Transcriptional Regulation of AQP-8, a Caenorhabditis elegans Aquaporin Exclusively Expressed in the Excretory System, by the POU Homeobox Transcription Factor CEH-6*§

Allan K. Mah‡1, Kristin R. Armstrong‡, Derek S. Chew‡, Jeffrey S. Chu‡, Domina K. Tu‡, Robert C. Johnsen‡, Nansheng Chen‡, Helen M. Chamberlin‡, and David L. Baillie‡§

From the ‡Department of Molecular Biology and Biochemistry, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada and the §Department of Molecular Genetics, Ohio State University, Columbus, Ohio 43210

Due to the ever changing environmental conditions in soil, regulation of osmotic homeostasis in the soil-dwelling nematode Caenorhabditis elegans is critical. AQP-8 is a C. elegans aquaporin that is expressed in the excretory cell, a renal equivalent tissue, where the protein participates in maintaining water balance. To better understand the regulation of AQP-8, we undertook a promoter analysis to identify the aqp-8 cis-regulatory elements. Using progressive 5’ deletions of upstream sequence, we have mapped an essential regulatory region to roughly 300 bp upstream of the translational start site of aqp-8. Analysis of this region revealed a sequence corresponding to a known DNA functional element (octamer motif), which interacts with POU homeobox transcription factors. Phylogenetic footprinting showed that this site is perfectly conserved in four nematode species. The octamer site’s function was further confirmed by deletion analyses, mutagenesis, functional studies, and electrophoretic mobility shift assays. Of the three POU homeobox proteins encoded in the C. elegans genome, CEH-6 is the only member that is expressed in the excretory cell. We show that expression of AQP-8 is regulated by CEH-6 by performing RNA interference experiments. CEH-6’s mammalian ortholog, Brn1, is expressed both in the kidney and the central nervous system and binds to the same octamer consensus binding site to drive gene expression. These parallels in transcriptional control between Brn1 and CEH-6 suggest that C. elegans may well be an appropriate model for determining gene-regulatory networks in the developing vertebrate kidney.

The Caenorhabditis elegans excretory system maintains osmotic homeostasis by expelling ionic and metabolic waste from the organism (1). The excretory system is comprised of four cells: the excretory duct cell, the binucleate excretory gland cell, the excretory pore cell, and the excretory (canal) cell. The excretory cell is a large fluid-filled cell consisting of a cell body, which is located ventral to the isthmus worm pharynx. From the cell body, two long canals extend posteriorly and two shorter canals extend anteriorly; the four canals are joined at the cell body forming an H-shape. The surface areas of the four canals are greatly increased by a system of canaliculi, which also provides exposure to the extracellular fluid-filled pseudocoelomic cavity (2). Developmental outgrowth of the excretory cell has been shown to be modulated by many of the same mechanisms that dictate neuronal guidance (3). Both of these tissues extend long cellular processes between their cell membranes and the epidermal basal lamina during development. Indeed, mutations in many genes, including lin-17, unc-5, unc-6, unc-34, unc-53, and unc-73, affect both neuronal and excretory cell development either by disrupting circumferential growth or by causing premature migrational termination of cellular processes (4). In addition, excretory cell tubulogenesis occurs in vitro using conditions intended for neuronal cell culturing (2).

Consistent with the excretory cell’s role in osmoregulation, previous experiments have shown that the excretory duct cell pumping rates in dauer larvae (a nematode diapause state induced by adverse environmental pressures) were inversely proportional to environmental osmotic pressures (5). Laser ablation of the excretory cell, duct cell, or pore cell leads to fluid retention within the worm (5). These two examples clearly demonstrate that the nematode excretory system is an osmoregulatory organ and therefore functionally analogous to the vertebrate kidney (1). The constrained developmental processes for such a unique structure and the physiological properties of the excretory cell make it an ideal tissue for studies on transcriptional regulation and how they relate to both renal and neuronal development. In addition, the excretory cell is the largest single cell in the worm making it ideal for studies of gross morphological defects.

To maintain fluid homeostasis between an organism and its environment and between tissues, members of the aquaporin (AQP) family of water channels are expressed in various tis-

---

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 Supported by a Natural Sciences and Engineering Reserach Council of Canada doctoral scholarship. To whom correspondence should be addressed: Dept. of Molecular Biology and Biochemistry, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia V5A 1S6, Canada. Tel.: 604-291-4597; Fax: 604-291-5583; E-mail: amaha@sfu.ca.

2 Supported by a Weyerhauser-Molecular Biology and Biochemistry fellowship and a Hemingway Nelson Architects Graduate Scholarship.

3 Supported by a grant from NSERC Canada and a Faculty start-up fund provided by Simon Fraser University.

4 Supported by National Science Foundation Grant IOB-0516554.

5 A Canada Research Chair in Genomics and supported by grants from NSERC and the Canadian Institutes of Health Research.

6 The abbreviations used are: AQP, aquaporin; GFP, green fluorescent protein; TESS, Transcriptional Element Search System; EMSA, electrophoretic mobility shift assay; PEST, degradation signal consisting of a Pro-Glu-Ser-Thr rich sequence; SAGE, serial analysis of gene expression; POU, pit, oct, and unci; RNAL, RNA interference; dsRNA, double strand RNA; EST, expressed sequence tag; TF, transcription factor.
sues and developmental stages in all forms of life to facilitate water movement across biological membranes. The existence of AQPs in virtually all cells allows for bidirectional passive flux of water across lipid bilayers, which in the absence of these proteins are essentially impermeable barriers. In addition, many cell types incorporate multiple different AQPs per cell and/or tissue. The spatial and temporal expression redundancy of AQPs may explain the relative lack of resultant gross phenotypes in AQP knock-out studies. AQPs have been discovered in all types of organisms, from mammals to a recently discovered 270-amino acid homolog, AQPV1, in the Chlorella virus MT325s (6).

The first AQP cloned, the 28-kDa protein AQP1, is a common protein in human red blood cell plasma membranes occurring at a level of ~120,000–160,000 copies per cell (7, 8). Besides transporting water, many AQP members have the capacity to transport other small non-charged molecules. Selectivity of the channel for different uncharged solutes is derived from a proton filter region in the protein characterized by a pore-associated arginine residue in association with neighboring aromatic amino acids (9). Due to these different channel specificities, AQPs are grouped into two functionally distinct classes: the aquaporins, which are exclusively water channels, and the aquaglyceroporins, which also have the ability to transport small non-ionic molecules such as glycerol and urea (10). AQPs contain a pair of signature domains, the NPA motifs (asparagine-proline-alanine, or in some cases, asparagine-proline-valine), which are essential for the pore structure and function (10). The locations of the NPA amino acid residues allow passage of water through the pore. Due to hydrogen bonding interactions, the molecules travel through the pore in a single file manner (11). Overall, each AQP1 unit has the capacity to allow the passage of three billion molecules of water per second (11). An additional requirement for a functional aquaporin is a conserved fold in the protein, which has been observed in both AQP1 and the bacterial aquaglyceroporin, GlpF (11). Some AQPs can be blocked by Hg\(^{2+}\) at a pore-associated cysteine residue (12).

