell, posterior neuronal
R13A5.10	2630	870	No	LA	M	Anterior neuronal, excretory cell,
	886					no expression
	860			LA	M	Anterior neuronal, excretory cell, possible hypoderm
F36H1.2	Zhao	824	No	LA	M	Excretory cell, intestine
	1075					no expression
	566					no expression
Y48A6B.8	1977	1523	No			Excretory cell, intestine
	1536					no expression
	1517					no expression
F29F11.6	3044	457	No	LA	H	Anterior neuronal, pharynx, nerve ring, excretory cell, intestine, reproductive system, vulva, seam cell, spermatheca, body wall muscle, hypoderm, posterior neuronal, anal sphincter
	466					Muscle, posterior neuronal, adult, low

Gene	5, end PCR primer	Octamer location	Octamer necessary for excretory cell expression	Stage	Level	Expression Pattern
	451					Anterior neuronal, 2nd bulb pharynx, intestine, muscle, spermatheca, posterior neuronal, larval/adult, high
B0334.4	3401	101	No	LA	M	Excretory cell, intestine, tail neurons, reproductive System, vulval muscle,
	119					no expression
	97					no expression
C02B8.6	3221	279	No			1st and 2nd pharyngeal bulbs, nerve ring, excretory cell, nerve cord, muscle, posterior neuron
	292					no expression
	275					no expression
H23N18.3	2899	601	No	LA	H	Anterior neuronal, pharynx, excretory cell, nervous system,
	626					no expression
	597					no expression
R13F6.3	1447,2822	679	No	LA	H	Anterior neuronal, excretory cell, intestine, rectal gland cells, posterior neuronal
	688					no expression
	675					no expression
Y53G8AR.3	3369	131	No	LA	H	Anterior neuronal, excretory cell, intestine, posterior neuronal
	1513					No expression
	1027					No expression
	620					No expression
	365					No expression
	102					No expression

The expression patterns driven by the default promoters are defined by the same criteria used previously in chapter 2 for defining the *aqp* promoter regions. Out of seventeen candidate promoter region, three contained upstream octamer elements that were required for excretory cell expression: *C01B12.3, ZC395.10*, and *hlh-8/C02B8.4.*

Previous expression pattern analysis of a larger promoter region of *ZC395.10* shows that the gene is expressed in the anterior neurons, intestine, and vulva (image not shown), however the once the promoter region is truncated to just include the octamer element expression was observed in the anterior neurons, excretory cell, intestine, and rectal epithelia (Figure 26 *A*). It would appear that between the large construct and the targeted truncation, a repressor for excretory cell and rectal epithelial expression was lost. Both the excretory cell and rectal epithelia overlap with the expression pattern of CEH-6. Once the octamer is lost, expression is lost in both of these tissues (Figure 26 *B*). *ZC395.10* encodes a gene that is orthologous to a co-chaperone that binds to and regulates Hsp90 family chaperones in *S. cerevisiae*. The ortholog of ZC395.10 in humans is annotated as a Prostaglandin E synthase 3. Since chaperones and prostaglandins are involved in a wide range of body functions it is unclear how this gene works in the excretory cell without further studies.

hlh-8/C02B8.4 encodes a helix-loop-helix transcription factor homologous the TWIST transcription factors originally identified in *Drosophila*. The gene is required for normal muscle development in *C. elegans*. Mutants of this gene have defects in defecation and egg-laying. The original ~3kb promoter construct drives expression in anterior neurons and the vulva, however, a 5’ truncation of the promoter region that cuts seven bases upstream of the octamer element results in a completely different expression

pattern. The octamer motif is located at -582bp, when the promoter construct's 5' end is at -589bp, expression is seen in the excretory cell and the second pharyngeal bulb (images not shown). It is possible that repressor elements that block excretory cell and pharyngeal expression are located within the large region lost between the two transgene constructs. There have been no previous reports of this gene being expressed in the excretory cell.

C01B12.3 encodes the *C. elegans* ortholog of bestrophin 3, a membrane protein that is involved in calcium dependent transport of chloride ions across cellular membranes. Expression driven by a 2.9kb promoter region was observed in the excretory cell, hypoderm, spermatheca, and the anal depressor muscle (Hunt-Newbury et al. 2007). An octamer element is located 1,055bp upstream of the translational start site. The expression level driven by the 2.9kb construct is strong in relation to the level of expression observed in a promoter construct lacking the upstream octamer element (Figure 26 *C*). A search for the octamer element in the 1,200bp upstream of its orthologs regions in *C. briggsae* and *C. remanei* reveals octamer element is conserved and located 1,108bp and 1,100bp upstream of their respective genes' translational start sites.

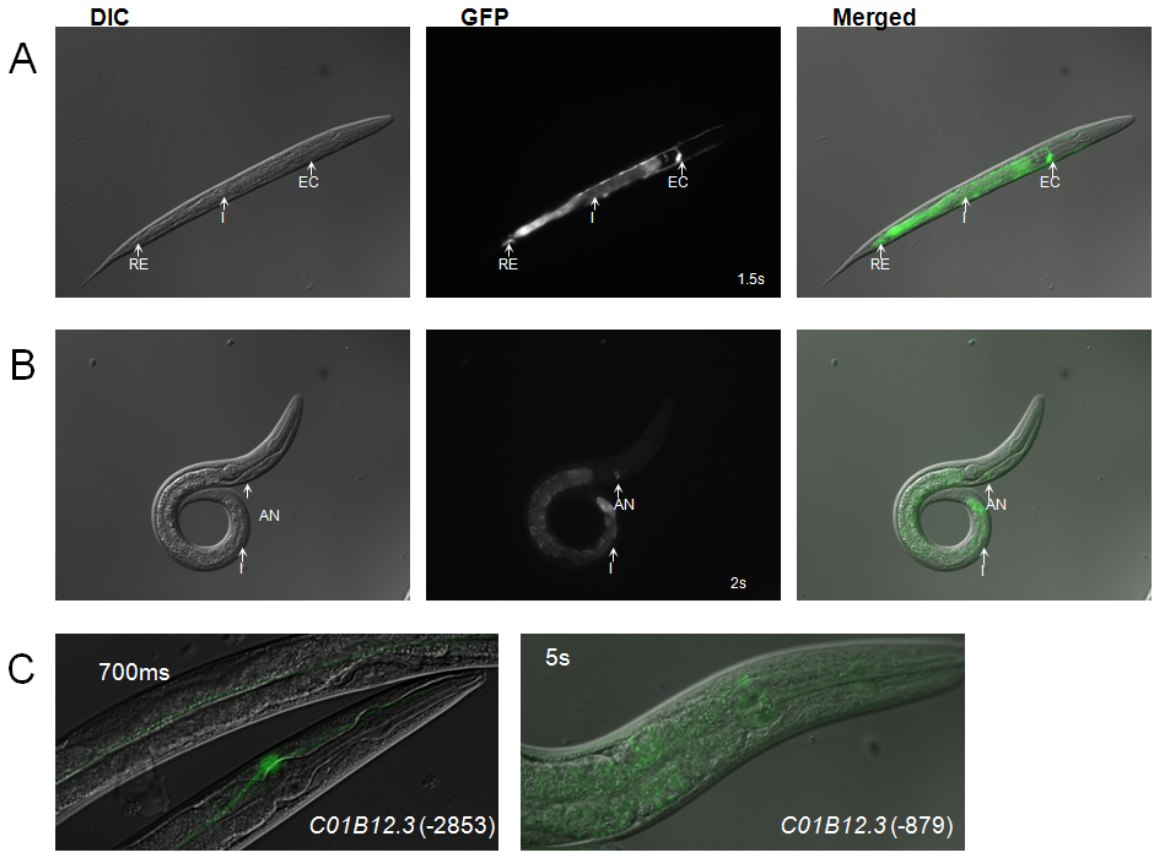

Figure 26. Changes in expression patterns and levels upon loss of the octamer element upstream of *ZC395.10* and *C01B12.3*

The octamer element is located 120bp upstream of *ZC395.10*'s translational start site.
A, A -135bp 5' truncation of *ZC395.10*'s promoter region drives expression in the excretory cell (EC) along with the rectal epithelia (RE) and intestine (I).
B, A -105bp 5' 5' truncation of *ZC395.10*'s promoter region can still drive expression in the intestine, but excretory cell and rectal epithelial expression is lost.
C, The excretory cell expression level of *C01B12.3* drops dramatically upon loss of the octamer element at -1055bp relative to the gene's translational start site.

Exposure times are indicated on the images.

5.3.2 Search all genes with conserved *cis*-regulatory element between three nematode species

With the confirmed interaction of CEH-6 with the octamer element, we searched for instances of octamer element conservation between the three nematode species: *C elegans*, *C. briggsae*, and *C. remanei* in order to determine other potentially co-regulated genes. Four sets of analyses were done according to different filtering criteria. The common criteria among all four sets were that the gene is orthologous in *C. elegans*, *C. briggsae*, and *C. remanei*; and that there is at least one octamer motif predicted in the upstream regulatory region (in directional conformation). The other criteria for each set are summarized in (Table 12). 107 genes were identified with perfect motif matches among the three genomes under the most relaxed condition and 44 genes under the strictest condition (Appendix 11). Of the candidate genes identified, *promoter::GFP* expression pattern data has been generated for 19 (relaxed condition; all) and ten (strictest condition; SE) of the upstream regulatory regions using *promoter::GFP* reporter constructs respectively (McKay et al. 2003) (Table 13). In order to determine whether octamer motifs are enriched in genes expressed in excretory cells, we carried out statistical analysis calculating the significance of observing three excretory cell expressions. We found that the probabilities were 0.3556 and 0.0857 for the most relaxed and most stringent conditions respectively (Table 12). For a set of genes with interspecies conserved upstream octamer elements using the same conditions, but using more recent versions of the *C. elegans*, *C. briggsae*, and *C. remanei* genome databases (WS170, CB25, and supercontig 2006 assembly respectively) refer to Appendix 12.

Table 12. Filtering criteria used for determination of genes with conserved octamer motifs in their 1kb

See appendix for lists of genes falling within each filtering criteria category

Set	Filtering Criteria	Number of predicted genes	Number of genes with existing expression data	Number of excretory cell expression genes	Probability
All	No additional criteria	107	19	3	0.3556
S	Octamer motifs are on the same strand as the downstream gene	54	13	3	0.1662
E	Octamer motifs are not overlapping with any upstream gene	83	16	3	0.2623
SE	Octamer motifs are on the same strand as the downstream gene and not overlapping with any upstream gene	44	10	3	0.0857

Table 13. Expression patterns of genes in *C. elegans* (All category, Table 12) that have upstream octamer sites. This preliminary expression pattern data was obtained from The Genome BC *C. elegans* Expression Pattern website (http://elegans.bcgsc.ca/perl/eprofile/index)

Sequence Name	Gene Name	Strain	Transgene ID	Expression Pattern
C01B12.5		BC20002	*sEX20002*	No expression
C15B12.7	*cdf-1*	BC14040	*sEX14040*	No expression
C16C10.4		BC14128	*sEX14128*	1st and 2nd pharyngeal bulbs, anterior neuron, posterior neuron, muscle, Intestine, vulva
C26F1.10	*flp-21*	BC12205	*sEX12205*	Intestine, anterior neuron
C38H2.1		BC13836	*sEX13836*	Nerve cord, anterior neuron, posterior neuron, amphids, nerve ring
C45G9.5		BC12539	*sEX12539*	Intestine, excretory cell, anterior neuron, posterior neuron, vulva, hypoderm
C54D10.1	*cdr-2*	BC15319	*sEX15319*	1st and 2nd pharyngeal bulbs, vulva, anterior neuron,
F14H12.1	*col-165*	BC16801	*sEX16801*	Hypoderm, seam cell
F39H11.3	*cdk-8*	BC14622	*sEX14622*	No expression
F53C11.3		BC14427	*sEX14427*	Intestine
F53E2.1	*tag-304*	BC10230	*sEX10230*	Pharynx, muscle, anterior neuron, posterior neuron, vulva
H24G06.1		BC15661	*sEX15661*	No expression
H43I07.3		BC15071	*sEX15071*	1st and 2nd pharyngeal bulbs, anterior neuron, posterior neuron, Nerve ring, pharynx,
K02G10.7	*aqp-8*	BC20052	*sEX20052*	Excretory cell,
R02F2.8		BC12904	*sEX12904*	No expression
R08B4.2	*alr-1*	BC16630	*sEX16630*	No expression
R10H1.2	*srab-14*	BC14834	*sEX14834*	Anterior neuron, nerve cord, excretory cell
R13A5.1	*cup-5*	BC10182	*sEX10182*	Pharynx, muscle
T05H10.3		BC14357	*sEX14357*	hypoderm
T10B5.5		BC12510	*sEX12510*	1st and 2nd pharyngeal bulbs, nerve cord, nerve ring
T12A2.9	*srg-8*	BC11603	*sEX11603*	Anterior neuron
T14G10.5	*tsp-12*	BC13695	*sEX13695*	No expression
T19A6.2	*ngp-1*	BC14682	*sEX14682*	Intestine,1st and 2nd pharyngeal bulbs,
W08D2.1	*egl-20*	BC17158	*sEX17158*	Muscle, anal depressor
Y43F8C.12	*mrp-7*	BC10031	*sEX865*	Intestine, muscle, neuron
Y54G2A.25	*lad-2*	BC13847	*sEX13847*	Nerve ring, posterior neuron
Y7A9A.1		BC11932	*sEX11932*	No expression
ZK512.9	*grl-11*	BC12881	*sEX12881*	No expression

5.3.3 Determine expression patterns for all genes with conserved upstream octamer motifs

Transgenic *promoter::GFP*-expressing *C. elegans* lines for each of the 107 gene-promoters identified in section 8.2, which did not already have existing strains, were generated and their resulting expression patterns were analyzed (Table 14) (except for *K01B6.2*, *K08F4.12*, *F16F9.4*, *Y51B9A.6* and *Y53C10A.4* of which transgenic strains could not be generated due to technical difficulties). The candidate set includes 25 promoter regions that drive expression in the excretory cell. Within this subset, 12 promoter regions drive expression only in the excretory cell during post-embryonic stages. These numbers represent an enrichment of genes expressed in the excretory cell compared to a large-scale *C. elegans* gene expression pattern dataset (Table 6, Chapter 2) (Hunt-Newbury et al. 2007).

Table 14. Expression patterns of 107 genes with octamer elements within 1kb upstream of the TSS.

Genes that express in the excretory cell are highlighted in grey. The third column represents the direction of the octamer element sequence. F: ATTTGCAT, R: ATGCAAAT.

Gene	Location of octamer	Octamer F / R	Expression Pattern	Strain ID
C05D12.1	205	F	Excretory cell; larval/adult; high	BC17548
C17G1.5	307	R	Excretory cell, unidentified tail; larval/adult; medium	BC17536
C07E3.10	884	F	Anterior neuronal, hypoderm; all stage; low	BC17549,BC17550
C26B9.5	339	F	Intestine; larval/adult; medium	BC17537,BC17538
C43G2.5	81	F	No expression	BC17633
C50F4.9	793	F	No expression	BC17546,BC17547
C54D10.1	234	F	2nd bulb pharynx, intestine; larval/adult; low	BC17642,BC17643
E04F6.4	62	R	No expression	BC17568
F01D5.6	819	F	No expression	BC17569
F13B6.1	265	R	2nd bulb pharynx, anal depressor; larval/adult; low	BC17570
F16F9.1	799	R	Excretory cell, 2nd bulb pharynx; larval/adult; low	BC17571,BC17572
F18C5.5	386	F	No expression	BC17555
F18C5.9	94	R	No expression	BC17556
F18G5.3	347	F	1st/2nd bulb pharynx, pharynx, hypoderm, seam cell; larval/adult; low	BC17599
F22F7.7	229	F	Hypoderm; larval/adult; low	BC17675,BC17676
F28F9.2	701	R	No expression	BC17620,BC17621
F29B9.8			Anterior neuronal, nerve ring, body wall muscle; adult; low	BC17573
F36F2.7	228	R	Excretory cell; larval/adult; medium	BC17574
F36H12.1	756, 31	R, F	No expression	BC17575,BC17576
F41E6.14	454	R	Anal depressor; adult; low	BC17622
F43B10.1	265	F	No expression	BC17577,BC17578
F44F4.3	967, 781, 336	R, R, F	No expression	BC17579
F49H12.3	329	F	Anterior neuronal, posterior neuronal, nerve cord; all stage high	BC17600
F55F3.4	557	F	Anterior neuronal, excretory cell; adult ; low	BC17580,BC17581
F56A4.10	798,316	R, R	No expression	BC17601

Gene	Location of octamer	Octamer F / R	Expression Pattern	Strain ID
H22K11.3	637	F	Excretory cell; adult; high	BC17602
K06A1.3	74	F	Anterior neuronal, posterior neuronal; adult; low	BC17623,BC17624
K08F4.4	104	R	Excretory cell; larval/adult; low	BC17644
K10C2.4	757	F	Excretory cell; adult; low	BC17645
K10C8.2	757	F	1st/2nd bulb pharynx, pharynx, anterior neuronal; larval/adult; low	BC17603
M03A8.3	657	R	1st/2nd bulb pharynx, excretory cell, vulva; adult; low	BC17634
M176.5	112	F	1st/2nd bulb pharynx, anterior neuronal, nerve ring, nerve cord, excretory cell, rectal epithelia; larval/adult; medium	BC17635
R12G8.2	79, 1091	F	Excretory cell; adult; low	BC17604,BC17605
T02C5.3	141	F	Anterior neuronal, vulva, spermatheca, anal depressor; embryo, adult; medium	BC17606
T10B5.4	436	F	1st/2nd bulb pharynx, hypoderm; larval/adult; medium	BC17618
T11F9.9	401	R	No expression	BC17636,BC17637
T16H12.9	565	F	Pharynx, excretory cell, anal depressor; adult; low	BC17619
Y105C5B.15	233, 214	R, F	No expression	BC17677,BC17678
Y19D10A.4	41	R	Excretory cell, muscle; larval/adult; low	BC17625
Y19D10A.5	175	R	No expression	BC17626
Y67A6A.2	697, 273	F, F	No expression	BC17627
ZC101.1	229	R	1st/2nd bulb pharynx, pharynx; larval/adult; high	BC17628
C07E3.2	414	R	Anterior neuronal, anal depressor; larval/adult	BC17672
C01B12.5	112	R	No expression.	BC20002
C15B12.7a	570	F	No expression.	BC14040
C16C10.4	109	R	1st/2nd bulb pharynx, anterior neuronal, intestine, muscle, posterior neuronal, vulva; larval/adult; medium	BC14128
C26F1.10	111	R	Anterior neuronal, intestine; larval/adult; medium	BC12205
C38H2.1	30	R	Anterior neuronal, nerve ring, amphids, nerve cord, posterior neuronal; larval/adult; medium	BC13836
C45G9.5	766	F	Anterior neuronal, 1st/2nd bulb pharynx, excretory cell, intestine, posterior neuronal, vulva, hypoderm; larval/adult; high	BC12539
C54D10.10			Anterior neuronal, 1st/2nd bulb pharynx, vulva; larval/adult; medium	BC15319
F14H12.1	124	F	Hypoderm, seam cell; larval/adult, high	BC16801
F39H11.3	612	F	No expression.	BC17886
F53C11.3	792	F	Intestine; larval/adult; medium	BC14427
F53E2.1	563	F	Pharynx; muscle; anterior neuronal; posterior neuronal; vulva.	BC10230
H24G06.1a	257		No expression.	BC15661
H43I07.3	533	R	1st/2nd bulb pharynx, anterior neuronal, posterior neuronal, nerve ring, pharynx;	BC15071
K02G10.7	268	F	Excretory cell; larval/adult; high	BC20052
R02F2.8	970	R	Excretory cell; larval/adult; medium	BC17709
R08B4.2	222	R	No expression.	BC16630

Gene	Location of octamer	Octamer F / R	Expression Pattern	Strain ID
R10H1.2	99	F	Anterior neuronal, nerve cord, excretory cell; larval/adult; high	BC14834
R13A5.1a	847	R	Pharynx, muscle; larval/adult; medium	BC10182
T05H10.3	974, 47	R, F	Hypoderm; larval/adult; high	BC14357
T10B5.5	123	R	1st/2nd bulb pharynx, nerve cord, nerve ring; larval/adult; medium	BC12510
T12A2.9	859	F	Anterior neuronal; larval/adult; low	BC11603
T14G10.5	788	R	No expression.	BC13695
T19A6.2a	879	F	1st/2nd bulb pharynx, intestine; larval/adult; medium	BC14682
W08D2.1	783, 170	P, R	Muscle, anal depressor; larval/adult; high	BC17158
Y43F8C.12	643	F	Intestine, muscle, neuronal; larval/adult; high	BC10031
Y54G2A.25	154	F	Nerve ring, posterior neuronal; larval/adult; low	BC13847
Y7A9A.1	120	F	No expression.	BC11932
ZK512.9	68	R	No expression.	BC12881
B0334.11	296	R	Excretory cell, intestine, muscle; larval/adult; low	BC17788
C09B8.7	349	R	No expression.	BC17700
C16B8.4	477	F	Anterior neuronal, muscle, posterior neuronal; larval/adult; medium	BC17717
C45G9.7	223	R	1st/2nd bulb pharynx, excretory cell, intestine, hypoderm, vulva, spermatheca; all stages; medium	BC17786,BC17787
D2096.8	707	F	Anterior neuronal, pharynx, body muscle; adult; low	BC8187
F09A5.4	561	F	Muscle, vulva, anal depressor; larval/adult; medium	BC17701
F11D5.6	610	R	Hypoderm, pharyngeal-intestinal valve, anal depressor	BC17702
F14B8.7	257	F	Excretory cell; pharyngeal-intestinal valve, intestine, anal depressor; larval/adult; medium	BC17704
F16F9.4	729	F	n/a	no strain
F18A11.2	178	R	Excretory cell; larval/adult; medium	BC17728
F28B3.6	526	R	No expression.	BC17707
F36H1.11	758, 31	R, F	No expression.	BC17770
F36H12.17	255	R	No expression.	BC17729
F43C9.4	837	R	No expression.	BC17730
F43E2.9	470	F	Muscle, anterior neuronal; adult; low	BC17771
F53G12.5	645	F	Anterior neuronal, intestine, posterior neuronal; adult; low	BC17773
K01B6.2	479	R	n/a	BC8188
K08F4.12	404	F	n/a	no strain
K09C8.2	141	R	No expression.	no strain
K09C8.8	424, 110	F, F	No expression.	BC17732
R07A4.2	497	F	No expression.	BC17734
R10E12.1	436	F	Anterior neuronal, 1st/2nd bulb pharynx, muscle, intestine, vulva, posterior neuronal; larval/adult; medium	BC17814

Gene	Location of octamer	Octamer F / R	Expression Pattern	Strain ID
R12C12.9	94	F	Excretory cell; adult; low	BC17753
T05A10.4	701	F	Anterior neuronal, hypoderm; adult; larval; medium	BC17774
T28H11.8	143	R	Excretory cell; adult/embryo; high	BC17754
W02F12.4	512	R	1st/2nd bulb pharynx, intestine, muscle; adult; medium	BC17755
Y102A11A.7	344	F	No expression.	BC17775
Y47D7A.14	451	F	No expression.	BC17756
Y51B9A.6	446, 98	R, F	n/a	no strain
Y53C10A.4	199	F	n/a	no strain
ZK1010.2	659	R	Anterior neuronal, intestine; adult; medium	BC8205
ZK1010.3	661	F	Excretory cell, muscle; adult; low	BC17776
ZK1248.2	318	R	No expression.	BC17777
ZK652.11	169	R	No expression.	BC17790
ZK688.1	179, 42	R, F	No expression.	BC17757
C18C4.2	967	R	Excretory cell; adult; low	BC16915,BC16916

5.3.4 Testing of excretory cell-expressing candidates by targeted deletion of upstream octamer motifs

The octamer elements were tested for a sample of promoter regions that drove excretory cell expression in the section 5.3.3/Table 14 by generating targeted 5' promoter truncation constructs that either included the motif or deleted it.

