Making Worm Guts: The Gene Regulatory Network of the Caenorhabditis elegans Endoderm

Morris F. Maduro and Joel H. Rothman
Department of MCD Biology and Neuroscience Research Institute, University of California, Santa Barbara, California 93106

The nematode Caenorhabditis elegans is a triploblastic ecdysozoan, which, although it contains too few cells during embryogenesis to create discernible germ "layers," deploys similar programs for germ layer differentiation used in animals with many more cells. The endoderm arises from a single progenitor, the E cell, and is selected from among three possible fates by a three-state combinatorial regulatory system involving intersecting cell-intrinsic and intercellular signals. The core gene regulatory cascade that drives endoderm development, extending from early maternal regulators to terminal differentiation genes, is characterized by activation of successive tiers of transcription factors, including a sequential cascade of redundant GATA transcription factors. Each tier is punctuated by a cell division, raising the possibility that intercession of one cell cycle round, or DNA replication, is required for activation of the next tier. The existence of each tier in the regulatory hierarchy is justified by the assignment of a unique task and each invariably performs at least two functions: to activate the regulators in the next tier and to perform one other activity distinct from that of the next tier. While the regulatory inputs that initiate endoderm development are highly divergent, they mobilize a gene regulatory network for endoderm development that appears to be common to all triploblastic metazoans. Genome-wide functional genomic approaches, including identification of > 800 transcripts that exhibit the same regulatory patterns as a number of endoderm-specific genes, are contributing to elucidation of the complete endoderm gene regulatory network in C. elegans. Dissection of the architecture of the C. elegans endoderm network may provide insights into the evolutionary plasticity and origins of this germ layer. © 2002 Elsevier Science (USA)

Key Words: gene network; endoderm; GATA factor; digestive tract.

Introduction

In the process of developing from a relatively featureless zygote into a complex multicellular organism with diverse tissue types and functions, metazoan embryos have evolved strategies to successively restrict developmental potential and assign fates to subsets of cells. During early embryonic development of most metazoans, cells are organized into three germ layers, each endowed with the capacity to engender distinct tissue and organ types. Precursor cells in each germ layer must coordinate cell movements with adjoining germ layers during gastrulation, activate a germ layer-specific transcriptional cascade, and repress genes that function in other germ layers. Further, all of these events must be coordinated within the context of active cell division. The means by which distinct tissue layers are directed to undergo specific programs of morphogenesis and differentiation must therefore be precisely controlled, while at the same time remaining flexible to evolutionary change. A paradigm for how gene regulatory networks are coordinated to dictate a restricted program of differentiation can be obtained by elaborating the network controlling the development of an entire germ layer. Moreover, as the invention of the germ layers (first, ectoderm and endoderm of diploblasts, and later, addition of the mesoderm in triploblasts) marked major evolutionary transitions during metazoan evolution, an understanding of the gene regulatory networks underlying germ layer development will also contribute important insights into the events that created these large evolutionary steps.

Based on a few representative examples, it appears that
the gene regulatory network controlling development of the innermost germ layer, the endoderm, is conserved across metazoan phylogeny. This is apparent even for an embryo in which too few cells are present at the time that the network becomes active for a "layer" to be evident. As recognized over 100 years ago, the endoderm "layer" in nematodes, members of the ecdysozoan clade of protostomes (Aguialdo et al., 1997), is established as a single cell, called E (Boveri, 1893, 1899). In the nematode Caenorhabditis elegans, this cell arises when there are only 7 cells in the embryo. The E cell undergoes no more than 5 rounds of division during embryogenesis, creating exclusively the 20 clonally derived cells of the juvenile intestine (Fig. 1) (Sulston et al., 1983). The detailed morphogenetic events leading to assembly of the intestine into a coherent organ, and its anatomical structure, have been well described (Leung et al., 1999). The regulatory program for endoderm-specific differentiation appears to be autonomous to the E cell: in the absence of all other embryonic cells, E can give rise to a full set of differentiated intestinal cells and even structural elements of a fully formed intestine (Pries and Thomson, 1987; Leung et al., 1999).

