Genome & Developmental Control

Multiple regulatory elements with spatially and temporally distinct activities control the expression of the epithelial differentiation gene lin-26 in C. elegans

Frédéric Landmann, Sophie Quintin, and Michel Labouesse*

Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, BP101423, F-67404 Illkirch Cedex, C.U. de Strasbourg, France

Received for publication 25 April 2003, revised 14 August 2003, accepted 8 September 2003

Abstract

Epithelial differentiation is a very early event during development of most species. The nematode Caenorhabditis elegans, with its well-defined and invariant lineage, offers the possibility to link cell lineage, cell fate specification and gene regulation during epithelial differentiation. Here, we focus on the regulation of the gene lin-26, which is required for proper differentiation of epithelial cells in the ectoderm and mesoderm (somatic gonad). lin-26 expression starts in early embryos and remains on throughout development, in many cell types originating from different sublineages. Using GFP reporters and mutant rescue assays, we performed a molecular dissection of the lin-26 promoter and could identify almost all elements required to establish its complex spatial and temporal expression. Most of these elements act redundantly, or synergistically once combined, to drive expression in cells related by function. We also show that lin-26 promoter elements mediate activation in the epidermis (hypodermis) by the GATA factor ELT-1, or repression in the foregut (pharynx) by the FoxA protein PHA-4. Taken together, our data indicate that lin-26 regulation is achieved to a large extent through tissue-specific cis-regulatory elements.

© 2003 Elsevier Inc. All rights reserved.

Keywords: C. elegans; Gene regulation; Promoter; Tissue specificity; Epithelial; Hypodermis

Introduction

Promoter sequences of genes that control important developmental decisions integrate the regulatory pathways active at a particular embryonic stage. Their analysis can provide clues as to which pathways play important roles in a given process. The molecular mechanisms that control the onset of epithelial differentiation remain poorly understood. For instance, despite a growing number of studies, no single class of transcription factor has emerged as the key player in regulating gene expression in the mammalian skin epithelium. Instead, evidence has been accumulating to suggest that the expression of terminal differentiation genes in keratinocytes is controlled by the combinatorial action of more broadly expressed transcription factors (Byrne, 1997; Fuchs and Raghavan, 2002; Sinha et al., 2000).

Instead of analyzing the promoter of a terminal differentiation gene, we turned our attention to a Caenorhabditis elegans gene, called lin-26, acting early in the process of epithelial differentiation. The LIN-26 protein is a presumptive zinc finger transcription factor expressed in all epithelial cells of the ectoderm, in the somatic gonad precursors and in the uterus. In lin-26 mutants, these epithelial cells are made but degenerate, suggesting that lin-26 is essential for their differentiation (den Boer et al., 1998; Labouesse et al., 1996, 1994). Importantly, ectopic expression of lin-26 in early blastomeres induces the expression of several epithelial markers (Quintin et al., 2001).

The nematode C. elegans stands apart from most other organisms in that its lineage is invariant (Sulston et al., 1983). Maternally expressed genes control the five rounds of asymmetric divisions giving rise to the six founder genes.
blastosomes (AB, MS, E, C, D and P₄), which differ in their properties, division rates and progeny (Sulston et al., 1983). In spite of the invariant lineage, there is no correlation between lineages and tissues except for E (from which the midgut originates), P₄ (the germ line) or D (muscles). Different models have been proposed to explain the relationship between cell fates and the lineage. One model is based on lineage-dependent mechanisms (Schnabel, 1996), whereby a refinement of differentiation potentials would occur at each cell division throughout embryogenesis through the asymmetric segregation of selected regulatory proteins, such as the HMG box protein POP-I (Lin et al., 1998). Another, not necessarily incompatible, model suggests that some “master regulatory” genes specify the identity of entire organs (for a review, see Labouesse and Mango, 1999). These organ identity genes are generally expressed in precursors derived from different lineages and control the expression of most organ-specific differentiation genes, irrespective of cell type. This is, for instance, the case for the C. elegans FoxA homologue pha-4, which is expressed in all pharyngeal cell types and specifies pharynx identity (Gaudet and Mango, 2002; Horner et al., 1998; Kalb et al., 1998). Similarly, Pax6–eyeless is thought to specify eye identity in the animal kingdom (for a review, see Gehring and Ikeo, 1999), or Pitr together with Gata2 ensures pituitary differentiation in vertebrates (Dasen et al., 1999).

Promoter studies performed in C. elegans have revealed cases of tissue-specific regulation, but mainly for terminal differentiation markers (Gilleard et al., 1997; Okkema et al., 1993). Regulatory genes whose promoters have been dissected so far appear to be controlled in part through lineage-dependent mechanisms. This is the case for unc-86, which specifies the identity of several neuroblasts originating from asymmetrically dividing mother cells, and the C. elegans MyoD homologue hlh-1 (Baumeister et al., 1996; Krause et al., 1994).

Cells that express lin-26 originate from the AB blastomere (most epidermal cells which are called hypodermal in C. elegans, sheath and socket support cells, rectal cells, epithelial cells of the excretory system), from the C blastomere (other hypodermal cells which are called hypodermal in C. elegans, sheath and socket support cells, rectal cells, epithelial cells of the excretory system), from the MS blastomere (most epidermal cells which are called hypodermal in C. elegans, sheath and socket support cells, rectal cells, epithelial cells of the excretory system), or from the C blastomere (other hypodermal cells). Transgenic lines and cell identification

Materials and methods

Genetics

Strains were maintained at 20°C as described (Brenner, 1974). We used the wild-type Bristol N2 strain and null mutant strains for the genes lin-26, pha-4 and elt-1: lin-26(mcl15) unc-4(e120)/mnC1 (den Boer et al., 1998); fog-2(q71) pha-4(q490)/stu-3(q265) rol-9(sc148) (Mango et al., 1994); elt-1(ze180) unc-43(e408)/unc-24(e138) dpy-20(e1282) (Page et al., 1997).