Out of the 23 candidates tested (including *aqp-8*; not included in Table 15), six promoter regions contain excretory cell expression-modulating octamer sequences. I found that octamer elements were required for driving excretory cell expression for the following genes:

1. *C05D12.1* encodes a predicted membrane protein containing DoH and Cytochrome b-561/ferric reductase transmembrane domains. The excretory cell expression level is lower upon deletion of the upstream octamer element at -205bp (Figure 27 *A*, *B*).
2. *twk-36/R12G8.2* is a member of the Twik family of tandem pore potassium channels.
3. *R02F2.8* is a predicted amino acid transporter similar to members of the SLC36 subfamily of the solute carrier transporter (SLC) superfamily.
4. *F16F9.1* is a gene encoding a predicted integral membrane protein which has homology to vertebrate lipopolysaccharide-induced tumor necrosis factor-alpha factor (LITAF).
5. *M176.5* is a novel gene only found in nematodes

A summary of the truncation tests are shown in (Table 15).

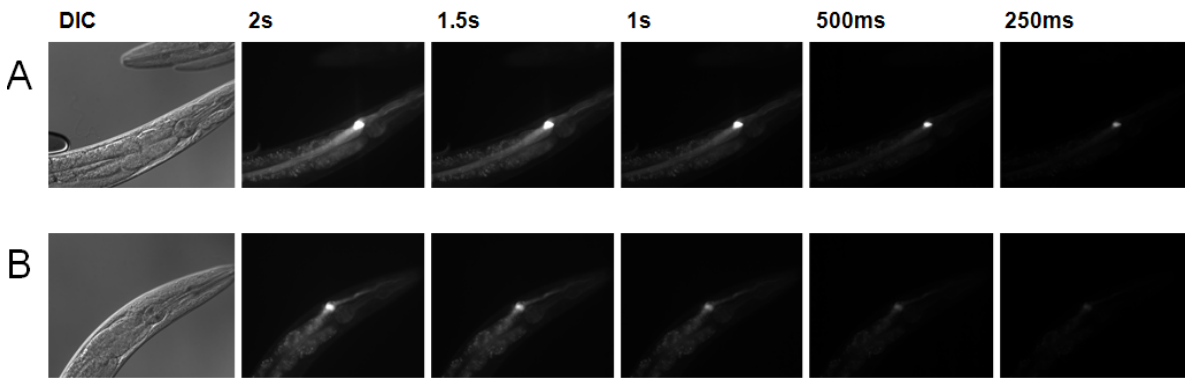

Figure 27. The level of excretory cell expression is decreased upon loss of the upstream octamer element in the promoter region of *C01B12.1*.

The octamer element is located 205bp upstream of the translational start site of *C01B12.1*.
A, A -247bp 5' truncation leads strong expression localized to the excretory cell.
B, A -141bp 5' truncation leads to a lower level of expression, but still localized to the excretory cell.

Table 15. Testing of upstream octamer elements in promoters that drive excretory cell expression.

The third column represents the direction of the octamer element sequence. F: ATTTGCAT, R: ATGCAAAT. Promoter regions that drove excretory cell expression were cut 5' truncated either in an octamer-element targeted manner. The 5' end of the PCR primer and octamer element location are relative to the genes' translational start sites. The stages of expression are designated as E: embryonic, L: larval, and A: adult. The expression levels are designated as L: low, M: medium, or H: high and are subjective.

Gene	Octamer	F/R	Octamer necessary for excretory cell expression	Stage	Level	Expression Pattern
ZK1010.3	648	F	No	A	L	Excretory cell, muscle
ZK1010.3 A'748						No expression
ZK1010.3 A''669				L	M	1st2nd bulb pharynx, anterior neuronal, nerve ring, vulva, anal depressor larval
ZK1010.3 A'''644				L	M	1st2nd bulb pharynx, anterior neuronal, nerve ring, vulva, anal depressor larval
Y19D10A.4	41	R	No	LA	L	Excretory cell, muscle
Y19D10A.4 A'143				LA	H	Muscle, intestine, vulva, vulva muscle, spermatheca; larval/adult; high
Y19D10A.4 A''65				ELA	H	Muscle, spermatheca; all stages; high
Y19D10A.4 A'''33				LA	H	Muscle
T28H11.8	143	R	No	AE	H	Excretory cell
T28H11.8.1 A'243				LA	M	Excretory cell
T28H11.8.1 A''168						No expression
T28H11.8.1 A'''141						No expression
T16H12.9	565	F	No	A	L	Pharynx, excretory cell, anal depressor
T16H12.9 A'667				LA	M	Anterior neuronal, pharynx, excretory cell, hypoderm, tail muscle
T16H12.9 A''598				LA	M	Anterior neuronal, pharynx, excretory cell, hypoderm, tail muscle
T16H12.9 A'''561				LA	M	Pharyngeal isthmus, Anterior neuronal, excretory cell, tail muscle
R12G8.2	79, 1091	F	Yes	A	L	Excretory cell
R12G8.2 A'174				LA	L	Excretory cell
R12G8.2 A''99				ELA	L	Anterior neuronal, excretory cell, anal depressor

Gene	Octamer	F/R	Octamer necessary for excretory cell expression	Stage	Level	Expression Pattern
R12G8.2 A'''70						No expression
R12C12.9	94	F	No	A	L	Excretory cell
R12C12.9 A'193						No expression
R12C12.9 A''112						No expression
R12C12.9 A'''82						No expression
R10H1.2	99	F	No	LA	H	Anterior neuronal, nerve cord, excretory cell
R10H1.2 A'201				LA	H	Anterior neuronal, excretory cell, intestine
R10H1.2 A''123				LA	L	Anterior neuronal, excretory cell
R10H1.2 A'''95				LA	M	Anterior neuronal, nerve ring, excretory cell, intestine, posterior neuronal
R02F2.8	970	R	Yes	LA	M	Excretory cell
R02F2.8 A'1072				ELA	H	Anterior neuronal, excretory cell, intestine, anal depressor
R02F2.8 A''990				ELA	H	Excretory cell
R02F2.8 A'''962						No expression
M176.5	112	F	Yes	LA	M	Anterior neuronal, 1st/2nd bulb pharynx, nerve ring, nerve cord, excretory cell, anal depressor
M176.5 A'163				A	M	Anterior neuronal, 2nd bulb pharynx, excretory cell, nerve cord, intestine, hypoderm, vulva, posterior neuronal, anal depressor
M176.5 A''133				A	M	Anterior neuronal, 2nd bulb pharynx, nerve ring, excretory cell, nerve cord, intestine, vulva, anal depressor
M176.5 A'''108						No expression
M03A8.3	657	F	No	A	L	1st2nd bulb pharynx, excretory cell, vulva
M03A8.3 A'720						No expression
M03A8.3 A''648				LA	M	Anterior neuronal, 2nd bulb pharynx
M03A8.3 A'''616				A	L	1st2nd bulb pharynx
K10C2.4	396	F	No	A	L	Excretory cell
K10C2.4 A'493				ELA	H	Anterior neuronal, intestine, hypoderm, vulva, tail muscle
K10C2.4 A''421				ELA	H	Anterior neuronal, amphids, CAN neuronal, anal depressor, tail muscle
K10C2.4 A'''392				ELA	H	Anterior neuronal, amphids, CAN neuronal, anal depressor, tail muscle
K08F4.4	104	R	No	LA	L	Excretory cell
K08F4.4 A'186				A	H	Excretory cell

Gene	Octamer	F/R	Octamer necessary for excretory cell expression	Stage	Level	Expression Pattern
K08F4.4 A''101				A	H	Excretory cell
K08F4.4 A'''72				A	H	Excretory cell
H22K11.3	637	F	No	A	H	Excretory cell
H22K11.3 A'735				LA	M	Anterior neuronal, nerve ring
H22K11.3 A''662				LA	H	Anterior neuronal, nerve ring, excretory cell, posterior neuronal
H22K11.3 A'''633				LA	H	Anterior neuronal, nerve ring, excretory cell, posterior neuronal
F55F3.4	557	F	No	A	L	Anterior neuronal, excretory cell
F55F3.4 A'668				ELA	L	Intestine
F55F3.4 A''578						No expression
F55F3.4 A'''553				E	H	Intestine
F36F2.7	228	R	No	LA	M	Excretory cell
F36F2.7 A'327				LA	H	Excretory cell
F36F2.7 A''252				LA	H	Excretory cell
F36F2.7 A'''225				LA	H	Excretory cell
F18A11.2	178	R	No	LA	M	Excretory cell
F18A11.2 A'279						No expression
F18A11.2 A''199						No expression
F18A11.2 A'''168				LA	L	1st2nd bulb pharynx, Anterior neuronal, muscle
F16F9.1	799	F	Yes	LA	L	Excretory cell, 2nd bulb pharynx
F16F9.1 A'890				A	L	Anterior neuronal, excretory cell, pharyngeal intestinal valve, anal depressor
F16F9.1 A''819				A	L	Anterior neuronal, excretory cell, anal depressor
F16F9.1 A'''787						No expression
F14B8.7	257	F	No	LA	M	Intestine, excretory cell pharyngeal-intestinal valve, anal depressor
F14B8.7 A'356						No expression
F14B8.7 A''282				ELA	L	Excretory cell, pharyngeal-intestinal valve, intestine, posterior muscle
F14B8.7 A'''248				ELA	M	Excretory cell, pharyngeal-intestinal valve, intestine, posterior muscle
C45G9.5	766	F	No	LA	H	Anterior neuronal, 1st2nd bulb pharynx, excretory cell, intestine, posterior neuronal, vulva, hypoderm

Gene	Octamer	F/R	Octamer necessary for excretory cell expression	Stage	Level	Expression Pattern
C45G9.5 A'865				LA	H	Pharynx, anterior neuron, excretory cell, pharyngeal-intestinal valve, nerve cord, developing vulva, intestine, posterior neuron, hypoderm
C45G9.5 A''789				LA	M	1st/2nd bulb pharynx, excretory cell, vulva, muscle, hypoderm
C45G9.5 A'''758				A	L	Excretory
C17G1.5	307	R	No	LA	M	Excretory cell, unidentified tail
C17G1.5 A'410				L	H	Anterior neuronal, excretory cell, posterior muscle, anal depressor larval
C17G1.5 A''326				L	H	Anterior neuronal, excretory cell, posterior muscle, anal depressor larval
C17G1.5 A'''199				ELA	H	1st2nd bulb pharynx, excretory cell, anal depressor
C05D12.1	205	F	Yes	LA	H	Excretory cell
C05D12.1 A'247				LA	H	Anterior neuronal, excretory cell
C05D12.1 A''168				LA	L	Anterior neuronal, excretory cell, intestine, vulval, spermatheca, tail
C05D12.1 A'''141				A	L	Excretory cell

5.4 Discussion:

I have identified an octamer motif upstream of *aqp-8*'s translation start site that in conjunction with the class III POU homeobox transcription factor, CEH-6, is responsible for *aqp-8*'s expression in the excretory cell. The *cis*-regulatory element is conserved among different species of nematodes in the gene-upstream regions of the *aqp-8* orthologs. In order to define more genes that are modulated by this transcriptional regulatory mechanism, I employed an initial approach of randomly screening the expression pattern of genes with upstream octamer motifs. Although cases were found where the octamer element participated in driving excretory cell expression, the low frequency of positive results in relation to the large pool of candidate promoter regions made testing all the candidates an inefficient means of finding excretory cell modulating *cis*-regulatory elements. However, via this screen, I identified three genes that are modulated in an octamer element-dependent manner.

ZC395.10 is related to HSP90 co-chaperone protein in *S. cerevisiae* and Prostaglandin E synthase 3 in mammals. The function of these two proteins in mammals and yeast are likely the same. Both classes of proteins are involved in wide ranges of functions. Prostaglandins are required in most tissues. Focusing on the kidney, prostaglandins have been found to regulate a wide range of functions including: hemodynamics, renin secretion, growth responses, tubular transport processes and cell fat (for a review see (Nasrallah, Clark, and Hebert 2007)). With such a wide range of possible functions, more studies of this gene's function must be performed.

C02B8.4/hlh-8 encodes a helix-loop-helix transcription factor homologous to the TWIST transcription factors originally identified in *Drosophila*. These transcription

factors dimerize to target E-box *cis*-regulatory elements. The TWIST genes are involved in targeting genes required for mesodermal development (Wang, Zhao, and Corsi 2006). The identification of a transcription factor gene target of CEH-6 indicates that there are additional levels of transcriptional regulation in the excretory cell. This provides interesting implications regarding transcriptional regulatory networks affecting the development of the excretory cell and to whether this regulatory relationship is preserved in other phyla.

C01B12.3 is a *C. elegans* homolog of bestrophin 3, a calcium dependent chloride transporter. *C01B12.3* is expressed in the excretory cell along with the hypoderm, spermatheca and anal depressor muscle (BC12593, (Hunt-Newbury et al. 2007)). Bestrophins are widely expressed in mammalian the plasma membranes of epithelia where they have been suggested to aid managing cellular volume (Fischmeister and Hartzell 2005). Most studies on bestrophins have focused on their association with macular degeneration. Recent studies have, however, pointed to the roles of these proteins in exocrine gland tissues (*e.*g. pancreas, lacrimal and salivary glands), lung, testis and kidney (Srivastava et al. 2008). The proposed role of bestrophin in exocrine glands is to facilitate trans-epithelial movement of chloride ions which leads to water and electrolyte movement (Srivastava et al. 2008).

Two of the genes, *ZC395.10* and *C02B4.8*, have not been previously identified to express in the excretory cell using transgene constructs each containing promoter regions spanning several kilobases (BC10796 and BC20206 respectively). Thus, the method of generating transcriptional reporter transgene constructs, which cut adjacent to cis-regulatory elements, might provide a clearer picture to the entire spectrum of expression

patterns that a gene is expressed in by removing the possible effects of upstream repressor sites.

Other genes in *C. elegans* have been identified to have upstream octamers elements that drive excretory cell expression. An experimentally validated octamer element is found upstream of the gene *nac-2.* This element is conserved in the orthologous promoter regions between *C. elegans* vs. *C. briggsae*. The 5' gene-upstream region of *clh-4* does not contain an octamer motif in either direction within 3000bp upstream of its translational start site. However, a derivative of the octamer element (CTTTGCAT) is located 130bp upstream of the translational start site. As mentioned in the previous chapter, this sequence has been shown to be able to bind POU homeobox transcription factors similarly to the traditional octamer motif (Schwarzenbach, Newell, and Matthias 1995). This alternate site is referred to as a "divergent octamer motif" or an "imperfect octamer motif".

Taking a more directed approach in identifying CEH-6 controlled genes; we identified 107 genes that have conserved octamer motifs in their promoter region (for a more recent version of the comparative analysis using later versions of the genomic sequences, see Appendix 12). Most of these candidate promoters have been assayed for their resulting expression patterns. The frequency of excretory cell expression of these genes is approximately four-fold over the background rate of excretory cell expression in a set of randomly selected promoter regions. More drastically, genes with excretory cell-exclusive expression increases by about ten-fold.

Targeted promoter truncation transgene constructs were used to identify genes that were being directly regulated by the octamer element. The genes that tested positive

for dependence upon the octamer element have roles that would be useful for this excretory cell. C05D12.1 is a homolog of SDR-2 a ferric reductase protein. SDR-2 expression has been observed in mouse brain (Ponting 2001) and kidney where it is suggested to aid in iron reabsorption (Ferguson et al. 2001). The orthologous upstream octamer motifs in the *Caenorhabditis* species are all situated in the same orientation (forward) and not located within an upstream gene model.

twk-36/R12G8.2, is a member of the Twik family, potassium ion channel encoding genes that are widely expressed in neuronal tissues and to a lesser extent in lung, kidney, and skeletal muscle (Lesage et al. 1997). A mouse TWIK protein, TWIK-1, is expressed in the tubular portions of the kidney (proximal tubule, ascending limbs, distal convoluted tubules, and medullary collecting duct) (Nie et al. 2005). The TWIK member, TASK, is sensitive to changes in extracellular pH, suggesting that these proteins have roles in cellular response to pH changes (Duprat et al. 1997). These proteins might also play a role in regulating cellular volume in response to osmotically induced increase in cell volume due to the conductance of these pores being osmotically modulated (Niemeyer et al. 2001). The *twk* gene class is expanded in *C. elegans* with forty-two members of these genes, but like in other organisms, their expression patterns are mostly restricted to neuronal cells (Salkoff et al. 2001). Of all the TWK proteins, only TWK-36/R12G8.2 is found in the excretory cell (Salkoff et al. 2001). Like case of *C05D12.1*, the upstream octamer motifs are all situated in the same orientation (forward) and not located within an upstream gene model among the nematode orthologs.

The best BlastP (Altschul et al. 1990) match of *R02F2.8* in mammals is SLC36A4, a member of the SLC superfamily of proteins. These proteins, like the ABC

transporters, conduct soluble organic molecules. Little is known about SLC36A4, but the other members of the SLC36 family SLC36A1 and SLC36A2 are expressed in the kidney, along with other tissues. Unfortunately, SLC36A3, like SLC36A4, has an unknown expression pattern. SLC36A2/PAT2, which has a specificity for transporting glycine, alanine, and proline (Boll et al. 2002), is transcriptionally regulated by the class III POU transcription factor, Oct-6 (TST-1/SCIP/POU3F1) which drives SLC36A2/PAT2 expression in the sciatic nerve, the longest and widest single nerve in the body. The upstream octamer motifs are located on either strand in the promoter regions of *R02F2.8* orthologs. In addition the entire gene and the octamer element are nested within the ninth intron of *R02F2.2* (Figure 28*A*). This theme is repeated in the *C. briggsae* genome with the *R02F2.8* ortholog nested within the intron of *R02F2.2*'s ortholog (Figure 28 *B*).

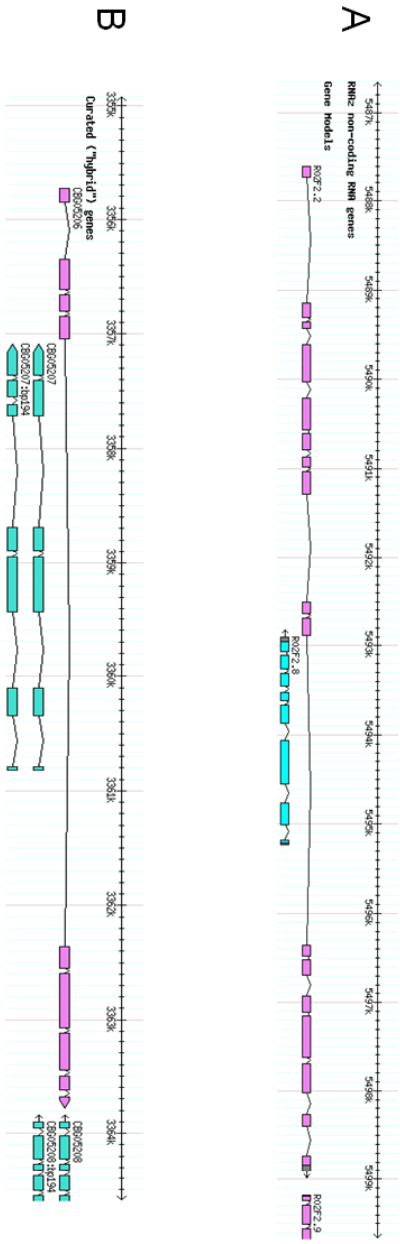

Figure 28. *R02F2.8* is nested within the intron of another gene.

A, The coding region of *R02F2.8* (top) is nested within the intron of *R02F2.2* (bottom) in *C. elegans.*
B, The coding region of the *C. briggsae* ortholog of *R02F2.8* (top) is nested within the intron of the *C. briggsae* ortholog of *R02F2.2* (bottom).
Images were obtained from WormBase.org (WS192).

F16F9.1 is a homolog of the mammalian protein LITAF (aka. Small Integral Membrane Protein of Lysosome/Late Endosome / SIMPLE). LITAF was discovered as a lipopolysaccharide-induced transcription factor that regulates the tumor necrosis factor alpha gene (Myokai et al. 1999). Mutant alleles of human LITAF leads to Charcot-Marie-Tooth disease, a heritable neuropathy characterized by loss of muscle tissue and touch sensation. LITAF is expressed in the sciatic nerve in addition to the other affected cell types in CMT1C (Bennett et al. 2004) where it plays a role in protein degradation (Saifi et al. 2005). The octamer elements are not located on the same strand upstream of the orthologous genes.

Finally, *M176.5* encodes a 211aa gene of unknown function with no homologs outside of nematodes. Domain analysis of the peptide sequence was initially carried out using SMART (Schultz et al. 1998) in order to gain insight into the protein structure and function of *M176.5*. Functional elements were not found using this computational approach (data not shown). Prior genome-wide analysis has demonstrated that this gene is male sex-enriched, but is localized to the pharynx and neurons in both sexes (Thoemke et al. 2005). In addition to the expression pattern assessed by the prior group, I found *M175.6* to be expressed in anterior neurons, the first and second bulb of the pharynx, nerve ring, ventral nerve cord, excretory cell, rectal epithelia, intestine, hypoderm, and vulva. The gene expression enrichment in male nematodes indicates that there should be also be male-specific enhancer element(s) regulating this gene's expression. This expression pattern of *M176.5* overlaps considerably with *ceh-6*'s expression pattern. A promoter truncation resulting in a removal of the octamer motif upstream of *M176.5* leads to a complete loss of reporter expression indicating that CEH-6 might be

responsible for the expression of *M176.5* in more than just the excretory cell. Although little is known about this protein except for its up-regulation in males, I show that it is co-regulated along with the other CEH-6 regulated genes. The closest protein matches (BlastP, (Altschul et al. 1990)) outside of nematodes are the much larger membrane-associated guanylate kinase (MAGUK) family of proteins. Relation between these two proteins is unlikely due to the extreme length differences between the two proteins. The orthologous upstream octamer motifs are all situated in the same orientation (forward) and not located within an upstream gene model.

In chapter 4, I demonstrated that the reverse complement of the octamer element, when fused upstream of the basal promoter sequences of the Δ*pes-10::GFP* construct, could not drive reporter gene expression. In this chapter, I show that the reverse conformation of the motif upstream of *F16F9.1* and *R02F2.8* are necessary for expression of these genes in the excretory cell and therefore caution must be taken when interpreting expression patterns of *cis*-regulatory elements fused to minimal promoter constructs.

Oct6/Tst-1/POU3F1 is a class III POU transcription factor that is required for the myelination of Schwann cells and is responsible for driving expression of R02F2.8's homolog SLC36A2/PAT2/tramdorin1 in the sciatic nerve (Bermingham et al. 2002). Coupled with the fact that both *F16F9*.2 and *R02F2.8* have homologous genes in mammals that are expressed in the sciatic nerve, a commonly assayed nerve of the peripheral nervous system due to its large size, it is possible that CEH-6 carries out the roles of each of the class III POU paralogs in mammals. However, further studies should

be carried out before, ascribing biological function to homologous genes and interactions across evolutionarily distant species.