In this review, we describe the emerging information regarding the gene regulatory network that directs development of the C. elegans endoderm, with emphasis on the cascade of early regulators. We describe the key regulatory factors constituting the backbone of the regulatory network that specifies the C. elegans endoderm and our initial understanding of the network through which they operate. The genome sequence (C. elegans Sequencing Consortium, 1998), a rapid reverse genetic method, RNA-interference (RNAi; Fire et al., 1998), procedures for detecting in vivo interactions between transcription factors and their targets in identified cells (e.g., Fukushige et al., 1999), and the ready access to transcriptional profiling with DNA microarrays (e.g., Reinke et al., 2000; Jiang et al., 2000; Kim et al., 2001), complement the developmental and genetic tools available for this organism and should make it possible to elaborate the entire regulatory network for endoderm-specific differentiation in this animal.

**Specification of the Mesendoderm Progenitor, EMS, by a Maternal-to-Zygotic Switch**

The events that establish the initial anteroposterior polarity of the C. elegans embryo and create differences between early descendant cells have been described in recent reviews (Bowerman, 1998; Bowerman and Shelton, 1999; Goldstein, 2000; Gotta and Ahringer, 2001).

The zygote divides asymmetrically into the two daughter cells, called AB and P₁, and embryonic polarity is manifested as the differential ability of these blastomeres to translate maternally provided mRNAs. P₁ divides to produce the mesodermal precursor, EMS, and the germline/mesodermal precursor, P₂ (Fig. 1A). The anterior daughter of EMS, called MS, gives rise to many mesodermal cell types, including body wall muscle and the posterior half of the feeding organ (the pharynx); the posterior daughter is the endodermal progenitor, E (Sulston et al., 1983). In the past decade, the molecular mechanisms by which the identity of the EMS cell is specified and endoderm fate is subsequently restricted to the E blastomere have been elucidated (Table 1). Two maternal regulatory pathways specify the identity of EMS and its daughters. Both pathways function within EMS itself: the first activates the genes that specify both MS and E fates, while the second, which makes the E cell different from MS, is part of a reiterative switching system that directs daughters of asymmetric cell divisions to acquire different transcriptional states.

The first regulator of early blastomere fate in C. elegans to be identified was SKN-1 (Bowerman et al., 1992), a composite bZIP/homeodomain transcription factor required maternally for EMS fates (Bowerman et al., 1992; Blackwell et al., 1994). While skn-1 mRNA is contributed maternally and is found throughout all cells of the early embryo, SKN-1 protein is translated asymmetrically (Bowerman et al., 1993; Seydoux and Fire, 1994), appearing at higher levels in P₁ descendants specifically. At the four-cell stage, maternal SKN-1 protein is present in the nuclei of both EMS and P₂. However, it is required for the identity of, and functions only in, the EMS blastomere. The level at which SKN-1 function is restricted exclusively to EMS was revealed when its zygotic targets, the med-1 and med-2 genes (for mesendoderm determination), a pair of unlinked though nearly identical target genes, were identified. In EMS, SKN-1 activates expression of the med genes, marking the switch from maternal to zygotic control in mesendoderm specification (Maduro et al., 2001). The med genes encode GATA-type transcription factors, named for the degenerate consensus DNA binding site HGATAR (Lowry and Atchley, 2000) to which they bind. Regulation of med-1 and -2 by SKN-1 appears to be direct: the med promoters contain clusters of SKN-1 binding sites (Blackwell et al., 1994) that bind SKN-1 protein in vitro and that are essential for reporter expression. At high levels, SKN-1 appears to be sufficient to activate med transcription: its widespread expression throughout the embryo results in ectopic activation of the med genes (Fig. 2), which in turn are able to convert non-EMS descendants into mesendoderm-generating cells.