Molecular biology

Deletion constructs were obtained either by restriction or by PCR amplification from the rescuing plasmid pML301 (den Boer et al., 1998) or its GFP derivative pML702 which does not rescue lin-26 null mutants (Bosher et al., 1999). PCR reactions were performed with the Ffu polymerase (Stratagene), or with the Expand Long Template PCR system (Roche Molecular Biochemicals) for larger fragments (size > 1 kb). Promoter fragments were cloned into the GFP reporter pPD97.82 containing the myo-2 minimal promoter and unc-54 3′ untranslated region (3′ UTR) (Fire Lab Vector Kit, 1997). Constructs rABCD + i1 to rABC + i3 were obtained with primers surrounding the deletion, which allowed us to copy the entire plasmid with the Expand Long Template PCR system (Roche Molecular Biochemicals). Construct pcm33 was obtained by exchanging the BamHI–AatII fragment of pML301 with a similar PCR-amplified fragment from homozygous lin-1(mcl33) embryos. The hormone response element or HRE (PuGGA- TAC) present in block CISc was replaced in pCmutHRE by a BamHI site (GGATCC), which was introduced using a PCR-based strategy. Primer extension was performed as described by Sambrook et al., (1989) with the primer 5′TGAAGTGCTTACTCGAATCCAC.

Transgenic lines and cell identification

Plasmids were injected into syncytial gonads of hermaphrodites at 10 ng/µl with prF4 as a transformation marker at 100–200 ng/µl (Mello and Fire, 1995). Several clones were tested alone or in coinjection; and in every case, at least three transgenic lines were established. Extrachromosomal arrays carrying eE1.3 and eC constructs were integrated either spontaneously (eE1.3, mcls33) or after X-ray treatment (eC, mcls32). Cell identification was performed in embryos or in larvae, depending on the timing of expression, by comparing Nomarski views of embryos or larvae with established fate maps (Sulston et al., 1983). For eE1 and eE1.2 constructs, an additional criterion to identify support cells was the GFP-labelled processes extending anteriorly, making them distinct from neurons which have processes extending anteriorly and posteriorly, and from hypodermal cells which do not extend processes. Transgenic
**Fig. 1.** *lin-26* expression pattern and cis-regulatory regions. (A) Scheme of an embryo at mid-embryogenesis adapted from Sulston et al. (1983). Cells expressing *lin-26* are color-coded; their lineage origins and numbers in each major branch of the lineage are shown on the right. (B) Organization of the *lir-1–lin-26* operon and position of the conserved intronic sequences (CISa–g) in *lir-1* intron 1 (Dufourcq et al., 1999); for simplicity, we will call *lir-1* exons 2–5 CISh, and the *lir-1–lin-26* intergenic region “i”. In subsequent figures, conserved blocks retained in plasmids will be designated with capital letters. The GFP coding sequence in pML702 is inserted in the middle of *lin-26* and does not rescue *lin-26* mutants. Other enhancer tests were performed with the GFP reporter pPD97.82 containing the *myo-2* minimal promoter and *unc-54* 3′-UTR. The expression cassettes used in each construct in this and other figures will be symbolized as indicated and are not drawn to scale. (C) *lin-26* expression pattern as observed by immunostaining with a LIN-26 antiserum and in an embryo transgenic for the GFP reporter pML702. Here and in subsequent figures (unless indicated otherwise), embryos will be approximately at the stage illustrated in A with anterior to the left, scale bar = 10 μm.
lines carrying the vector pPD97.82 alone did not induce GFP expression.

Immunofluorescence

Immunostaining with polyclonal antibodies against LIN-26 (Labouesse et al., 1996), the monoclonal antibody MH27 recognizing adherens junctions (Francis and Waterston, 1991) or polyclonal antibodies against GFP (Clontech, diluted 1:500) was performed as described (Chanal and Labouesse, 1997).

RNA interference (RNAi)

Double-stranded RNA (dsRNA) was obtained from PCR products as described elsewhere (Bosher et al., 1999). dsRNA targeted exons 6–8 of nhr-23, exons 4–7 of nhr-25, exons 1 and 2 of nob-1, exons 2 and 3 of php-3, exons 2 and 3 of ceh-13, exons 3–6 of elt-3, exons 1 and 2 of elt-5, exons 3–5 of elt-6. These regions do not show strong similarity to other C. elegans genes and the corresponding RNAi phenotypes were as expected (Brunschwig et al., 1999; Gilleard et al., 1999; Gissendanner and Sluder, 2000; Koh and Rothman, 2001; Kostrouchova et al., 1998; Van Auken et al., 2000). Since elt-3 RNAi does not give any obvious phenotype, the dsRNA was first tested in a integrated elt-3::gfp strain in which the GFP was shut down after the RNAi treatment.

Dye-filling experiments

Adult animals were stained with 3,3′-dioctadecyloxacarbocyanine perchlorate (DiO) as described by Herman and Hedgecock (1990).