The truncation constructs produced the unintentional side effects of uncovering enhancer and repressor elements-containing regions that are unrelated to the octamer element. These regions are represented as promoter sections, that when deleted, produce an expression pattern different from that of a larger promoter region (Tables 11 and 15). Although these changes were not analyzed in this study, they could prove to be useful in future studies.

6: DISCUSSION AND CONCLUSIONS

In this thesis, I identify and analyze a novel transcriptional regulatory mechanism governing excretory cell gene expression in *C. elegans* and identify other genes controlled by the same factors. These objectives were carried out via: determining the expression patterns of each member of a candidate gene family (aquaporins), mapping of a *cis*-regulatory element located in the gene-upstream region of *aqp-8*, determining the cognate transcription factor for the upstream DNA element, and testing the *cis*-regulatory element by selecting a set of candidate promoter regions to assay for their ability to drive excretory cell expression.

I chose to study the nematode excretory cell due to its distinct function (osmoregulatory) and large size of the cell. These properties mean that the cell likely has unique developmental and functional requirements. The excretory cell consists of two pairs of tubular processes running parallel almost the entire length of the worm connected at a cell body. These tubular processes are exposed to the worm's closest equivalent to a circulatory system, the fluid-filled pseudocoelomic cavity. The structure of the cell is well suited for its function to maintain osmotic homeostasis.

Aquaporins are highly conserved group of integral membrane proteins found in all forms of life. They are responsible for osmotic and hydrostatic pressure driven transport of water across biological membranes. Animal AQPs are prevalent in renal tissue where they are required for recapture of water lost via primary renal filtration. Due

to their prevalence in renal tissue, I examined the expression patterns of each member of this gene family in *C. elegans* to find appropriate promoter region(s) for further analysis of excretory cell enhancing *cis*-regulatory regions.

I discovered that *aqp-8* is dependent on an upstream octamer element that interacts with the class III POU homeobox transcription factor CEH-6, for expression in the excretory cell. The octamer element upstream of *aqp-8* was discovered via serial 5' truncations the promoter region. Comparative analysis of this sequence against the orthologous region in five nematode species show that the octamer is conserved perfectly over the course of 31 million years separating the most distant two species of the group (Cutter 2008). A directed 5' promoter truncation cutting within the octamer led to loss the promoter fragment's ability to drive excretory cell expression. Although nucleosomes generally have a 175bp periodicity along the genome, the octamer site has a low degree nucleosome representation in mixed stage animals at the sequence corresponding to the octamer sequence (Valouev et al. 2008) indicating that the region is open for transcription factor access throughout much of the worm's life. I also found that the octamer element is able to drive excretory cell expression (in addition to some anterior neurons) when synthetically placed upstream of other promoter elements (Δ*pes-10* basal promoter (Kelly et al. 1997) and the *vit-2* promoter(MacMorris et al. 1992)). These tests demonstrated that the octamer element, by itself, is sufficient to drive expression independently of other conserved sequences in the *aqp-8* promoter region. A consequence of this study is I show that generating targeted synthetic promoter regions is very simple in *C. elegans* due to the simple modular structure of the individual *cis*-regulatory elements.

It is likely that the other conserved regions downstream of the octamer in *aqp-8*'s promoter region are involved in transcriptional regulation. A repressor element is required to suppress the octamer element's influence on anterior neuron expression that is seen in other genes that require CEH-6/octamer modulation such as *M176.5*. In addition, the ability to cope with fluctuations in osmotic pressures is important for terrestrial nematodes due to their large surface to volume ratio and their dependence on turgor for their motion. Determining *cis*-regulatory sites affecting conditional response would be more difficult to assess at this point due to vast number of circumstances that could be tested to understand the function of the DNA elements, but hypertonic stress has been shown to induce expression of *aqp-8* mRNA eight-fold (*pers. comm.* Lamitina T. and Strange K). Although the analysis of these additional conserved sequences in *aqp-8*'s promoter region were not in the scope of this thesis, their presences shows that there must be combinatorial participation of other *cis*-regulatory elements that modulate repression, activity, conditional response, and perhaps even temporal timing of *aqp-8* gene expression.

POU homeobox transcription factors, bipartite DNA binding proteins, were considered the best candidates for binding proteins to the *cis*-regulatory element upstream of *aqp-8* based upon transcription factor database searches using the putative *cis*-regulatory element. With the discovery that POU homeoboxes as the cognate transcription factor, I found that the sequence corresponded to a well studied *cis*-regulatory element, the octamer motif.

I verified that CEH-6, a class III POU transcription factor that is essential for excretory cell development, is the cognate transcription factor for the octamer element

via super shift EMSAs. The binding of antibody::CEH-6 to the octamer was represented by a gel band in a lane containing protein pre-incubated with anti-CEH-6 antibody that migrated slower than a gel band in a lane lacking the antibody pre-incubation. To confirm the association of CEH-6 with *aqp-8*'s promoter region *in vivo,* RNAi was targeted against *ceh-6* in an AQP-8::GFP; *eri-1(RNAi)* sensitized background. The result of the double RNAi was a loss excretory cell reporter expression, again, indicating that CEH-6 is responsible for *aqp-8* excretory cell expression.

The site-directed mutagenesis of octamer elements upstream of *aqp-8* shows that the octamer motif could tolerate a range of nucleotide substitutions and still retain the ability to drive expression. This is despite the fact that the motif is conserved perfectly among five nematode species in the *aqp-8* promoter region. The ability of the octamer element to tolerate the synthetically introduced degeneracy demonstrated that the *cis*-regulatory sequence could be better represented by a degenerate motif pattern. Binding of Oct1 to a range of core octameric sequences has been observed *in vitro* using a SELEX-based method, with even a difference in flanking sequences affecting the relative binding affinity of the protein to the nucleotide sequence (Bendall et al. 1993). The flanking regions of excretory cell-modulating octamer motifs found in this study also have A-T rich flanking regions possibly representing an expanded POU TF binding motif. This characteristic is observed in an alignment of the octamer elements along with limited flanking sequences in the promoter regions of genes with octamer elements demonstrated to drive excretory cell expression (FIG 29 *A-D*).

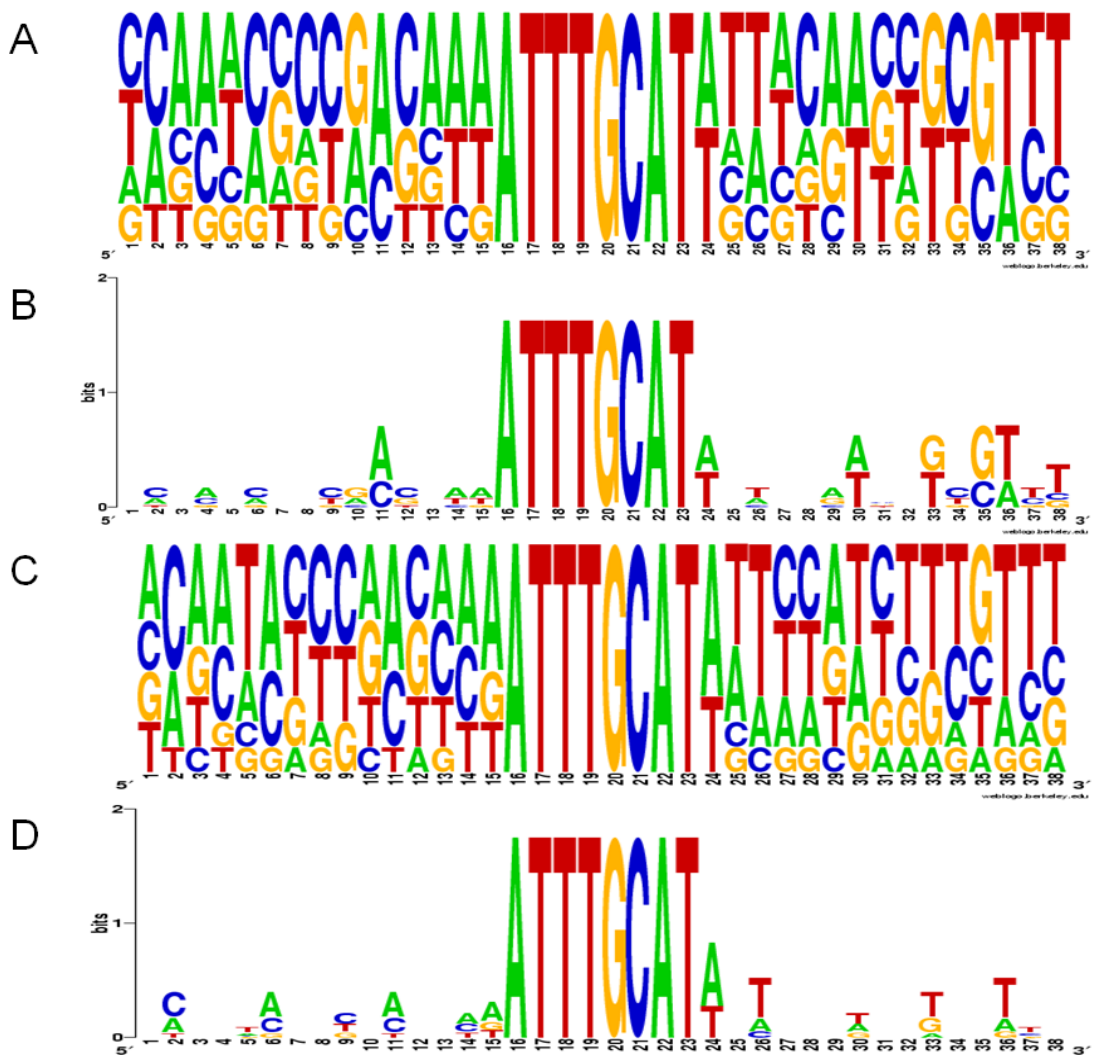

Figure 29. Alignment of octamer and flanking regions of octamer elements responsible for excretory cell expression reveal that flanking residues are A-T rich.

15bp upstream and 15bp downstream flanking regions of the octamer element were used for the WebLogo alignment (http://weblogo.berkeley.edu/) (Schneider and Stephens 1990).

*A,*Percentage-based WebLogo representing the octamer element and flanking regions from promoters with interspecies conserved octamer elements.

*B,*Bit-score-based WebLogo representing the octamer element and flanking regions from promoters with interspecies conserved octamer elements.

*C,*Percentage-based WebLogo representing the octamer element and flanking regions from all promoters with octamer elements necessary for driving excretory cell expression.

*D,*Bit-score-based WebLogo representing the octamer element and flanking regions from from all promoters with octamer elements necessary for driving excretory cell expression.

The A-T rich flanking region is also seen in the alignment of the octamer element upstream of the *aqp-8* orthologs in which there is a flanking adenosine residue on either end (Figure 12). These results indicate that CEH-6 is able to recognize a range of octameric sequence variants and that the element would be more appropriately represented by an octamer-centred degenerate motif of a yet unknown sequence length.

A previous study has shown that *pgp-12* expression is regulated by DCP-66 binding to the upstream enhancer site Ex-1 (CCATACATTA). Loss of either the *cis*-regulatory element or the DNA-binding protein leads to loss of excretory cell expression by the *pgp-12* gene-upstream region. Ex-1 is located 238bp upstream *pgp-12*'s translational start site (Zhao et al. 2005). *In vitro* and genetic analyses confirmed the DCP-66's involvement with the Ex-1 element. Upon further examination of the *pgp-12* promoter region, I detected an octamer-like sequence (ATTTCCAT) partially overlapping the EX-1 element located 241bp upstream *pgp-12*'s translational start site and conserved between *C. elegans* and *C. briggsae*. Analysis of this region using TESS (Schug J 1996) predicts that the octamer-like sequence is a potential target for octamer binding proteins. This variant of the octamer sequence has been shown to have affinity for Oct1 *in vitro* (Bendall et al. 1993).

In the study by Zhao *et al*. (2005), a promoter fragment including 286bp upstream of the TSS drove strong EC expression a GFP reporter in all developmental stages. In contrast, promoter region defined by a -238bp 5' end (cutting within the octamer-like sequence) led to loss of reporter intensity in adult and larval worms and with embryonic expression almost eliminated. A -228bp 5' end construct was not able to drive expression of GFP at all. Therefore, it appears that loss of the octamer-like element at -241bp

possibly plays a part in regulating *pgp-12* expression and that this regulation might work in concert with the DCP-66/Ex-1 mechanism or act as an alternate mode of transcription. Mutagenesis of the canonical octamer sequence in *aqp-8*'s promoter region to the octamer-like sequence found upstream of *pgp-12* led to a loss of the promoter region's ability to drive excretory cell expression, this might indicate that differences between elements downstream of the octamer element in the promoter regions of *aqp-8* and *pgp-12* lead to differential ability to recruit CEH-6.

The *C. elegans* genome contains almost 2,000 genes with an octamer sequences within 1,200bp upstream of the translational start site. Random sampling of these promoter regions proved to be an exhausting method for finding for CEH-6 interacting octamer motifs with only four promoters out of the thirteen regions sampled having octamer motifs that were necessary for excretory cell expression. By taking advantage of the sequenced genomes of *C. elegans, C. briggsae* and *C. remanei*, *promoter::reporter* transgenic constructs were generated for promoter regions of genes that have conserved octamer elements. The overall group of expression patterns resulting from these constructs biased towards enrichment in the excretory cell. The excretory cell-expressing genes from this selected set were then further dissected via octamer motif-targeted 5' promoter truncation constructs. From this test, six promoter regions (including *aqp-8*'s) were found to have conserved octamer motifs that are required to drive expression in the excretory cell. Many of the genes were found to have functions consistent with roles in maintaining osmotic balance or related to epithelial tissue development which would be important in the excretory cell. One of the genes, *F16F9.1*, encodes a transcription factor that homologous to LITAF, a lipopolysaccharide induced transcription factor that

regulates TNFα and other cytokines. LITAF is expressed predominantly in lymphoid tissues (Myokai et al. 1999) indicating a likely role for the protein in immunological response mechanisms. The expression of F16F9.1 in the excretory cell, a tissue exposed to the environment, might indicate that this protein is required for response to invasion from foreign bodies.

The octamer element functions in either its forward or reverse complement orientation. The forward orientation of the octamer was found in the promoters of: *K02G10.7*, *M176.5*, *R12G8.2*, and *C05D12.1*, while the reverse conformation was found in the promoters of: *F16F9.4* and *R02F2.8*. It is possible that the directional preference of the motif might have functional implications as there is a difference in the orientation of the element upstream of immunoglobulin light chain genes (ATTTGCAT) and immunoglobulin heavy chain genes (ATGCAAT) (Parslow et al. 1984). In addition, human Oct1 is auto-regulated by two upstream octamer sites. The proximal element consists of an octamer in the forward orientation and the distal octamer in the reverse condition. Both sites can recruit octamer binding proteins with equivalent affinity, but each site has a different effect on *Oct1* expression levels (Pankratova, Sytina, and Polanovsky 2006). A larger sample size must be assessed to determine the significance of the octamer element's conformation in relation to its modulatory effects in *C. elegans*.

The candidate promoter regions, which tested positive for excretory cell expression modulating octamer elements, are linked to downstream gene products that have functions that are consistent with expected requirements of the excretory cell. The fact that they are co-expressed indicates that these transcripts could be required in a similar capacity and during similar periods in development. It is therefore possible to start

extrapolating functions of genes with limited prior analysis such as *M176.5*, a gene that only known to be expressed at a higher level in male nematodes.

Most of the promoter regions carrying interspecies conserved octamer elements did not appear to be regulated by CEH-6. For some of these promoter candidates, it might be a coincidence that the octamer motif is found in each of the orthologous promoters. A simple calculation using the assumption that *C. elegans* intergenic GC-content is 31% (Webb et al. 2002), however, shows that the probability of an octamer element occurring is one incidence per 25kb (or 12kb if looking at both strands). With this in mind, it is unlikely that these sequences would remain conserved under neutral selection over millions of years of separation.

The conservation of the element in these promoters therefore raises the prospect that these sites are indeed functional. Therefore, it still is possible that these sites are binding sites for CEH-6. Some scenarios that might be occurring are:

1. The octamer/CEH-6 mechanism drives excretory cell expression along with another transcriptional regulatory mechanism in a redundant manner (in the cases where excretory cell expression was not lost upon its deletion).
2. The octamer/CEH-6 mechanism drives excretory cell expression along with another transcriptional regulatory mechanism, which is required (in the cases where excretory cell expression was not seen in any of the targeted 5' truncation constructs).
3. The octamer/CEH-6 mechanism drives expression in tissues other than the excretory cell (not the focus of my assay).

4. The octamer/CEH-6 mechanism was repressed.
5. The octamer/CEH-6 mechanism is acting as a repressor.
6. The octamer element is functioning to recruit octamer binding proteins other than CEH-6.

Further study into these sites could be carried out to determine if any of the above listed scenarios are taking place.

The gene-upstream region of *ceh-6* contains a reverse octamer sequence located at -1,167bp in *C. elegans*, and one at -1,255bp upstream of the *C. briggsae* ortholog. The sequence located within a predicted non-coding RNA gene (class RNAz). The RNAz class is based upon predictions via the program RNAz which predicts structurally conserved and thermodynamically stable RNA secondary structures in multiple sequence alignments (Washietl, Hofacker, and Stadler 2005). A possible avenue for further study is to determine whether CEH-6 is auto-regulated as this is a common theme among POU transcription factors in other organisms (Pankratova, Sytina, and Polanovsky 2006; Chen et al. 1990; Trieu et al. 2003).

Vertebrate genomes encode four different class III POU transcription factors. Some studies have focused on these transcription factors relative to their function in the kidney (Lan et al. 2007; Hauptmann and Gerster 2000; Nakai et al. 2003). The study of POU III TFs in the kidneys is largely over-looked due to the majority of the studies centring on POU III TF effects on nervous system tissues where more severe effects from mutants are generally observed. CEH-6 is the only member of the class III POU TFs in *C. elegans.* This lack of POU III TF paralogs in nematodes provides the ability to apply a

reductionist approach for analyzing transcriptional targets. This system would help in understanding how the class III POU TFs are involved in kidney tissue function (in addition to its functions in neuronal tissue).

One of the CEH-6-regulated genes identified in this study, *R02F2.8*, is homologous to the SLC36 family of soluble organic molecule transporter. SLC36A2/PAT2/Tramdorin1, is regulated by the mammalian class III POU TF Oct-6/POU3F1(Bermingham et al. 2002). This parallel in transcriptional regulation between nematodes and mammals is arguably irrelevant, but should not be ignored due to other examples of transcription factor/binding motif association being preserved over long evolutionary distances. For example, the GATA-4 transcription factor in mammals and ELT-2 in *C. elegans* (BlastP e-value 4^{-24} over 36.4% of ELT-2's length, WormBase, WS195) bind the same consensus sequence (WGATAR) (Yamagata et al. 1995; Hawkins and McGhee 1995) indicating an evolutionary constraint to maintain these binding relationships. Understanding of the target genes of these transcriptional binding associations in nematodes should provide a framework for the understanding of expected binding associations and target genes of these interactions in other organisms.

APPENDICES

Appendix 1. Sequences used for multiple alignments of AQP proteins

>AQP-1 (a) CE29304 [Caenorhabditis elegans]
MLLRFIRKIMTAEEDTLPERLRFHGVHTNILARNLIAEFFGTFLLCFIGLSIVFQFHAGGGKTTEWIGVNIGWGFAIMFAVMATARMSG
GHLNPAVSLLLWSLGHLKLAWVPLYAIAQTAGAFVASLGMYSYYYEQFNAFDGGNRTILGATGTAGCFASYPSPNLGVWGPYIDQCVGT
GVLAYFLCVVIDERNQIPKIWHPMFFGFLVMMIGTGFGMNIGYPINPARDLGPRLFSYFIYGPGVFHSPYPNYWLAPAIAPFVGALVGG
WFYHFSLGMHNPDIEEADDIFVQQPPKSVEQQKLLQA
>AQP-2 (a) CE00863 [Caenorhabditis elegans]
MILDKLRAKFHIRKELLRAVLAEFTGTYLLCLIGLSVVAQKVLPRPEVNEFIGVNVGFGIAIVFGVAVSAKLSGGHINPAVSFAFLSVG
QITIVQFIAYFVAQFFGAFFGAATVYAVYNDAINVFDGGVRTVGGPKDTAGIFASYPAPHLGLVNGFVDQFVATAVFVFLIAHIVDKRN
SYPTWLQPILVGTGFVAIGAAFGYNCGYPVNPARDFAPRLFTSIFYGGAVFTKWFWVPIVGPFVGAVVGIWLYYFLIGFHTPQDAEEKY
VVLTGNQELKPLTAKETVDEEAA
>AQP-3 CE22814 [Caenorhabditis elegans]
MLSDSSCSSSDRSFKFPFDTTDALSIHELAVKDLPKPDAENPFSVAMHSPPGSPPFAVDRKSVDNSVVAVTDTPFEFAPSQKSSQHTNR
PPPFVKPEEEMMYINHVERLKPKFAIGNELIRAFLAELFCTGFLVFGGECVNAQYVLSQGKNNEWIGISVGWGLVLMLAVLMGSKISGA
HLNPAVSFFQLTQGKINLIRFLVYAVAQNIGAFLGAFGVFCVYYDAINVFEGGNRTVTGPTATASIFATYPGPFLGTFNAIVDQIAGTL
VLCLGVAAITDRRNGIPAFLQPAWIGALLAFLGMSLALNAGYAINPARDFAPRLFNLCAGYGWEVFSYRNYKWFWIPIICPMIGGVLGA
WLYEFFIGFHIQDEDAVSLDSESDKQLKTMIDNMVDIENQLPEYTDKKQLSDIASIHQNPSLRNI
>AQP-4 CE31345 [Caenorhabditis elegans]
MVSPYEEDSRPPYMSSYAEETWGQPATTNRKSSYTSRKKEYSLLTKCVAEFLGDLTFVYVGTMQASLFQYADGILHAAFAHGFTIFILV
TAFGHISGGHFNPAVSWAIAGAGKMPIFHLPFYVVSQLLGGICGAFLTAAVLSQEQLTSCEAGATLLSPGSQWWQGLIAETVVTFFLVH
TILITAADTDTVTLAPLAIGLTLSIDILSTGSITGASMNPARSLGPSIIGSIFATQKTSFYWNNHYIYWAGPLLGSTIALCIYKLFESR
EERIVR
>AQP-5 CE05359 [Caenorhabditis elegans]
MSMNSQKTSTVRPYNLISRCYAEFLGTFIFIFSGTMQANVYDISQPVGLTHAALTHGLATIVVIAVFGKISGGHFNPVVSWAMVLCQKL
HPFELPFYMFSQFFGGFAGNLLSACLQRKRDFLNWEDYSSIRYPLPTASIEYGYDKVHNSTLEKTILLTTQLAATTSGITHLGENHEWW
EGLISETITTYFFVTVILMNVVNNEPSEATPFIIGMMVIVNIFATASITGTAMNPVRALSPNIVGEIVLSSSSLPPNFWTYHYIYWAGP
YLGSTIAVIGFKLLLSKTDRLIP
>AQP-6 CE05345 [Caenorhabditis elegans]
MVEDEKDYTIYSKCAAEFIAVLLFVYIGSMQAAGVFLHDGVLHAAFAHGVAIFVLAATFGGVSGAHINPAVTFGIALVGRISPIHAVCY
VVSQLLGSVFGALLVRISLPYKMYNVISAGATLCGKGYNWQEGLTAEIVTTYILVQTVLLCAVDTDKNRLAPLAIGFSLIIEILAAGAI
SGASMNPARSFGPNIMGQVFLKPEHLDAQYMYWNYHWIYYIGPIIGAFIAAGVYRMFFARDYRVLA
>AQP-7 CE34058 [Caenorhabditis elegans]
MAAELERTEQVRAKIQIKNPLLRNALSEFFGTFLLLFIGIGIVMQFILSNEKLNTWININLGWGLAIAFTVYTCSKTSGGHFNPAVSIA
FLTLGKLPFKDFLVYCVVQTIGAALGSAAAFGLYYDQFVKFAGAYRTILGPKATAGCFCSYPALHVSNTTAFFDQFAGTALLVLFVCVV
IDKRNGIPGAAHPLLFGLVVMMIGTAYGMNLGYPINPARDLGPRLFSFFIYGSGVFSYHSYYFWIPVIAPLFGAIFGAWSYTFFVGAHI
PDQRETTYVLVDEANQPLKLATDA
>AQP-8 CE34331 [Caenorhabditis elegans]
MGVFQDKVASILRIEDQQFTRELLAECIGTFFLLLIGNAANIQAAVAVGGNSTSCHIAWGIGFMFAVYLAASVSGGHLNPAISVAQSIL
GNLPPWKIIPYAIAQVIGAFLGAAVAYFGHHDDLWKLDGGIRQVTGGQATAGLFTTFPPDHMSVWGSLLDQIIGTAMLSGLVCLITDKR
HQIPTGVVPVLAGSIMSMVAMTFGANGGFAINPARDFGPRVFCLCAGYGWEVFSAHGYYFWIPIVGALIGSIIGAWIYKIFVGLHGMNE
SLDIQPAKGFNVSVKVDREYSDSSGSY
>AQP-9 CE41143 [Caenorhabditis elegans]
MRIWVASLIFYGSVFAICELLRFLVIKSFDNSKRLSALLILEFIGTLQICVPMFDVGTILDNYGLLGVFVEITVIELANCYFQRDAVAH
PCPLVTNCYRKSKAIRRGIYVFLVQLAAAYLSYFVARLFWSIGVHPIHLELLDAESCSSDLTVAITTGCIIEGVATFVAKWFEKYVDER
YDGETKLCSIANCLFSGLLCAIGINYTGMYANPIVAWACTFNCLGVSHAGHLFVYWLSPLIAWYFAEIVFGSEDVLEEESEEQEKDTKK
KE
>AQP-10 CE01710 [Caenorhabditis elegans]
MEAVSSEYYFPLYSALGYFALVFGIGEIARIITAKYVSPRGNSQLFLYELIGTIQMCTCVYENGIIFKNYGFPAIFICVALLLTAGNIF
NRGAMTNCAPIFEQFVFGNLGSSKFLTILSAQLIGATFASKFAYLIWNITAPYSTAHLENASNLECILHYKQTAGIVIGFEIVGAFVVR
IVVAQLLARPALIKLIPFAISAYLSLALYVVGVPGLNPIVATARLYGCRGIDNSSFFILYWFCPVLGWLTGAYVVGQKSPSKKSAKDVK
AEKKAKAAAKKSD
>AQP-11 CE33909 [Caenorhabditis elegans]
MEISGPLTDALIYYLTVILVCEGARHVADHLFDKKGSVHRFIIEFLGTLQVTTTIYENAVIDIYLGRQAFAITLFSTGLIFALCNRTAF
CSPLAPIEQYLFGKLRLGELLQTLAAQFTAGYFAFSFARTIWLRAYSTTDAHSYVMGLMESCGFNHPYPIYYHLAIELIGTFIVRHVLT
RATSEARDSRVRFVFPALFMAAVFTGTVTFVGDQALDPLVASTLFFGCRGLNFENYMLVYWIAPTIGWMASAYWDSTGEESSKKKAAKE
KKAEKKRAKKNE
>AQP-12 CE35113 [Caenorhabditis elegans]