In humans, 7 of 13 identified AQPs are expressed in various parts of the kidney to maintain osmotic balance and to prevent excessive fluid loss (13). The C. elegans genome contains eleven AQPs (aaps 1-11) (14). Like their mammalian counterparts, C. elegans AQP-4 (F40F9.9) and AQP-2 (C01G6.1), have been shown to be involved in fluid homeostasis. These results were obtained by physiological experiments determining changes in water flux as a result of the insertion of AQPs into Xenopus laevis oocyte membranes (15–17). In addition, AQP-4 was shown to be inhibited by Hg\(^{2+}\), much like other aquaporins (12, 16, 17). The expression patterns for the C. elegans AQPs 1–8 have been determined previously (15). Three of the eight AQPs studied were demonstrated to be expressed in the excretory cell (aap-2, aap-3, and aap-8) (15).

Prior studies have identified DNA regions containing functional cis-regulatory elements of various AQPs. Areas containing positive acting cis-regulatory elements for human AQP1 have been determined by analyzing the transcriptional activity of various length gene-upstream fragments (18). Regions containing negative transcriptional regulators of mouse AQP2 have also been determined (19). In addition, human AQP4, which has two splice variants that lead to isoforms with distinct N termini and pore permeability properties, has been shown to be under the control of alternate upstream regulatory sequences directly upstream of each splice variant (20). Studies in plants have also revealed regions containing cis-regulatory elements, which modulate AQPs. An analysis of the AthH2 (PIP1b), an Arabidopsis AQP, upstream regulatory region revealed two phytohormone-induced enhancer-containing regions (21). In this study we have determined a cis-regulatory element that is required for the expression of an aquaporin in the excretory cell of C. elegans. Information pertaining to AQP expression and regulation in the excretory system of C. elegans will provide complementary information to previous studies on transcriptional regulation of aquaporins and to also perhaps provide a basis for determining mechanisms controlling transcription in mammalian renal and neuronal tissues.

**EXPERIMENTAL PROCEDURES**

Nematode Strains and Maintenance—Strains were maintained at 20 °C, and all manipulations were conducted using standard methods (22). Strains used in these studies are shown in Table 1. We generated a stable green fluorescent protein (GFP)-expressing line by subjecting (BC6835) aqp-8–711::GFP-PEST P0 worms to 1,500-rad X-irradiation. Spontaneous transgene integrants were isolated by selecting for F3 lines, which produced 100% rescued dpy-5 progeny thus producing (BC7032) dpy-5(e907);sIs1241 rces[K02G10.7(−711):GFP-PEST + pCeh361].

Transgene Construction—DNA constructs were generated via fusion PCR as previously described (23). 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), pAF207 (GFP-PEST), and pPD97.78 (Δpes-10::GFP). All GFP variants used are modified by the addition of a 5′-nuclease localization, 3′-unc-54-untranslated region, S65C mutation, and additional synthetic introns. Site-directed mutagenesis of the motif site was carried out via nucleotide substitutions corresponding to residue changes in the motif in the forward primer. In all cases except the Δpes-10::GFP fusion construct and the AQP-8::GFP translational fusion construct, the reverse primer for upstream regulatory region amplification is aqp-8R: agtcgacctgcaggcatgcaagcttagaaacggatcgcagaaaa. The forward primers used for amplification of deletional constructs are shown in supplemental Table S1. The forward primers used for mutagenesis of the cis-regulatory element are as follows (mutated residues are underlined): aqp-8 oct G aaps 1-11 (14). Like their mammalian counterparts, C. elegans AQP-4 (F40F9.9) and AQP-2 (C01G6.1), have been shown to be involved in fluid homeostasis. These results were obtained by physiological experiments determining changes in water flux as a result of the insertion of AQPs into Xenopus laevis oocyte membranes (15–17). In addition, AQP-4 was shown to be inhibited by Hg\(^{2+}\), much like other aquaporins (12, 16, 17). The expression patterns for the C. elegans AQPs 1–8 have been determined previously (15). Three of the eight AQPs studied were demonstrated to be expressed in the excretory cell (aap-2, aap-3, and aap-8) (15).

Prior studies have identified DNA regions containing functional cis-regulatory elements of various AQPs. Areas containing positive acting cis-regulatory elements for human AQP1 have been determined by analyzing the transcriptional activity of various length gene-upstream fragments (18). Regions containing negative transcriptional regulators of mouse AQP2

Transcriptional Regulation of AQP-8 by CEH-6
ward primer used for the translational construct corresponds to aqp-8−1.6kb (supplemental Table S1).

PCR constructs were co-injected with the Dpy-5 rescuing construct, pCeh361 (24), into the synciatal gonad of late L4 dpy-S(e907) worms. Wild-type F1 worms were plated individually. Wild-type F2 worms were selected to start the lines. In the case of generating multiple independent lines, each were analyzed separately and designated as individual segregants (Table 1).

**Microscopy**—A Zeiss Axioscope equipped with a QImaging camera and the appropriate optical filter sets were used for GFP expression pattern analysis. Worms were immobilized with 100 mM sodium azide (in water) immediately prior to imaging. All images were taken at 400× with identical camera settings for all images (exposure times are indicated in the figures). Images were captured using QCapture software and processed using Adobe Photoshop CS.

**Sequence Analysis**—DNA and peptide sequence alignments were carried out using ClustalX (25) with default settings. The DNA sequence spanning the bases −283 to −234 upstream of the *C. elegans* aqp-8 translational start site was used as a query in the Transcriptional Element Search System (TESS) to identify potential conserved transcription factor binding sites. Default parameters were used.

**Electrophoretic Mobility Shift Assay**—Nuclear and cytoplasmic extracts were isolated from N2 worms harvested in M9 buffer. The extracts were prepared using the NE-PER Nuclear and Cytoplasmic Extraction Reagents Kit (Pierce). The synthetic biotinylated oligonucleotides used in this study include the consensus octamer oligonucleotide, 5′-TTTGCCAAAT-TTGCACTTGGAAT-3′ and its complement 5′-ATCCAG-TATGCAAATTTTGGCAAT-3′.

