Dissection of cis-regulatory elements in the C. elegans Hox gene egl-5 promoter

Yingqi Teng\textsuperscript{a}, Lisa Girard\textsuperscript{b}, Henrique B. Ferreira\textsuperscript{a,1}, Paul W. Sternberg\textsuperscript{b}, Scott W. Emmons\textsuperscript{a,}\textsuperscript{*}

\textsuperscript{a}Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, NY 10461, United States
\textsuperscript{b}HHMI/Division of Biology, 156-29, California Institute of Technology, Pasadena, CA 91125, United States

Received for publication 12 July 2004, revised 7 September 2004, accepted 8 September 2004

Abstract

Hox genes are highly conserved segmental identity genes well known for their complex expression patterns and divergent targets. Here we present an analysis of cis-regulatory elements in the Caenorhabditis elegans Hox gene egl-5, which is expressed in multiple tissues in the posterior region of the nematode. We have utilized phylogenetic footprinting to efficiently identify cis-regulatory elements and have characterized these with gfp reporters and tissue-specific rescue experiments. We have found that the complex expression pattern of egl-5 is the cumulative result of the activities of multiple tissue or local region-specific activator sequences that are conserved both in sequence and near-perfect order in the related nematode Caenorhabditis briggsae. Two conserved regulatory blocks analyzed in detail contain multiple sites for both positively and negatively acting factors. One of these regions may promote activation of egl-5 in certain cells via the Wnt pathway. Positively acting regions are repressed in inappropriate tissues by additional negative pathways acting at other sites within the promoter. Our analysis has allowed us to implicate several new regulatory factors significant to the control of egl-5 expression.

© 2004 Elsevier Inc. All rights reserved.

Keywords: cis-Regulatory elements; C. elegans; egl-5; Hox; Phylogenetic footprinting

Introduction

The Hox genes encode a highly conserved set of transcription factors that are essential for specifying patterns along the anterior–posterior axis of all metazoans studied. Consistent with their role in establishing regional identity, the Hox genes are expressed in a series of consecutive domains along this axis. Mutations in these genes usually result in transformation phenotypes in which body regions are transformed into the identity of adjacent ones. Hox genes are situated in clusters, with the order of the genes in the cluster generally colinear with their order of expression along the anterior–posterior axis. In Caenorhabditis elegans, a quasi Hox gene cluster contains a homolog of the anterior group genes ceh-13, which is most similar to labial in the Drosophila HOM-C complex; two genes homologous to medial-group genes lin-39 (sex-combs reduced, deformed, proboscipedia), mab-5 (fushi tarazu), and a homolog of the posterior group egl-5 (Abdominal-B) (for the assignment of C. elegans Hox genes to orthologous groups, see Aboobaker and Blaxter, 2003) (Fig. 1A). Additionally, two more posterior homologs not contained within the quasicluster, nob-1 and php-3, have been characterized (Van Auken et al., 2000).

Early studies of C. elegans Hox genes have emphasized their unique features, for example, the establishment of Hox gene expression pattern by lineal cues (Cowing and Kenyon, 1996). However, despite the unique, lineal way by which the C. elegans body plan is established, common mechanisms have been found that are shared by C. elegans and other organisms, including autoregulation (Streit et al., 2002).
genetic interactions between Hox genes (Chow and Emmons, 1994; Ferreira et al., 1999; Salser et al., 1993), and the presence of a global repression mechanism (Ross and Zarkower, 2003; Zhang et al., 2003). To gain further insight into the regulation of \textit{C. elegans} Hox genes, we have carried out a promoter analysis of the \textit{C. elegans} Hox gene \textit{egl-5}.

\textit{egl-5} is involved in posterior patterning. It is expressed in a group of cells that, during late embryogenesis, are situated around the future tail region (Ferreira et al., 1999). These cells include the rectal epithelial cells K, F, B, U, and Y, the most posterior neuroectoblast cell P12, body wall muscle cells, the PLM mechanosensory neurons and PVC interneurons, and the hermaphrodite-specific neuron (HSN). In males, \textit{egl-5} is also expressed in the sex-specific mesoblast cell M, cells in the male gonad, and also in the V6 seam cell lineage, which gives rise to a subset of rays. Defects seen in \textit{egl-5} mutants are consistent with this expression pattern (Chisholm, 1991). In \textit{egl-5} hermaphrodites, the fate of rectal epithelial cells is misspecified, Y fails to differentiate into PDA, K fails to divide, and DVB is not formed. Additionally, HSN defects result in an “Egl-defective egg-laying phenotype. Differentiation of the neuroectoblast cell P12 is also defective, resulting in the transformation of P12 to its anterior neighbor P11. In \textit{egl-5} males, sensory rays are abnormal due to cell cell defects, and the male mating structures, the spicules, gubernaculum, and the proctodeal chamber, are missing or largely abnormal due to defects in B, Y, U, and F lineages. \textit{egl-5} mutants also demonstrate mechanosensory defects as insensitivity to both light touch and heavy touch, which is probably due to the defects of PLM and PVC neurons.

Expression of \textit{egl-5} is under the control of distinct regulatory pathways in different cell types. In P12, genetic evidence suggests that \textit{egl-5} expression is controlled by both the Wnt and the EGF pathways, which function synergistically through \textit{egl-5} to specify P12 fate (Jiang and Stemberg, 1998). In specification of the male rays, \textit{mab-5} appears to be involved (Ferreira et al., 1999). In the T lineage that gives rise to the touch cell PLM, \textit{egl-5} seems to be under the negative regulation of \textit{spalr}-like gene \textit{sem-4} (Toker et al., 2003).