MDDELAKSDIKHSQFHNLLIRPNIGEFLGAVIFSFLACFAGQYQRSNDLVYPFLSAFSLYIARCLVSHLTPAHLNPAISLLQWLRNEIP
LVLLITFCFVQLIGFLFGVTLFRALVTQTEFNDYIVMYEIVAVDGTRKINRLQAFLLEVVLSMIFFMANALEDRQEPTVAATWGFIQFV
SYPLYGFTSNISLLLVTSTVSYIFSPLTTPSFLLLYLNVFASLLAVMLAWCIDIISRPSPAAIGE
>AQP0 gi|6912506|ref|NP_036196.1| major intrinsic protein of lens fiber [Homo sapiens]
MWELRSASFWRAIFAEFFATLFYVFFGLGSSLRWAPGPLHVLQVAMAFGLALATLVQSVGHISGAHVNPAVTFAFLVGSQMSLLRAFCY
MAAQLLGAVAGAAVLYSVTPPAVRGNLALNTLHPAVSVGQATTVEIFLTLQFVLCIFATYDERRNGQLGSVALAVGFSLALGHLFGMYY
TGAGMNPARSFAPAILTGNFTNHWVYWVGPIIGGGLGSLLYDFLLFPRLKSISERLSVLKGAKPDVSNGQPEVTGEPVELNTQAL
>AQP1 gi|37694062|ref|NP_932766.1| aquaporin 1 [Homo sapiens]
MASEFKKKLFWRAVVAEFLATTLFVFISIGSALGFKYPVGNNQTAVQDNVKVSLAFGLSIATLAQSVGHISGAHLNPAVTLGLLLSCQI
SIFRALMYIIAQCVGAIVATAILSGITSSLTGNSLGRNDLADGVNSGQGLGIEIIGTLQLVLCVLATTDRRRRDLGGSAPLAIGLSVAL
GHLLAIDYTGCGINPARSFGSAVITHNFSNHWIFWVGPFIGGALAVLIYDFILAPRSSDLTDRVKVWTSGQVEEYDLDADDINSRVEMK
PK
>AQP2 gi|685001|gb|AAB31999.1| water-channel aquaporin 2; AQP2 [Homo sapiens]
MWELRSIAFSRAVFAEFLATLLFVFFGLGSALNWPQALPSVLQIAMAFGLGIGTLVQALGHISGAHINPAVTVACLVGCHVSVLRAAFY
VAAQLLGAVAGAALLHEITPADIRGDLAVNALSNSTTAGQAVTVELFLTLQLVLCIFASTDERRGENPGTPALSIGFSVALGHLLGIHY
TGCSMNPACSLAPAVVTGKFDDHWVFWIGPLVGAILGSLLYNYVLFPPAKSLSERLAVLKGLEPDTDWEEREVRRRQSVELHSPQSLPR
GTKA
>AQP3 gi|49457003|emb|CAG46822.1| AQP3 [Homo sapiens]
MGRQKELVSRCGEMLHIRYRLLRQALAECLGTLILVMFGCGSVAQVVLSRGTHGGFLTINLAFGFAVTLGILIAGQVSGAHLNPAVTFA
MCFLAREPWIKLPIYTLAQTLGAFLGAGIVFGLYYDAIWHFADNQLFASGPNGTAGIFATYPSGHLDMINGFFDQFIGTASLIVCVLAI
VDPYNNPVPRGLEAFTVGLVVLVIGTSMGFNSGYAVNPARDFGPRLFTALAGWGSAVFTTGQHWWWVPIVSPLLGSIAGVFVYQLMIGC
HLEQPPPSNEEENVKLAHVKHKEQI
>AQP4 gi|1680710|gb|AAB26958.1| aquaporin 4 [Homo sapiens]
MVAFKGVWTQAFWKAVTAEFLAMLIFVLLSLGSTINWGGTEKPLPVDMVLISLCFGLSIATMVQCFGHISGGHINPAVTVAMVCTRKIS
IAKSVFYIAAQCLGAIIGAGILYLVTPPSVVGGLGVTMVHGNLTAGHGLLVELIITFQLVFTIFASCDSKRTDVTGSIALAIGFSVAIG
HLFAINYTGASMNPARSFGPAVIMGNWENHWIYWVGPIIGAVLAGGLYEYVFCPDVEFKRRFKEAFSKAAQQTKGSYMEVEDNRSQVET
DDLILKPGVVHVIDVDRGEEKKGKDQSGEVLSSV
>AQP5 gi|49456997|emb|CAG46819.1| AQP5 [Homo sapiens]
MKKEVCSVAFLKAVFAEFLATLIFVFFGLGSALKWPSALPTILQIALAFGLAIGTLAQALGPVSGGHINPAITLALLVGNQISLLRAFF
YVAAQLVGAIAGAGILYGVAPLNARGNLAVNALNNNTTQGQAMVVELILTFQLALCIFASTDSRRTSPVGSPALSIGLSVTLGHLVGIY
FTGCSMNPARSFGPAVVMNRFSPAHWVFWVGPIVGAVLAAILYFYLLFPNSLSLSERVAIIKGTYEPDEDWEEQREERKKTMELTTR
>AQP6 gi|86792455|ref|NP_001643.2| aquaporin 6 [Homo sapiens]
MDAVEPGGRGWASMLACRLWKAISRALFAEFLATGLYVFFGVGSVMRWPTALPSVLQIAITFNLVTAMAVQVTWKASGAHANPAVTLAF
LVGSHISLPRAVAYVAAQLVGATVGAALLYGVMPGDIRETLGINVVRNSVSTGQAVAVELLLTLQLVLCVFASTDSRQTSGSPATMIGI
SVALGHLIGIHFTGCSMNPARSFGPAIIIGKFTVHWVFWVGPLMGALLASLIYNFVLFPDTKTLAQRLAILTGTVEVGTGAGAGAEPLK
KESQPGSGAVEMESV
>AQP7 gi|115527728|gb|AAI19674.1| AQP7 protein [Homo sapiens]
MNAAVTFANCALGRVPWRKFPVYVLGQFLGSFLAAATIYSLFYTAILHFSGGQLMVTGPVATAGIFATYLPDHMTLWRGFLNEAWLTGM
LQLCLFAITDQENNPALPGTEALVIGILVVIIGVSLGMNTGYAINPSRDLPPRIFTFIAGWGKQVFRYCPCPGPFL
>AQP8 gi|26251901|gb|AAH40630.1| AQP8 protein [Homo sapiens]
MCEPEFGNDKAREPSVGGRWRVSWYERFVQPCLVELLGSALFIFIGCLSVIENGTDTGLLQPALAHGLALGLVIATLGNISGGHFNPAV
SLAAMLIGGLNLVMLLPYWVSQLLGGMLGAALAKAVSPEERFWNASGAAFVTVQEQGQVAGALVAEIILTTLLALAVCMGAINEKTKGP
LAPFSIGFAVTVDILAGGPVSGGCMNPARAFGPAVVANHWNFHWIYWLGPLLAGLLVGLLIRCFIGDGKTRLILKAR
>AQP9 gi|49457007|emb|CAG46824.1| AQP9 [Homo sapiens]
MQPEGAEKGKSFKQRLVLKSSLAKETLSEFLGTFILIVLGCGCVAQAILSRGRFGGVITINVGFSMAVAMAIYVAGGVSGGHINPAVSL
AMCLFGRMKWFKLPFYVGAQFLGAFVGAATVFGIYYDGLMSFAGGKLLIVGENATAHIFATYPAPYLSLANAFADQVVATMILLIIVFA
IFDSRNLGAPRGLEPIAIGLLIIVIASSLGLNSGCAMNPARDLSPRLFTALAGWGFEVFRAGNNFWWIPVVGPLVGAVIGGLIYVLVIE
IHHPEPDSVFKAEQSEDKPEKYELSVIM
>AQP10 gi|55663099|emb|CAH70483.1| aquaporin 10 [Homo sapiens]
MVFTQAPAEIMGHLRIRSLLARQCLAEFLGVFVLMLLTQGAVAQAVTSGETKGNFFTMFLAGSLAVTIAIYVGGNVSGAHLNPAFSLAM
CIVGRLPWVKLPIYILVQLLSAFCASGATYVLYHDALQNYTGGNLTVTGPKETASIFATYPAPYLSLNNGFLDQVLGTGMLIVGLLAIL
DRRNKGVPAGLEPVVVGMLILALGLSMGANCGIPLNPARDLGPRLFTYVAGWGPEVFSAGNGWWWVPVVAPLVGATVGTATYQLLVALH
HPEGPEPAQDLVSAQHKASELETPASAQMLECKL
>AQP11 gi|27370565|ref|NP_766627.1| aquaporin 11 [Homo sapiens]
MSPLLGLRSELQDTCTSLGLMLSVVLLMGLARVVARQQLHRPVAHAFVLEFLATFQLCCCTHELQLLSEQHPAHPTWTLTLVYFFSLVH
GLTLVGTSSNPCGVMMQMMLGGMSPETGAVRLLAQLVSALCSRYCTSALWSLGLTQYHVSERSFACKNPIRVDLLKAVITEAVCSFLFH
SALLHFQEVRTKLRIHLLAALITFLVYAGGSLTGAVFNPALALSLHFMCFDEAFPQFFIVYWLAPSLGILLMILMFSFFLPWLHNNHTI
NKKE
>AQP12 gi|47115836|sp|Q8IXF9.1|AQ12A_HUMAN Aquaporin-12A (AQP-
12)MAGLNVSLSFFFATFALCEAARRASKALLPVGAYEVFAREAMRTLVELGPWAGDFGPDLLLTLLFLLFLAHGVTLDGASANPTVSL
QEFLMAEQSLPGTLLKLAAQGLGMQAACTLMRLCWAWELSDLHLLQSLMAQSCSSALRTSVPHGALVEAACAFCFHLTLLHLRHSPPAY
SGPAVALLVTVTAYTAGPFTSAFFNPALAASVTFACSGHTLLEYVQVYWLGPLTGMVLAVLLHQGRLPHLFQRNLFYGQKNKYRAPRGK
PAPASGDTQTPAKGSSVREPGRSGVEGPHSS
>GlpF ecoli gi|89110102|ref|AP_003882.1| glycerol facilitator [Escherichia coli W3110]
MSQTSTLKGQCIAEFLGTGLLIFFGVGCVAALKVAGASFGQWEISVIWGLGVAMAIYLTAGVSGAHLNPAVTIALWLFACFDKRKVIPF
IVSQVAGAFCAAALVYGLYYNLFFDFEQTHHIVRGSVESVDLAGTFSTYPNPHINFVQAFAVEMVITAILMGLILALTDDGNGVPRGPL
APLLIGLLIAVIGASMGPLTGFAMNPARDFGPKVFAWLAGWGNVAFTGGRDIPYFLVPLFGPIVGAIVGAFAYRKLIGRHLPCDICVVE
EKETTTPSEQKASL
>AqpZ ecoli gi|26107299|gb|AAN79482.1|AE016758_86 Aquaporin Z [Escherichia coli CFT073]

MDMFRKLAAECFGTFWLVFGGCGSAVLAAGFPELGIGFAGVALAFGLTVLTMAFAVGHISGGHFNPAVTIGLWAGGRFPAKEVVGYVIA
QVVGGIVAAALLYLIASGKTGFDAAASGFASNGYGEHSPGGYSMLSALVVELVLSAGFLLVIHGATDKFAPAGFAPIAIGLALTLIHLI
SIPVTNTSVNPARSTAVAIFQGGWALEQLWFFWVVPIVGGIIGGLIYRTLLEKRD

Appendix 2. List of *aqp*^{promoter}*::reporter* constructs.

Common name	Gene Locus	Construct size Mah size / primer location	Strain GFP	Strain PEST
aqp-1	*F32A5.5*	2952bp / (-)2968 - (-)16	BC20117	BC20019, BC20025
aqp-2	*C01G6.1*	2897 / (-)2887 - (+)10	BC13767, BC16348, BC16349, BC20099, BC20100	BC20039, BC20040
aqp-3	*Y69E1A.7*	2987bp / (-)2986 - (+)11	BC20096, BC20348	BC20026
aqp-4	*F40F9.9*	2111bp / (-)2157 - (-)47	BC20097	BC20027, BC20028
aqp-5	*C35A5.1*	2925bp / (-)2951 - (-)27	BC20115	BC20034, BC20041
aqp-6	*C32C4.2*	2925bp / (-)2951 - (+)15	BC20050	BC20035
aqp-7	*M02F4.8*	2968bp / (-)2982 - (+)5	BC20051	BC20029, BC20037
aqp-8	*K02G10.7*	1556bp / (-)1575 - (-20)	BC20052	BC20018
aqp-9	*K07A1.16*	2072bp / (-)2061 - (+)10	BC20116	BC20030, BC20031
aqp-10	*ZK1321.3*	921bp / (-)933 - (+)13	BC20053	BC20036
aqp-11	*ZK525.2*	2952bp / (-)2937 (+)14	BC20098	BC20032, BC20033
aqp-12	*Y57A10A.35*	2507bp / (-)2563 - (-)56	BC16792	n/a

Appendix 3. Primers used to for amplification of promoter regions of *aqp*^{promoter}*::GFP* transgene constructs.

Common name	Gene locus	(-) TSS (bp)	Forward primer	Reverse primer
aqp-1	*F32A5.5*	1756	CTGATTTTCTTTGGTTCGTCG	AGTCGACCTGCAGGCATGCAAGCTCCTGATGATTTTCTGAAATG
aqp-2	*C01G6.1*	2882	GACGGAACGAGAGAAATTGG	AGTCGACCTGCAGGCATGCAAGCTCCAAGATGATTTTGACCTGGA
aqp-3	*Y69E1A.7*	3082	AATTGCAGATTCCAATTTTCG	AGTCGACCTGCAGGCATGCAAGCTTCGGACAAGATTGAAATGAAGA
aqp-4	*F40F9.9*	2276	TTCATAAACGTGGATGCCTTC	AGTCGACCTGCAGGCATGCAAGCTAGTTTCAAATGCAAAACGTGG
aqp-5	*C35A5.1*	3096	CATTTCGGGGATTTCTAGTCC	AGTCGACCTGCAGGCATGCAAGCTTTTACGACGTTGCTGGAAAAC
aqp-6	*C32C4.2*	2965	TCAGAGATGAAAAAGGAAAGCA	AGTCGACCTGCAGGCATGCAAGCTCTCATCTTCAACGATTTTCGG
aqp-7	*M02F4.8*	3094	CGAACTTTAAATGCGTCCCTT	AGTCGACCTGCAGGCATGCAAGCTAACCGAAAGATGAAAAGCGAT
aqp-8	*K02G10.7b*	1572	GATGTTATCTGAATTGGATGCTC	AGTCGACCTGCAGGCATGCAAGCTTAGAAACGGATCGCAGAAAA
aqp-9	*K07A1.16*	2012	CGAAAAACATATGGATGCGTTA	AGTCGACCTGCAGGCATGCAAGCTCCAAATTCGGATTTTTCAAATC
aqp-10	*ZK1321.3*	1043	ATTTATTTGTGCGATCCTTGC	AGTCGACCTGCAGGCATGCAAGCTAAACTGCTTCGATTTTTGGCT
aqp-11	*ZK525.2*	3076	GCAACACGGCATAAACAGATT	AGTCGACCTGCAGGCATGCAAGCTCCCGAGATTTCGATAACTTAAAA
aqp-12	*Y57A10A.35*	4000	GAGGTTTTTGACGTGGCAAT	AGTCGACCTGCAGGCATGCAAGCTCCAATTCATCGTCGATTTTGT

Appendix 4. Sequences from Wormbase (WS190) used for multiple sequence alignment of promoter regions.

Non-coding regions are in lowercase and coding regions are in uppercase.

>Ce aqp-8 -550,+50(aqp-8)

gaattcatattttgcactcattattattcttacgaaaaataaacattcctctgaactagttgaatccgaactggaattgaacttttaga
tgttgcgcataattgaatttgatacaaaaaagtcatgtgaaaggtattcgaaaaaattctttgaaaatttccaaaagtcattttagaca
caattaggttatagccatgaaaatgtaacttgagcacccctaaatgtttatagtagttgtatatgattcactgttgaatgcgtaaccgc
aaaatgattgccaaaatttgcatactggaattgcgagtcaatcaatctggattatgataagtagttcccaaacccaaaaacttttcacc
cgcctctgttttatctatttttcgtgttagtagaatatcagggcctaccgtaacatcccgcccattctgaagtttcgacagtgtccact
tccctatttaaaatcagcgcacgccttctagataaggacattttccgtccattaccatttctttttttttctgcgatccgtttctataat
ttatttatttttagatATGGGTGTCTTTCAAGACAAAGTTGCCAGTATATTGCGAATTGAAGATCA

>cbre aqp-8 -550,+50(aqp-8)

ttctttttttcaacccgtctaggaatgtgacacaaagatcagttgctttttgtggtccagttcagaaaactgaactcgtgaatgttgc
gtataattaaattcatttattacagaaaagtcatgcgaaatgtatacgaaaacgctctttgaaaattgtctgaagtaacttcaaatcta
attatagtcccaaagatccggtgaaactttattctatgttccccgttaaacattttgcagacatgattcatgtaaccgttttgccatgg
ttctagttgtcaaaatttgcatattgaaactgagtcaatcaatctggattatgataagtattctcatttttcaaccgaagcaaaaaccc
gcctccaaatgtattcgtatcttctcttagtttgtgtcaccggctaccgtaacttcccgcccattctgaagttctgccgtgtccacttc
cctatttaaaatgagcgcacgccttctcgataaggacattttccgtccgttgatattgtgttgtccggctagggtgcctaattttattc
tttatttttttcagatATGGGTGTCTTTCAAGACAAAGTCGCCAGCATACTGCGAATTGAAGATCA

>Cjap aqp-8 -550,+50(aqp-8)

agactcaattctgcgcataattttatcaacggaaaaaagaaaaaaaaaagcttgtgacctatacgaaaatgattttcgaaagtttttta
ttctgcaaacatttacgttactaaatgcgtaaattaaattattcaaattggataaataactctaattctccaatttgaaaagatcattg
ccaagatatttcgactttcaaacgcctagtgtttttccgtccccctgactcataaaacgtattgttgctctgcaaaatcaattcatttt
gaaactgaaatttgcatagttttgtatctatcaaatcattaccataagtcgtttttcaccgaaaacccgccttcgtccgattttctctc
aaagttttcacaactaccgtaacttcccgcccattctaaagtttcgaccaagtctactctcctatttaaaatagagaaatgcttcctga
ataggttcattttgcggtcgatttgtcgttttctttttttgctttgttttgtgattattttaacttttgctcttcaagtctttaatct
acaatatcctttagatATGGGTGTGCTTCAAGACAAAATGGCGAGCATACTTCGAATTGAAGATCA

>Crem aqp-8 -550,+50(aqp-8)
tttccaacttagttcataaaaacaagttgaaagtgaacccaagaatgttgcgcataatgtaatttgctacaaaaagtcatgtgaaaact
ttacgacagcgctctttgaaaattactataagtcatacttcggaatacaaattcaattatagtcttctgctgataaatgtttcagagaa
ctcagcagtatgttccctgaaaacttttttgttttattcagacatgaatcactgcacaatatgtcatgacagtaactgtgtaaccatga
aactcggaatttgcataccaaactcaattgagaccatcaaccaggattatcataagtattttcatttttcaaccggtgaaacccgcctc
ggacgcattcgtatcttctcttagtttgtgtcaccggctaccgtaacatcccgcccattctgaagtttcgccgtgtccagttccctatt
taaaatgagcgcacgccttctagataaggacattctgcgtccgccactcttctttttttttaattcttaggagattttatttcaatttg
tgagaattttcaggttATGGGTGTCTTTCAAGACAAGGTAGCCAACATTCTTCGAATCGAAGATCA

>Cbrig aqp-8 -550,+50(aqp-8)
aaattggaattcaaccggagaacctgcatcgttaatttaattttcatatcatgtgaatacttctgcgctctttgaaaattacgaaaagt
catactagaagataactgagttatagtcaggtacgaatgtttcaggtcacacttccgactttacgatttttcagaattcatggtttgtg
tgtccagaatttttagtccctctgcgtagacagtgacacattgtcatgcatgattcacggttgaagttaaaacattgacaactctgaa
atttgcatagaattgaaaattcggtccggattatcatactaagcagtttttcaattttatcaaccgaaaccgacataaaactccccgcc
ttcaaatcaacgcattcgcatcatctcatcgaaatttttcttgtcacccgctaccgtaacttcccgcccattttgaccgtgtccggttg
cctatttaaaataagcgacacgccttctagataaggacattctgcgtccactgttgttttttcccaagaatttttgctttaattttac
taattatttcagataATGGGTGTCTTTCAAGACAAAATAGCCAGCATACTTCGAATTGAAGATCA

Appendix 5. *aqp-8* 5' promoter region truncation primers (left) and resulting strains.