**List of *C. elegans* strains analyzed**

| Strain     | Genotype                      |
|------------|-------------------------------|
| N2         | N2 (Bristol) *C. elegans* wild type |
| CB9007     | dpy-S(e907)                   |
| BC07032    | dpy-S(e907)x1241 rCe[K02G10.7(-711)]-GFP-PEST + pCeh361 |
| BC20018    | dpy-S(e907)x20018 rCe[K02G10.7(-1.6kb)]-GFP-PEST + pCeh361 |
| BC20052    | dpy-S(e907)x20052 rCe[K02G10.7(-1.6kb)]-GFP + pCeh361 |
| TM1919     | aqp-8(tm1919), kindly provided by S. Mitani |
| BC06835    | dpy-S(e907)x1241 rCe[K02G10.7(-711)]-GFP-PEST + pCeh361(SegI) |
| BC06836    | dpy-S(e907)x1242 rCe[K02G10.7(-711)]-GFP-PEST + pCeh361(SegI) |
| BC06837    | dpy-S(e907)x1243 rCe[K02G10.7(-567)]-GFP-PEST + pCeh361(SegI) |
| BC06838    | dpy-S(e907)x1244 rCe[K02G10.7(-567)]-GFP-PEST + pCeh361(SegI) |
| BC07226    | dpy-S(e907)x1252 rCe[K02G10.7(-434)]-GFP-PEST + pCeh361 |
| BC06839    | dpy-S(e907)x1245 rCe[K02G10.7(-342)]-GFP-PEST + pCeh361 |
| BC07227    | dpy-S(e907)x1524 rCe[K02G10.7(-315)]-GFP-PEST + pCeh361 |
| BC06876    | dpy-S(e907)x1278 rCe[K02G10.7(-297)]-GFP + pCeh361 |
| BC06877    | dpy-S(e907)x1279 rCe[K02G10.7(-279)]-GFP + pCeh361(SegI) |
| BC06878    | dpy-S(e907)x1280 rCe[K02G10.7(-279)]-GFP + pCeh361(SegI) |
| BC06892    | dpy-S(e907)x1314 rCe[K02G10.7(-272)]-GFP + pCeh361(SegI) |
| BC06892    | dpy-S(e907)x1315 rCe[K02G10.7(-272)]-GFP + pCeh361(SegI) |
| BC06896    | dpy-S(e907)x1312 rCe[K02G10.7(-267)]-GFP + pCeh361 |
| BC06879    | dpy-S(e907)x1281 rCe[K02G10.7(-261)]-GFP + pCeh361(SegI) |
| BC06887    | dpy-S(e907)x1289 rCe[K02G10.7(-261)]-GFP + pCeh361(SegI) |
| BC07229    | dpy-S(e907)x1526 rCe[K02G10.7(-228)]-GFP-PEST + pCeh361 |
| BC06916    | dpy-S(e907)x1312 rCe[K02G10.7(-207)]-GFP + pCeh361(SegI) |
| BC06840    | dpy-S(e907)x1246 rCe[K02G10.7(-207)]-GFP-PEST + pCeh361(SegI) |
| BC06841    | dpy-S(e907)x1247 rCe[K02G10.7(-207)]-GFP-PEST + pCeh361(SegI) |
| BC06842    | dpy-S(e907)x1248 rCe[K02G10.7(-207)]-GFP-PEST + pCeh361(SegI) |
| BC06843    | dpy-S(e907)x1249 rCe[K02G10.7(-207)]-GFP-PEST + pCeh361(SegI) |
| BC06925    | dpy-S(e907)x1318 rCe[K02G10.7(-207)]-GFP-PEST + pCeh361(SegI) |
| BC06926    | dpy-S(e907)x1319 rCe[K02G10.7(-207)]-GFP-PEST + pCeh361(SegI) |
| BC06929    | dpy-S(e907)x1345 rCe[4xATCTGGATATA]::pes-10::GFP + pCeh361(SegI) |
| BC07030    | dpy-S(e907)x1387 rCe[4xATCTGGATATA]::pes-10::GFP + pCeh361(SegI) |
| BC07285    | dpy-S(e907)x1358 rCe[aqp-8 oct G→A]:GFP + pCeh361(SegI) |
| BC07286    | dpy-S(e907)x1355 rCe[aqp-8 oct G→A]:GFP + pCeh361(SegI) |
| BC07312    | dpy-S(e907)x1574 rCe[aqp-8 oct C→G]:GFP + pCeh361 |
| BC01300    | dpy-S(e907)x10300 rCe[JK724.1]:GFP + pCeh361 |
their expression patterns by searching a *C. elegans* GFP expres-
sion data base.7 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 was calculated by count-
ing 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. Mathemati-
cally, this can be represented by letting $v$ be the observed value and let-
ting $N$ be the set of excretory cell observations from random selec-
tions. $S$ is the largest subset of $N$ such that $\forall \sigma \in S$, $\sigma \geq v$. The resulting
probability is $|S|/|N|$.

RESULTS

Identification of an Excretory System Expressed AQP—Gene transcrip-
tional patterns were determined for each of the eleven *C. elegans* aqp
members (aqps 1–11) by assaying expression patterns derived from
promoter::GFP fusions in vivo (data not shown). Consistent with previous
evidence (15), aqp-8 is the only AQP that had its expression localized
exclusively to the excretory system of the worm (Fig. 1). The intergenic
region between aqp-8 and its closest upstream gene neighbor,
K02G10.1, is 2,220 bp. Our initial aqp-8 upstream region was defined as the
1,556-bp region from $1575$ to $20bp$ immediately upstream of the
K02G10.7 translational start site (Fig. 1A). Expression of aqp-8 also appears
to be localized to an additional cell. We presume, by the location of the
additional cell, that this cell may be the excretory gland cell. Because
many genes are transiently expressed for specific physiological functions
and processes, we assayed for the temporal activity of the upstream reg-
ulatory region of aqp-8 using the GFP-PEST reporter (27) (Fig. 1B).