Results

Experimental strategy

Previous experiments suggested that lin-26 and the upstream genes, lir-2 and lir-1, form two overlapping transcriptional operons (Dufourcq et al., 1999). According to this model, the first operon includes lir-2 and long lir-1 isoforms, while the second operon includes short lir-1 isoforms, starting at or after exon 2, and lin-26. Sequence comparison between the C. elegans and Caenorhabditis briggsae lir-1–lin-26 operons identified long stretches of highly conserved sequences (78% on average) within the first lir-1 intron (a 9-kb-long intron), which we previously

Fig. 2. lin-26 expression in rectal rep cells and the somatic gonad. (A) Conserved regions are as in Fig. 1. Three plasmids containing the CISa–b (above the map) or CISc (below the map) regions displayed enhancer activity in the rectal rep cells or in the somatic gonad. Nucleotides are numbered relative to the lin-26 initiation codon; H3, HindIII site. Open triangles mark the positions of conserved RGTTCA sites (putative hormone response elements, HRE). The sequence alignment represents the 100 most-conserved bases between C. elegans and C. briggsae in block CISc. The black triangle indicates a mutated HRE introduced in the block CISc of pML301 (see Fig. 3). (B) L1 larva transgenic for the plasmid eAB; the GFP expressed in rep cells (ventral view). (C) Top: wild-type pretzel embryos transgenic for the plasmid eC; the GFP is expressed in the two somatic gonad precursors Z1 and Z4. Bottom: late embryo carrying an integrated construct eC after RNAi against the nhr-23 and nhr-25 genes; the GFP was still expressed in Z1–Z4 (N = 65 embryos). None of the reporter constructs described in subsequent figures was affected by RNAi against nhr-23 and nhr-25 (data not shown).
called conserved intronic sequences or CIS (Fig. 1B) (Dufourcq et al., 1999). In separate experiments, we showed that lir-1 intron 1 and exons 2–5 (11 kb) are required to obtain an expression pattern identical to what is observed with the LIN-26 antiserum and rescue the lethal phenotype of lin-26 null mutants (Figs. 1B,C) (den Boer et al., 1998).

To understand how the lin-26 expression pattern is established, we dissected its promoter elements using enhancer tests and complementation rescue experiments. We performed enhancer tests by examining whether promoter fragments corresponding to one or more CIS blocks could drive GFP expression in specific cells using either the native lin-26 promoter or the well-characterized myo-2 minimal promoter (Fire et al., 1990; Krause et al., 1994). Conversely, we tested whether removal of these fragments from the whole promoter would affect GFP expression in cells that normally express lin-26, or the ability of lin-26 transgenes to rescue lin-26 mutations (see Materials and methods). Hereafter, we use the following nomenclature to call plasmids: the first letter refers to the plasmid backbone (p, r or e for pML301, pML702 or pPD97.82, respectively) and the subsequent letters refer to CIS blocks retained in the construct (capital letters A–H, with the small case letter “i” to designate the weakly conserved lir-1–lin-26 intergenic region). For the sake of clarity, we will introduce constructs affecting a given cell type together, beginning with cells that initiate lin-26 expression last.

**Expression in rectal rep cells**

There are two major rectal cell types, those that make the anus (called B, F, U, Y, K, K') and those that make the interface between the intestine and the anus (called repD, repVL and repVR). LIN-26 antibodies detect the protein in the direct precursors of anal cells, but much later at the end of embryogenesis in the rectal interfacial rep cells (Chanal and Labouesse, 1997; Labouesse et al., 1996). The most distal promoter that we isolated is a 3595-bp element corresponding to CISa–b, which promoted GFP expression in rep cells (construct eAB; Figs. 2A,B). The GFP appeared at the end of embryogenesis, long after rep cells differentiate and were maintained until the adult stage, which is consistent with LIN-26 antibody staining. We did not further dissect this promoter element, as a construct lacking CISab gave rise to lin-26(mc15) adults able to defecate (data not shown; see also Fig. 3). Hence, lin-26 expression in these rectal cells does not seem to be essential for proper differentiation or function of rep cells. The activation of lin-26 in other rectal cells will be introduced further below.

**Expression in the somatic gonad**

LIN-26 expression starts in somatic gonad precursors 3 h after their appearance. It disappears from somatic gonad precursors at the end of the L2 larval stage and reappears in uterine cells during the L4 larval stage (den Boer et al., 1998). The second element that we identified contains CISc. Previous studies showed that a 452-bp region including CISc is required for normal expression of lin-26 in the somatic gonad; deletion of this fragment prevents normal gonad development, leading to sterility (den Boer et al., 1998). The constructs eC and eC1 containing CISc could drive GFP expression in somatic gonad precursors Z1 and Z4 (Figs. 2A,C). Expression started at the two-fold stage, with a maximal intensity at the pretzel stage and in young L1 larvae, and subsequently vanished. This expression pattern is very similar to that observed with the LIN-26 antiserum. Occasionally, the GFP was maintained for one or two rounds of divisions in the Z1–Z4 daughter cells (except in distal tip cells). Later, very weak expression could be seen in the uterus of L4 animals.

In a search for transcription factors that could control lin-26 expression in the somatic gonad, we turned our attention...
to the orphan nuclear hormone receptors NHR-25 and NHR-23. Indeed, inactivation of NHR-23 or NHR-25 leads to defects reminiscent of those resulting from loss of lin-26 expression (Asahina et al., 2000; Gissendanner and Sluder, 2000; Kostrouchova et al., 1998). Moreover, NHR-23 was shown to bind direct repeats or more weakly a monomeric consensus hormone response element (HRE) (Kostrouchova et al., 1998), a copy of which is found in the conserved CISC element (Fig. 2A). We tested whether nhr-25 and nhr-23 are required for lin-26 expression by performing RNA interference (RNAi) experiments against these genes and by mutating the HRE in CISC. We found that RNAi against nhr-23 and nhr-25 did not prevent GFP expression in somatic gonad precursors driven by the eC construct (Fig. 2C). In addition, we found that a GGCTCA to GGATCC mutation in the HRE of CISC (construct pCmutHRE, Fig. 3A) did not affect the potential of an otherwise wild-type lin-26 transgene to rescue the null allele lin-26(mc15) to viability and fertility. We conclude that nhr-23 and nhr-25 are unlikely to be essential for lin-26 expression in the somatic gonad.