Number following aqp-8 in the primer name designates the 5' end of the amplicon relative to the translational start site of *aqp-8.*

Primer Name	Sequence (5' - 3')
aqp-8-711	atgacctgtcggtgtgtgaa
aqp-8-567	agatgttgcgcataattgaa
aqp-8-354	gaaaatgtaacttgagcacccc
aqp-8-434	aaagtcatgtgaaaggtattcg
aqp-8-342	gaaaatgtaacttgagcacccc
aqp-8-315	tgtatatgattcactgttgaa
aqp-8-297	gaatgcgtaaccgcaaaatga
aqp-8-279	tgattgccaaaatttgcatac
aqp-8-272	caaaatttgcatactggaatt
aqp-8-267	tttgcatactggaattgcgag
aqp-8-243	atcaatctggattatgataag
aqp-8-228	gataagtagttcccaaaccca
aqp-8-207	aaaactttcacccgcctct
aqp-8-80	agcgcacgccttctagataa.

Strain	Genotype
BC06835	*dpy-5(e907);sEx1241 rCes[K02G10.7(-711)::GFP-PEST+pCeh361](SegI)*
BC06836	*dpy-5(e907);sEx1242 rCes[K02G10.7(-711)::GFP-PEST+pCeh361](SegII)*
BC06837	*dpy-5(e907);sEx1243 rCes[K02G10.7(-567)::GFP-PEST+pCeh361](SegI)*
BC06838	*dpy-5(e907);sEx1244 rCes[K02G10.7(-567)::GFP-PEST+pCeh361](SegII)*
BC07228	*dpy-5(e907);sEx1525 rCes[K02G10.7(-434)::GFP-PEST+pCeh361]*
BC07226	*dpy-5(e907);sEx1523 rCes[K02G10.7(-354)::GFP-PEST+pCeh361]*
BC06839	*dpy-5(e907);sEx1245 rCes[K02G10.7(-342)::GFP-PEST+pCeh361]*
BC07227	*dpy-5(e907);sEx1524 rCes[K02G10.7(-315)::GFP-PEST+pCeh361]*
BC06876	*dpy-5(e907);sEx1278 rCes[K02G10.7(-297)::GFP+pCeh361]*
BC06877	*dpy-5(e907);sEx1279 rCes[K02G10.7(-279)::GFP+pCeh361](SegI)*
BC06878	*dpy-5(e907);sEx1280 rCes[K02G10.7(-279)::GFP+pCeh361](SegII)*
BC06921	*dpy-5(e907);sEx1314 rCes[K02G10.7(-272)::GFP+pCeh361](SegI)*
BC06922	*dpy-5(e907);sEx1315 rCes[K02G10.7(-272)::GFP+pCeh361](SegII)*
BC06916	*dpy-5(e907);sEx1312 rCes[K02G10.7(-267)::GFP+pCeh361]*
BC06879	*dpy-5(e907);sEx1281 rCes[K02G10.7(-261)::GFP+pCeh361](SegI)*
BC06887	*dpy-5(e907);sEx1289 rCes[K02G10.7(-261)::GFP+pCeh36](SegII)*
BC07229	*dpy-5(e907);sEx1526 rCes[K02G10.7(-228)::GFP-PEST+pCeh361]*
BC06916	*dpy-5(e907);sEx1312 rCes[K02G10.7(-207)]::GFP+pCeh361](SegI)*
BC06840	*dpy-5(e907);sEx1246 rCes[K02G10.7(-207)::GFP-PEST+pCeh361](SegII)*
BC06841	*dpy-5(e907);sEx1247 rCes[K02G10.7(-207)::GFP-PEST+pCeh361](SegIII)*
BC06842	*dpy-5(e907);sEx1248 rCes[K02G10.7(-80)::GFP-PEST+pCeh361](SegI)*
BC06843	*dpy-5(e907);sEx1249 rCes[K02G10.7(-80)::GFP-PEST+pCeh361](SegII)*

Appendix 6. Mutagenized octamer strains and oligos used for construction.

The underlined residue(s) in the sequence column designates the changed base(s).

Primer Name	Sequence (5' - 3')
MutOctG→A	TTGCCAAAATTTACATACTGGAAT
MutOctG→T	TTGCCAAAATTTTCATACTGGAAT
MutOctG→C	TTGCCAAAATTTCCATACTGGAAT
MutOctC→G	TTGCCAAAATTTGGATACTGGAAT
MutOctC→A	TTGCCAAAATTTGAATACTGGAAT
MutOctC→T	TTGCCAAAATTTGTATACTGGAAT
MutOctGC→AG	TTGCCAAAATTTAGATACTGGAAT

Strain	Genotype
BC07485	*dpy-5(e907);sEx1583 rCes[aqp-8 oct mut G-T::GFP+pCeh361]*
BC07501	*dpy-5(e907);sEx1593 rCes[aqp-8 oct mut G-T::GFP+pCeh361]*
BC07312	*dpy-5(e907);sEX1574rCes[aqp-8 oct mut C-G::GFP+pCeh361]*
BC07486	*dpy-5(e907);sEx1584 rCes[aqp-8 oct mut G-C::GFP+pCeh361]*
BC07487	*dpy-5(e907);sEx1585 rCes[aqp-8 oct mut G-C::GFP+pCeh361]*
BC07488	*dpy-5(e907);sEx1586 rCes[aqp-8 oct mut G-C::GFP+pCeh361]*
BC07489	*dpy-5(e907);sEx1587 rCes[aqp-8 oct mut C-A::GFP+pCeh361]*
BC07552	*dpy-5(e907);sEX1599 rCes[aqp-8 oct mut C-T::GFP+pCeh361]*
BC07283	*dpy-5(e907);sEX1556 rCes[aqp-8 oct mut G-A::GFP+pCeh361]*
BC07284	*dpy-5(e907);sEX1557 rCes[aqp-8 oct mut G-A::GFP+pCeh361]*
BC07285	*dpy-5(e907);sEX1558 rCes[aqp-8 oct mut GC-AG::GFP+pCeh361]*
BC07286	*dpy-5(e907);sEX1559 rCes[aqp-8 oct mut GC-AG::GFP+pCeh361]*
BC07287	*dpy-5(e907);sEX1560 rCes[aqp-8 oct mut GC-AG::GFP+pCeh361]*
BC07288	*dpy-5(e907);sEX1561 rCes[aqp-8 oct mut GC-AG::GFP+pCeh361]*

Appendix 7. Left oligos used for 5' synthetic addition of octamer sequence to various lengths of *vit-2* promoter regions

For genotype information refer to the corresponding BC number in Baillie Laboratory strain log). In general, the genotype conforms to the the model: *dpy-5(e907);sEx(####) rCes[vit2oct(-distance upstream of the translational start site of vit-2)]::GFP+pCeh361]*.

Oligo name/position	Strain	Oligo Sequence (5' - 3')
vit2oct258	BC8199	AATTTGCATAAATTTGCATAAATTTGCATAGATCAAACTG TATTACTTGAAAC
vit2oct299	BC8200	AATTTGCATAAATTTGCATAAATTTGCATACACCTCATCG TTAAAAAGTCATG
vit2oct347	BC8201	AATTTGCATAAATTTGCATAAATTTGCATACTTTGTTTTA AATACCATGTG
vit2oct392	BC8235	AATTTGCATAAATTTGCATAAATTTGCATACTATCCTGTC GGTCACAATGC
vit2oct448	BC8015	AATTTGCATAAATTTGCATAAATTTGCATAGGTTGCGTTT TAGGTGCCTAC
vit2oct497	BC8202	AATTTGCATAAATTTGCATAAATTTGCATAGGATTCAGTT TTAGTTTTTGTG
vit2oct556	BC8240	AATTTGCATAAATTTGCATAAATTTGCATAGGTTTTTAAT GTTTAGTTGTT
vit2oct600	BC8203,BC8204	AATTTGCATAAATTTGCATAAATTTGCATAGTTAGAATAA TGTCAAAACATC
vit2oct652	BC8241	AATTTGCATAAATTTGCATAAATTTGCATAGCCGCGGATT TCCAACTAATCA
vit2oct700	BC8263	AATTTGCATAAATTTGCATAAATTTGCATACAACCGAAT AATACGCAGAAC

Appendix 8. Left oligos used for 5' truncations of promoter regions that contain non-interspecies conserved octamer elements.

For genotype information refer to the corresponding BC number in Baillie Laboratory strain log). In general, the genotype conforms to the the model: *dpy-5(e907);sEx(####) rCes[gene name(-distance upstream of the translational start site of gene)]::GFP+pCeh361]*.

GENE	5' PCR PRIMER	PRIMER SEQUENCE	STRAIN
C01B12.3	C01B12.3A*	TTTCTATTGGCTCAATTTCACGTA	BC12593
	C01B12.3A4	TTTGTGGGTATTTTGCTTACGTT	Lost strain
	C01B12.3A3	CCCATATTGTGTCATCCGATA	BC6845
	C01B12.3A2	AACCATTTCCCTCATTTCCA	BC6883
	C01B12.3A1	TGCTTAACAAATTTACACGAGTTT	BC6882
ZC395.10	ZC395.10A*	TGGGAAATGAAAATACAATCACAG	BC10796
	ZC395.10A'	GCTTTATTTTCGTCAATTTGCAT	BC8208,BC8209
	ZC395.10A''	GCTAATTTCAACTGTGCGCTTAT	BC8210
C02B8.4	C02B8.4A*	AAAGCTTTCAAAATCCCAAACA	BC12188
	C02B8.4A'	CAAACCAATTTGCATAATCTTGA	BC8213
	C02B8.4A''	CATAATCTTGATTTTGTATCATG	BC8214,BC8215
R10H1.2	R10H1.2A*	CGCGACCTCAGTTTTTGAG	BC14834
	R10H1.2A'	CACTCCTCCGGAAGTCTCCATT	BC7969
	R10H1.2A''	GCATAATCCACTCTTCGCACGT	BC7970,BC7971
Y69E1A.7	Y69E1A.7	AATTGCAGATTCCAATTTTCG	BC20026
	Y69E1A.7A4	CGAGTTTGCGGTTTCACATT	BC7974,BC7975
	Y69E1A.7A3	AAAAATGCAAATCAATGCAGTA	BC7945
	Y69E1A.7A2	AGTACTCGTGCACCGTTTCC	BC7946,BC7947
	Y69E1A.7A1	GAACGTGGTCCAGGCTAATC	BC7948
F58B4.1	F58B4.1A*	TCCTAAACGAAGAGCATGAGTTC	BC13573
	F58B4.1A4	TTTGGAGCATTGCAAAGGTAT	BC6833
	F58B4.1A3	GCCTGAAGAAGTTCAAGTCTGC	BC6834
	F58B4.1A2	AACCACGGTTTGAATTGTGA	BC6844
	F58B4.1A1	TGCACGTGCTTATAAAACAACA	Lost strain
C45G9.5	C45G9.5A*	CATGCGGAGACAATTCAAAAT	BC12540
	C45G9.5A'	CTCTCATGGCCGAATTTGCATAC	BC7972
	C45G9.5A''	CATTTTTACTCAATTCCGTTCGG	BC7973
ZK470.5	ZK470.5A*	CCATTTCAAAAACTCCATCACC	BC10105
	ZK470.5A5	CATTCCAATGCATTATTTCGG	BC8001
	ZK470.5A4	TAAAACTTTCCCCATTTGCG	BC7993,BC7994
	ZK470.5A3	GCCCGCGTTATGCTACTTAT	BC7999,BC8000
	ZK470.5A2	GTCAGTGCGCCTTCTTTTTC	BC7992
	ZK470.5A1	TTTCCCACTTGTTCTCCTCG	BC7998,BC8053
R13A5.10	R13A5.10A*	CCATTTGATGATTCTTCACCATT	BC12561
	R13A5.10A'	GAATTATAAATTCAGAAATTTG	BC8192,BC8193
	R13A5.10A''	GAATGACGCACATTTTAAAAGT	BC7302,BC8194
F36H1.2	F36H1.2A*	TCTCGCTTCTCTTTATCACCTTG	BC11724
	F36H1.2A4	AAAATTCAACGTGGTTTCGG	BC6852
	F36H1.2A3	AATACCGCAGATTAAAGGCG	BC6853,BC6854
Y48A6B.8	Y48A6B.8A*	TCGATCGCATTAAGTCTCTTGA	BC15694
	Y48A6B.8A'	GGTTTTGGCAGTAAATTTGCAT	BC8195,BC8196
	Y48A6B.8A''	CATAAATATGTATGCGAATTAT	BC8242
F29F11.6	F29F11.6A*	GGACAGCTGATTCAATGTTCTTC	BC14231
	F29F11.6A''	CAAATTCAAAATTTGCATATCG	BC8216
	F29F11.6A'	CATATCGATTGTTAAAAACCAT	BC8217

GENE	5' PCR PRIMER	PRIMER SEQUENCE	STRAIN
B0334.4	B0334.4A*	CGAGACCGGATACAGTAACTTTG	BC10392
	B0334.4A'	GCGTACGCGAGAAATTTGCATTC	BC8249,BC8250
	B0334.4A''	GCATTCATTATTTTCTTTTCCAC	BC8251,BC8252
C02B8.6	C02B8.6A*	CACTCATTTCAAGCCTTCCAG	BC20206
	C02B8.6A'	CAAACCAATTTGCATAATCTTGA	BC8253
	C02B8.6A''	CATAATCTTGATTTTGTATCATG	BC8254
H23N18.3	H23N18.3A*	AGGTTCCAGAGATCAATAAAGTCG	BC15565
	H23N18.3A'	GTTTTTAATATAAATATTGCGTGT	BC8255,BC8256
	H23N18.3A''	GCATGCTACTTTCAGTTTTGCGAA	BC8257
R13F6.3	R13F6.3A*	TGGCAAACGTCTTCCTTTATTT	BC14593
	R13F6.3A'	CTGTATTTTATTTGCATTGTATTA	BC8260,BC8261
	R13F6.3A''	GCATTGTATTATCATTTTTAATGA	BC8262
Y53G8AR.3	Y53G8AR.3A*	GGAAAATTTGCTGAAAATCTACTT	BC14906
	Y53G8AR.3A5	TTTCGAGTAAACCAGCCGAT	BC7991
	Y53G8AR.3A4	GAAATCCCCCGTAATTTTCC	BC8052
	Y53G8AR.3A3	CGAGTTTTTAAATTTTCGGCA	BC8004
	Y53G8AR.3A2	ATTCAACTGATTTTCACAATTTTT	BC7990
	Y53G8AR.3A1	AATTTTGCAATTTCCCGGAT	BC7989

Appendix 9. 5' primers used for generation of *promoter::GFP* constructs for promoter regions that contain conserved octamer elements.

For genotype information refer to the corresponding BC number in Baillie Laboratory strain log). In general, the genotype conforms to the the model: *dpy-5(e907);sEx(BC number) rCes[gene name::GFP+pCeh361]*.

GENE	5' PRIMER SEQUENCE	STRAIN
C05D12.1	ACCCAAGCAAAACAAAGCTC	BC17548
C17G1.5	CTGTTAAATCATTGACGTTGTTT	BC17536
C07E3.10	TGGTGGGTGTGGAGGTCTAT	BC17549,BC17550
C26B9.5	TCTGGTTCACCATTGAATTTGT	BC17537,BC17538
C43G2.5	AAGGCAACGCATCCAATAAG	BC17633
C50F4.9	GCGACGACAATACTTGCATA	BC17546,BC17547
C54D10.1	CAACCAAAAATAAATTCCTGACG	BC17642,BC17643
E04F6.4	TTGTGACTGTCCGTTCCTGA	BC17568
F01D5.6	ATAGCAGAAGCAGCCGACAT	BC17569
F13B6.1	TCGATTTGAGAACGAAATTGG	BC17570
F16F9.1	GTTTTTAATATTCCTTGAACCCC	BC17571,BC17572
F18C5.5	AAAAATTCAACGTTTTCTGAGAT	BC17555
F18C5.9	CGATGTCACGCGGGTATATT	BC17556
F18G5.3	AGTGATGTGCACCACCGTAA	BC17599
F22F7.7	TAAACGGGTGCGGTCTTATC	BC17675,BC17676
F28F9.2	TTGTAAACTGTTTCTGAGATTGC	BC17620,BC17621
F29B9.8	AACACGTGACAAATCCCACC	BC17573
F36F2.7	CTTCCCTCTCGTTGCTTTTT	BC17574
F36H12.1	TGCTAGGTCATTGCAAAACAA	BC17575,BC17576
F41E6.14	TTTTTGCAGGTGTCCTTTTA	BC17622
F43B10.1	TGTATCAAAACTGCAAAGCCA	BC17577,BC17578
F44F4.3	AAGCGGTAGAGTAGGGCCAG	BC17579
F49H12.3	CAGCCATTTTGCACTTATTCA	BC17600
F55F3.4	TGCACTCTTAAAAAGTTCGTTG	BC17580,BC17581
F56A4.10	AAGGATGGAGAGCGATTCCT	BC17601
H22K11.3	AAATGTTCCGAATTGGTGGA	BC17602
K06A1.3	TTTGCAAGCATTTTTGGTCA	BC17623,BC17624
K08F4.4	TCTTATCGGCAAATCCTTCG	BC17644
K10C2.4	TCAATAGTTCATCGCTGTTGGT	BC17645
K10C8.2	TTAGTCTTGTATTGGCAGACGG	BC17603
M03A8.3	GTACATTGGCGGTTTCGTTC	BC17634
M176.5	TTCGTTATTCCAATGACCCAG	BC17635
R12G8.2	TTTCCCGAAAGTTCGATTAAAA	BC17604,BC17605
T02C5.3	GCACCAGAAATAAATAGTTGGAA	BC17606
T10B5.4	GGGTCAAAATGGTTTAAAATCG	BC17618
T11F9.9	CTGGTACCACCAAAACCAGG	BC17636,BC17637
T16H12.9	TTCAAGCAGGGAAATTGAAA	BC17619
Y105C5B.15	ATGTATTTCGCGTTTCCGTC	BC17677,BC17678
Y19D10A.4	GGAATGTTGCTGAACGATTG	BC17625
Y19D10A.5	TCGAGATCCCGTAAATCGAC	BC17626
Y67A6A.2	TCGATAGCAAGCCTTACCAA	BC17627
ZC101.1	CATTAGAGCCCCCTCTCACA	BC17628
C07E3.2	CCTTGTCTATTCATTCCGTTAGAT	BC17672
C01B12.5	TGGCCTGAACATGAAAATCA	BC20002
C15B12.7a	TTCGGTTTGAAGTGCCTAAAAT	BC14040
C16C10.4	ATGCCATATTGGTAGCTGTGG	BC14128
C26F1.10	TCTAGTTGTTCATCGCTTCTTTTT	BC12205
C38H2.1	GAACTTTTCCATCTGTTTGTTCG	BC13836
C45G9.5	CATGCGGAGACAATTCAAAAT	BC12539
C54D10.10	TGCATTTTTACTCGTCTCGCT	BC15319

GENE	5' PRIMER SEQUENCE	STRAIN
F14H12.1	CCCTGAACCAGTAAAGACGG	BC16801
F39H11.3	GAAATGGCGGAAGCTACTGT	BC17886
F53C11.3	TGATCAGAAGATTCCAACGTTTT	BC14427
F53E2.1	TGAACGGACTCTGAACAATGA	BC10230
H24G06.1a	AGACATCTGCGTCTCCTTTCTAGT	BC15661
H43I07.3	ATCAAAATTGTTAAAGGTCACACG	BC15071
K02G10.7	GATGTTATCTGAATTGGATGCTC	BC20052
R02F2.8	TCAAAATTGAAACCACTTTTCAGA	BC17709
R08B4.2	ATAGAATTGCCATAAACCCCG	BC16630
R10H1.2	CGCGACCTCAGTTTTTGAG	BC14834
R13A5.1a	TTCTTGGCTTCTTCGATTGC	BC10182
T05H10.3	ACATCGTTCATTTGAGAAGTTTGA	BC14357
T10B5.5	GATCCTAGCTGCTTTGGTCCT	BC12510
T12A2.9	GCGCAATGGTACTTCGATATT	BC11603
T14G10.5	TTGAGATTTTGAGCGAAACAGA	BC13695
T19A6.2a	ATAGTTAAATGGGAGCGGTGTG	BC14682
W08D2.1	ATTTTGCTTTATCTCGTTTATCGC	BC17158
Y43F8C.12	ATCACTGTTGAGCCGTTTTC	BC10031
Y54G2A.25	ATTGTACTTCCGATTGATGCG	BC13847
Y7A9A.1	TCTGAGAATTCTTTTCACCTTTTG	BC11932
ZK512.9	CCCATGCATTGACATAGCAC	BC12881
B0334.11	ATTTCTGGCTGTGCTCGTTT	BC17788
C09B8.7	ATGCTTTTTCGGTCCAAGATT	BC17700
C16B8.4	TTTTGTTGGGAATCTCGCTC	BC17717
C45G9.7	TGAAACACAAAACAATTGAAAGC	BC17786,BC17787
D2096.8	AACTTGAGCATTAGGCACGC	BC8187
F09A5.4	CGCGTAAAAAGAATGAACCC	BC17701
F11D5.6	GGTTAGTGCACGACGTTTCA	BC17702
F14B8.7	TGCAACTTTGTACCCAGCAA	BC17704
F18A11.2	CAAGCCTTTCTTCGATGACC	BC17728
F28B3.6	TTTCACAAATCGACCGACAA	BC17707
F36H1.11	GCTCCAAAATGCAGATAAAACA	BC17770
F36H12.17	GCTCCAAAATGCAGATAAAACA	BC17729
F43C9.4	TCCACATACGTTCCCAAGTG	BC17730
F43E2.9	TCCACATACGTTCCCAAGTG	BC17771
F53G12.5	TCCACATACGTTCCCAAGTG	BC17773
K09C8.2	AAATATGTTCGGCCTCTCCTC	BC17731
K09C8.8	CAAAAATCACCAAGCTGATCC	BC17732
R07A4.2	CCCGTGGCGGAGAGTTAAA	BC17734
R10E12.1	AATCGAATCGCGTAACCAAC	BC17814
R12C12.9	AGAGTTGCGCAGAATTTTTGT	BC17753
T05A10.4	TGCCAAAATGTTCTATTTGTGTT	BC17774
T28H11.8	CAATAAAAGCTACCCCGCAA	BC17754
W02F12.4	CTTCCGAAAGCAGTCTCACC	BC17755
Y102A11A.7	GTTGAAGGATTGAAGGACGG	BC17775
Y47D7A.14	CCATGCAAAATCAGTGGAGA	BC17756
ZK1010.2	TCAGTTTTCCCCCTAATTCC	BC8205
ZK1010.3	CTATCGAATTCAGTGTGCGCC	BC17776
ZK1248.2	ACGGGGATGAAATGTTGAAA	BC17777
ZK652.11	AGTGCACAAGTCCGAAGTGA	BC17790
ZK688.1	GCAGACATTAGCAATACTTGGG	BC17757
C18C4.2	CCATCGTATTGCATGTATTGCT	BC16915,BC16916

Appendix 10. Left oligos used for 5' truncations of promoter regions that contain interspecies conserved octamer elements.