The mechanism that prevents SKN-1-dependent activation of the med genes in P₂, the sister of EMS, was revealed when it was found that a maternally provided transcription factor, PIE-1, acts as a global repressor of transcription throughout the germline (P) lineage (Mello et al., 1996; Seydoux et al., 1996; Batchelder et al., 1998). As such, PIE-1 blocks activation of med-1/2 by SKN-1 in the P₁ lineage (Maduro et al., 2001). While the med genes are first expressed exclusively in EMS and its early descendants, in pie-1(–) embryos, med-1/2 are activated by SKN-1 in both EMS and, inappropriately, in P₂, causing both EMS and P₂ to adopt EMS-like fates (Mello et al., 1992; Maduro et al., 2001; Teneval et al., 2001).
FIG. 1. Origin of the C. elegans digestive tract. (A) The embryonic fate map at the four-cell stage is diagrammed. The ventralmost cell, EMS, gives rise to the endoderm (E) precursor and a mesodermal precursor (MS), which produces primarily body wall muscle and the posterior half of the feeding organ, the pharynx. The anterior half of the pharynx and the entire rectum are produced by ABa and ABp, respectively. The sister of EMS, P2, gives rise to body wall muscle, hypodermis, and the germline. (B) The C. elegans intestine is clonally derived from the E blastomere. The embryonic E lineage (adapted from Sulston et al., 1983) is shown, along with a time scale marking embryonic development at 20°C. Horizontal lines indicate a cell division, while vertical lines indicate an undividing cell. Differential interference contrast (DIC) images overlaid with E-lineage-specific GFP fluorescence are shown for particular stages (indicated by the
In promoting EMS fate, it is apparent that MED-1/2 must also repress genes that direct the differentiation of other tissue types. In skn-1 and med-1/2 mutants, MS and E produce body wall muscle and hypodermis, tissue types that also arise from the somatic daughter of P2, called C (Sulston et al., 1983; Bowerman et al., 1992; Maduro et al., 2001). Specification of C fate requires the maternal CAUDAL-like transcription factor PAL-1, a protein found in all early P1 descendants, including C, MS, and E (Hunter and Kenyon, 1996). The body wall muscle produced inappropriately by MS and E in skn-1 and med-1/2 mutant embryos also requires PAL-1 function (Hunter and Kenyon, 1996; Maduro et al., 2001). The mesectodermal C-like program is therefore the default state for MS and E in the absence of mesendoderm-specifying information, owing to the presence of PAL-1. In the wild-type EMS lineage, PAL-1 C-promoting activity (and therefore acquisition of C fate) is overridden by SKN-1/MED-1/2. While it is not known at what level this inhibition of PAL-1-specific transcription functions, it seems likely that it is not the MED proteins, but rather their targets, that directly interfere with activation of PAL-1 targets, since elimination of the MED targets, end-1 and end-3, also enables PAL-1 to activate C development in the E blastomere (see below).

Like EMS, the C cell contains high levels of both SKN-1 and PAL-1. Unlike EMS, however, C produces mesectoderm rather than mesendoderm. What distinguishes C from EMS? This mechanism operates at the level of med transcription: the med genes are not expressed in C irrespective of the presence of SKN-1. The repression of SKN-1-

### TABLE 1
Genes Involved in Endoderm Formation

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product</th>
<th>EMS lineage phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos-1</td>
<td>CCCH finger</td>
<td>E, MS → C</td>
<td>Tabara et al., 1999</td>
</tr>
<tr>
<td>spin-4</td>
<td>RNA recognition motif</td>
<td>E, MS → C</td>
<td>Gomes et al., 2001</td>
</tr>
<tr>
<td>pie-1</td>
<td>CCCH finger</td>
<td>P, → EMS</td>
<td>Mello et al., 1992</td>
</tr>
<tr>
<td>sgg-1</td>
<td>GSK-3β kinase</td>
<td>C → EMS</td>
<td>Maduro et al., 2001</td>
</tr>
<tr>
<td>skn-1</td>
<td>bZIP/homeodomain TF</td>
<td>E, MS → C</td>
<td>Bowerman et al., 1992</td>
</tr>
<tr>
<td>med-1/2</td>
<td>GATA-type TF</td>
<td>E, MS → C</td>
<td>Maduro et al., 2001</td>
</tr>
<tr>
<td>end-1/3</td>
<td>GATA-type TF</td>
<td>E → C</td>
<td>Zhu et al., 1997; M.F.M. and J.H.R., unpublished observations</td>
</tr>
<tr>
<td>pop-1</td>
<td>TCF/LEF homolog</td>
<td>MS → E</td>
<td>Lin et al., 1995a</td>
</tr>
<tr>
<td>wrm-1</td>
<td>β-catenin</td>
<td>E → MS</td>
<td>Rocheleau et al., 1997</td>
</tr>
<tr>
<td>lit-1</td>
<td>Nemo-like kinase</td>
<td>E → MS</td>
<td>Rocheleau et al., 1999</td>
</tr>
<tr>
<td>mom-2</td>
<td>Wnt</td>
<td>E → MS</td>
<td>Thorpe et al., 1997</td>
</tr>
<tr>
<td>elt-2</td>
<td>GATA-type TF</td>
<td>Loss of gut integrity</td>
<td>Fukushige et al., 1998</td>
</tr>
<tr>
<td>elt-7</td>
<td>GATA-type TF</td>
<td>Unknown; redundant with elt-2</td>
<td>K. Strohmaier and J.H.R., unpublished observations</td>
</tr>
<tr>
<td>elt-4</td>
<td>GATA-type TF</td>
<td>Unknown</td>
<td>T. Fukushige and J. McGhee, personal communication</td>
</tr>
</tbody>
</table>

* Phenotypes due to loss of function.