Support cells and minor hypodermal cells: three partially redundant elements

LIN-26 is detected in the direct precursors of tail epithelial cells, which include two support cells, minor hypodermal cells and rectal cells (for rectal repD/V cells, see above). In the 36 head support cells, in the minor hypodermal cells (hyp1–3) and in epithelial cells of the excretory system, LIN-26 expression starts immediately after their appearance between the 400- and 500-cell stages, when ventral enclosure begins (lima bean stage) (Chanal and Labouesse, 1997; Labouesse et al., 1996).

Systematic enhancer tests identified three regions able to drive GFP expression in support cells and minor hypodermal cells, which displayed different temporal activities. Specifically, we found that the first half of the CISE region (construct eE1; Fig. 4A) promoted GFP expression in head and tail support cells associated with chemosensory neurons (marked “s” and “Phsh” in Fig. 4B), the excretory cell and minor hypodermal cells (hyp1–3 in the head and hyp8–10 in the tail; marked “hyp” in Figs. 4Bb). GFP expression from construct eE1 started at the lima bean stage, to disappear at the end of the L1 larval stage. Thus, GFP expression driven by construct eE1 recapitulates what is observed after staining with LIN-26 antibodies. Hence, the first half of CISE is sufficient to drive lin-26 expression in support cells and minor hypodermal cells.

Two other regions that do not overlap with CISE, as they correspond to CISa–d and CISf–h, could also drive GFP expression in support cells when coupled to the lir-1–lin-26 intergenic region (constructs eABCD + i, eFGHi and eFG5V).

Fig. 4. Two short overlapping sequences promote lin-26 expression in support cells and minor hypodermal cells. (A) Four constructs that can drive GFP expression in support cells and minor hypodermal cells are shown under the map of the lin-26 promoter region. Constructs derived from the CISE region are shown at an enlarged scale below the dotted line. (B) Nomarski pictures (a, c, e) and GFP expression patterns (b, d, f) in embryos transgenic for the indicated constructs. In transgenic lines carrying the construct eE1.3, some anterior cells expressing a weak GFP were sometimes observed. For constructs eE1 and eE1.3, some additional posterior cells could express a faint GFP, at the onset of expression, but the signal was too weak and transient to allow their identification. Symbols for GFP-positive cells are given in the box at the bottom; hyp, minor hypodermal cell; RV, EcoRV site; BHI, BamHI site, AatII, AatII site.
In contrast to CISe, these other regions were poorly active if at all in embryos. We do not think that the promoter element carried by each of these constructs is active in a different subset of support cells, because they triggered on average GFP expression in a comparable number of cells, and because we could identify GFP expression in each support cell in separate animals. In addition, the construct eABCD + i could drive GFP expression in minor hypodermal cells (hyp1–3 in the head; hyp8–11 in the tail) and in rectal cells (B, F, Y, U, K, K').

We tried to reduce the size of the element present in eE1, which is the first turned on in support cells and minor hypodermal cells. We could identify different fragments driving GFP expression either in the head (eE1.2, Figs. 4A, Bc) or in the tail regions (eE1.3, Figs. 4A, Be). However, GFP expression in the excretory cell and in hyp10 was lost with these shorter constructs. As the eE1.2 and eE1.3 constructs overlap through 129 bp, it suggests that expression in anterior and posterior parts of the embryo is controlled through many common regulators.

To test if deletion of the promoter elements defined above affected lin-26 expression in support cells, we used an indirect test in larvae. The lipophilic dye DiO normally stains amphid and phasmid sensory neurons, but only if support cells are well differentiated and provide the necessary opening to allow neurons to become exposed to the environment. For instance, in animals carrying the weak and viable mutation lin-26 (n156), approximately 50% of amphid neurons take up DiO (Labouesse et al., 1996). We examined lin-26 (mc15) null animals carrying two non-overlapping constructs (pABCDE5 + i and pE3FGHi), or the construct pD(E5) which lacks the first half of CISe (Fig. 3A). Animals carrying these constructs could reach adulthood,
implying that none of the deleted elements is essential for viability. The number of sensory neurons able to take up DiO was normal in animals carrying pABCDE5 + i, slightly reduced in animals carrying pD(E5V) (P < 0.05) and 50% lower in animals carrying pE3FGHi (P < 0.001) (Fig. 3B). These results are consistent with the enhancer tests described above and indicate that these elements play a functionally redundant role. In summary, we have identified three sequence elements driving LIN-26 expression in support cells and minor hypodermal cells.

**Major hypodermal cells: three functionally redundant promoter elements**

LIN-26 is first expressed in precursors of the major hypodermis at the 100-cell stage when the pharynx, epidermis, intestine, neuronal and mesodermal territories have just formed after the onset of gastrulation (reviewed by Labouesse and Mango (1999)). The major hypodermal cells, which originate from different AB and C sublineages (see Fig. 1A), are produced at the 350-cell stage, slightly after the end of gastrulation. Once on, LIN-26 remains expressed throughout development in all ectodermal epithelial cells.

To identify promoter elements controlling lin-26 expression in the hypodermis, we initially looked at the ability of deletion constructs to rescue the embryonic lethality caused by the null allele lin-26(mc15). We found that constructs simultaneously lacking the CISd, CISe and CISf elements did not rescue lin-26(mc15) (pGHi and pAB + i, Fig. 3A). In contrast, several large non-overlapping deletion constructs carrying at least one of these three elements (pABCDE5V+i, pABCD+i, pE3VFGHi) could all rescue. Interestingly, we had previously observed that the small deletion lir-1(mc33), which removes most of CISf–g and lir-1 exons 2–4, did not affect lin-26 expression (Bosher et al., 1999), arguing that the chromosomal copy of CISf–g and lir-1 exons is not essential for lin-26 expression either. We reproduced this observation by showing that a deletion construct corresponding to lir-1(mc33) could still rescue (construct pmc33, Fig. 3A).