For genotype information refer to the corresponding BC number in Baillie Laboratory strain log). In general, the genotype conforms to the the model: dpy-5(e907);sEx(####) rCes[gene name(-distance upstream of the translational start site of gene)]::GFP+pCeh361].

5' PCR PRIMER	PRIMER SEQUENCE	STRAIN
F14B8.7 A'356	caacagtgcagaaaaaagtgtcga	BC8022
F14B8.7 A"282	cggcgtccaccacgctcactcacat	BC8069
F14B8.7 A"'248	cttgatcggtacgccgataaagaga	BC8070
T28H11.8.1 A'243	gtcacttcaaaaaatcaaaacgg	BC8020,BC8021
T28H11.8.1 A"168	ctactcactcgagactcactgactt	BC8019
T28H11.8.1 A"'141	GCAAATtcaaatttttagccagtg	BC8017,BC8018
F18A11.2 A'279	cttatgagaataaagtaccaaat	BC8071
F18A11.2 A"199	tttcgcgaaaaaaattcccatATG	BC8072
F18A11.2 A"'168	gctctaacgaagctacaaagccaca	BC8109
R12C12.9 A'193	cagtctcaattgtgtgtacataaa	BC8165
R12C12.9 A"112	cgtcccccattttgactgATTTGC	BC8110
R12C12.9 A"'82	ctttctgtctcttttttaactcatt	BC8111
Y19D10A.4 A'143	gttcaaatattcaaaaattccga	BC8073
Y19D10A.4 A"65	cggaagagctcacagaaccagcga	BC8074
Y19D10A.4 A"'33	cgggcctagaatttgatacaaaaa	BC8128
C45G9.5 A'865	gaaacgcacacatacaacatgagg	BC8075
C45G9.5 A"789	catcttcctactctcatggccga	BC8076,BC8077
C45G9.5 A"'758	acatttttactcaattccgttcgg	BC8078,BC8079
R10H1.2 A'201	catattttctgatcttttgaaaatg	BC8090
R10H1.2 A"123	ctagtcactcctccggaagtctc	BC8166
R10H1.2 A"'95	GCATaatccactcttcgcacgttt	BC7970,BC7971,BC8117
ZK1010.3 A'748	gattccctccttgtcttggat	BC8091
ZK1010.3 A"669	cccgtcagagtcttgacgaagAT	BC8092,BC8093
ZK1010.3 A"'644	GCATgactgaaattatataatttg	BC8094,BC8095
F36F2.7 A'327	gacgtactgctttgacgttctttt	BC8112,BC8113
F36F2.7 A"252	gggaaaaaatgaaataaaaaac	BC8114
F36F2.7 A"'225	CAAATccgatcagtgtggcaaacac	BC8115,BC8116
M176.5 A'163	tacacataattcaacactctttt	BC8171
M176.5 A"83	cagccgtcccgatagaagttt	BC8107,BC8108
M176.5 A"'67	GCATttctcctccttctttttctt	BC8178
C17G1.5 A'410	caatgcgttccttgtaatttctcgg	BC8124
C17G1.5 A"326	ggtttcggggagcggacaacATGC	BC8125
C17G1.5 A"'199	tcgtccgtccgcatcgcatacatt	BC8126,BC8127
R12G8.2 A'174	gtgtaaaaataaacttttcac	BC8129
R12G8.2 A"99	gttttagtctgaaaacccccacca	BC8160
R12G8.2 A"'70	GCATtttaaatcgtcgtccgttg	BC8161
M03A8.3 A'720	gagaccatcaaacacgaacggtca	BC8146,BC8147
M03A8.3 A"648	caaaaattggtgggaatctgagtg	BC8148,BC8149
M03A8.3 A"'616	GCAAATagattcgacgtggtagt	BC8150
T16H12.9 A'667	gagaaaaaaaggaaagtacatttt	BC8151
T16H12.9 A"598	ccaaacattaaaaagtttagaata	BC8152,BC8153
T16H12.9 A"'561	GCATttttgacaatttaaaatttg	BC8130
F55F3.4 A'668	gtcttttgaaaaaatgtggaaatt	BC8122
F55F3.4 A"578	gaagttgtccagtttcaaatcATT	BC8123
F55F3.4 A"'553	GCATaaaattcaatttggtatttc	BC8137
F16F9.1 A'890	catcacctctcaacacgcacacatt	BC8154,BC8155
F16F9.1 A"819	gacgaaaagacgcttgaaatATGC	BC8162,BC8163

5’ PCR PRIMER	PRIMER SEQUENCE	STRAIN
F16F9.1 A'"787	gtggggacagatgagagagagacc	BC8164
H22K11.3 A'735	gaacttcacttcaaaacacagaaa	BC8167
H22K11.3 A"662	cgtattgctcagcgtttgctcgc	BC8168,BC8169
H22K11.3 A'"633	GCATacgtaataaggaaacggcta	BC8170
K08F4.4 A'186	cacatataagtcagctttaaatt	BC8142,BC8143
K08F4.4 A"101	catttcaatgtcaactgcgtcagc	BC8186
K08F4.4 A'"72	GCAAATttcagacggcactagaat	BC8144,BC8145
C05D12.1 A'247	gtgatcaaatgtatatgtagttc	BC8138
C05D12.1 A"168	ctttctctctgcccgccgacaaaA	BC8139
C05D12.1 A'"141	GCATtgaccgattgcgacgattcc	BC8140,BC8141
K10C2.4 A'493	catgaaacagtatcattgccaatg	BC8175
K10C2.4 A"421	catcacgtgactgtttcgcaattc	BC8133,BC8134
K10C2.4 A'"392	GCATccaatgtagttttctctct	BC8135,BC8136
R02F2.8 A'1072	gtgatcaatgtttccatttttctc	BC8176,BC8177
R02F2.8 A"990	gagtcaaaccatcacattttATGC	BC8156,BC8157
R02F2.8 A'"962	cgcctcacctttttgggagttattg	BC8158,BC8159
F18G5.3A'450	gtcatcatttgtttccgtgctaaaa	BC8172,BC8173
F18G5.3A"371	cctctcccgtgtctactgtttgca	BC8206,BC8207
F18G5.3A'"343	GCATtttcgtatttcgcttgattt	BC8174
C05D12.1A'	AATCCTTCCCTTTCTCTCTGC	BC7964
C05D12.1A"	GACCGATTGCGACGATTC	BC7962,BC7963
C17G1.5A'	AGCGGACAACATGCAAATTC	BC7965,BC7966
C17G1.5A"	TCCGCATCGCATACATTACA	BC7997

Appendix 11. List of genes with upstream conserved octamer elements sorted according to the different condition sets described in Table 12.

Set A		Set E		Set S		Set SE	
Sequence Name	Gene Name	Sequence Name	Gene Name	Sequence Name	Gene Name	Sequence Name	Gene Name
B0334.11	*ooc-3*	*C01B12.5*		*C05D12.1*		*C05D12.1*	
C01B12.5		*C05D12.1*		*C07E3.10*		*C07E3.10*	
C05D12.1		*C07E3.10*		*C15B12.7*	*cdf-1*	*C15B12.7*	*cdf-1*
C07E3.10		*C09B8.7*	*pak-1*	*C16B8.4*		*C16B8.4*	
C07E3.2	*pro-2*	*C15B12.7*	*cdf-1*	*C43G2.5*	*try-3*	*C43G2.5*	*try-3*
C09B8.7	*pak-1*	*C16B8.4*		*C45G9.5*		*C45G9.5*	
C15B12.7	*cdf-1*	*C17G1.5*		*C54D10.1*	*cdr-2*	*F01D5.6*	
C16B8.4		*C26B9.5*		*D2096.8*		*F14B8.7*	
C16C10.4		*C26F1.10*	*flp-21*	*F01D5.6*		*F14H12.1*	*col-165*
C17G1.5		*C38H2.1*		*F14B8.7*		*F16F9.4*	
C26B9.5		*C43G2.5*	*try-3*	*F14H12.1*	*col-165*	*F18G5.3*	*gpa-12*
C26F1.10	*flp-21*	*C45G9.5*		*F16F9.4*		*F22F7.7*	
C38H2.1		*C50F4.9*		*F18C5.5*		*F29B9.8*	
C43G2.5	*try-3*	*E04F6.4*		*F18G5.3*	*gpa-12*	*F43B10.1*	
C45G9.5		*F01D5.6*		*F22F7.7*		*F43E2.9*	
C45G9.7		*F09A5.4*		*F29B9.8*		*F44F4.3*	
C50F4.9		*F11D5.6*		*F39H11.3*	*cdk-8*	*F53C11.3*	
C54D10.1	*cdr-2*	*F13B6.1*		*F43B10.1*		*F53E2.1*	*tag-304*
D2096.8		*F14B8.7*		*F43E2.9*		*F53G12.5*	*mex-3*
E04F6.4		*F14H12.1*	*col-165*	*F44F4.3*		*H22K11.3*	
F01D5.6		*F16F9.1*		*F53C11.3*		*H24G06.1*	
F09A5.4		*F16F9.4*		*F53E2.1*	*tag-304*	*K02G10.7*	*aqp-8*
F11D5.6		*F18A11.2*		*F53G12.5*	*mex-3*	*K06A1.3*	
F13B6.1		*F18G5.3*	*gpa-12*	*F57F5.5*	*pkc-1*	*K08F4.12*	
F14B8.7		*F22F7.7*		*H22K11.3*		*K09C8.8*	
F14H12.1	*col-165*	*F28F9.2*		*H24G06.1*		*K10C2.4*	
F16F9.1		*F29B9.8*		*K02G10.7*	*aqp-8*	*K10C8.2*	
F16F9.4		*F36H1.11*		*K06A1.3*		*M176.5*	
F18A11.2		*F36H12.17*		*K08F4.12*		*R07A4.2*	
F18C5.5		*F41E6.14*		*K09C8.8*		*R10H1.2*	*srab-14*
F18C5.9		*F43B10.1*		*K10C2.4*		*R12C12.9*	
F18G5.3	*gpa-12*	*F43C9.4*	*mig-13*	*K10C8.2*		*R12G8.2*	*twk-36*
F22F7.7		*F43E2.9*		*M176.5*		*T02C5.3*	*igcm-3*
F28B3.6		*F44F4.3*		*R07A4.2*		*T05A10.4*	
F28F9.2		*F49H12.3*		*R10E12.1*	*alx-1*	*T05H10.3*	
F29B9.8		*F53C11.3*		*R10H1.2*	*srab-14*	*T14G10.6*	*tsp-12*
F36F2.7		*F53E2.1*	*tag-304*	*R12C12.9*		*T19A6.2*	*ngp-1*
F36H1.11		*F53G12.5*	*mex-3*	*R12G8.2*	*twk-36*	*W08D2.1*	*egl-20*
F36H12.17		*F55F3.4*		*T02C5.3*	*igcm-3*	*Y102A11A.7*	
F39H11.3	*cdk-8*	*F56A4.10*		*T05A10.4*		*Y53C10A.4*	*rga-2*
F41E6.14		*H22K11.3*		*T05H10.3*		*Y54G2A.25*	*lad-2*
F43B10.1		*H24G06.1*		*T10B5.5*		*Y67A6A.2*	*nhr-62*

Set A		Set E		Set S		Set SE	
Sequence Name	**Gene Name**	**Sequence Name**	**Gene Name**	**Sequence Name**	**Gene Name**	**Sequence Name**	**Gene Name**
F43C9.4	*mig-13*	*H43I07.3*		*T12A2.9*	*srg-8*	*Y7A9A.1*	
F43E2.9		*K01B6.2*	*srx-45*	*T14G10.6*	*tsp-12*	*ZK863.6*	*dpy-30*
F44F4.3		*K02G10.7*	*aqp-8*	*T19A6.2*	*ngp-1*		
F49H12.3		*K06A1.3*		*W08D2.1*	*egl-20*		
F53C11.3		*K08F4.12*		*Y102A11A.7*			
F53E2.1	*tag-304*	*K09C8.8*		*Y51B9A.6*			
F53G12.5	*mex-3*	*K10C2.4*		*Y53C10A.4*	*rga-2*		
F55F3.4		*K10C8.2*		*Y54G2A.25*	*lad-2*		
F56A4.10		*M03A8.3*		*Y67A6A.2*	*nhr-62*		
F57F5.5	*pkc-1*	*M176.5*		*Y7A9A.1*			
H22K11.3		*R07A4.2*		*ZK1010.3*			
H24G06.1		*R08B4.2*	*alr-1*	*ZK863.6*	*dpy-30*		
H43I07.3		*R10H1.2*	*srab-14*				
K01B6.2	*srx-45*	*R12C12.9*					
K02G10.7	*aqp-8*	*R12G8.2*	*twk-36*				
K06A1.3		*R13A5.1*	*cup-5*				
K08F4.12		*T02C5.3*	*igcm-3*				
K08F4.4	*glt-3*	*T05A10.4*					
K09C8.2		*T05H10.3*					
K09C8.8		*T14G10.5*					
K10C2.4		*T14G10.6*	*tsp-12*				
K10C8.2		*T16H12.9*					
M03A8.3		*T19A6.2*	*ngp-1*				
M176.5		*T28H11.8*					
R02F2.8		*W02F12.4*					
R07A4.2		*W08D2.1*	*egl-20*				
R08B4.2	*alr-1*	*Y102A11A.7*					
R10E12.1	*alx-1*	*Y105C5B.15*					
R10H1.2	*srab-14*	*Y19D10A.4*					
R12C12.9		*Y19D10A.5*					
R12G8.2	*twk-36*	*Y43F8C.12*	*mrp-7*				
R13A5.1	*cup-5*	*Y47D7A.14*					
T02C5.3	*igcm-3*	*Y53C10A.4*	*rga-2*				
T05A10.4		*Y54G2A.25*	*lad-2*				
T05H10.3		*Y67A6A.2*	*nhr-62*				
T10B5.4		*Y7A9A.1*					
T10B5.5		*ZC101.1*					
T11F9.9	*col-157*	*ZK512.9*	*grl-11*				
T12A2.9	*srg-8*	*ZK652.11*	*cuc-1*				
T14G10.5		*ZK688.1*					
T14G10.6	*tsp-12*	*ZK863.6*	*dpy-30*				
T16H12.9							
T19A6.2	*ngp-1*						
T28H11.8							
W02F12.4							
W08D2.1	*egl-20*						

Set A		Set E		Set S		Set SE	
Sequence Name	**Gene Name**	**Sequence Name**	**Gene Name**	**Sequence Name**	**Gene Name**	**Sequence Name**	**Gene Name**
Y102A11A.7							
Y105C5B.15							
Y19D10A.4							
Y19D10A.5							
Y43F8C.12	*mrp-7*						
Y47D7A.14							
Y51B9A.6							
Y53C10A.4	*rga-2*						
Y54G2A.25	*lad-2*						
Y67A6A.2	*nhr-62*						
Y7A9A.1							
ZC101.1							
ZK1010.2							
ZK1010.3							
ZK1248.2	*col-74*						
ZK512.9	*grl-11*						
ZK652.11	*cuc-1*						
ZK688.1							
ZK863.6	*dpy-30*						

Appendix 12. An updated set of gene-upstream regions (within 1,000bp of the translational start site) that contain interspecies conserved octamer elements between *C. elegans*, *C. briggsae*, and *C. remanei*

The sequences from all three nematode genomes were isolated from Wormbase WS195. The top list includes all genes found using these latest versions of the database.

The bottom list shows the genes that were removed from the old list due to changes in the new databases.

Genes with conserved upstream octamer elements

B0244.6	*F49H12.3*	*C45G9.7*	*T05H10.3*
B0334.11	*F53C11.2*	*C50F4.9*	*T10B5.4*
B0334.4	*F53C11.3*	*C53D6.11*	*T10B5.5*
B0496.1	*F53E2.1*	*C54D10.1*	*T11F9.9*
B0496.7	*F53G12.5*	*E04D5.3*	*T12A2.9*
B0511.6	*F55C10.3*	*E04F6.4*	*T14G10.5*
C01B12.5	*H22K11.3*	*F01D5.6*	*T14G10.6*
C01B4.7	*H24G06.1*	*F02D10.1*	*T16H12.9*
C01B4.8	*H43I07.3*	*F09A5.4*	*T19A6.2*
C01G6.8	*K01B6.2*	*F10F2.1*	*T20B5.1*
C03A7.14	*K02G10.7*	*F11F1.2*	*T22B7.1*
C03C10.3	*K03D10.1*	*F11F1.4*	*T22F7.1*
C03G5.9	*K03H1.2*	*F11F1.5*	*T26H2.9*
C05A2.1	*K04B12.2*	*F11F1.8*	*T28H11.8*
C05D12.1	*K06A1.3*	*F11H8.4*	*W02F12.4*
C06E7.1	*K07C11.4*	*F13E9.4*	*W03F9.11*
C07A4.1	*K08F4.12*	*F14B8.7*	*W04G3.8*
C07E3.10	*K08F4.4*	*F14H12.1*	*W08D2.1*
C07E3.2	*K09C8.8*	*F16F9.1*	*Y102A11A.7*
C09B8.7	*K09F5.6*	*F16F9.4*	*Y105C5B.15*
C09F12.2	*K10C2.4*	*F17C11.13*	*Y19D10A.4*
C10G8.8	*K10C8.2*	*F18C5.5*	*Y19D10A.5*
C14H10.4	*K12G11.2*	*F18C5.9*	*Y39B6A.2*
C15B12.7	*M03A8.3*	*F18G5.3*	*Y39B6A.27*
C16B8.4	*M176.5*	*F22H10.2*	*Y39B6A.6*
C16C10.4	*M88.1*	*F28B3.6*	*Y51A2D.15*
C17G1.5	*R02F2.8*	*F28F9.2*	*Y51B9A.6*
C18C4.1	*R03D7.1*	*F29F11.6*	*Y53C10A.4*
C18C4.2	*R03H4.6*	*F31F7.2*	*Y53C12A.4*
C25A11.4	*R07A4.2*	*F36A4.8*	*Y54G2A.25*
C25B8.4	*R08B4.2*	*F36F2.7*	*Y67A6A.2*
C26B9.5	*R107.1*	*F36H1.11*	*Y7A9A.1*
C26F1.10	*R10E12.1*	*F36H12.17*	*ZC101.1*

C32H11.6	*R10H1.2*	*F38H12.5*	*ZC250.4*
C35A5.11	*R12C12.9*	*F39H11.3*	*ZK1010.2*
C36B7.5	*R12E2.6*	*F41E6.14*	*ZK1010.3*
C41G11.3	*R12G8.2*	*F42F12.10*	*ZK1248.13*
C43G2.5	*T02C5.3*	*F42F12.6*	*ZK1248.2*
C45B2.7	*T04C12.5*	*F43B10.1*	*ZK512.9*
C45G9.5	*T05A10.4*	*F43C9.4*	*ZK652.11*
F44F4.3	*ZK829.9*	*F43D2.4*	*ZK688.1*
	ZK863.6		

Genes that are removed from the gene set in the previous analysis due to updates	
C38H2.1	octamer element no longer upstream in *C. briggsae*
D2096.8	No ortholog in *C. remanei* in WS195
F11D5.6	distance of element and gene increased in *C. briggsae* to > 1kb
F13B6.1	octamer element no longer upstream in *C. briggsae*
F18A11.2	octamer element no longer upstream in *C. briggsae*
F22F7.7	octamer element not found in *C. remanei* ortholog
F29B9.8	octamer element >1kb upstream in *C. elegans*
F43E2.9	octamer element not found in *C. remanei* ortholog
F55F3.4	octamer element not found in *C. remanei* ortholog
F56A4.10	no ortholog in *C. briggsae* WS195
F57F5.5	octamer element not found in *C. briggsae* ortholog
K09C8.2	octamer element not found in *C. remanei* ortholog
R13A5.1	octamer element not found in *C. remanei* ortholog
Y43F8C.12	octamer element not found in *C. remanei* ortholog
Y47D7A.14	octamer element not found in *C. remanei* ortholog

REFERENCE LIST

Genome sequence of the nematode C. elegans: a platform for investigating biology. 1998. *Science* 282 (5396):2012-8.

Agre, P., M. Bonhivers, and M. J. Borgnia. 1998. The aquaporins, blueprints for cellular plumbing systems. *J Biol Chem* 273 (24):14659-62.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* 215 (3):403-10.

Alvarez-Bolado, G., M. G. Rosenfeld, and L. W. Swanson. 1995. Model of forebrain regionalization based on spatiotemporal patterns of POU-III homeobox gene expression, birthdates, and morphological features. *J Comp Neurol* 355 (2):237-95.

Assa-Munt, N., R. J. Mortishire-Smith, R. Aurora, W. Herr, and P. E. Wright. 1993. The solution structure of the Oct-1 POU-specific domain reveals a striking similarity to the bacteriophage lambda repressor DNA-binding domain. *Cell* 73 (1):193-205.

Aurora, R., and W. Herr. 1992. Segments of the POU domain influence one another's DNA-binding specificity. *Mol Cell Biol* 12 (2):455-67.

Banerjee-Basu, S., D. W. Sink, and A. D. Baxevanis. 2001. The Homeodomain Resource: sequences, structures, DNA binding sites and genomic information. *Nucleic Acids Res* 29 (1):291-3.

Bao, Z., Z. Zhao, T. J. Boyle, J. I. Murray, and R. H. Waterston. 2008. Control of cell cycle timing during C. elegans embryogenesis. *Dev Biol* 318 (1):65-72.

Baranek, C., E. Sock, and M. Wegner. 2005. The POU protein Oct-6 is a nucleocytoplasmic shuttling protein. *Nucleic Acids Res* 33 (19):6277-86.

Bendall, A. J., R. A. Sturm, P. A. Danoy, and P. L. Molloy. 1993. Broad binding-site specificity and affinity properties of octamer 1 and brain octamer-binding proteins. *Eur J Biochem* 217 (3):799-811.

Bennett, C. L., A. J. Shirk, H. M. Huynh, V. A. Street, E. Nelis, L. Van Maldergem, P. De Jonghe, A. Jordanova, V. Guergueltcheva, I. Tournev, P. Van Den Bergh, P. Seeman, R. Mazanec, T. Prochazka, I. Kremensky, J. Haberlova, M. D. Weiss, V. Timmerman, T. D. Bird, and P. F. Chance. 2004. SIMPLE mutation in demyelinating neuropathy and distribution in sciatic nerve. *Ann Neurol* 55 (5):713-20.

Bermingham, J. R., Jr., S. Shumas, T. Whisenhunt, E. E. Sirkowski, S. O'Connell, S. S. Scherer, and M. G. Rosenfeld. 2002. Identification of genes that are downregulated in the absence of the POU domain

transcription factor pou3f1 (Oct-6, Tst-1, SCIP) in sciatic nerve. *J Neurosci* 22 (23):10217-31.

Bernstein, E., A. A. Caudy, S. M. Hammond, and G. J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409 (6818):363-6.

Bill, R. M., K. Hedfalk, S. Karlgren, J. G. Mullins, J. Rydstrom, and S. Hohmann. 2001. Analysis of the pore of the unusual major intrinsic protein channel, yeast Fps1p. *J Biol Chem* 276 (39):36543-9.

Bird, A.F., and J Bird. 1991. The structure of nematodes. *2nd*

edn. 21 (13):1653-74.