Here, we present an analysis of the \textit{egl-5} promoter. Our analysis was facilitated by the method of phylogenetic
footprinting. *Caenorhabditis elegans* and the related nematode species *Caenorhabditis briggsae* diverged an estimated 50–120 million years ago (Coghan and Wolfe, 2002; Stein et al., 2003), and the DNA sequences of both genomes are available. Sequence comparisons between these two species have already revealed *cis*-regulatory regions for a number of genes (for examples, see Cui and Han, 2003; Gilleard et al., 1997; Kirouac and Sternberg, 2003; Krause et al., 1994; Xue et al., 1992). We describe informative comparisons between *C. elegans* and *C. briggsae* promoter sequences that enabled us to identify the *cis*-regulatory elements for individual tissues. We investigated the functions of these regulatory elements with reporters and tissue-specific rescue experiments. We also present detailed dissections of two tissue-specific enhancer regions within the egl-5 promoter and provide evidence for a previously unknown activation pathway that operates in the most posterior branch of the V6 lineage.

### Materials and methods

#### Nematode strains and cultures

Nematodes were cultured as described in Brenner (1974). Bristol (N2) and CB4088 (*him-5(e1490)V*), which has increased frequency of the male population, are used as reference wild-type strains. Strains were maintained at 20–22°C. Reporter-bearing strains containing the temperature sensitive mutation *pha-1(e2123)* III as selectable marker were maintained at 25°C. The following mutations were used: LG III: *pal-1(e2091)*, *mab-5(e2039)*, egl-5*(u202)*, *pes-10* (gift from Dr. Morris Maduro) animals.

#### Construction of egl-5::gfp reporters EM#285 and EM#286

Clone EM#285 contains the 21.5-kb NruI fragment of cosmid C08C3 (Wang et al., 1993) with gfp inserted into the homeodomain of *egl-5*, rendering the gene inactive (for details, see Toker et al., 2003). It contains cosmid C08C3 nucleotides 23,448–43,981 and includes 16,027 nucleotides of 5' upstream sequence. This reporter was co-injected with pJM23 (Huang et al., 1994) into strain *him-5(e1490)*; *lin-15(n765)*. The extrachromosomal array was subsequently integrated to generate bxIs12.

EM#286 was constructed by introducing restriction sites into PCR primers. About 7 kb 5' region, coding region, and 3' UTR of egl-5 was cloned into pBluescript (Stratagene). GFP (pPD102.33, A. Fire laboratory vector, 1999) was inserted near the C terminal of EGL-5. This reporter covers cosmid C08C3 nucleotides 32,429–43,981 and includes 7046 nucleotides of upstream sequence. It failed to rescue an egl-5 mutant despite the intention to make it a functional reporter. EM#286 was co-injected with pJM23 into strain *him-5(e1490); lin-15(n765)*. The extrachromosomal array was subsequently integrated to generate bxIs13.

Generation of egl-5 promoter dissection constructs and transgenic animals

#### Transgenic arrays with bx numbers

The egl-5::gfp reporters used in the promoter dissection study were generated by PCR using EM#285 as template. The *Apes-10:*::gfp reporters were PCR products using a PCR-fusion-based method (Hobert, 2002). The *Apes-10:*::gfp and unc-54 3'UTR regions were from pPD122.53 (A. Fire laboratory vector, 1999). The point mutations in the putative TRA-1A site (in *bxEx90*) were introduced with mismatches in the PCR primer. Primer sequences are available upon request.

Transgenic nematodes were generated by microinjection as described by Mello et al. (1991). Unless otherwise stated, PCR products (10 ng/μl) are co-injected with pBX1 (100 ng/μl) (Granato et al., 1994) into strain *pha-1(e2123ts); him-5(e1490)*. Animals with transgenes were maintained at 25°C. Transgenes were introduced into different mutant strains by crossing the wild-type transgenic animals to strains that contain the *pha-1(e2123)* mutation in the background in addition to the desired mutations. Transgenic *tra-1(e1099)* animals were generated by injecting PCR products with pRF4 (100 ng/μl) (Mello et al., 1991) into strain CB2590: *tra-1(e1099)/dpy-18(e1096)*, and selecting for roller males.


#### Transgenic arrays with sy numbers

Specified regions of the egl-5 promoter were amplified using PCR. The amplified products were cloned into the pes-10 minimal promoter (pPD 107.94, A. Fire laboratory vector, 1997) using restriction sites on the ends of the PCR primers. The template used for these constructs was EM#285. For the creation of deletion constructs, the constructs were first PCR amplified in two pieces with one primer spanning the deletion site and containing the desired deletion in its sequence. The two products obtained in these reactions were then gel purified and used as the template in a third PCR reaction utilizing primers that amplified from the ends of the two pieces, resulting in a single product containing the specified deletion. Primer sequences are available upon request.

Promoter constructs were microinjected into the gonads of *unc-119(ed4)* (gift from Dr. Morris Maduro) animals.
(Mello et al., 1991) at a concentration of 100 ng/μl with 50–100 ng/μl of PvuII-linearized genomic DNA and 50 ng/μl unc-119 rescuing plasmid pDP#MM016B (gift from Dr. Morris Maduro). Animals stably transmitting the array were selected by isolating non-unc F1’s and examining their progeny for transmission.


Microscopy of transgenic animals

Animals from a minimum of two lines per construct were mounted on 5% agarose/noble agar pads and examined under Nomarski optics at ×400 or ×1000. Fluorescence was observed at ×1000 with a Zeiss 487905 filter set/a Chroma High Q GFP LP filter set.

Results

Sequence comparison of the intergenic region around egl-5 in C. elegans and C. briggsae reveals multiple high-scoring pairs

We took the C. elegans and C. briggsae egl-5 genomic sequence as the input of a NCBI Blast 2 Searches sequence using the default parameters. We found that the coding sequences of egl-5 are well conserved, while, as expected, the noncoding sequences have generally diverged. However, within the 5’ and 3’ regions of the egl-5 gene, there are sequences (22–66 bp) conserved between the two species. Additional shorter or less well-conserved elements were identified when the size of the input sequences was significantly shortened. We hypothesize that the conserved DNA segments represent functional sequences that govern the expression pattern of egl-5. As a starting framework, we numbered the longest sequences 1–15 from left to right (Table 1).