Using enhancer tests, we could then identify three redundant promoter elements driving GFP expression in the major hypodermis. We first showed that the GFP fusions corresponding to some of the rescuing constructs described above could promote GFP expression in the major hypodermis (see rABCD + i, eABCD + i and rE3FGHi; Fig. 5A). However, we found that constructs carrying CISd, CISf or CISg alone failed to give any expression (see rD + i, eD, ef, eg; Figs. 5A and 6A), prompting us to assay CIS sequences in combination. Indeed, we found that the CISf–h elements could drive GFP expression in major hypodermal cells of the head and body (hyp4–7, seam cells, P cells) and in hyp10 nuclei of the tail (see eFGHi; Figs. 5A,Be–h). This expression was first detected in hypodermal precursors two rounds of divisions before hypodermal cells are born, as observed after LIN-26 antibody staining. We could isolate two non-overlapping fragments that were both able to drive GFP expression in all major hypodermal cells (eFG5 and eHi; Fig. 5B). There were temporal differences among all these elements, in that some could initiate GFP expression in hypodermal precursors (eFGHi and eHi), while others could maintain GFP expression, in seam cells and vulval cells (eFGHi and eFG5). (Fig. 5A; see also Fig. 4 for other cell types). We conclude that these partially redundant promoter elements can activate lin-26 expression in the major hypodermis.

Finally, we tested whether sequences located immediately upstream of the lin-26 initiation codon or lin-26 transcribed sequences played a regulatory role. We found that

---

**Fig. 6. Synergies among different enhancers elements. (A)** The 559-bp element corresponding to the beginning of CISe (symbolized by a black arrow; construct eE.1 in Fig. 4) was cloned either downstream of CISd (eD+ E.1) or upstream of CISf (eE.1 + F) and CISg (eE.1 + G). In addition to the pattern observed with construct eE.1 alone, the expression was visible in the cells listed on the right (symbols are given in the box below). (B) Embryos transgenic for the constructs eE.1 + G or eE.1rev + G, showing that in both orientations the 559-bp element can synergize with CISg.
removal of the lir-1–lin-26 intergenic region (the “i” region) abolished GFP expression (compare rABCD with eABCD or rABCD + i, Fig. 5A), showing that the “i” region contains essential sequences. Tested on their own, the corresponding 251 nucleotides do not have any promoter activity (Fig. 5A construct “ei”). To further map essential sequences within this region, we created a set of three progressive deletions (rABCD + i1 to rABCD + i3; Fig. 5A, bottom). Each of the deletions abolished GFP expression, raising the possibility that multiple sequences within this region are important for expression. We conclude that the 251-bp sequence located upstream of lin-26 coding sequences can synergize with CISa–d to drive expression in body hypodermal cells. Finally, we observed that GFP expression in seam cells was weaker or absent when driven from the pPD97.82 cassette (compare eABCD + i and rABCD + i, Fig. 5A), suggesting that sequences within lin-26 exons, introns or 3′ UTR could have a role in maintaining expression.

In summary, we identified at least three distinct promoter elements driving GFP expression in the major hypodermis: one is loosely defined within CISa–d and the other two correspond to CISf–g and CISh. We note that none of these promoter elements appears to work in a lineage- or sublineage-specific manner, since we never isolated an element that would promote expression in one AB sublineage (for instance ABplaa which produces hyp4–7, H0L and H1L, or ABplap which produces left P cells, V3L, Q/V5L, TL and one hyp7). Consistent with this idea, GFP expression triggered by a given construct was generally weaker or stronger in cells related by function (seam cells versus hyp7) rather than by lineage.

A 559-bp element acts synergistically with conserved sequences

As outlined above, although some combinations of the CIS block promoted GFP expression in the major hypodermis (i.e., eFG5′), none of them did so on their own. One possibility to explain this surprising result could be that sequences within different blocks bind transcription factors that act in a synergistic manner when brought together within a single promoter. We initially made the serendipitous observation that linking the first part of the CISe block with a fragment corresponding to the second half of CISg and lir-1 exons strongly promoted GFP expression in the major hypodermis (data not shown). It suggested that the first part of CISe might reveal a weak promoter activity present in other sequences. We therefore systematically assayed this 559-bp element in combination with CISd, CISf or CISg (Fig. 6A). In addition to the pattern observed with the first part of CISe alone (see eE1, Figs. 4A,C), we observed GFP expression in all or a subset of rectal cells with CISd, CISf and CISg (Fig. 6; see eD + E1, eE1 + F, eE1 + G and eE1rev + G). The 559-bp fragment from CISe has all the hallmarks of an enhancer sequence since it acted in both orientations relative to CISg to induce GFP expression in the ventral P cells in about 50% of transgenic embryos and occa-

![Fig. 7. PHA-4 represses while ELT-1 activates lin-26 expression through its promoter. (A) Positions along plasmid pML702 of the PHA-4 response elements of high affinity (TGTTTGC or TGTTTAC), medium affinity (TGTTTGT) or low affinity (TATTTGT) (Gaudet and Mango, 2002), either in sense (above the scale bar) or in reverse (under the scale bar) orientation. GATA sites are indicated in construct eFGHi by black arrows. (B) Wild-type (a, c) and pha-4(q490) null mutant (b, d) embryos transgenic for the plasmid pML702; embryos were stained with the monoclonal antibody MH27 (c, d) which recognizes AJM-1 in the apical junctions and with an anti-GFP antibody (a, b). (C) Wild-type animal (a, b) and elt-1(zu180) null mutant (c, d) transgenic for the construct eFGHi. We scored at mid-embryogenesis 48 ± 5 GFP positive cells (n = 10) in wild-type embryos expressing eFGHi, and in elt-1(zu180) embryos 3.6 ± 1.5 GFP positive cells (n = 23) for one transgenic line, and 3.3 ± 1.5 (n = 20) in an independent transgenic line.](https://www.sciencedirect.com/science/article/pii/S109592430090702X)
sionally in Pnp cells until the L3 larval stage (Fig. 6B; eE1 + G and eE1rev + G). Attempts to map the CISc sequences that would synergize with CISg failed to define a shorter region (i.e., eE1.5 + G, Fig. 6). The second half of CISc could also drive GFP expression in P cells and in most rectal cells when placed in front of CISg (data not shown). We conclude that the first part of CISc can synergize with other sequences to drive GFP expression in additional cells, and that it does so in cells related by function (rectal cells, P cells) and not by lineage.