Bohmann, D., W. Keller, T. Dale, H. R. Scholer, G. Tebb, and I. W. Mattaj. 1987. A transcription factor which binds to the enhancers of SV40, immunoglobulin heavy chain and U2 snRNA genes. *Nature* 325 (6101):268-72.

Bok, S. W. 1982. Hyaluronidase, from wound healing to cancer (II). *Med Hypotheses* 8 (5):455-9.

Boll, M., M. Foltz, I. Rubio-Aliaga, G. Kottra, and H. Daniel. 2002. Functional characterization of two novel mammalian electrogenic proton-dependent amino acid cotransporters. *J Biol Chem* 277 (25):22966-73.

Borgnia, M. J., and P. Agre. 2001. Reconstitution and functional comparison of purified GlpF and AqpZ, the glycerol and water channels from Escherichia coli. *Proc Natl Acad Sci U S A* 98 (5):2888-93.

Borgnia, M. J., D. Kozono, G. Calamita, P. C. Maloney, and P. Agre. 1999. Functional reconstitution and characterization of AqpZ, the E. coli water channel protein. *J Mol Biol* 291 (5):1169-79.

Borgnia, M., S. Nielsen, A. Engel, and P. Agre. 1999. Cellular and molecular biology of the aquaporin water channels. *Annu Rev Biochem* 68:425-58.

Botfield, M. C., A. Jancso, and M. A. Weiss. 1992. Biochemical characterization of the Oct-2 POU domain with implications for bipartite DNA recognition. *Biochemistry* 31 (25):5841-8.

Brabletz, T., I. Pfeuffer, E. Schorr, F. Siebelt, T. Wirth, and E. Serfling. 1993. Transforming growth factor beta and cyclosporin A inhibit the inducible activity of the interleukin-2 gene in T cells through a noncanonical octamer-binding site. *Mol Cell Biol* 13 (2):1155-62.

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

Brown, C. T., A. G. Rust, P. J. Clarke, Z. Pan, M. J. Schilstra, T. De Buysscher, G. Griffin, B. J. Wold, R. A. Cameron, E. H. Davidson, and H. Bolouri. 2002. New computational approaches for analysis of cis-regulatory networks. *Dev Biol* 246 (1):86-102.

Brown, C. T., Y. Xie, E. H. Davidson, and R. A. Cameron. 2005. Paircomp, FamilyRelationsII and Cartwheel: tools for interspecific sequence comparison. *BMC Bioinformatics* 6:70.

Buechner, M. 2002. Tubes and the single C. elegans excretory cell. *Trends Cell Biol* 12 (10):479-84.

Buechner, M., D. H. Hall, H. Bhatt, and E. M. Hedgecock. 1999. Cystic canal mutants in Caenorhabditis elegans are defective in the apical membrane domain of the renal (excretory) cell. *Dev Biol* 214 (1):227-41.

Burglin, T. R., and G. Ruvkun. 2001. Regulation of ectodermal and excretory function by the C. elegans POU homeobox gene ceh-6. *Development* 128 (5):779-90.

Cederberg, A., M. Hulander, P. Carlsson, and S. Enerback. 1999. The kidney-expressed winged helix transcription factor FREAC-4 is regulated by Ets-1. A possible role in kidney development. *J Biol Chem* 274 (1):165-9.

Chalfie, M., and M. Au. 1989. Genetic control of differentiation of the Caenorhabditis elegans touch receptor neurons. *Science* 243 (4894 Pt 1):1027-33.

Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward, and D. C. Prasher. 1994. Green fluorescent protein as a marker for gene expression. *Science* 263 (5148):802-5.

Chavez, M., C. Landry, S. Loret, M. Muller, J. Figueroa, B. Peers, F. Rentier-Delrue, G. G. Rousseau, M. Krauskopf, and J. A. Martial. 1999. APH-1, a POU homeobox gene expressed in the salt gland of the crustacean Artemia franciscana. *Mech Dev* 87 (1-2):207-12.

Chen, R. P., H. A. Ingraham, M. N. Treacy, V. R. Albert, L. Wilson, and M. G. Rosenfeld. 1990. Autoregulation of pit-1 gene expression mediated by two cis-active promoter elements. *Nature* 346 (6284):583-6.

Consortium, C. elegans Genome Sequencing. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* 282 (5396):2012-8.

Cutter, A. D. 2008. Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. *Mol Biol Evol* 25 (4):778-86.

Danilition, S. L., R. M. Frederickson, C. Y. Taylor, and N. G. Miyamoto. 1991. Transcription factor binding and spacing constraints in the human beta-actin proximal promoter. *Nucleic Acids Res* 19 (24):6913-22.

Davey, K. G., and S. P. Kan. 1968. Molting in a parasitic nematode, Phocanema decipiens. IV. Ecdysis and its control. *Can J Zool* 46 (5):893-8.

Deen, P. M., and C. H. van Os. 1998. Epithelial aquaporins. *Curr Opin Cell Biol* 10 (4):435-42.

Deen, P. M., M. A. Verdijk, N. V. Knoers, B. Wieringa, L. A. Monnens, C. H. van Os, and B. A. van Oost. 1994. Requirement of human renal water channel aquaporin-2 for vasopressin-dependent concentration of urine. *Science* 264 (5155):92-5.

Denslow, S. A., and P. A. Wade. 2007. The human Mi-2/NuRD complex and gene regulation. *Oncogene* 26 (37):5433-8.

Dudziak, K., N. Mottalebi, S. Senkel, E. L. Edghill, S. Rosengarten, M. Roose, C. Bingham, S. Ellard, and G. U. Ryffel. 2008. Transcription factor HNF1beta and novel partners affect nephrogenesis. *Kidney Int* 74 (2):210-7.

Duprat, F., F. Lesage, M. Fink, R. Reyes, C. Heurteaux, and M. Lazdunski. 1997. TASK, a human background K+ channel to sense external pH variations near physiological pH. *Embo J* 16 (17):5464-71.

Dupuy, D., N. Bertin, C. A. Hidalgo, K. Venkatesan, D. Tu, D. Lee, J. Rosenberg, N. Svrzikapa, A. Blanc, A. Carnec, A. R. Carvunis, R. Pulak, J. Shingles, J. Reece-Hoyes, R. Hunt-Newbury, R. Viveiros, W. A. Mohler, M. Tasan, F. P. Roth, C. Le Peuch, I. A. Hope, R. Johnsen, D. G. Moerman, A. L. Barabasi, D. Baillie, and M. Vidal. 2007. Genome-scale analysis of in vivo spatiotemporal promoter activity in Caenorhabditis elegans. *Nat Biotechnol* 25 (6):663-8.

Duverger, O., and M. I. Morasso. 2008. Role of homeobox genes in the patterning, specification, and differentiation of ectodermal appendages in mammals. *J Cell Physiol* 216 (2):337-46.

Echevarria, M., and A. A. Ilundain. 1998. Aquaporins. *J Physiol Biochem* 54 (2):107-18.

Efimenko, E., K. Bubb, H. Y. Mak, T. Holzman, M. R. Leroux, G. Ruvkun, J. H. Thomas, and P. Swoboda. 2005. Analysis of xbx genes in C. elegans. *Development* 132 (8):1923-34.

Endeward, V., R. Musa-Aziz, G. J. Cooper, L. M. Chen, M. F. Pelletier, L. V. Virkki, C. T. Supuran, L. S. King, W. F. Boron, and G. Gros. 2006. Evidence that aquaporin 1 is a major pathway for CO2 transport across the human erythrocyte membrane. *Faseb J* 20 (12):1974-81.

Etchberger, J. F., and O. Hobert. 2008. Vector-free DNA constructs improve transgene expression in C. elegans. *Nat Methods* 5 (1):3.

Etchberger, J. F., A. Lorch, M. C. Sleumer, R. Zapf, S. J. Jones, M. A. Marra, R. A. Holt, D. G. Moerman, and O. Hobert. 2007. The molecular signature and cis-regulatory architecture of a C. elegans gustatory neuron. *Genes Dev* 21 (13):1653-74.

Ferguson, C. J., M. Wareing, D. T. Ward, R. Green, C. P. Smith, and D. Riccardi. 2001. Cellular localization of divalent metal transporter DMT-1 in rat kidney. *Am J Physiol Renal Physiol* 280 (5):F803-14.

Finney, M., and G. Ruvkun. 1990. The unc-86 gene product couples cell lineage and cell identity in C. elegans. *Cell* 63 (5):895-905.

Finney, M., G. Ruvkun, and H. R. Horvitz. 1988. The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. *Cell* 55 (5):757-69.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello. 1998. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* 391 (6669):806-11.

Fischer, U., J. Huber, W. C. Boelens, I. W. Mattaj, and R. Luhrmann. 1995. The HIV-1 Rev activation domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell* 82 (3):475-83.

Fischmeister, R., and H. C. Hartzell. 2005. Volume sensitivity of the bestrophin family of chloride channels. *J Physiol* 562 (Pt 2):477-91.

Fogel, G. B., D. G. Weekes, G. Varga, E. R. Dow, H. B. Harlow, J. E. Onyia, and C. Su. 2004. Discovery of sequence motifs related to coexpression of genes using evolutionary computation. *Nucleic Acids Res* 32 (13):3826-35.

Forrester, W. C., and G. Garriga. 1997. Genes necessary for C. elegans cell and growth cone migrations. *Development* 124 (9):1831-43.

Frand, A. R., S. Russel, and G. Ruvkun. 2005. Functional genomic analysis of C. elegans molting. *PLoS Biol* 3 (10):e312.

Friedrich, R. P., B. Schlierf, E. R. Tamm, M. R. Bosl, and M. Wegner. 2005. The class III POU domain protein Brn-1 can fully replace the related Oct-6 during schwann cell development and myelination. *Mol Cell Biol* 25 (5):1821-9.

Froger, A., J. P. Rolland, P. Bron, V. Lagree, F. Le Caherec, S. Deschamps, J. F. Hubert, I. Pellerin, D. Thomas, and C. Delamarche. 2001. Functional characterization of a microbial aquaglyceroporin. *Microbiology* 147 (Pt 5):1129-35.

Fu, D., A. Libson, L. J. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. 2000. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science* 290 (5491):481-6.

Fukuta, M., K. Matsuno, C. C. Hui, T. Nagata, S. Takiya, P. X. Xu, K. Ueno, and Y. Suzuki. 1993. Molecular cloning of a POU domain-containing factor involved in the regulation of the Bombyx sericin-1 gene. *J Biol Chem* 268 (26):19471-5.

Gaudet, J., and S. E. Mango. 2002. Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. *Science* 295 (5556):821-5.

Gaudet, J., S. Muttumu, M. Horner, and S. E. Mango. 2004. Whole-genome analysis of temporal gene expression during foregut development. *PLoS Biol* 2 (11):e352.

Gehring, W. J. 1987. Homeo boxes in the study of development. *Science* 236 (4806):1245-52.

Gehring, W. J., and Y. Hiromi. 1986. Homeotic genes and the homeobox. *Annu Rev Genet* 20:147-73.

Gilleard, J. S., J. D. Barry, and I. L. Johnstone. 1997. cis regulatory requirements for hypodermal cell-specific expression of the Caenorhabditis elegans cuticle collagen gene dpy-7. *Mol Cell Biol* 17 (4):2301-11.

Givens, M. L., R. Kurotani, N. Rave-Harel, N. L. Miller, and P. L. Mellon. 2004. Phylogenetic footprinting reveals evolutionarily conserved regions of the gonadotropin-releasing hormone gene that enhance cell-specific expression. *Mol Endocrinol* 18 (12):2950-66.

Gorelick, D. A., J. Praetorius, T. Tsunenari, S. Nielsen, and P. Agre. 2006. Aquaporin-11: a channel protein lacking apparent transport function expressed in brain. *BMC Biochem* 7:14.

Greenstein, D., S. Hird, R. H. Plasterk, Y. Andachi, Y. Kohara, B. Wang, M. Finney, and G. Ruvkun. 1994. Targeted mutations in the Caenorhabditis elegans POU homeo box gene ceh-18 cause defects in oocyte cell cycle arrest, gonad migration, and epidermal differentiation. *Genes Dev* 8 (16):1935-48.

Grishok, A. 2005. RNAi mechanisms in Caenorhabditis elegans. *FEBS Lett* 579 (26):5932-9.

Hanba, Y. T., M. Shibasaka, Y. Hayashi, T. Hayakawa, K. Kasamo, I. Terashima, and M. Katsuhara. 2004. Overexpression of the barley aquaporin HvPIP2;1 increases internal CO(2) conductance and CO(2) assimilation in the leaves of transgenic rice plants. *Plant Cell Physiol* 45 (5):521-9.

Harfe, B. D., and A. Fire. 1998. Muscle and nerve-specific regulation of a novel NK-2 class homeodomain factor in Caenorhabditis elegans. *Development* 125 (3):421-9.

Haskins, W. T., and P. P. Weinstein. 1957. Nitrogenous excretory products of Trichinella spiralis larvae. *J Parasitol* 43 (1):19-24.

Hauptmann, G., and T. Gerster. 2000. Combinatorial expression of zebrafish Brn-1- and Brn-2-related POU genes in the embryonic brain, pronephric primordium, and pharyngeal arches. *Dev Dyn* 218 (2):345-58.

Hawkins, M. G., and J. D. McGhee. 1995. elt-2, a second GATA factor from the nematode Caenorhabditis elegans. *J Biol Chem* 270 (24):14666-71.

Hazama, A., D. Kozono, W. B. Guggino, P. Agre, and M. Yasui. 2002. Ion permeation of AQP6 water channel protein. Single channel recordings after Hg2+ activation. *J Biol Chem* 277 (32):29224-30.

Hedgecock, E. M., J. G. Culotti, D. H. Hall, and B. D. Stern. 1987. Genetics of cell and axon migrations in Caenorhabditis elegans. *Development* 100 (3):365-82.

Herman, R. K. 1984. Analysis of genetic mosaics of the nematode Caneorhabditis elegans. *Genetics* 108 (1):165-80.

Herr, W., and M. A. Cleary. 1995. The POU domain: versatility in transcriptional regulation by a flexible two-in-one DNA-binding domain. *Genes Dev* 9 (14):1679-93.

Herr, W., R. A. Sturm, R. G. Clerc, L. M. Corcoran, D. Baltimore, P. A. Sharp, H. A. Ingraham, M. G. Rosenfeld, M. Finney, G. Ruvkun, and et al. 1988. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. *Genes Dev* 2 (12A):1513-6.

Hikima, J., M. L. Lennard, M. R. Wilson, N. W. Miller, and G. W. Warr. 2006. Regulation of the immunoglobulin heavy chain locus expression at the phylogenetic level of a bony fish: transcription factor interaction with two variant octamer motifs. *Gene* 377:119-29.

Hillier, L. W., A. Coulson, J. I. Murray, Z. Bao, J. E. Sulston, and R. H. Waterston. 2005. Genomics in C. elegans: so many genes, such a little worm. *Genome Res* 15 (12):1651-60.

Hobert, O. 2002. PCR fusion-based approach to create reporter gene constructs for expression analysis in transgenic C. elegans. *Biotechniques* 32 (4):728-30.

Hodgkin, J., H. R. Horvitz, and S. Brenner. 1979. Nondisjunction Mutants of the Nematode CAENORHABDITIS ELEGANS. *Genetics* 91 (1):67-94.

Holland, P. W., H. A. Booth, and E. A. Bruford. 2007. Classification and nomenclature of all human homeobox genes. *BMC Biol* 5:47.

Huang, C. G., P. Agre, K. Strange, and T. Lamitina. 2006. Isolation of C. elegans deletion mutants following ENU mutagenesis and thermostable restriction enzyme PCR screening. *Mol Biotechnol* 32 (1):83-6.

Huang, C. G., T. Lamitina, P. Agre, and K. Strange. 2007. Functional analysis of the aquaporin gene family in Caenorhabditis elegans. *Am J Physiol Cell Physiol* 292 (5):C1867-73.

Huang, P., E. D. Pleasance, J. S. Maydan, R. Hunt-Newbury, N. J. O'Neil, A. Mah, D. L. Baillie, M. A. Marra, D. G. Moerman, and S. J. Jones. 2007. Identification and analysis of internal promoters in Caenorhabditis elegans operons. *Genome Res* 17 (10):1478-85.

Hunt-Newbury, R., R. Viveiros, R. Johnsen, A. Mah, D. Anastas, L. Fang, E. Halfnight, D. Lee, J. Lin, A. Lorch, S. McKay, H. M. Okada, J. Pan, A. K. Schulz, D. Tu, K. Wong, Z. Zhao, A. Alexeyenko, T. Burglin, E. Sonnhammer, R. Schnabel, S. J. Jones, M. A. Marra, D. L. Baillie, and D. G. Moerman. 2007. High-throughput in vivo analysis of gene expression in Caenorhabditis elegans. *PLoS Biol* 5 (9):e237.

Hwang, B. J., and P. W. Sternberg. 2004. A cell-specific enhancer that specifies lin-3 expression in the C. elegans anchor cell for vulval development. *Development* 131 (1):143-51.

Hyde-DeRuyscher, R. P., E. Jennings, and T. Shenk. 1995. DNA binding sites for the transcriptional activator/repressor YY1. *Nucleic Acids Res* 23 (21):4457-65.

Ilia, M., E. Bazigou, and J. Price. 2003. Expression of the POU domain transcription factor, Oct-6, is attenuated in the adult mouse telencephalon, but increased by neurotoxic damage. *Exp Neurol* 181 (2):159-69.

Ingraham, H. A., S. E. Flynn, J. W. Voss, V. R. Albert, M. S. Kapiloff, L. Wilson, and M. G. Rosenfeld. 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. *Cell* 61 (6):1021-33.

Ishibashi, K. 2006. Aquaporin superfamily with unusual npa boxes: S-aquaporins (superfamily, sip-like and subcellular-aquaporins). *Cell Mol Biol (Noisy-le-grand)* 52 (7):20-7.

Ishibashi, K., M. Kuwahara, and S. Sasaki. 2000. Molecular biology of aquaporins. *Rev Physiol Biochem Pharmacol* 141:1-32.

Itoh, T., T. Rai, M. Kuwahara, S. B. Ko, S. Uchida, S. Sasaki, and K. Ishibashi. 2005. Identification of a novel aquaporin, AQP12, expressed in pancreatic acinar cells. *Biochem Biophys Res Commun* 330 (3):832-8.

Jensen, M. O., E. Tajkhorshid, and K. Schulten. 2003. Electrostatic tuning of permeation and selectivity in aquaporin water channels. *Biophys J* 85 (5):2884-99.

John, B., A. J. Enright, A. Aravin, T. Tuschl, C. Sander, and D. S. Marks. 2004. Human MicroRNA targets. *PLoS Biol* 2 (11):e363.

Jones, S. J., and D. L. Baillie. 1995. Characterization of the let-653 gene in Caenorhabditis elegans. *Mol Gen Genet* 248 (6):719-26.

Jung, J. S., G. M. Preston, B. L. Smith, W. B. Guggino, and P. Agre. 1994. Molecular structure of the water channel through aquaporin CHIP. The hourglass model. *J Biol Chem* 269 (20):14648-54.

Kabat, J. L., S. Barberan-Soler, P. McKenna, H. Clawson, T. Farrer, and A. M. Zahler. 2006. Intronic alternative splicing regulators identified by comparative genomics in nematodes. *PLoS Comput Biol* 2 (7):e86.

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

Kelly, W. G., and A. Fire. 1998. Chromatin silencing and the maintenance of a functional germline in Caenorhabditis elegans. *Development* 125 (13):2451-6.

Kelly, W. G., S. Xu, M. K. Montgomery, and A. Fire. 1997. Distinct requirements for somatic and germline expression of a generally expressed Caernorhabditis elegans gene. *Genetics* 146 (1):227-38.

Kennedy, S., D. Wang, and G. Ruvkun. 2004. A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. *Nature* 427 (6975):645-9.

Kiontke, K., and D. H. Fitch. 2005. The phylogenetic relationships of Caenorhabditis and other rhabditids. *WormBook*:1-11.

Kitamoto, T., and P. M. Salvaterra. 1995. A POU homeo domain protein related to dPOU-19/pdm-1 binds to the regulatory DNA necessary for vital expression of the Drosophila choline acetyltransferase gene. *J Neurosci* 15 (5 Pt 1):3509-18.

Krek, A., D. Grun, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel, and N. Rajewsky. 2005. Combinatorial microRNA target predictions. *Nat Genet* 37 (5):495-500.

Kuroda, T., M. Tada, H. Kubota, H. Kimura, S. Y. Hatano, H. Suemori, N. Nakatsuji, and T. Tada. 2005. Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. *Mol Cell Biol* 25 (6):2475-85.

Kutay, U., and S. Guttinger. 2005. Leucine-rich nuclear-export signals: born to be weak. *Trends Cell Biol* 15 (3):121-4.

Kuwahara, M., T. Asai, K. Sato, I. Shinbo, Y. Terada, F. Marumo, and S. Sasaki. 2000. Functional characterization of a water channel of the nematode Caenorhabditis elegans. *Biochim Biophys Acta* 1517 (1):107-12.

Kuwahara, M., K. Ishibashi, Y. Gu, Y. Terada, Y. Kohara, F. Marumo, and S. Sasaki. 1998. A water channel of the nematode C. elegans and its implications for channel selectivity of MIP proteins. *Am J Physiol* 275 (6 Pt 1):C1459-64.

Lan, L., M. Liu, Y. Liu, and R. He. 2007. Expression and antibody preparation of POU transcription factor qBrn-1. *Protein Pept Lett* 14 (1):7-11.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H.

Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, and Y. J. Chen. 2001. Initial sequencing and analysis of the human genome. *Nature* 409 (6822):860-921.

Lesage, F., I. Lauritzen, F. Duprat, R. Reyes, M. Fink, C. Heurteaux, and M. Lazdunski. 1997. The structure, function and distribution of the mouse TWIK-1 K+ channel. *FEBS Lett* 402 (1):28-32.

Lewis, E. B. 1978. A gene complex controlling segmentation in Drosophila. *Nature* 276 (5688):565-70.

Li, W., A. Streit, B. Robertson, and W. B. Wood. 1999. Evidence for multiple promoter elements orchestrating male-specific regulation of the her-1 gene in Caenorhabditis elegans. *Genetics* 152 (1):237-48.

Li, X., X. Zhao, Y. Fang, X. Jiang, T. Duong, C. Fan, C. C. Huang, and S. R. Kain. 1998. Generation of destabilized green fluorescent protein as a transcription reporter. *J Biol Chem* 273 (52):34970-5.

Li, Y. J., X. H. Fu, D. P. Liu, and C. C. Liang. 2004. Opening the chromatin for transcription. *Int J Biochem Cell Biol* 36 (8):1411-23.

Llimargas, M., and J. Casanova. 1997. ventral veinless, a POU domain transcription factor, regulates different transduction pathways required for tracheal branching in Drosophila. *Development* 124 (17):3273-81.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* 389 (6648):251-60.

MacMorris, M., S. Broverman, S. Greenspoon, K. Lea, C. Madej, T. Blumenthal, and J. Spieth. 1992. Regulation of vitellogenin gene expression in transgenic Caenorhabditis elegans: short sequences required for activation of the vit-2 promoter. *Mol Cell Biol* 12 (4):1652-62.

Matsunami, K., H. Kokubo, K. Ohno, and Y. Suzuki. 1998. Expression pattern analysis of SGF-3/POU-M1 in relation to sericin-1 gene expression in the silk gland. *Dev Growth Differ* 40 (6):591-7.