We have confirmed our finding with an additional algorithm FamilyRelations/SeqComp (Brown et al., 2002) (Fig. 1B). This program identified all the conserved sequences we identified by NCBI Blast search. Furthermore, it revealed five more conserved sites around egl-5, which we termed F1–F5 (Table 1). In general, the orders and the orientations of the conserved sequences are also conserved between the two species, with the exception of sequences F4 and F5, which are positioned 3’ of the elegans egl-5 gene, yet 5’ of the briggsae egl-5 gene, and the orientation of F4 is reversed. These conserved sequences exist in clusters, for example, 7 and 8, 9 and 10, and 11 and 12 (also see Table 1). The most probable reason for this is that the blast search detected different parts of the cis-regulatory regions separately, likely due to gaps within these regions that cannot align between the two species.

Deletion analysis suggests that the conserved regions are necessary for egl-5 expression in individual tissues

To define tissue-specific cis-regulatory regions, we undertook a deletion analysis of the egl-5 promoter region using our longest egl-5::gfp reporter, EM#285, as template (Fig. 2A). This reporter carries 17 kb of DNA sequences 5’ of the egl-5 gene. Transgenic animals bearing integrated EM#285 (bxIs12) show the same expression pattern as EGL-5 antibody staining (Ferreira et al., 1999). We designed primers to successively delete the conserved sequence clusters, generating shorter egl-5::gfp reporters, and observed the resulting expression pattern changes in the transgenic lines generated with the PCR products (Fig. 2A).

Transgenic bxIs12 animals show expression in all the tissues we have studied, namely P12, rectal epithelial cells, body wall muscles, sex muscles, male tail V6 lineage, and PLM touch cells (Figs. 2B, C, and E). In addition, it is expressed in the male tail hypodermis during the late L4 larval stage (Fig. 2D), a site of egl-5 expression that was not reported previously. The late L4 larval stage is a brief period during which the tail hypodermis undergoes rapid, extensive remodeling (retraction), a process that is defective in egl-5 mutants.

Deletion of conserved sequences 1–6 (bxEx108) had no observed effect on the expression pattern. These sequences may represent cis-regulatory elements controlling the

### Table 1

<table>
<thead>
<tr>
<th>Conserved region</th>
<th>Start (C08C3 No.)</th>
<th>End (C08C3 No.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25552</td>
<td>25578</td>
</tr>
<tr>
<td>2</td>
<td>25619</td>
<td>25655</td>
</tr>
<tr>
<td>3</td>
<td>26249</td>
<td>26280</td>
</tr>
<tr>
<td>4</td>
<td>28219</td>
<td>28245</td>
</tr>
<tr>
<td>5</td>
<td>28312</td>
<td>28333</td>
</tr>
<tr>
<td>6</td>
<td>28390</td>
<td>28439</td>
</tr>
<tr>
<td>7</td>
<td>31655</td>
<td>31676</td>
</tr>
<tr>
<td>8</td>
<td>32050</td>
<td>32083</td>
</tr>
<tr>
<td>9</td>
<td>34080</td>
<td>34106</td>
</tr>
<tr>
<td>10</td>
<td>34278</td>
<td>34306</td>
</tr>
<tr>
<td>11</td>
<td>36069</td>
<td>36114</td>
</tr>
<tr>
<td>12</td>
<td>36184</td>
<td>36249</td>
</tr>
<tr>
<td>13</td>
<td>37497</td>
<td>37533</td>
</tr>
<tr>
<td>14</td>
<td>39319</td>
<td>39340</td>
</tr>
<tr>
<td>15</td>
<td>41955</td>
<td>41977</td>
</tr>
<tr>
<td>F1</td>
<td>30397</td>
<td>30415</td>
</tr>
<tr>
<td>F2</td>
<td>32791</td>
<td>32809</td>
</tr>
<tr>
<td>F3</td>
<td>37703</td>
<td>37720</td>
</tr>
<tr>
<td>F4</td>
<td>41977</td>
<td>41994</td>
</tr>
<tr>
<td>F5</td>
<td>42130</td>
<td>42149</td>
</tr>
</tbody>
</table>
expression of the adjacent gene C08C3.4. Deletion of the region containing conserved sequences F1, 7, 8, and F2 (bxIs13) results in loss of expression in P12, rectal epithelial cells, and body wall muscles. Further deletion through the region containing conserved sequences 9 and 10 (bxEx85) results in loss of expression in tail hypodermis and sex muscle. Finally, deletion through the region containing sequences 11 and 12 (bxEx87) results in loss of expression in the V6 lineage. In this shortest construct, there is still expression in touch cell PLM.

In addition to loss of expression in various tissues, these deletions also resulted in expression of the reporters in ectopic, non-EGL-5-expressing tissues, revealing the existence of negative elements, as discussed further below.
Enhancer activity assays with \(\text{Apes}-10::\text{gfp}\) reporter verify the sufficiency of the conserved regions as tissue-specific cis-regulatory elements

Having established the correlation between the conserved sequences and the tissues in which their activity is required, we then wanted to test the sufficiency of these putative regulatory elements. We used promoter regions containing these elements to drive \(\text{Apes}-10::\text{gfp}\) reporters specialized for enhancer activity assays (Fig. 2A) to ask if these elements could recapitulate the expression lost upon deletion. We tested short DNA segments around 7 and 8, 9 and 10, and 11 and 12. We also tested 13 and F3 for their ability to drive expression in PLM touch cells (conserved sequence 14 is very close to the ATG of \(\text{egl}-5\), has AT rich sequences, and is thus likely to be the basal promoter of the \(\text{egl}-5\) gene).