Ectopic GFP expression and potential negative regulatory sequences

Some of the constructs based on the myo-2 minimal promoter and unc-54 3’ UTR (vector pPD97.82) lead to ectopic GFP expression in cells that normally do not express lin-26. This was particularly prominent in the endoderm with the constructs eE1.2 (Fig. 4), eF (Fig. 6) and eHi (Fig. 5). In addition, we frequently observed ectopic GFP expression in the pharyngeal-intestinal valve cells (vpi cells) with some constructs containing CISa–b or in the pharynx (constructs eAB and eABCD). Since we never observed GFP expression in the endoderm or in the pharynx when the vector pPD97.82 was injected alone, we suggest that ectopic expression might result from the loss of certain inhibitory sequences within the

lin-26 promoter or 3’ UTR. The precise location of these sequences is not clear, although the presence of the CISg block seems to prevent endoderm expression.

lin-26 expression is repressed by the FoxA–HNF3 factor PHA-4 and activated by the GATA factor ELT-1

We and others previously showed that pha-4, the Drosophila forkhead homolog, specifies pharynx identity (Horner et al., 1998; Kalb et al., 1998). In particular, we showed that in pha-4 mutants pharyngeal precursors adopt an ectodermal fate to produce LIN-26 positive cells; and conversely, that ectopic pha-4 expression dramatically reduces the number of LIN-26-positive cells (Horner et al., 1998). An attractive hypothesis was that pha-4 represses lin-26 by binding to its promoter. To test this idea, we injected the complete GFP reporter (pML702) into the pha-4(q490) genetic background (Figs. 7A,B). We observed ectopic expression of the transgene where pharyngeal cells should normally be found, raising the possibility that repression might involve direct binding of PHA-4 to the lin-26 promoter. Computer-assisted searches for potential PHA-4 binding sites (Gaudet and Mango, 2002) revealed the presence of several sites along the promoter (see Fig. 7 legend), but we could not further define which particular site(s) was (were) more important.

The other clearly identified lin-26 regulator is the GATA factor ELT-1, which is expressed in the major

![Fig. 8. Summary of the main lin-26 promoter elements. The main lin-26 promoter regions are represented under the map of the area. The major cell types in which a given promoter (lines) is active are represented by the color code indicated in the bottom box (see Fig. 1 for positions in the embryo). The first promoter module to be active is the most proximal one relative to the lin-26 AUG codon, while modules promoting later expression are located more distally. The module CISabcd is active when linked to the lir-1–lin-26 intergenic region denoted “i”. Combining the beginning of CISe (thick scale bar) with CISd, CISf or CISg revealed dormant promoter activities not apparent when each promoter was tested on its own.](http://example.com/fig8.png)
hypodermis. In elt-1(zu180) null embryos, hypodermal cells adopt neuronal and muscle fates (Page et al., 1997). When we injected the construct eFGHi in elt-1(zu180) null mutants (Figs. 5A, 7A and Cc–d), the number of GFP-positive cells was dramatically reduced to three or four cells compared to about 48 GFP-positive cells in wild-type embryos (Figs. 7Ca–b). ELT-1 might directly bind to sequences found within eFGHi since several GATA sites are perfectly conserved between C. elegans and C. briggsae in this region (Fig. 7A).

Since three other GATA factors are known to act in the hypodermis (Gilleard and McGhee, 2001), we similarly checked whether ELT-3 or ELT-5–ELT-6, which are expressed in different major hypodermal cells, could be required to initiate or maintain lin-26 expression. The gene elt-3 is expressed in hyp7 and P cells during embryonic morphogenesis, and its ectopic expression in very early embryos can trigger lin-26 expression (Gilleard et al., 1999). The genes elt-5 and elt-6, which form a transcriptional operon, are required for the development of seam cells; elt-5–elt-6 inactivation leads to ELT-3 up-regulation in seam cells (Koh and Rothman, 2001). We performed RNA interference against elt-3 or against elt-5 and elt-6 together, but did not observe any change in GFP expression driven by the constructs containing hypodermal promoters (eABCD + i, eFGHi, and eE1) at any developmental stage. We conclude that elt-3, elt-5 and elt-6 are unlikely to play a major role in controlling lin-26 expression.

Discussion

The gene lin-26 is an early marker and a key regulator of all epithelial cells originating from the ectoderm and the mesoderm in C. elegans, but which is not expressed in epithelial cells of the foregut (pharynx) or midgut (intestine). Our analysis shows that lin-26 expression is controlled through several cis-regulatory elements spread over 11 kb. Since we used different functional assays, we are confident that elt-3 and elt-5 or elt-6 together, but did not observe any change in GFP expression driven by the constructs containing hypodermal promoters (eABCD + i, eFGHi, and eE1) at any developmental stage. We conclude that elt-3, elt-5 and elt-6 are unlikely to play a major role in controlling lin-26 expression.