* Among the "Mom" genes, we have only included mom-2 by way of illustration. A full description of genes involved in the endoderm-including Wnt/MAPK pathway is found in Rocheleau et al. (1997), Thorpe et al. (1997), and Thorpe et al., (2000).
dependent activation of the med genes requires SGG-1, a GSK-3β homolog (Maduro et al., 2001). In sgg-1(−) embryos, C inappropriately expresses the med genes and adopts an EMS-like fate. It is conceivable that this kinase might directly phosphorylate SKN-1, thereby abrogating its activation function; however, the well-described role for GSK-3β in the Wnt signaling pathway suggests an alternative mechanism that might involve Wnt signaling, which is required positively for E cell fate in the EMS lineage (see below).

med-1/2 are not the exclusive targets of skn-1 in EMS. While both skn-1(−) and med-1/2(−) embryos fail to specify EMS as a mesendoderm progenitor, skn-1 mutants are also defective in production of the secondary (induced) mesoderm engendered by A Ba (Bowerman et al., 1992). This effect results from the failure of MS to express an unidentified Delta-like ligand for the GLP-1 receptor that signals descendants of A Ba to make mesoderm (Priess et al., 1987; Bowerman et al., 1992). Like the med genes, this ligand is apparently expressed in both E and MS (Lin et al., 1995a). Thus, SKN-1 not only initiates mesendoderm specification in EMS by activating med-1/2, but also induces secondary mesoderm in another lineage through activation of another target. A number of genes encoding Delta-like ligands are present in the C. elegans genome, several of which contain SKN-1 consensus binding sites; these are strong candidates for other SKN-1 targets.

FIG. 2. Regulatory hierarchies demonstrated by ectopic expression experiments. Overexpression of endoderm regulators directs nonendodermal precursors to express downstream targets, demonstrating their potency as activators. Images are confocal micrographs showing expression of GFP reporters. (A) Normal expression of a med-1 reporter at the eight-cell stage in MS and E. (B) Ectopic expression of skn-1 driven by a heat-shock promoter leads to widespread med expression. (C) Expression of an end-1 reporter at the E4 stage. (D) Widespread expression of end-1 when med-1 is ectopically expressed. (E) Expression of a ges-1 reporter in the embryonic intestine. (F) Overexpression of med-1 in the early embryo leads to misexpression of ges-1 in many lineages.
A Molecular Switch Distinguishes E from Its Sister MS

Early embryological experiments provided evidence that the fate of E is specified by cell-autonomous factors that are segregated through the early lineage (Lauffer et al., 1980; Edgar and McGhee, 1986). However, while E alone appears to contain the cell-intrinsic information necessary to produce differentiated intestine, the ability of EMS to engender an endoderm-producing E cell was shown to require a cell–cell interaction between it and P2, which contacts EMS on its posterior side (Schierenberg, 1987; Goldstein, 1992). This signal induces an asymmetry in EMS that results in the adoption of endoderm fate by the daughter derived from the part of EMS that contacted the inducing cell (Goldstein, 1992). In the absence of this interaction, EMS divides symmetrically into two daughters that both exhibit an MS-like fate. The signal elicited by this interaction is mediated by intersecting Wnt and MAPK signaling pathways (Brunner et al., 1997; Shin et al., 1997; Rocheleau et al., 1997; Me- neghini et al., 1999). Deletion of any of these components results in the production of two MS-like cells by EMS. Transduction of the endoderm-inducing Wnt/MAPK signals ultimately influences the state of maternally supplied POP-1, a member of the TCF/LEF class of HMG box proteins, which serve as the terminal transcription factors in Wnt signaling pathways (Brunner et al., 1997; Korswagen and Clevers, 1999). Transduction of Wnt/MAPK signaling results in a change in nuclear POP-1 that is evident as a difference in levels: in MS, which does not receive the endoderm-inducing signal, POP-1 levels are high, while in an E cell that receives the signal, nuclear POP-1 is at low levels (Lin et al., 1995a, 1998).