The expression patterns generated by the constructed reporters are summarized in Fig. 2A. We found that most of the regulatory elements tested can direct expression in the tissues suggested by analysis of the corresponding deletions. Further examination has shown that in most of the tissues the timing and the detailed patterns of these reporters also reflect the wild-type \(\text{egl}-5\) expression pattern (data not shown). One exception is conserved sequences 13 and F3 (bxEx110), which fail to drive reporter expression in the PLM touch neurons. The regulatory region of PLM may be in the 3' region of \(\text{egl}-5\). To further confirm that the conservation in DNA sequence has functional significance, we also tested the region around conserved sequences 11 and 12 from the \(\text{C. briggsae}\) \(\text{egl}-5\) promoter in a \(\text{Apes}-10::\text{gfp}\) assay (bxEx111) (Fig. 6B). It drives an identical expression pattern in the V6 lineage as the corresponding region in \(\text{C. elegans}\), suggesting that the conserved sequences in the region around 11 and 12 are sufficient to perform the regulatory function in the V6 lineage. For convenience, we have named this region around 11 and 12 V6CRE, for the V6 lineage cis-regulatory element.

Two of the apparent tissue-specific elements were chosen for further study: V6CRE (bxEx88) and the region containing conserved sequences 7 and 8 (syEx 615), which drives expression in P12, the rectal epithelium, and body wall muscles.

Tissue-specific rescue experiments provide further evidence for the independence of cis-regulatory elements

To correlate \(\text{egl}-5\) expression with function, as well as to further test the autonomy of the regulatory elements, we performed tissue-specific rescue experiments. \(\text{egl}-5\) mutant males have retraction defects in the tail region as well as sensory ray identity transformations (Fig. 3B) (Chisholm, 1991; Chow and Emmons, 1994). Although, in most cases, the correlation between \(\text{egl}-5\) expression pattern and mutant phenotype is apparent, how \(\text{egl}-5\) affects the retraction process has not yet been established. Given our observation that the tail hypodermis expression of \(\text{egl}-5\) reporters starts at late L4, right before the retraction initiates, we consider this expression of \(\text{egl}-5\), which is driven by sequence around 9 and 10, likely to be directing the process of retraction. To test this hypothesis, we drove expression of \(\text{egl}-5\) with only conserved region around 9 and 10. A small piece of DNA containing conserved sequences 9 and 10 (C08C3 34,036–34,350) was placed 5' of the putative basal promoter, coding sequence, and 3' UTR of \(\text{egl}-5\) gene (C08C3 37,621–43,928). The resulting construct (Fig. 3E, a) rescues the male tail retraction defects in an \(\text{egl}-5\) mutant (57%, \(n = 28\)), but does not rescue the male sensory ray transformation defects (0%, \(n = 26\)) (Fig. 3C). Another construct (Fig. 3E, b) with both tail hypodermis and V6 lineage regulatory elements (C08C8 34,036–44,011) gives significant rescue of both male tail retraction (53%, \(n = 15\)) and ray identity transformation (39%, \(n = 28\)) defects (Fig. 3D). These results also associate ray development rescuing activity with V6CRE. Rescue of V6 lineage defects alone by the region containing conserved sequences 11 and 12 could not be tested, because without proper retraction, it is difficult to score the identity of V6 sensory rays.

Two types of repressor elements in the native promoter prevent inappropriate activation of cis-regulatory elements

Autoregulation is an important characteristic of many Hox genes. To investigate the importance of autoregulation in defining the expression pattern of \(\text{egl}-5\), we crossed \(\text{egl}-5::\text{gfp}\) reporter bxIs13, which does not rescue an \(\text{egl}-5\) mutant, into a strain containing the null allele \(\text{egl}-5(\text{u202})\). In the V6 lineage, gfp expression was lost in the \(\text{egl}-5(\text{u202})\) background, but in other tissues, including PLM neurons, male tail hypodermis, and sex muscles, expression persists without functional EGL-5 protein (Figs. 4A–B, and data not shown). Thus, expression of bxIs13 in the V6 lineage requires \(\text{egl}-5\) gene function, suggesting that in the context of a native promoter, \(\text{egl}-5\) is autoregulatory in the V6 lineage.

This positive autoregulation of \(\text{egl}-5\) in the V6 lineage is controlled by elements outside of the region capable of driving V6 lineage expression in enhancer assays. In the wild-type background, V6CRE drives \(\text{Apes}-10::\text{gfp}\) (V6CRE::gfp) in the wild-type expression pattern of \(\text{egl}-5\) in the V6 lineage (Fig. 4C). However, unlike the translational \(\text{egl}-5::\text{gfp}\) reporters (i.e., bxIs13) that fail to be expressed in \(\text{egl}-5(\text{u202})\), V6CRE::gfp is still expressed in the \(\text{egl}-5(\text{u202})\) background (90%, \(n = 31\)) (Fig. 4D). Thus, V6CRE does not require the presence of EGL-5 for its continuous expression. We conclude that negative elements lie elsewhere in the \(\text{egl}-5\) promoter that repress V6CRE in the absence of EGL-5 function.

A second, more general kind of negative regulation is also utilized to prevent ectopic expression of \(\text{egl}-5\). In some \(\text{egl}-5::\text{gfp}\) constructs from our deletion analysis (i.e., bxIs13), we have observed ectopic expression of the gfp reporter in some head neurons (Fig. 4E), which is not a domain of expression of endogenous \(\text{egl}-5\). Similarly, we have observed instances
of ectopic expression of Δpes-10::gfp reporter constructs. For example, the region around 9 and 10 drives gfp expression not only in tail hypodermis and sex muscles, but also in many cells in the head (data not shown). V6CRE gives strong gut and head neuron expression in addition to expression in V6.ppp and its descendants (data not shown). We assume the ectopic expression of these reporters is due to loss of certain promoter sequences, partially mapped in our analysis, which can recruit repressors to prevent the assembly or function of activator complexes on cis-regulatory elements like V6CRE in inappropriate tissues.