The PEST sequence (Pro-Glu-Ser-Thr) is a signal for rapid degradation
of proteins, which was discovered as a conserved sequence from multiple alignments of known short-
lived proteins (28). The coding sequence for the PEST sequence was inserted directly upstream of the 3′-untrans-
lated region toward the C-terminal portion of GFP, produc-
ing a GFP variant with an in vivo half-life of less than 1 h (27). The aqp-8p::GFP-PEST-expressing worms displayed an
identical spatial pattern to the worms carrying the usual aqp-8::GFP construct, but due to the short half-life of the GFP-PEST construct, we were able to determine that aqp-8 is tran-

FIGURE 1. GFP expression pattern analysis of AQP-8. A, the aqp-8 1.6-kb upstream fragment drove GFP
expression in the excretory cell (EC) and a secondary cell ( ) (L2). B, expression of a transgenic strain carrying
the aqp-8p::GFP-PEST construct. Spatial expression patterns of the aqp-8p::GFP-PEST containing strains were
effectively identical to the expression of strains carrying promoter::GFP constructs. C, expression of a stabilized
transgene (genome-integrated tandem DNA array) leads to expression patterns that are identical to non-
stabilized transgenic lines. D, closer analysis of an adult worm expressing the AQP::GFP construct (translational
reporter fusion) shows that the GFP is localized to the outer walls of the excretory cell canal. The expression
level of the translational construct was slightly lower than that of the transcriptional constructs (L3). E, expres-
sion of a stabilized transgene consisting of a 711-bp upstream fragment fused to GFP-PEST. All images were
captured at 400× magnification. Camera conditions were the same for all images. Exposure times are as
indicated on the GFP images (left images).
Transcriptional Regulation of AQP-8 by CEH-6

scribed only in the interval between the first larval stage and early adulthood. The relative levels of expression in the excretory cell and the excretory gland cell appeared to be similar to each other (Fig. 1, A and B). Expression patterns derived from extrachromosomal arrays may be confounded by somatic loss of the transgene (leading to mosaically expressing transgenes). Therefore, we confirmed the expression pattern of aqp-8 by generating a genome-integrated aqp-8::GFP transgenic line to prevent the sporadic loss of the transgene in somatic tissue (Fig. 1C). These expression data are consistent with the C. elegans developmental Serial Analysis of Gene Expression (SAGE) profiles of aqp-8 (29), which show aqp-8 mRNA production starting in L1 larval stage and intensifying until L4 when the level of transcription tapers off. SAGE data, corresponding to various libraries derived from fluorescence-activated cell sorting-derived C. elegans embryonic cells, revealed that aqp-8 is expressed at a low but detectable level in each of the purified oocyte, embryonic, and sorted AFD neuron mRNA libraries (29). The aqp-8 gene locus, K02G10.7, encodes two splice variants that differ by an alternatively spliced fourth exon. The smaller isoform, K02G10.7a, encodes a 258-amino acid protein, whereas K02G10.7b encodes a slightly larger 294-amino acid protein. Both AQP-8 isoforms have been confirmed by expressed sequence tags (30). K02G10.7a, contains five possible membrane-spanning regions, whereas the larger b variant contains six membrane-spanning regions.

The closest mammalian homolog of C. elegans AQP-8 is AQP10, an aquaglyceroporin predominantly expressed in the jejunal glands (Crypts of Lieberkühn), and the duodenum epithelia. The jejunal glands function to secrete various digestive enzymes and contain mitotically active stem cells for the purpose of epithelial regeneration. A possible role for AQP10 is to regulate osmolarity in these regions of the small intestine that are known to be subject to considerable changes in solute concentrations. In particular, stringent regulation of duodenum osmolytes may be due to large changes in solute concentrations as undiluted stomach contents are passed directly through to it (31). Like AQP-8, vertebrate AQP10 has two isoforms with five (30 kDa) and six (35 kDa) transmembrane domains, respectively (32). Though bona fide transcripts have been identified for the shorter splice variants of both AQP-8 and AQP10, it is possible that these isoforms are non-functional. Unlike AQP-8, AQP10 can conduct both water and glycerol (15, 32). It has been suggested, however, that AQP-8 may be important for adaptation to osmotic stress, because its expression levels have been observed to be induced significantly when worms are placed under hypotonic stress (15). The two NPA domains in K02G10.7b are located between the transmembrane segments II/III and V/VI, which correspond to the locations of the NPA domains in human AQP1. A null mutant of aqp-8 (tm1919) does not show any obvious structural defects in the excretory cell or any assayable response to changes in osmotic pressures (data not shown). Likewise, treatment of nematodes with RNAi corresponding to aqp-8, clone sjj_K02G10.7 (33), did not result in any obvious change in response from that of wild-type worms when placed in media containing different levels of salinity (data not shown). Although loss of AQP-8 itself does not lead to an observable phenotype, quadruple mutants of aqp-2, aqp-3, aqp-4, and aqp-8 have been shown to lead to worms with impaired mobility when subjected to hypotonic environments (15). To determine whether AQP-8 remains in the excretory cell after translation, we generated a K02G10.7 translational GFP fusion. This construct consisted of GFP (pPD 95.75) fused in-frame at the C terminus of a PCR product consisting of aqp-8’s 1.6-kb upstream region and coding sequence. Localization of AQP-8 was identical to the expression pattern revealed by the transcriptional GFP fusions albeit displaying a lesser level of fluorescence than worms carrying the promoter::GFP (and promoter::GFP-PEST) constructs (Fig. 1, D and E). The lower fluorescence level of the translational reporter construct relative to the transcriptional reporter constructs may be attributable to a higher protein turnover of the AQP-8::GFP protein relative to untagged GFP.

Determination of Upstream Regions Required for Excretory Cell Expression of aqp-8—The initial expression pattern analysis of aqp-8 was examined for two constructs in vivo using the 1.6-kb fragment fused to both GFP and GFP-PEST. To map the upstream DNA elements responsible for the excretory system-specific expression of aqp-8 in C. elegans, a series of fragments consisting of progressive 5’ deletions of the upstream regulatory region of aqp-8 were fused to the GFP or GFP-PEST coding cassettes. A cis-regulatory element was initially localized to a region spanning the nucleotides −342 to −207. A further round of deletions within the defined window resolved the cis-regulatory element containing region to an interval between −279 and −261. GFP expression levels and patterns of the transgenics were consistent with the original 1.6-kb aqp-8 upstream constructs’ expression pattern until loss of expression in −261 constructs and all subsequent shorter constructs (Fig. 2). From the deletion analysis of the upstream regulatory region, we have determined that aqp-8 expression is modulated by at least one cis-regulatory element located within the 19-bp interval spanning −279 to −261bp relative to the translational start site of aqp-8.

Phylogenetic Footprinting of the aqp-8 Gene Upstream Region—Although the morphologies of the two nematode species, C. briggsae and C. elegans, are similar, analysis of the mutation rates in gene ortholog pairs have revealed that the two species have diverged ~80–110 million years ago (34). Because their general body plan and developmental programs have changed very little over the long evolutionary distance, most DNA coding regions and functional non-coding DNA elements are likely to be under purifying selection. With the availability of both of their genomic sequences, we can compare orthologous non-coding genomic regions to identify conserved functional nucleotide regions. In addition to the extensively curated C. elegans (14) and annotated C. briggsae (34) genome sequences, the recent availability of two other closely related nematode genome sequences, C. remanei and C. brenneri, allows for a multiple alignment of the four species’ orthologous upstream regions. A ClustalX alignment revealed a perfectly conserved 10-bp region between the four nematode species (AATTTGCATA) that falls within the region (−279 bp to −261 bp), in C. elegans, determined by the upstream regulatory region-deletional analysis (Fig. 3). The distances between the start of the motif and the translational start sites were fairly well conserved
with the positions in *C. remanei* (−277 bp) and *C. briggsae* (−283 bp). The similarity in the upstream distance of the motifs indicates that the position of the element relative to the translational start site may be important for the ability of the element to modulate gene expression. Positional preference of cis-regulatory elements upstream of the gene translational start sites in *C. elegans* has been observed with the X-box, E-box, SMAD, and CdxA (36) transcription factor binding motifs.