lin-26 is regulated by multiple promoter elements acting in different ways

Our experiments could distinguish four major classes of regulatory sequences. First, we found that a 251-bp sequence (denoted “i” in figures), which is located immediately upstream of lin-26, is essential for expression. We propose that this sequence can provide different signals, and that lin-26 transcription might be initiated in different ways. We previously suggested that lin-26 forms an operon with lir-1, whose transcription should be initiated near lir-1 exon 2 (see Fig. 1B) (Dufourcq et al., 1999). Accordingly, this 251-bp element corresponds to the lir-1–lin-26 intergenic region and should provide signals to simultaneously trans-splice lin-26 transcripts to SL2–SL4 and polyadenylate lir-1 transcripts. This model, however, is in apparent contradiction with the observation that deletions removing lir-1 exons 2–5 and part of lir-1 first intron do not prevent lin-26 expression as long as this 251-bp region is present (Figs. 3 and 5). We do not think that this observation reflects an artifact linked to the use of transgenes, because the 3.3 kb deletion lir-1(mc33) with an endpoint in lir-1 exon 5 does not prevent endogenous lin-26 expression. One possibility would be that the 251-bp region can also act as a minimal promoter to initiate transcription downstream of lir-1. In support of this hypothesis, we identified by primer extension products with a 5′ end located 50 bases downstream of the lir-1 polyadenylation signal (i.e., 70 bases upstream of lin-26; data not shown). However, all available lin-26 cDNAs, except one, start after the 3′ acceptor splice site found right before the lin-26 initiation codon (http://www.wormbase.org) (Dufourcq et al., 1999; Labouesse et al., 1994). Therefore, if these products do correspond to transcripts initiated in the lir-1/lin-26 intergenic area, and not to intermediates generated during processing of the lir-1/lin-26 operon, they would have to be very rapidly trans-spliced to SL2–SL4.

Second, we identified for each major ectodermal epithelial cell type three distinct cis-regulatory regions, which appeared to play partially redundant functional roles in complementation assays. Although these promoter sequences became active at different developmental stages, only their simultaneous deletion affected rescue. This was the case in major hypodermal lineages, where sequences corresponding to lir-1 exons and introns became active in precursors, whereas the promoter elements CISd, CISf and CISg became active in recently born hypodermal cells. Similarly, in support cells and minor hypodermal cells, sequences corresponding to the promoter element CISe became active before CISd, CISf and CISg.

Third, the onset of lin-26 expression in some cells appeared to involve the synergistic activity of distinct sequences that were inactive on their own. This was the case in ventral hypodermal P cells and in most rectal cells by combining the first half of CISe and the CISg blocks. Examples of synergy at the level of promoter sequences are common outside C. elegans. It can involve chromatin effects or proteins that bridge and stabilize the interaction among distinct transcription factors. An example is provided by the Drosophila LIM-binding protein Chip, which interacts with the GATA-factor Pannier and the bHLH heterodimer achaete–daughterless to promote the formation of macrochaetes in a restricted area of the wing imaginal disc (Ramain et al., 2000). By analogy, a CHIP-like protein might stabilize the simultaneous binding of distinct transcription factors on CISe and CISg to drive lin-26 expression in P cells. The occurrence of

Discussion

The gene lin-26 is an early marker and a key regulator of all epithelial cells originating from the ectoderm and the mesoderm in C. elegans, but which is not expressed in epithelial cells of the foregut (pharynx) or midgut (intestine). Our analysis shows that lin-26 expression is controlled through several cis-regulatory elements spread over 11 kb. Since we used different functional assays, we are confident that they play a significant role during expression of the endogenous gene. Altogether, we identified several distinct sequence elements recapitulating the entire expression pattern of the LIN-26 protein as established with specific antibodies, including its temporal aspects.

lin-26 is regulated by multiple promoter elements acting in different ways

Our experiments could distinguish four major classes of regulatory sequences. First, we found that a 251-bp sequence (denoted “i” in figures), which is located immediately upstream of lin-26, is essential for expression. We propose that this sequence can provide different signals, and that lin-26 transcription might be initiated in different ways. We previously suggested that lin-26 forms an operon with lir-1, whose transcription should be initiated near lir-1 exon 2 (see Fig. 1B) (Dufourcq et al., 1999). Accordingly, this 251-bp element corresponds to the lir-1–lin-26 intergenic region and should provide signals to simultaneously trans-splice lin-26 transcripts to SL2–SL4 and polyadenylate lir-1 transcripts. This model, however, is in apparent contradiction with the observation that deletions removing lir-1 exons 2–5 and part of lir-1 first intron do not prevent lin-26 expression as long as this 251-bp region is present (Figs. 3 and 5). We do not think that this observation reflects an artifact linked to the use of transgenes, because the 3.3 kb deletion lir-1(mc33) with an endpoint in lir-1 exon 5 does not prevent endogenous lin-26 expression. One possibility would be that the 251-bp region can also act as a minimal promoter to initiate transcription downstream of lir-1. In support of this hypothesis, we identified by primer extension products with a 5′ end located 50 bases downstream of the lir-1 polyadenylation signal (i.e., 70 bases upstream of lin-26; data not shown). However, all available lin-26 cDNAs, except one, start after the 3′ acceptor splice site found right before the lin-26 initiation codon (http://www.wormbase.org) (Dufourcq et al., 1999; Labouesse et al., 1994). Therefore, if these products do correspond to transcripts initiated in the lir-1/lin-26 intergenic area, and not to intermediates generated during processing of the lir-1/lin-26 operon, they would have to be very rapidly trans-spliced to SL2–SL4.

Second, we identified for each major ectodermal epithelial cell type three distinct cis-regulatory regions, which appeared to play partially redundant functional roles in complementation assays. Although these promoter sequences became active at different developmental stages, only their simultaneous deletion affected rescue. This was the case in major hypodermal lineages, where sequences corresponding to lir-1 exons and introns became active in precursors, whereas the promoter elements CISd, CISf and CISg became active in recently born hypodermal cells. Similarly, in support cells and minor hypodermal cells, sequences corresponding to the promoter element CISe became active before CISd, CISf and CISg.