**Defining cis-regulatory elements for P12 and rectal epithelial cells**

The rectal epithelial cells K, F, B, and U, and the most posterior neuroectoblast cell P12, are a group of positionally related cells located adjacent to the tail region. P12 undergoes a series of divisions that result in the production of three ventral cord neurons and an epidermal cell P12.p. In the hermaphrodite, the rectal epithelial cells F, B, and U do not divide, and K divides once, generating the DVB motorneuron and a rectal hypodermal cell. In the male, these cells differentiate to form essential mating structures and male-specific neurons (Sulston et al., 1980). egl-5 expression is visible in rectal epithelial cells at hatching, while egl-5 expression in P12 does not become apparent until the first larval stage (Ferreira et al., 1999). The egl-5 phenotype is consistent with the gene playing a role in the differentiation of these cells (see Introduction).

To begin to ask what factors are involved in regulating egl-5 expression in these cell types, we defined cis-regulatory elements sufficient to drive expression in these cells. A reporter construct pLG1 (syEx609), containing...
approximately 13 kb of egl-5 promoter (Fig. 2A) fused to the heterologous basal pes-10 promoter driving gfp, was sufficient to direct expression in K, F, B, U, and P12. EM#286 (bxIs13, Fig. 2A), which contains the most proximal 7 kb of promoter, did not drive expression in these cell types. We then tested the 6-kb differential fragment between the two (pLG2/syEx610, Figs. 2A and 5A) in our pes-10 assay and found that it was sufficient to drive gfp expression in the same cell types.

We next examined a series of egl-5 promoter subfragments in order to narrow down the region further (Fig. 5A). We identified a 1.3-kb subfragment, pLG7, which was sufficient to direct expression in K, F, B, U, and P12 p (Figs. 2A and 5A). Within pLG7, we identified a 469-bp piece pLG12 sufficient to drive gfp expression in B, but not K, F, U, and P12, and a 446-bp piece pLG11 sufficient to drive expression in K, but not F, B, U, or P12. We also identified a 734-bp subfragment pLG17 sufficient to direct expression in both K and B. Interestingly, pLG13, which contains sequences both necessary and sufficient for expression in K, drives only extremely weak expression in this cell, potentially due to a repressive element contained in the additional, distal sequences this construct contains relative to pLG17. While we have been able to delineate regions necessary and sufficient for expression in K and B; F-, U-, and P12-specific elements have proven more difficult to distinguish. Each of the constructs described was also

Fig. 4. Comparison of expression of an egl-5::gfp reporter gene containing 7046 nucleotides of upstream sequence (bxIs13) to one driven only by V6CRE, the cis-regulatory element that supports expression in the rays (V6CRE::gfp). The results show sequences outside of V6CRE contribute to a more nearly wild-type expression pattern. (A) Expression of bxIs13 in the seam cell V6.ppp in an L2 larval male. V6.ppp is the progenitor of rays 2–6 (see Fig. 7D). Like the chromosomal egl-5 gene (see Fig. 7D), bxIs13 is expressed, beginning in V6.ppp, throughout the lineages leading to rays 4–6 and also in ray 3 (data not shown). (B) Expression of bxIs13 in the seam requires wild-type egl-5 gene function. The transgene is not expressed in V6.ppp or its descendants in egl-5(u202). Here, lack of expression in V6.pppp is shown. (C) Like bxIs13, in wild-type V6CRE::gfp (bxEx88) is expressed in seam lineages leading to rays 3–6, starting in seam cell V6.ppp. (D) Unlike bxIs13, V6CRE expression in the V6 lineage does not require wild-type egl-5 gene function. In egl-5(u202), V6CRE::gfp is expressed in the V6 lineage; here, expression in V6.ppp is shown; expression continues later in the lineage (data not shown). It is concluded that promoter sequences present in bxIs13 but absent from V6CRE::gfp block the activity of V6CRE in the absence of EGL-5 protein. (E) Both bxIs13 and V6CRE::gfp (data not shown) are expressed in head neurons (white arrows) where EGL-5 protein is not detected in wild type. This observation implies the presence of additional negative elements outside of the regions covered by these reporters.
examined on the endogenous egl-5 promoter and found to drive gfp expression in the identical cell types as those seen in our pes-10 assay (data not shown).

Phylogenetic footprinting reveals important regulatory sites for P12 and rectal epithelial cells

The 1.3-kb subfragment (pLG7; Figs. 2A and 5A) that drives gfp expression in K, F, B, U, and P12 is highly conserved between C. elegans and C. briggsae. Within this region, there are six high-scoring pairs (denoted here in distal to proximal order as a–f) in the range of 14–33 bp each (Fig. 5B). The six sites are arranged in two quasiclusters, with approximately 300 bp between the clusters. We generated small deletions of these short conserved sequences to study their function (Fig. 5C). Although deletions of single and double (with the exception of Δa,b) sites seem to have little effect, combined deletions of either a, b, and c (conserved region 7) or d, e, and f (conserved region 8) abolish expression in K, F, U, and P12 (reporter is still on in B and body wall muscles), suggesting that sequences in both clusters may be important for expression in these regions (Fig. 5C). One caveat of these types of experiments is that deletions could alter spacing in the promoter, disturbing enhancer–promoter interactions. Constructs that specifically deleted c, d, and e, removing a similar number of base pairs as the Δa, b, c and Δd, e, f constructs, did not affect expression in K, F, U, and P12, suggesting that the interaction of the cis-regulatory sequences is not affected by deletions of this magnitude.

The Wnt pathway is a likely candidate for direct regulation of egl-5 expression in the positionally related rectal epithelial and P12 cells

There is genetic evidence that egl-5 is regulated by both the Wnt and the EGF pathway in the neuroectoblast cell P12. Epistasis analysis indicates that egl-5 functions downstream of the two pathways and is required for correct P12 fate specification (Jiang and Sternberg, 1998).