**Determination of Conjugate Transcription Factors**—To determine whether there were any previously defined DNA motifs from other organisms located within the 19-bp window defined by the transgenic constructs, the sequence spanning the bases −283 to −234 was used as a query in TESS. The search results revealed an 8-nucleotide non-palindromic sequence located at the bases spanning −268 to −261 relative to the translational start site of *aqp-8*, again falling within the window defined earlier. This site corresponded to a POU (pit, oct, and unc) transcription factor consensus DNA binding site commonly referred to as the octamer motif (ATTTGCAT).

The POU homeobox transcription factors have a C-terminal bipartite DNA-binding region that consists of a POU-specific (POUs) domain, a flexible linker region, and a downstream homeobox (POUh) domain (37). The POUs region associates with the 5'-half of the DNA motif, whereas the POUh associates with the 3'-half (38). Several POU transcription factors have been shown to regulate important developmental processes in the vertebrate embryo. The entire *C. elegans* genome encodes only three POU transcription factors, *unc-86*, *ceh-6*, and *ceh-18* (14). Expression pattern analyses and functional characterizations of each *C. elegans* POU transcription factor member have been carried out in previous studies. One POU transcription factor, *UNC-86*, is a nuclear protein that plays a role in neuronal development (39). *UNC-86* has been shown to be expressed predominantly in mechanosensory, odorsensory, and chemosensory neurons where it controls neuroblast specification (40). The consensus *UNC-86* binding site sequence in *C. elegans* is CATnnnT/AAAT, which is identical to the binding site of its mammalian ortholog Brn3 (41). Another POU transcription factor, *CEH-18*, is required for directing proper gonadal sheath development and function. Loss of *CEH-18* leads to defective oocyte maturation (42, 43). Among the three *C. elegans* POU transcription factors, *CEH-6*, a class III POU, is the only one that is expressed in the excretory cell. *CEH-6* has been shown to be expressed early in the development of the
excretory cell in addition to four pairs of head neurons, the SABV motorneuron cell pair (SABVL and SABVR) and several cells in the vulva and tail (44). The expression of ceh-6 not only overlaps the expression of aqp-8 but also precedes the expression of aqp-8, thus fulfilling the spatial and temporal expression pattern criteria as a modulator of aqp-8 expression. Null mutants of ceh-6 and post-transcriptional disruption by RNAi of ceh-6 lead to phenotypes that resemble mutants with impaired development of the excretory cell (33, 45). Two additional aqp-8::GFP fusion constructs were generated with the 5′-end of the construct terminating at 279 (containing the whole octamer site within the construct) and another with the 5′-end of the construct terminating at 261 (containing the whole octamer site within the construct) to confirm the function of the site. Although the −279 construct drove expression of the GFP in the excretory cell, the −261 construct failed to drive expression of the GFP-coding cassette (Fig. 4A). This gives strong evidence that the site is required for recruiting trans-acting factors to mediate gene expression in the excretory cell. To further characterize this site, we constructed a promoter::GFP fusion construct containing mutated octamer sites. The first mutagenized construct consisted of a G → A change at position −264 (Fig. 4, A and B). The single residue change did not lead to a change in the GFP expression level. Previous studies have shown that the site, ATTTACAT, and/or its reverse complement, ATGTAAAT, are functional POU transcription factor binding sites (46, 47). Changing the adjacent downstream residue in the octamer site (ATTTGAT → ATTTGGAT) lead to a complete loss of GFP expression (Fig. 4, A and B). Crystallographic studies of the Oct1 POU domain bound to an octamer motif containing DNA strand have shown that the −263 G → C change affects a DNA binding site amino acid in the POU region of the transcription factor (38). The POU homeobox DNA-interacting amino acid residues, which contact this region of the DNA octamer motif, are highly conserved among POU transcription factor homologs in both mammals and C. elegans (data not shown).

Expression Specificity of the POU TF Binding Site—Because POU binding sites have been shown to mediate expression in a variety of vertebrate cells, we tested the ability of the octamer motif to drive expression of the GFP reporter independently of other cis-linked downstream factors associated with aqp-8. To this end, we used the ∆pes-10::GFP cassette (pPD97.78, kindly provided by A. Z. Fire, Stanford University School of Medicine). The ∆pes-10::GFP cassette is composed of the minimal promoter from C. elegans pes-10 fused to a GFP reporter. Alone, the ∆pes-10::GFP reporter construct has minimal transcriptional activity. The minimal promoter can be activated by the
presence of upstream enhancers for the purpose of determining the transcriptional activities of the introduced cis-linked elements. Using the Δpes-10::GFP construct, we tested for the ability of the putative cis-regulatory element to act as an excretory system enhancer. We fused four tandem repeats of the 10-bp nematode conserved sequence (AATTTGCATA) to the 5′-end of the Δpes-10::GFP cassette (Fig. 4C). The resulting GFP fluorescence, driven by the tandem repeats fused to the basal promoter, was observed in the excretory cell beginning at L1 and continuing into adulthood much like the expression pattern of the aqp-8::GFP constructs, albeit at a much lower level than the initial GFP-expressing constructs (Figs. 4C and 5). We did not detect expression in the additional cell identified earlier as possibly the excretory gland cell. This may be due to expression of

FIGURE 4. Validation of the putative cis-regulatory element. A, the ability of the promoter fragment to drive GFP expression in the C. elegans excretory system was abrogated when the terminal nucleotide of the octamer motif was excluded from the transgene construct. B, A → 264 G → A mutation in the octamer site leads to an expression pattern identical to the original 1.6-kb aqp-8 promoter fragment, however, upon mutating the adjacent upstream residue in the octamer site (-263 C → G) led to complete loss of GFP expression (images not shown). C, a heterologous reporter construct consisting of four tandem repeats of (AATTTGCATA) fused to Δpes-10::GFP was sufficient to drive expression in an excretory cell-specific manner. D, the Δpes-10::GFP construct expresses at a low level in the excretory cell in addition to the AUA and the AVH neuronal cell pairs.
Transcriptional Regulation of AQP-8 by CEH-6

GFP in the additional cell being below the detection level of the microscope configuration used or that expression in the gland cell is controlled by a separate cis-regulatory element. Additional GFP fluorescence arising from this construct was detected in two anterior neurons. The expression in the anterior neurons indicates that the octamer motif may also be responsible for recruiting transcription factor(s) responsible for driving expression in those neurons. The lower excretory cell expression level can be explained by the possibility of additional expression enhancing cis-regulatory elements, which exist downstream of the octamer element that were not included in the sequence fused to the Δpes-10::GFP cassette. Another possible explanation is that the expression level may be dependent upon the distance between the cis-regulatory element and the translational start site.