McKay, S. J., R. Johnsen, J. Khattra, J. Asano, D. L. Baillie, S. Chan, N. Dube, L. Fang, B. Goszczynski, E. Ha, E. Halfnight, R. Hollebakken, P. Huang, K. Hung, V. Jensen, S. J. Jones, H. Kai, D. Li, A. Mah, M. Marra, J. McGhee, R. Newbury, A. Pouzyrev, D. L. Riddle, E. Sonnhammer, H. Tian, D. Tu, J. R. Tyson, G. Vatcher, A. Warner, K. Wong, Z. Zhao, and D. G. Moerman. 2003. Gene expression profiling of cells, tissues, and developmental stages of the nematode C. elegans. *Cold Spring Harb Symp Quant Biol* 68:159-69.

Mello, C. C., J. M. Kramer, D. Stinchcomb, and V. Ambros. 1991. Efficient gene transfer in C.elegans: extrachromosomal maintenance and integration of transforming sequences. *Embo J* 10 (12):3959-70.

Messina, D. N., J. Glasscock, W. Gish, and M. Lovett. 2004. An ORFeome-based analysis of human transcription factor genes and the construction of a microarray to interrogate their expression. *Genome Res* 14 (10B):2041-7.

Mori, I., and Y. Ohshima. 1995. Neural regulation of thermotaxis in Caenorhabditis elegans. *Nature* 376 (6538):344-8.

Morishita, Y., T. Matsuzaki, M. Hara-chikuma, A. Andoo, M. Shimono, A. Matsuki, K. Kobayashi, M. Ikeda, T. Yamamoto, A. Verkman, E. Kusano, S. Ookawara, K. Takata, S. Sasaki, and K. Ishibashi. 2005. Disruption of aquaporin-11 produces polycystic kidneys following vacuolization of the proximal tubule. *Mol Cell Biol* 25 (17):7770-9.

Munke, M., D. R. Cox, I. J. Jackson, B. L. Hogan, and U. Francke. 1986. The murine Hox-2 cluster of homeo box containing genes maps distal on

chromosome 11 near the tail-short (Ts) locus. *Cytogenet Cell Genet* 42 (4):236-40.

Myokai, F., S. Takashiba, R. Lebo, and S. Amar. 1999. A novel lipopolysaccharide-induced transcription factor regulating tumor necrosis factor alpha gene expression: molecular cloning, sequencing, characterization, and chromosomal assignment. *Proc Natl Acad Sci U S A* 96 (8):4518-23.

Nakai, S., Y. Sugitani, H. Sato, S. Ito, Y. Miura, M. Ogawa, M. Nishi, K. Jishage, O. Minowa, and T. Noda. 2003. Crucial roles of Brn1 in distal tubule formation and function in mouse kidney. *Development* 130 (19):4751-9.

Narlikar, L., R. Gordan, and A. J. Hartemink. 2007. A nucleosome-guided map of transcription factor binding sites in yeast. *PLoS Comput Biol* 3 (11):e215.

Nasrallah, R., J. Clark, and R. L. Hebert. 2007. Prostaglandins in the kidney: developments since Y2K. *Clin Sci (Lond)* 113 (7):297-311.

Nelson, F. K., P. S. Albert, and D. L. Riddle. 1983. Fine structure of the Caenorhabditis elegans secretory-excretory system. *J Ultrastruct Res* 82 (2):156-71.

Nelson, F. K., and D. L. Riddle. 1984. Functional study of the Caenorhabditis elegans secretory-excretory system using laser microsurgery. *J Exp Zool* 231 (1):45-56.

Nie, X., I. Arrighi, B. Kaissling, I. Pfaff, J. Mann, J. Barhanin, and V. Vallon. 2005. Expression and insights on function of potassium channel TWIK-1 in mouse kidney. *Pflugers Arch* 451 (3):479-88.

Niemeyer, M. I., L. P. Cid, L. F. Barros, and F. V. Sepulveda. 2001. Modulation of the two-pore domain acid-sensitive K+ channel TASK-2 (KCNK5) by changes in cell volume. *J Biol Chem* 276 (46):43166-74.

Nozaki, K., D. Ishii, and K. Ishibashi. 2008. Intracellular aquaporins: clues for intracellular water transport? *Pflugers Arch* 456 (4):701-7.

P.P., Weinstein. 1952. Regulation of water balance and function of the excretory system of the filariform larvae of Nippostrongylus muris and Ancylostoma caninum. *Experimental Parasitology*:363-376.

Pankratova, E., E. Sytina, and O. Polanovsky. 2006. Autoregulation of Oct-1 gene expression is mediated by two octa-sites in alternative promoter. *Biochimie* 88 (10):1323-9.

Pao, G. M., L. F. Wu, K. D. Johnson, H. Hofte, M. J. Chrispeels, G. Sweet, N. N. Sandal, and M. H. Saier, Jr. 1991. Evolution of the MIP family of integral membrane transport proteins. *Mol Microbiol* 5 (1):33-7.

Park, J. H., and M. H. Saier, Jr. 1996. Phylogenetic characterization of the MIP family of transmembrane channel proteins. *J Membr Biol* 153 (3):171-80.

Parslow, T. G., D. L. Blair, W. J. Murphy, and D. K. Granner. 1984. Structure of the 5' ends of immunoglobulin genes: a novel conserved sequence. *Proc Natl Acad Sci U S A* 81 (9):2650-4.

Phillips, K., and B. Luisi. 2000. The virtuoso of versatility: POU proteins that flex to fit. *J Mol Biol* 302 (5):1023-39.

Ponting, C. P. 2001. Domain homologues of dopamine beta-hydroxylase and ferric reductase: roles for iron metabolism in neurodegenerative disorders? *Hum Mol Genet* 10 (17):1853-8.

Poulin, G., Y. Dong, A. G. Fraser, N. A. Hopper, and J. Ahringer. 2005. Chromatin regulation and sumoylation in the inhibition of Ras-induced vulval development in Caenorhabditis elegans. *Embo J* 24 (14):2613-23.

Prabhu, A., D. P. O'Brien, G. L. Weisner, R. Fulton, and B. Van Ness. 1996. Octamer independent activation of transcription from the kappa immunoglobulin germline promoter. *Nucleic Acids Res* 24 (23):4805-11.

Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier. 1992. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene* 111 (2):229-33.

Preston, G. M., and P. Agre. 1991. Isolation of the cDNA for erythrocyte integral membrane protein of 28 kilodaltons: member of an ancient channel family. *Proc Natl Acad Sci U S A* 88 (24):11110-4.

Preston, G. M., T. P. Carroll, W. B. Guggino, and P. Agre. 1992. Appearance of water channels in Xenopus oocytes expressing red cell CHIP28 protein. *Science* 256 (5055):385-7.

Quinn, C. C., D. S. Pfeil, E. Chen, E. L. Stovall, M. V. Harden, M. K. Gavin, W. C. Forrester, E. F. Ryder, M. C. Soto, and W. G. Wadsworth. 2006. UNC-6/netrin and SLT-1/slit guidance cues orient axon outgrowth mediated by MIG-10/RIAM/lamellipodin. *Curr Biol* 16 (9):845-53.

Ramachandra, N. B., R. D. Gates, P. Ladurner, D. K. Jacobs, and V. Hartenstein. 2002. Embryonic development in the primitive bilaterian Neochildia fusca: normal morphogenesis and isolation of POU genes Brn-1 and Brn-3. *Dev Genes Evol* 212 (2):55-69.

Reece-Hoyes, J. S., B. Deplancke, J. Shingles, C. A. Grove, I. A. Hope, and A. J. Walhout. 2005. A compendium of Caenorhabditis elegans regulatory transcription factors: a resource for mapping transcription regulatory networks. *Genome Biol* 6 (13):R110.

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

Rehberg, S., P. Lischka, G. Glaser, T. Stamminger, M. Wegner, and O. Rosorius. 2002. Sox10 is an active nucleocytoplasmic shuttle protein, and shuttling

is crucial for Sox10-mediated transactivation. *Mol Cell Biol* 22 (16):5826-34.

Reizer, J., A. Reizer, and M. H. Saier, Jr. 1993. The MIP family of integral membrane channel proteins: sequence comparisons, evolutionary relationships, reconstructed pathway of evolution, and proposed functional differentiation of the two repeated halves of the proteins. *Crit Rev Biochem Mol Biol* 28 (3):235-57.

Ren, X. C., S. Kim, E. Fox, E. M. Hedgecock, and W. G. Wadsworth. 1999. Role of netrin UNC-6 in patterning the longitudinal nerves of Caenorhabditis elegans. *J Neurobiol* 39 (1):107-18.

Riddle, Donald L., Thomas Blumenthal, Barbara J. Meyer, and James R. Priess. 1997. *C. Elegans II*. Edited by D. L. Riddle, T. Blumenthal, B. J. Meyer and J. R. Priess. II ed. Vol. II. COLD SPRING HARBOR: LABORATORY PRESS.

Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 234 (4774):364-8.

Rose, K. L., V. P. Winfrey, L. H. Hoffman, D. H. Hall, T. Furuta, and D. Greenstein. 1997. The POU gene ceh-18 promotes gonadal sheath cell differentiation and function required for meiotic maturation and ovulation in Caenorhabditis elegans. *Dev Biol* 192 (1):59-77.

Rual, J. F., J. Ceron, J. Koreth, T. Hao, A. S. Nicot, T. Hirozane-Kishikawa, J. Vandenhaute, S. H. Orkin, D. E. Hill, S. van den Heuvel, and M. Vidal. 2004. Toward improving Caenorhabditis elegans phenome mapping with an ORFeome-based RNAi library. *Genome Res* 14 (10B):2162-8.

Ruvkun, G., and M. Finney. 1991. Regulation of transcription and cell identity by POU domain proteins. *Cell* 64 (3):475-8.

Saifi, G. M., K. Szigeti, W. Wiszniewski, M. E. Shy, K. Krajewski, I. Hausmanowa-Petrusewicz, A. Kochanski, S. Reeser, P. Mancias, I. Butler, and J. R. Lupski. 2005. SIMPLE mutations in Charcot-Marie-Tooth disease and the potential role of its protein product in protein degradation. *Hum Mutat* 25 (4):372-83.

Sakube, Y, A Hirao, S Sasaki, and K Ishibashi. 2003. Aquaporin genes in Caenorhabditis elegans. *International Worm Meeting 2003* (5):421-9.

Salkoff, L., A. Butler, G. Fawcett, M. Kunkel, C. McArdle, G. Paz-y-Mino, M. Nonet, N. Walton, Z. W. Wang, A. Yuan, and A. Wei. 2001. Evolution tunes the excitability of individual neurons. *Neuroscience* 103 (4):853-9.

Sarin, S., M. M. O'Meara, E. B. Flowers, C. Antonio, R. J. Poole, D. Didiano, R. J. Johnston, Jr., S. Chang, S. Narula, and O. Hobert. 2007. Genetic screens for Caenorhabditis elegans mutants defective in left/right asymmetric neuronal fate specification. *Genetics* 176 (4):2109-30.

Schneider, T. D., and R. M. Stephens. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 18 (20):6097-100.

Schreiber, J., J. Enderich, E. Sock, C. Schmidt, C. Richter-Landsberg, and M. Wegner. 1997. Redundancy of class III POU proteins in the oligodendrocyte lineage. *J Biol Chem* 272 (51):32286-93.

Schug J, Overton GJ. 1996. Chapter 2.6. *Current Protocols in Bioinformatics*.

Schultz, J., F. Milpetz, P. Bork, and C. P. Ponting. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95 (11):5857-64.

Schwarzenbach, H., J. W. Newell, and P. Matthias. 1995. Involvement of the Ets family factor PU.1 in the activation of immunoglobulin promoters. *J Biol Chem* 270 (2):898-907.

Shakir, M. A., J. S. Gill, and E. A. Lundquist. 2006. Interactions of UNC-34 Enabled with Rac GTPases and the NIK kinase MIG-15 in Caenorhabditis elegans axon pathfinding and neuronal migration. *Genetics* 172 (2):893-913.

Sheps, J. A., S. Ralph, Z. Zhao, D. L. Baillie, and V. Ling. 2004. The ABC transporter gene family of Caenorhabditis elegans has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. *Genome Biol* 5 (3):R15.

Sibler, A. P., J. Courtete, C. D. Muller, G. Zeder-Lutz, and E. Weiss. 2005. Extended half-life upon binding of destabilized intrabodies allows specific detection of antigen in mammalian cells. *Febs J* 272 (11):2878-91.

Sijen, T., J. Fleenor, F. Simmer, K. L. Thijssen, S. Parrish, L. Timmons, R. H. Plasterk, and A. Fire. 2001. On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107 (4):465-76.

Smale, S. T., and J. T. Kadonaga. 2003. The RNA polymerase II core promoter. *Annu Rev Biochem* 72:449-79.

Smit, R. B., R. Schnabel, and J. Gaudet. 2008. The HLH-6 transcription factor regulates C. elegans pharyngeal gland development and function. *PLoS Genet* 4 (10):e1000222.

Sock, E., J. Enderich, M. G. Rosenfeld, and M. Wegner. 1996. Identification of the nuclear localization signal of the POU domain protein Tst-1/Oct6. *J Biol Chem* 271 (29):17512-8.

Sonnichsen, B., L. B. Koski, A. Walsh, P. Marschall, B. Neumann, M. Brehm, A. M. Alleaume, J. Artelt, P. Bettencourt, E. Cassin, M. Hewitson, C. Holz, M. Khan, S. Lazik, C. Martin, B. Nitzsche, M. Ruer, J. Stamford, M. Winzi, R. Heinkel, M. Roder, J. Finell, H. Hantsch, S. J. Jones, M. Jones, F. Piano, K. C. Gunsalus, K. Oegema, P. Gonczy, A. Coulson, A. A. Hyman, and C. J. Echeverri. 2005. Full-genome RNAi profiling of early embryogenesis in Caenorhabditis elegans. *Nature* 434 (7032):462-9.

Srivastava, A., V. G. Romanenko, M. Gonzalez-Begne, M. A. Catalan, and J. E. Melvin. 2008. A variant of the Ca2+-activated Cl channel Best3 is expressed in mouse exocrine glands. *J Membr Biol* 222 (1):43-54.

Stein, L. D., Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent, N. Chen, A. Chinwalla, L. Clarke, C. Clee, A. Coghlan, A. Coulson, P. D'Eustachio, D. H. Fitch, L. A. Fulton, R. E. Fulton, S. Griffiths-Jones, T. W. Harris, L. W. Hillier, R. Kamath, P. E. Kuwabara, E. R. Mardis, M. A. Marra, T. L. Miner, P. Minx, J. C. Mullikin, R. W. Plumb, J. Rogers, J. E. Schein, M. Sohrmann, J. Spieth, J. E. Stajich, C. Wei, D. Willey, R. K. Wilson, R. Durbin, and R. H. Waterston. 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. *PLoS Biol* 1 (2):E45.

Stein, L. D., C. Mungall, S. Shu, M. Caudy, M. Mangone, A. Day, E. Nickerson, J. E. Stajich, T. W. Harris, A. Arva, and S. Lewis. 2002. The generic genome browser: a building block for a model organism system database. *Genome Res* 12 (10):1599-610.

Stringham, E., N. Pujol, J. Vandekerckhove, and T. Bogaert. 2002. unc-53 controls longitudinal migration in C. elegans. *Development* 129 (14):3367-79.

Sulston, J. E., and H. R. Horvitz. 1977. Post-embryonic cell lineages of the nematode, Caenorhabditis elegans. *Dev Biol* 56 (1):110-56.

———. 1981. Abnormal cell lineages in mutants of the nematode Caenorhabditis elegans. *Dev Biol* 82 (1):41-55.

Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev Biol* 100 (1):64-119.

Swoboda, P., H. T. Adler, and J. H. Thomas. 2000. The RFX-type transcription factor DAF-19 regulates sensory neuron cilium formation in C. elegans. *Mol Cell* 5 (3):411-21.

Tanaka, M., and W. Herr. 1990. Differential transcriptional activation by Oct-1 and Oct-2: interdependent activation domains induce Oct-2 phosphorylation. *Cell* 60 (3):375-86.

Thacker, C., J. A. Sheps, and A. M. Rose. 2006. Caenorhabditis elegans dpy-5 is a cuticle procollagen processed by a proprotein convertase. *Cell Mol Life Sci* 63 (10):1193-204.

Tharakaraman, K., O. Bodenreider, D. Landsman, J. L. Spouge, and L. Marino-Ramirez. 2008. The biological function of some human transcription factor binding motifs varies with position relative to the transcription start site. *Nucleic Acids Res* 36 (8):2777-86.

Thoemke, K., W. Yi, J. M. Ross, S. Kim, V. Reinke, and D. Zarkower. 2005. Genome-wide analysis of sex-enriched gene expression during C. elegans larval development. *Dev Biol* 284 (2):500-8.

Thompson, J. D., T. J. Gibson, F. Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25 (24):4876-82.

Trieu, M., A. Ma, S. R. Eng, N. Fedtsova, and E. E. Turner. 2003. Direct autoregulation and gene dosage compensation by POU-domain transcription factor Brn3a. *Development* 130 (1):111-21.

Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J. A. Malek, G. Costa, K. McKernan, A. Sidow, A. Fire, and S. M. Johnson. 2008. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. *Genome Res* 18 (7):1051-63.

van Leeuwen, H. C., M. J. Strating, M. Rensen, W. de Laat, and P. C. van der Vliet. 1997. Linker length and composition influence the flexibility of Oct-1 DNA binding. *Embo J* 16 (8):2043-53.

van Os, C. H., E. J. Kamsteeg, N. Marr, and P. M. Deen. 2000. Phsyiological relevance of aquaporins: luxury or necessity? *Pflugers Arch* 440 (4):513-20.

Velculescu, V. E., L. Zhang, B. Vogelstein, and K. W. Kinzler. 1995. Serial analysis of gene expression. *Science* 270 (5235):484-7.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C.

Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu. 2001. The sequence of the human genome. *Science* 291 (5507):1304-51.

Verrijzer, C. P., M. J. Alkema, W. W. van Weperen, H. C. Van Leeuwen, M. J. Strating, and P. C. van der Vliet. 1992. The DNA binding specificity of the bipartite POU domain and its subdomains. *Embo J* 11 (13):4993-5003.

Wallace, I. S., and D. M. Roberts. 2004. Homology modeling of representative subfamilies of Arabidopsis major intrinsic proteins. Classification based on the aromatic/arginine selectivity filter. *Plant Physiol* 135 (2):1059-68.

Wang, F., X. C. Feng, Y. M. Li, H. Yang, and T. H. Ma. 2006. Aquaporins as potential drug targets. *Acta Pharmacol Sin* 27 (4):395-401.

Wang, L., and J. C. Way. 1996. Promoter sequences for the establishment of mec-3 expression in the nematode Caenorhabditis elegans. *Mech Dev* 56 (1-2):183-96.

Wang, P., J. Zhao, and A. K. Corsi. 2006. Identification of novel target genes of CeTwist and CeE/DA. *Dev Biol* 293 (2):486-98.

Ward, S., N. Thomson, J. G. White, and S. Brenner. 1975. Electron microscopical reconstruction of the anterior sensory anatomy of the nematode Caenorhabditis elegans.?2UU. *J Comp Neurol* 160 (3):313-37.

Washietl, S., I. L. Hofacker, and P. F. Stadler. 2005. Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* 102 (7):2454-9.

Webb, C. T., S. A. Shabalina, A. Y. Ogurtsov, and A. S. Kondrashov. 2002. Analysis of similarity within 142 pairs of orthologous intergenic regions of Caenorhabditis elegans and Caenorhabditis briggsae. *Nucleic Acids Res* 30 (5):1233-9.

Wegner, M., D. W. Drolet, and M. G. Rosenfeld. 1993. POU-domain proteins: structure and function of developmental regulators. *Curr Opin Cell Biol* 5 (3):488-98.

Weig, A., C. Deswarte, and M. J. Chrispeels. 1997. The major intrinsic protein family of Arabidopsis has 23 members that form three distinct groups with functional aquaporins in each group. *Plant Physiol* 114 (4):1347-57.

Wen, W., J. L. Meinkoth, R. Y. Tsien, and S. S. Taylor. 1995. Identification of a signal for rapid export of proteins from the nucleus. *Cell* 82 (3):463-73.

Witzgall, R., E. O'Leary, R. Gessner, A. J. Ouellette, and J. V. Bonventre. 1993. Kid-1, a putative renal transcription factor: regulation during ontogeny and in response to ischemia and toxic injury. *Mol Cell Biol* 13 (3):1933-42.

Wu, L., and A. Berk. 1988. Constraints on spacing between transcription factor binding sites in a simple adenovirus promoter. *Genes Dev* 2 (4):403-11.

Xie, Z., E. Allen, N. Fahlgren, A. Calamar, S. A. Givan, and J. C. Carrington. 2005. Expression of Arabidopsis MIRNA genes. *Plant Physiol* 138 (4):2145-54.

Xu, P. X., M. Fukuta, S. Takiya, K. Matsuno, X. Xu, and Y. Suzuki. 1994. Promoter of the POU-M1/SGF-3 gene involved in the expression of Bombyx silk genes. *J Biol Chem* 269 (4):2733-42.

Yakata, K., Y. Hiroaki, K. Ishibashi, E. Sohara, S. Sasaki, K. Mitsuoka, and Y. Fujiyoshi. 2007. Aquaporin-11 containing a divergent NPA motif has normal water channel activity. *Biochim Biophys Acta* 1768 (3):688-93.

Yamagata, T., J. Nishida, R. Sakai, T. Tanaka, H. Honda, N. Hirano, H. Mano, Y. Yazaki, and H. Hirai. 1995. Of the GATA-binding proteins, only GATA-4 selectively regulates the human interleukin-5 gene promoter in interleukin-5-producing cells which express multiple GATA-binding proteins. *Mol Cell Biol* 15 (7):3830-9.

Yamamoto, T., and S. Sasaki. 1998. Aquaporins in the kidney: emerging new aspects. *Kidney Int* 54 (4):1041-51.

Yasui, M., A. Hazama, T. H. Kwon, S. Nielsen, W. B. Guggino, and P. Agre. 1999. Rapid gating and anion permeability of an intracellular aquaporin. *Nature* 402 (6758):184-7.

Yuh, C. H., E. R. Dorman, and E. H. Davidson. 2005. Brn1/2/4, the predicted midgut regulator of the endo16 gene of the sea urchin embryo. *Dev Biol* 281 (2):286-98.

Zhao, Z., T. J. Boyle, Z. Bao, J. I. Murray, B. Mericle, and R. H. Waterston. 2008. Comparative analysis of embryonic cell lineage between Caenorhabditis briggsae and Caenorhabditis elegans. *Dev Biol* 314 (1):93-9.

Zhao, Z., L. Fang, N. Chen, R. C. Johnsen, L. Stein, and D. L. Baillie. 2005. Distinct regulatory elements mediate similar expression patterns in the excretory cell of Caenorhabditis elegans. *J Biol Chem* 280 (46):38787-94.

Zucker-Aprison, E., and T. Blumenthal. 1989. Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. *J Mol Evol* 28 (6):487-96.

Zupkovitz, G., J. Tischler, M. Posch, I. Sadzak, K. Ramsauer, G. Egger, R. Grausenburger, N. Schweifer, S. Chiocca, T. Decker, and C. Seiser. 2006. Negative and positive regulation of gene expression by mouse histone deacetylase 1. *Mol Cell Biol* 26 (21):7913-28.

Zwilling, S., A. Annweiler, and T. Wirth. 1994. The POU domains of the Oct1 and Oct2 transcription factors mediate specific interaction with TBP. *Nucleic Acids Res* 22 (9):1655-62.