Third, the onset of lin-26 expression in some cells appeared to involve the synergistic activity of distinct sequences that were inactive on their own. This was the case in ventral hypodermal P cells and in most rectal cells by combining the first half of CISe and the CISg blocks. Examples of synergy at the level of promoter sequences are common outside C. elegans. It can involve chromatin effects or proteins that bridge and stabilize the interaction among distinct transcription factors. An example is provided by the Drosophila LIM-binding protein Chip, which interacts with the GATA-factor Pannier and the bHLH heterodimer achaete–daughterless to promote the formation of macrochaetes in a restricted area of the wing imaginal disc (Ramain et al., 2000). By analogy, a CHIP-like protein might stabilize the simultaneous binding of distinct transcription factors on CISe and CISg to drive lin-26 expression in P cells. The occurrence of
synergistic activities among different cis-regulatory sequences complicates the analysis and might explain why we could not identify sequences active in support cells but not in minor hypodermal cells, only in dorsal hypodermal cells or only in seam cells.

Finally, although sequences corresponding to lin-26 exons, introns and 3′-UTR do not play an essential role, we noticed that their presence made GFP expression less mosaic, suggesting a potential role in the maintenance of expression. Likewise, our identification of lir-1 exons as a major hypodermal promoter shows that exons can hide important transcription promoter elements in C. elegans.

It is reassuring that we could identify promoter elements that recapitulate the temporal activation of the endogenous LIN-26 protein in major classes of cells. LIN-26 expression first starts in the grandparent of major hypodermal cells and so does the promoter element present in the construct eFGH. LIN-26 is then activated at the time of birth in minor hypodermal cells and support cells and so does the promoter element present in the construct eC. LIN-26 is first expressed around the two-fold stage in somatic gonad precursors and so does the construct eC. Last, LIN-26 is switched on in the rectal epithelial cells repD/V at the end of embryogenesis and so is the construct eAB. Intriguingly, there is a correlation between the activation time of the various modules and their position along the chromosome upstream of lin-26 (Fig. 8). It would be interesting to determine which proteins make the promoter sequentially accessible and activated in a 3′ to 5′ direction during development.

Tissue-specific and lineage-specific regulation of lin-26

As observed with other C. elegans genes encoding proteins that influence rather than respond to cell fate decisions, lin-26 expression is regulated in a modular and piecemeal manner. This has been clearly established for unc-86, which controls the differentiation of a subset of neurons, or hlh-1, which controls and maintains the differentiation of all body wall muscles and of six glial-like cells (Baumeister et al., 1996; Krause et al., 1994). Many regulatory sequences controlling the expression of unc-86 and hlh-1 were described as acting at the level of specific lineages. For instance, there is upstream of hlh-1, a promoter element active mostly in muscles derived from the C blastomere, and another in muscles derived from the MS and D blastomeres (Krause et al., 1994). Similarly, the asymmetric regulation of unc-86 is controlled to a large extent through enhancers active in distinct cell lineages.

A surprising outcome of our analysis is that we did not identify promoter elements active in major AB or C sublineages. Instead, we found cis-regulatory elements active in cells that belong to the same organ, are functionally related or that have similar positions along the body (Fig. 8). For instance, we identified sequences active in the major hypodermis, the somatic gonad precursors, others in anterior or posterior support cells and minor hypodermal cells. Likewise, we found that a particular combination of sequences promoted expression in P cells. Although P cells do originate from two major AB sublineages (ABplap and ABprap), there are other hypodermal cells originating from these sublineages (W/G2, V3L/R and TL/R) in which the same cis-regulatory sequences were inactive. It is not clear why lin-26 would be regulated in a different way than hlh-1, as they occupy similar positions in the genetic hierarchy that controls epithelial and myogenic differentiation, respectively. One possibility could be that our analysis was not refined enough. Indeed, attempts to narrow down promoter elements resulted in seemingly inactive sequences when tested alone. Alternatively, epitheliogenesis and myogenesis might be regulated by different regulatory mechanisms. Owing to the wide variety of epithelial cells, lin-26 expression might be primarily controlled by genes that specify tissue identity.

Consistent with this notion, we found that the GATA-factor ELT-1 is essential for expression in the major hypodermis. Although we do not know whether this regulation is direct or indirect, we note that there are several WGATAR sequences in CISf, CISg and lir-1 exons, which can promote hypodermal expression (Fig. 8). We also observed that wild-type activity of the FoxA protein PHA-4, which specifies pharynx identity, is required to prevent expression of a lin-26::gfp transgene in pharyngeal cells (again, it is unclear how direct is PHA-4 activity). In addition to tissue-specific regulation, we suggest that lin-26 expression is controlled by transcription factors that determine the identity of specific regions in the embryo. While dissecting the CIS enhancer (Fig. 8), we showed that two partially overlapping fragments were active either in the most anterior or in the most posterior support cells and minor hypodermal cells. We do not know the identity of the genes that are responsible for regionalized expression. RNAi experiments indicate that the anterior and posterior Hox genes ceh-13 and nob-1/php-1 are not essential (data not shown).

In conclusion, the most salient features of the surprisingly complex lin-26 promoter include a prominence of tissue-specific regulatory modules, an abundance of redundant signals and the sequential activation of gene expression in a 3′ to 5′ direction. Epithelial differentiation in general, and epidermal differentiation in particular, are attracting considerable attention (Fuchs and Raghavan, 2002). Future studies should reveal to what extent the promoters of other key players in epithelial differentiation display similar properties.

Acknowledgments

We thank the Caenorhabditis Genetics Center and Susan Mango for providing strains. This work was supported by funds from the CNRS, INSERM, HUS, and grants to M.L. from the ARC and the EEC-TMR program. F.L. and S.Q. were supported by fellowships from the Ministère de la Recherche.
References