Binding of a Nuclear Protein to Motif Fragment—Previous studies have shown that the Brn3, a POU homebox transcription factor, complexes with its cognate DNA recognition motif stably even in non-physiological conditions (48). This property allows for the complex to be resolvable by in vitro binding reactions. To investigate whether the putative POU site is able to bind C. elegans nuclear proteins, an EMSA was conducted using the complementary biotinylated 25-bp oligonucleotides containing the octamer site along with flanking sequence as probes. The probes, when incubated with C. elegans mixed-stage nuclear protein extract, led to band shifts when run on a non-denaturating acrylamide gel. To determine if the binding was specific, unlabeled competitor oligonucleotides with the identical sequences were co-incubated with the biotinylated probes in separate binding reactions. The presence of a 1000-fold excess of identical unlabeled probe led to a decrease in the amount of protein bound to the biotinylated probe (Fig. 6, A and B). Band shifts were also observed in reactions using cytosolic extracts. Although the shifted band appeared to be the same size, implying that the bound protein is most likely the same, the amount of DNA recruited in the cytosolic fraction appeared to be significantly more than in reactions with nuclear extract. The presence of a 1000-fold excess probe in the cytosolic fraction also leads to a decrease in the amount of protein bound by the biotinylated probe (Fig. 6, A and B). We also performed an EMSA reaction containing nuclear protein extracts in the presence of the same octamer-containing biotinylated oligonucleotides in addition to a 250-fold excess of an unrelated dsDNA probe (5′-TTTGTCCCTCCTGGGAGACACAT annealed to its complementary sequence, 3′-ATGTGTCTCAGACGAGGACAAA-5′). The presence of the excess unrelated dsDNA did not affect the octamer-site probe/nuclear protein binding interaction (data not shown). These results suggest that the element and some flanking sequence are sufficient to recruit a trans-acting factor that is present in both the cytoplasm and the nucleus to the DNA sequence.

Confirmation of the CEH-6-Octamer Interaction—To verify the dependence of aqp-8 transcription on CEH-6, we knocked down CEH-6 in an AQP-8::GFP-expressing background using RNAi (49). We performed a double RNAi experiment using both aqp-8 and eri-1 dsRNA. eri-1 encodes an small interfering RNase, which expresses in C. elegans gonadal and nervous tissue. Knocking down ERI-1 leads to a pronounced RNAi effect in the tissues, in which ERI-1 is expressed (50). Treatment of the AQP-8::GFP-expressing worms with eri-1 dsRNA (n = 30) failed to down-regulate AQP-8::GFP expression in any worms, however, the double RNAi treatment of AQP-8::GFP-expressing worms with both eri-1 and ceh-6 dsRNA (n = 30) led to a consistent complete elimination of GFP expression in the excretory cell when scored 48 h post-injection (Fig. 5). The double dsRNA treatment led to developmental arrest at the L2 stage larva as a result of knocking down ceh-6 expression. This phenotype is consistent with the phenotype of the ceh-6(md60) null mutant showing that the double dsRNA treatment is effectively knocking down ceh-6 expression. Developmental arrest was not observed for worms injected with eri-1 dsRNA alone (scored at 72 h post treatment, data not shown). Taken together, we show that CEH-6 is the POU transcription factor that regulates aqp-8 via binding to its cognate octameric POU homeobox transcription factor binding site.

Determination of Other Candidate Genes Modulated by CEH-6—With the confirmed interaction of CEH-6 with the octamer element, we searched for instances in which the octamer motif was conserved between these three nematode species: C. elegans, C. briggsae, and C. remanei 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. The other criteria specific for each set are summarized in Table 2. 107 genes were identified with perfect motif matches among the three genomes under the most relaxed condition, and 44 genes were identified under the strictest condition (supplemental Table S2). Of the candidate genes identified, promoter::GFP

FIGURE 5. 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 h post dsRNA injection. Camera conditions and exposure times were the same for all images.
Transcriptional Regulation of AQP-8 by CEH-6

AQP-8, but we have discovered that expression is found in another cell, which we presume to be the excretory gland cell. Moreover, using a reporter protein with a limited half-life, we have determined a precise temporal window of expression for aqp-8. Here we present a novel model of transcriptional regulation, which is the interaction of the POU homeobox transcription factor, CEH-6, with an octamer motif in the upstream regulatory region of aqp-8.

The binucleate excretory gland cell is the product of a cell fusion. Although the gland cell has been proposed to function as a secretory organ, based on its morphology, it is the only non-vital cell in the excretory system. Laser ablation of the gland cell does not lead to any obvious shortcomings in worm development or function under standard laboratory conditions (5). Clues as to the excretory gland cell's role in the worm can be inferred by virtue of its possible co-developmental regulation with the excretory cell.

Expression pattern analysis using GFP-based reporters generally agreed with SAGE profiles derived from either staged worms or fluorescence-activated cell sorting-based isolation of certain cell types. Surprisingly, aqp-8 message was also detected in AFD-specific libraries (29). Analyses of mutants and cell ablations of the AFDs have determined that the ciliated neurons AFDL and AFDR are required for thermosensation in C. elegans (52, 53). The developmental lineages of the excretory cell and the AFD neurons diverge at the four-cell stage, a stage that has not been used for deriving the AFD SAGE library (54). A possible reason for the SAGE tag arising in the AFD-specific library is that the tag might represent message derived from background foreign cell contamination during the fluorescence-activated cell sorting stage.