We were interested to explore the possibility that egl-5 functions as a Wnt-responsive promoter in P12. The region of the egl-5 promoter that is required for expression of egl-5 in P12 contains three consensus sites for Tcf/POP-1, a transcription factor usually found as a downstream component of the Wnt pathway (Fig. 5B). In order to ask if these sites were functionally significant, we made deletions of the three sites within the context of the 1.3-kb pLG7 subfragment. These deletions eliminated expression in K, F, B, U as well as P12 (data not shown). This implies that egl-5 may be coordinately regulated in these positionally related cell types and the Wnt pathway is a plausible candidate for this regulation.

Dissection of the V6 cis-regulatory element (V6CRE) implicates redundant regulation in the seam cell V6 lineage

The expression of egl-5 in the V6 lineage starts during the L2 larval stage in the seam cell V6.ppp and continues in the three ray precursor cells this cell generates, R4, R5, and R6. Later at Rn stage (during the L3 larval stage, when the ray precursor cells are born), R3 also begins to express egl-5. The expression of egl-5 persists in the four ray sublineages until the divisions of the ray precursor cells are final (Ferreira et al., 1999; Fig. 7D, green lineages in wild type).

To define the factors that regulate egl-5 expression in the V6 lineage, we carried out a more detailed dissection of V6CRE. This region drives robust expression of the Δpec-10::gfp reporter in the V6 lineage (Figs. 2A and 8A). The alignment of the V6CRE region between C. elegans and C. briggsae is shown in Fig. 6A. In addition to the conserved sequences 11 and 12, which were identified in the original blast search (Table 1), there is a shorter piece of DNA situated between 11 and 12, which is also conserved between the two species.

To define a minimal enhancer for the V6 lineage, we carried out a deletion analysis of V6CRE. Representative deletion constructs are shown in Fig. 6B. Similar to the results described above with the regulatory regions for the rectal epithelium and P12, dissection of V6CRE revealed considerable redundancy of sites across the region. Sequences sufficient for normal or near-normal expression lie in both sequences 11 and 12. For example, in the absence of conserved sequence 11, conserved sequence 12 plus the central region are sufficient (construct 7). Sites at the right end of 12 are necessary for this expression (compare constructs 7, 9, and 10). However, in the presence of 11, this region is dispensable (construct 3), indicating redundancy with sites in conserved sequence 11. Similarly, these sites in 11 along with the central region are themselves sufficient (construct 4). Neither conserved sequence 11 nor 12 alone is sufficient without the central region (constructs 6 and 8), and likewise the central region alone is not sufficient (i.e., construct 10). Thus the arrangement of regulatory sites within V6CRE is complex, with multiple, interacting sites spread across the region.

Fig. 5. Analysis of the cis-regulatory region that drives expression in rectal epithelial cells and P12. (A) Dissection of sequences necessary for egl-5 expression in rectal epithelial cells and the posterior neuroectoblast cell P12. Sequences necessary for expression in K are depicted schematically in green; B in blue; and P12, F, and U are shown in lined region. Plusses and minuses indicate the presence or absence, respectively, of gfp expression in the specified cell type. Wavy lines represent weak or erratic expression. (B) Sequence alignment of the region around conserved sequences 7 and 8. Conserved sites a–f and Tcf consensus sites are displayed in rectangles. (C) Deletion analysis of conserved blocks a–f in pLG7 suggests that the conserved sites bear a certain degree of redundancy. Reporters with deletions covering the regions indicated were generated and analyzed. Plusses and minuses indicate the presence or absence, respectively, of gfp expression in the specified cell type.
A regulatory pathway unique to the most posterior branch of the V6 lineage

Given the evidence presented above that V6CRE::gfp (bxEx88) does not require egl-5 to maintain its expression, we asked whether this region contains target sites for other regulators known to be required for egl-5 expression in the V6 lineage. Another Hox gene, mab-5, is a V6 lineage identity gene (Kenyon, 1986). Without mab-5, the fate of V6 is transformed to that of anterior seam cells V1–V4 (Fig. 7D). It has been confirmed by antibody staining that egl-5 is not expressed in the V6 lineage in a mab-5 null mutant (Ferreira et al., 1999). This regulatory requirement is also reflected by reporter genes such as bxIs13, which are not activated in V6.ppp in a mab-5(e1239) background. However, we did observe a small percent of worms expressing bxIs13 in V6.pppp and its daughter cells in L3–L4 males (Fig. 7C, arrow) (70% n = 34).
58). The timing of this expression is distinct from the expression of egl-5 in V6.ppp in wild-type animals. It starts at late Rn stage (as judged by the T lineage, which is not affected by mab-5; indicated by arrow B in Fig. 7D). Similar results were obtained with animals bearing a regulatory mutation e2039 in pal-1, another V6 lineage identity gene upstream of mab-5 (67%, n = 91; only animals with complete V6 to V1–V4 transformation were scored) (Hunter et al., 1999).

In many egl-5::gfp reporters and Apes-10::gfp reporters driven by part of V6CRE, expression in the R5 sublineage is significantly stronger than in other egl-5-positive ray sublineages. For example, bxEx113 (Fig. 6B, construct 4) shows wild-type egl-5 expression pattern (Fig. 7A). Similar results were obtained with animals bearing a regulatory mutation e2039 in pal-1, another V6 lineage identity gene upstream of mab-5 (67%, n = 91; only animals with complete V6 to V1–V4 transformation were scored) (Hunter et al., 1999).
that, unlike its *elegans* counterpart, this piece only drives expression in R5 and its descendants starting at late Rn stage (Fig. 7B).

We propose that there is a separate pathway, independent of known V6 lineage identity genes, that has the potential to activate egl-5 through V6CRE in the most posterior cell in the V6 lineage at the Rn stage (R5 in wild type). The relevance of this pathway to activation of egl-5 in wild type is unknown.