By assaying for the expression activity of progressive deletions of the aqp-8 upstream regulatory region, we have delimited a region important for the expression of aqp-8. We used sequenced closely related rhabditid genomes to facilitate interspecies comparisons of the upstream regulatory region of aqp-8. The alignments, in conjunction with the window derived from the promoter truncation analysis, allowed us to identify a single cis-regulatory element required for expression of AQP-8. The cis-regulatory element corresponds to the octamer motif, a DNA sequence known to recruit POU homeobox transcription factors for activation of downstream genes. We demonstrate that the transcription factor responsible for aqp-8 excretory cell expression is CEH-6 using a double RNAi strategy, which enhances the RNAi effects in certain tissue types. We find that CEH-6 is found in both the cytosolic and nuclear protein extracts. The presence of a cytosolic binding partner for the

FIGURE 6. Electrophoretic mobility shift assay. The DNA fragment consisting of 5’-ATTGCCAAAATTTTGCATACTGGAAAT-3’ and its complement 5’-ATCCAGTATCGAAATTTTGCATAT-3’ were incubated with C. elegans cytosolic and nuclear protein extracts in vitro. A, the labeled probe was able to bind nuclear proteins in vitro. A greater degree of protein binding was observed when the probe was incubated with cytosolic protein extracts. The specificity of the octamer motif for the nuclear binding protein was determined by competitive binding of 1000× molar excess of unlabeled competition probe. In the presence of excess unlabeled probe, loss of protein binding to the labeled probe was observed in both the nuclear protein fraction. B, the specificity of the octamer motif for the cytosolic binding protein was determined by competitive binding of 1000× molar excess of unlabeled competition probe. In the presence of excess unlabeled probe, loss of protein binding to the labeled probe was observed in the cytosolic protein fraction. The dividing line (white) between the cytosolic and cytosolic + 1000 XS lanes indicates that two lanes are from different parts of the same gel.

eexpression pattern data have been generated for 19 (relaxed condition; all) and ten (strictest condition; S.E.) of the upstream regulatory regions using promoter::GFP reporter constructs, respectively (29) (Table 2). Three genes that contain excretory cell expression are consistently observed for all gene sets (Table 3). 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 2). Of the 28 gene candidates identified with the least stringent conditions (S.E.) (Table 3), 17 genes have associated gene ontology terms (supplemental Table S3) and thus provide a starting point for determining the physiological role of AQP-8 and its orthologs.

DISCUSSION

Previous work has shown that ~12% of C. elegans excretory expressing genes may be transcriptionally regulated by the EX-1 cis-regulatory motif in conjunction with its cognate transcription factor, DCP-66 (51). DCP-66 is a widely expressed transcription factor and has been shown to be expressed in neurons, the pharynx, body wall muscle, excretory cell, and vulva (51). Close homologs of DCP-66 generally act as transcriptional repressors, although in the case presented by Zhao et al. (51), DCP-66 clearly acted as a positive regulator of gene expression. In addition, Zhao et al. (51) reported two other cis-regulatory elements that mediate gene expression in the excretory cell.

Although the expression of AQP-8 is tightly regulated in the excretory system, all of the transcriptional regulatory elements that contribute to its expression pattern have not been elucidated. Previous studies have reported the expression pattern of
octamer site is not surprising due to the fact that POU homeobox transcription factors have highly conserved basic nuclear localization and leucine-rich hydrophobic nuclear export signals (55, 56). An alignment of the human POU homeobox TFs, Oct6 and Brn1, against all three C. elegans members reveals that the two localization signal sequences are conserved in the nematode POU proteins (data not shown). To verify the cytosolic localization of CEH-6, we assayed for the expression pattern of a CEH-6 transglutaminase fusion construct. The construct revealed that the protein is intracellularly localized both to the nucleus and the cytoplasm of the excretory cell (data not shown). The nuclear export signal found in POU homeobox transcription factors have highly conserved basic nuclear localization and leucine-rich hydrophobic nuclear export signals (55, 56). An alignment of the human POU homeobox TFs, Oct6 and Brn1, against all three C. elegans members reveals that the two localization signal sequences are conserved in the nematode POU proteins (data not shown). To verify the cytosolic localization of CEH-6, we assayed for the expression pattern of a CEH-6 transglutaminase fusion construct. The construct revealed that the protein is intracellularly localized both to the nucleus and the cytoplasm of the excretory cell (data not shown). The nuclear export signal found in POU homeobox TFs has been demonstrated to act in a CRM1/Exp1-dependent manner. 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::GFP was too weak to be detected, and therefore we could not determine if CEH-6 and IMB-4 are cellu- 
larly co-localized. The EMSA results also indicate a greater abundance of the DNA-interacting protein in the cytoplasm than in the nucleus. This may indicate a nuclear export rate for CEH-6 that exceeds its nuclear import rate. The existence of these localization signals in CEH-6 likely facilitates rapid transient transcriptional modulation of target genes via nucleocytoplasmic shuttling with the cytosol acting as a repository.

Analysis of the expression pattern derived from the tandem POU motif repeat fused to the $\Delta p$es-10::GFP reporter revealed that the motif also drove expression in the AUA (AUAL and AUAR) and AVH (AVHL and AVHR) neuron pairs (Fig. 4D). The AUA neurons are involved with integrating environmental cues to dictate social versus solitary feeding choices (57).

Although AQP-8 was not detected in the AUA neurons, AQP-8 may play a role in integrating osmoregulatory cues in conjunction with these cells. The function of the AVH neurons is unknown, but it appears that both the AUA-type and AVH-type neurons co-express the glutamate receptor GLR-4 and a tyramine receptor SER-2 (58). The motif repeats fused to the minimal reporter was sufficient to drive excretory cell expression of the reporter in vivo, albeit at a lower expression level than that of the aqp-8 promoter-reporter constructs, suggesting that there are other elements downstream of the octamer element that are important for fine-tuning the levels of mRNA production but which themselves do not necessarily modulate the spatial pattern of AQP-8. This leads to a combinatorial model for gene expression of aqp-8. This is supported by multiple alignments of the region between the POU

### TABLE 2

| 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                                     | 109                        | 19                                            | 3                                        | 0.3556      |
| S     | POU motifs are on the same strand as the downstream gene   | 54                         | 13                                            | 3                                        | 0.1662      |
| E     | POU motifs are not overlapping with any upstream gene      | 83                         | 16                                            | 3                                        | 0.2623      |
| SE    | POU motifs are on the same strand as the downstream gene and not overlapping with any upstream gene | 44                         | 10                                            | 3                                        | 0.0857      |

### TABLE 3

| Sequence name | Gene name | Strain | Transgene ID | Expression pattern |
|---------------|-----------|--------|--------------|--------------------|
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | 1st and 2nd pharyngeal bulbs; anterior neuron; posterior neuron; muscle; intestine; vulva |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Nerve cord; anterior neuron; posterior neuron; amphts; nerve ring |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Intestine; excretory cell; anterior neuron; posterior neuron; vulva; hypoderm |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | 1st and 2nd pharyngeal bulbs; vulva; anterior neuron |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Hypoderm; seam cell |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Intestine          |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Pharynx; muscle; anterior neuron; posterior neuron; vulva |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | 1st and 2nd pharyngeal bulbs; anterior neuron; posterior neuron; nerve ring; pharynx |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Excretory cell     |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Anterior neuron; nerve cord; excretory cell |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Pharynx; muscle    |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | 1st and 2nd pharyngeal bulbs; nerve cord; nerve ring |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Anterior neuron    |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | 1st and 2nd pharyngeal bulbs; anterior neuron; posterior neuron; nerve ring |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Intestine;1st and 2nd pharyngeal bulbs |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Hypoderm; seam cell |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Intestine; muscle; neuron |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | Nerve ring; posterior neuron |
| F39H11.3      | cdk-8     | BC13856| sEX13856     | No expression      |

The expression pattern data was taken from The Genome BC C. elegans Gene Expression Consortium webpage.
motif and the translational start site of four nematode species, which show seven blocks of perfect conservation for sequences greater than six residues (Fig. 3). Another possible explanation for the lower observed GFP expression level as a result of the Δpse-10::GFP-based construct is that the motif may have an optimal effectiveness at a specific distance relative to the translational start site. Further studies should be carried out to determine which if not both of these hypotheses are applicable in this situation. An alternative, osmotic balance-controlled model of gene regulation has been also suggested for aqp-8 (15). It is possible that some of the other conserved regions upstream of aqp-8 may be responsible for this aspect of its regulation.

The vertebrate ortholog of CEH-6 is the class III POU homeobox protein, Brn1. Members of the class III POU transcription factors play important roles in the development of the nervous system (59). Zebrafish Brn1 has been localized via whole mount in situ expression patterning to neuronal tissue (60). Brn1 expression has also been detected in the gastrointestinal tract of embryonic sea urchins (61). The Brn1 ortholog in quail has been observed, by whole mount in situ hybridization, to be localized in neuronal tissue and in the mesodermal sections of the developing kidney in 5-day-old embryonic quail. In addition, it has been detected in as early as 2-day-old embryonic tissue sections by Northern blot analysis (62). Homozygous mBrn1-deficient mice die within 24 h of birth due to renal complications. Dissection of 2-h-old mice revealed that mBrn1−/− mice had significantly lower volumes of urine compared with their wild-type counterparts. mBrn-1 was observed to be localized to the macula densa, the distal convoluted tubule, and the Loop of Henle. Closer inspection of these tissues revealed a shortened loop of Henle and suppressed differentiation in all three tissues (63). The CEH-6 ortholog in the crustacean Artemia franciscana, APH-1, is expressed in the salt gland, which like the C. elegans excretory cell and mammalian kidney, is an osmoregulatory organ. Reverse transcription-PCR analysis of APH-1 reveals that, like CEH-6, the transcription factor is expressed predominantly during development (64). Because the excretory cell phenotype of ceh-6(mg60) manifests early in development (44), and due to the general role of POU homeobox TFs in modulating gene expression in early development, we presume that many of the genes transcriptionally regulated by CEH-6 are required for morphogenesis of renal and neuronal tissues.

To determine genes that may be co-regulated with aqp-8, we searched for genes in which the octamer motif was perfectly conserved in the upstream region of three nematode species (C. elegans, C. briggsae, and C. remanei), and, using publicly available expression pattern data (29), we determined the frequency that the motif arises in the upstream region of excretory cell-expressing genes. We did not observe a high level of significance when determining whether these genes were more likely than not to be expressed in the excretory cell. The low significance was likely caused by the lack of expression data for many of the genes predicted providing a small sample size. We expect that the significance level of the data would increase if the expression pattern for a larger group of the bioinformatically predicted candidates was available. Upon pattern analysis of other genes in the most relaxed set (Table 2, all) we found 10 of the 19 expressing genes show expression in neuronal cells, a tissue that also expresses CEH-6 (Table 2).

Studies pertaining to transcriptional regulation in vertebrates can be difficult due to the lack of sequenced genomes, the tissue and physiological complexity of the systems, and problems with determining complete expression patterns due to long developmental time courses. The intergenic spacing in the C. elegans genome is relatively compact; therefore, studies of long range regulation are usually not required for the identification of single-gene cis-regulatory elements, although long range studies may identify islands of co-regulated gene clusters due to factors such as higher order chromatin structure. In addition, many studies rely on expression profile correlations and/or determining over-representative motifs in the promoter-containing regions using whole genome approaches. One of the problems of these expression pattern correlation studies is that tissue co-expression does not always imply gene co-regulation as we have shown in our study.

Although the experimentally identified octamer sequence was perfectly conserved and functional in the upstream regulatory region of aqp-8, the octamer motif did not necessarily drive excretory cell expression. Due to these results, we have concluded that the octamer motif, although probably a functional DNA region in the many cases in which it is perfectly conserved between nematode species, is not sufficient in all cases to drive expression in the excretory cell. We intend to study the expression pattern of the other candidate promoter regions to develop a better understanding of which tissues and at what frequencies the octamer motif modulates expression.

Because this work and the previous study by Zhao et al. (51) were not exhaustive searches for cis-regulatory elements that modulate gene expression in the excretory cell, there are still other transcription factor binding sites that affect excretory cell expression. In this study, we have revealed a conserved relationship between a transcription factor and its cognate DNA binding locus, which is relevant to both renal and neuronal development in nematodes and in other higher organisms.

Acknowledgments—We thank the members of our laboratories (D. L. B., H. M. C., and N. S. C.) for productive discussions pertaining to this project. We also thank the Mitani Laboratory for kindly providing the strain TM1919.

REFERENCES

1. Nelson, F. K., Albert, P. S., and Riddle, D. L. (1983) J. Ultrastruct. Res. 82, 156–171
2. Buechner, M., Hall, D. H., Bhatt, H., and Hedgecock, E. M. (1999) Dev. Biol. 214, 227–241
3. Buechner, M. (2002) Trends Cell Biol. 12, 479–484
4. Hedgecock, E. M., Culotti, J. G., Hall, D. H., and Stern, B. D. (1987) Development 100, 365–382
5. Nelson, F. K., and Riddle, D. L. (1984) J. Exp. Zool. 231, 45–56
6. Gazzarrini, S., Kang, M., Epimashko, S., Van Ettten, J. L., Dainty, J., Thiel, G., and Moroni, A. (2006) Proc. Natl. Acad. Sci. U. S. A. 103, 5355–5360
7. Denker, B. M., Smith, B. L., Kuhajda, F. P., and Agre, P. (1988) J. Biol. Chem. 263, 15634–15642
8. Agre, P., Preston, G. M., Smith, B. L., Jung, S., Raina, S., Moon, C., Guggino, W. B., and Nielsen, S. (1993) Am. J. Physiol. 265, F463–F476
9. de Groot, B. L., and Grubmuller, H. (2001) Science 294, 2353–2357
10. Ishibashi, K., Sasaki, S., Fushimi, K., Uchida, S., Kuwahara, M., Saito, H.,