A repressor motif in V6CRE contributes to the sexual, temporal, and spatial specificity of egl-5 expression in the V6 lineage

An 18-base pair sequence in V6CRE closely matches (11/15 conserved bases) the binding site for the zinc-finger protein TRA-1A, a repressor in the sex determination pathway (Zarkower and Hodgkin, 1993) (Fig. 6A). TRA-1A is active in hermaphrodites to prevent the expression of male-specific genes and is thought to have no activity in the male soma (Conradt and Horvitz, 1999; Hodgkin, 1987; Yi et al., 2000). On assumption that this motif might contribute to the male-specific expression of egl-5 in the V6 lineage, we generated a reporter with the site deleted (Fig. 6B, construct 1). Deletion of the TRA-1A consensus sequence not only caused expression in the V6 lineage in hermaphrodites, but also caused ectopic expression in the anterior seam cells and ventral cord cells in both hermaphrodites and males (Figs. 8B and C). Ectopic expression starts as early as V6.p stage instead of V6. ppp stage. A point mutation (CC to TG, underlined in Fig. 6A) that has been shown in vitro to abolish TRA-1A binding (Fig. 6B, construct 12) (Yi et al., 2000) has similar effects on reporter expression (Fig. 8D). However, when we introduced *V6CRE::gfp* into the putative null allele *tra-1(e1091)*, we did not observe any ectopic expression in the anterior seam cells (n = 44) (data not shown). Therefore, loss of TRA-1A binding alone cannot account for the ectopic expression. Some unidentified protein with similar binding specificity to TRA-1A must bind to the motif in V6CRE and act as a repressor.

We propose one possible scenario to explain the function of the TRA-1A-like repressor in Fig. 8E. V6CRE contains binding motifs for activators existing in all seam cells (complex 1). The presence of a repressor with similar binding specificity to TRA-1A (complex 2) prevents these activators from activating reporter expression. Expression in the V6 lineage is achieved by a V6. ppp-specific combination of activators (complex 3) that can block or bypass this repressor.

Discussion

Structure of the egl-5 promoter

We have characterized the overall structure of the promoter region of the *C. elegans* posterior Hox gene egl-5 and dissected two tissue-specific presumptive enhancer regions in greater detail. Like other genes of the Hox family, egl-5 is expressed in multiple, diverse tissues and cells within a localized body region, endowing these tissues and cells with their region-specific differentiated characteristics (Chisholm, 1991; Ferreira et al., 1999). Some 18.5 kb of DNA sequence lies between the start of the egl-5 coding sequence and the coding region of the next 5′ gene. This is a rather large potential regulatory region for a *C. elegans* gene. We have found that throughout this region are dispersed a number of sequence blocks conserved in the related species *C. briggsae*. Some of these conserved blocks, contained within an almost 10-kb region, appear to represent regulatory sequences that direct the expression of egl-5. For several we tested, we found they had the capability to drive the expression of an enhancerless reporter gene in non-overlapping subsets of tissues where the chromosomal egl-5 gene is normally expressed (Ferreira et al., 1999). Thus, the detailed expression pattern of egl-5 is built up from the combined activities of multiple, autonomous regulatory sequences making up a complex regulatory control region.

This complex promoter structure contrasts with that previously demonstrated for some other tissue-specific genes in *C. elegans*. Each of these genes is responsive to a key transcription factor that drives its expression through interaction with an important regulatory site (e.g., Cui and Han, 2003; Gaudet and Mango, 2002; Swoboda et al., 2000; Way et al., 1991). Although Hox genes are often thought to have a unitary function to convey regional cell identity, apparently in *C. elegans*, posterior Hox gene expression is not controlled by any unitary “posterior” signaling pathway. Rather, for egl-5, it appears that transcription at least in part is under the control of multiple, separate tissue-specific or possibly local signal-specific regulatory proteins. Such a promoter structure may correlate with the evolution in nematodes toward a greater degree of cell autonomy in fate specification within developmental cell lineages (Cowing and Kenyon, 1996; Sulston and White, 1980).

In addition to positively acting regulatory sequences, the egl-5 promoter contains negative regulatory elements that we only broadly located within certain general intergenic regions. We found that upon deletion of these regions, reporter genes not only lost expression in tissues associated with the positive elements contained therein, but also gained expression in cells in inappropriate parts of the body, for example within neurons in the head. Moreover, when single positive elements were used to drive an enhancerless reporter gene, they promoted expression not only in one or more egl-5-appropriate tissues but also in additional, inappropriate
tissues. Thus, positive elements are silenced in certain tissues by additional sequences lying in other parts of the promoter regulatory region. The V6CRE enhancer that supported expression in a hypodermal cell lineage, for example, also supported strong expression in the gut and other cells. This element must be responsive to positive factors present in the gut as well as those in the hypodermis. However, in the context of a full-length promoter, V6CRE activity is silenced by presumptive negative factors present in the gut but not in the hypodermis. Such a multifold, positive plus negative mechanism may be necessary to accurately control a gene such as a Hox gene whose activity is required in many, diverse tissues, resulting in a complex promoter structure. Apparently, the information contained within single positive enhancer elements is insufficient to specify a cell identity uniquely, and this can only be achieved by taking advantage of the combined effects of both positive and negative cell-specific factors.

Multiple negative regulatory pathways

We found evidence for several negative pathways acting on the egl-5 promoter. One type prevented ectopic expression in a wild-type genetic background. When regions of the promoter apparently containing the target sites of such pathways were deleted, reporter genes were
expressed in inappropriate cells or tissues, as discussed above. A second type acted to enforce an autoregulatory requirement for the egl-5 gene product. This pathway was revealed by examining an egl-5 mutant. In the context of a full-length promoter, expression of reporter genes in the V6 hypodermal seam cell lineage required the function of the endogenous egl-5 gene, indicating the presence of a positively acting autoregulatory feedback loop. However, this requirement was lost when the V6 lineage enhancer V6CRE was isolated upstream of an enhancerless reporter. This raises the possibility that autoactivation by EGL-5 occurs by blocking the function of a negative pathway acting at a site outside of V6CRE. Yet another negative factor apparently acts within V6CRE itself. When the site for this factor was deleted, V6CRE promoted expression throughout the seam and with incorrect timing. Thus, the promoter analysis leads to evidence for a complex interplay of multiple positive and negative pathways governing the expression of the egl-5 gene.

In other animals, Polycomb Group (PcG) proteins play a prominent role in the negative regulation of Hox genes, preventing their expression in ectopic locations. In C. elegans, PcG genes of the Esc/Ez class are known and have some role in Hox gene repression (Ross and Zarkower, 2003). PcG genes of the PRC1 class have not so far been identified in the C. elegans genome. However, Hox genes are strongly regulated by a C. elegans-specific protein SOP-2 that shares a protein interaction motif (SAM domain) with certain PRC1 PcG proteins in other animals, suggesting a possible common mechanism of repression (Zhang et al., 2003). Esc/Ez PcG proteins and SOP-2 are candidates for some of the negative pathways we have implicated in the silencing of egl-5 reporter genes in ectopic locations. Their possible sites of action have not been defined. Another negatively acting factor, the Zn-finger Sal family protein sem-4, acts at a site near to or within V6CRE (Toker et al., 2003). In a sem-4 mutant, both the endogenous egl-5 gene and egl-5 reporter genes are ectopically expressed in the T cell lineage, the postembryonic seam cell lineage lying adjacent to the V6 lineage on the posterior side. As a result of this ectopic egl-5 expression, T cell descendants inappropriately assume properties of touch cell neurons.

Positive control regions as complex sites of enhanceosome assembly

For two conserved blocks of regulatory sequences we studied, our more detailed analysis revealed that they have similar features. One of these blocks spans 328 bp and drives expression in rectal epithelial cells and a ventral neuroectoblast cell P12 (conserved sequences 7 and 8). The other spans 180 bp and drives expression in the V6 hypodermal seam cell lineage (V6CRE). Both of these regions are to a degree bipartite, that is, they have separate halves each of which is capable of supporting expression in some target cells. Requirements for specific subregions differed for different cells. Therefore, both regions were complex, containing multiple sites that can independently and to a degree redundantly support expression in multiple target cells. In neither case was it possible to identify a single key sequence that could represent the target site for an essential, controlling transcriptional activator.

In order to identify factors that might act at sites within these two regulatory regions, we attempted to relate their activities to pathways previously demonstrated to regulate egl-5. In the case of the rectal epithelium and P12, egl-5 is downstream of the EGF and Wnt pathways (Jiang and Sternberg, 1998). Regions 7 and 8 contain several putative target sites for the downstream Wnt effector POP-1. Deletion of these sequences resulted in loss of expression of a reporter gene, suggesting that these are indeed functional POP-1 target sites and that one of the signaling pathways integrated by this region is the Wnt pathway. All of the cells in which egl-5 expression is activated by regions 7 and 8 are clustered near one another in a localized region. A possible function of the clustering of regulatory sites in regions 7 and 8 might therefore be to integrate the activities of cell-specific factors with a localized Wnt signal.

In the case of the V6 seam cell lineage, egl-5 is thought to be part of a transcriptional cascade in which the caudal homolog pal-1 turns on the Hox gene mab-5 in V6 and mab-5 turns on egl-5 in V6.ppp (Ferreira et al., 1999; Hunter et al., 1999; Salsers and Kenyon, 1996). Expression of the endogenous egl-5 gene and full-length egl-5 reporter genes requires the functions of pal-1 and mab-5 (Ferreira et al., 1999; this work). Activation of the V6 lineage regulatory region V6CRE in a minimal promoter did not occur in V6.ppp in these backgrounds, consistent with this region containing the target site for response to activation by the PAL-1–MAB-5 pathway. However, to our surprise, V6CRE::gfp was strongly expressed in the most posterior branch of the V6 lineage starting at the V6.pppp stage in mab-5 and pal-1 mutants. This suggests the possible role of additional unknown factors in some lineage branches in activation via this element.

The complex structures of the two regulatory regions we dissected, each containing multiple putative sites for the binding of both positive and negative regulators, suggest that these sequences act as the sites of assembly of enhanceosomes whose function is to integrate multiple signals relevant to the activation of egl-5 in a particular subset of cells. Such an enhanceosome could be assembled at the time of transcription activation or it could be assembled over several cell generations and maintained epigenetically (Cavalli and Paro, 1999; Cosma et al., 1999; Soutoglou and Talianidis, 2002). In the latter case, transcriptional activation might occur via release of a negative mechanism that holds a preassembled complex in check. We obtained evidence in an earlier study for activation of the caudal homolog pal-1 by such a mechanism. We found that a component of the Mediator complex, presumably present in a transcriptional initiation complex preassembled at the
promoter, blocked transcription of pal-1 until its activity was released by a factor acting through a downstream site in an intron (Zhang and Emmons, 2000). A mutation in another component of the Mediator complex also resulted in activation of egl-5 in inappropriate branches of the V6 lineage (Zhang and Emmons, 2001). Such a release mechanism might allow for the high fidelity and apparent precision of timing of initiation of egl-5 transcription in the cell lineage and might also in part account for the numerous negative pathways that apparently act in egl-5 regulation.

Acknowledgments

We thank the C. elegans Genetics Center for strains; A. Fire for plasmids; and members of the Emmons and Sternberg laboratories for helpful discussions. This research was supported by NIH Grant R01 GM39353 to S. W. E and the Siegfried Ullmann Professor of Molecular Genetics.

References