Elimination of maternal POP-1 reveals its most conspicuous role in the mesendoderm: that of a repressor of endoderm in MS (Lin et al., 1995a). The Wnt/MAPK signal blocks the endoderm-repressing activity of POP-1 in the E cell: hence, this signaling system apparently induces endoderm by inactivating a repressor. Repression by POP-1 involves a complex containing the Groucho-like molecule UNC-37 and a histone deacetylase, HDA-1 (Calvo et al., 2001). Specification of the endoderm, therefore, is directed by the combined action of a positive regulator, SKN-1 (through its zygotic targets, MED-1/2), and the switching mechanism provided by Wnt/MAPK signaling through POP-1. The combination of SKN-1 and POP-1 input, combined with the default activity of PAL-1, provides a mechanism by which, via these three transcriptional regulatory inputs, the E blastomere can adopt one of three fates (Fig. 3).

Recent findings indicate that POP-1 does not act exclusively in endoderm specification as a repressor. While mutants lacking either maternal SKN-1 or zygotic MED-1/2 appear to be completely deficient in specification of the MS blastomere, only ~50% of med-1/2(RNAi) and 80% of skn-1(−) embryos lack intestine (Bowerman et al., 1992; Maduro et al., 2001). Similarly, ≤60% of embryos lacking maternal MOM-2, the Wnt ligand, do not make intestine (Thorpe et al., 1997; Rocheleau et al., 1997). However, both skn-1;mom-2 and med-1/2;mom-2 embryos completely lack endoderm, suggesting the existence of a Wnt-dependent, and SKN-1/MED-1/2-independent, input that activates endoderm development (Rocheleau et al., 1997; Maduro et al., 2001). As TCF/LEFs function as activators downstream of Wnt signaling in other systems (reviewed in Korswagen and Clevers, 1999), one candidate for such a second positive activator of E fate is POP-1. Indeed, embryos lacking both SKN-1 and POP-1 show a much more penetrant loss of intestine, compared with skn-1 mutants alone (J. Kasmir, J. Zhu, M. F. M., and J. H. R., unpublished observations). Moreover, a Lef-1-like consensus site is required for expression of a minimal E-specific promoter from the endoderm-promoting end-1 gene (see below). Both observations suggest that, rather than blocking the repressive function of POP-1 in MS per se, Wnt/MAPK signaling in the E lineage converts POP-1 from a transcriptional repressor to a transcriptional activator. This toggling of POP-1 between repressing and activating states may occur throughout the entire development of the animal: POP-1 is a component of a molecular switch used reiteratively throughout C. elegans development to establish differences between sister cells that are born by A/P divisions (Lin et al., 1998); moreover, a positive contribution of POP-1 has been implicated during larval development (Herman, 2001; Jiang and Sternberg, 1999).

 Establishment of Endoderm by end-1 and end-3

How do the MED-1/2 activators and the Wnt/MAPK/POP-1 molecular switch collaborate to activate the endoderm-specific gene network? The notion that there must exist a zygotic gene responsive to the regulators that directs E cell identity was confirmed when deletions of an interval on chromosome V (the endoderm determining region, or “EDR”) were identified that invariably eliminate endoderm. As in skn-1 and med-1/2 mutant embryos, these deletions cause E to adopt a C-like fate (Zhu et al., 1997). Several screens for zygotic mutations that block endoderm formation failed to identify any penetrant gene-specific point mutations; the consequent inference that multiple genes in the EDR might act redundantly to specify endoderm was shown to be correct when each of two nearby genes in the EDR, end-1 and end-3, was found to be individually capable of restoring endoderm formation in embryos lacking the EDR (Zhu et al., 1997; our unpublished observations). Like med-1/2, both end-1 and -3 encode GATA factors. Transcripts from both genes are first detected in E shortly after its birth (Zhu et al., 1997; our unpublished observations), indicating that the end genes are the earliest expressed genes known in the endoderm lineage. end-1 and end-3 have each been conserved (M. F. M. and J. H. R., unpublished observations) over the 20–40 million years since C. elegans and a close relative, C. briggsae, diverged (Kennedy et al., 1993), indicating selective pressure for the presence of both genes. Indeed, while the end
genes overlap in function, they are not completely redundant for endoderm specification. The only zygotic point mutation identified that leads to (impenetrant) lack of intestine alters a residue in the zinc finger of the END-3 protein (M.F.M., R. Hill, J. Priess, J. Zhu, and J.H.R., unpublished results). Moreover, while simultaneous elimination of both gene functions leads to a large fraction of embryos without endoderm, a minor fraction depleted for end-3 function alone also lack endoderm (E. Witze, M.F.M., and J.H.R., unpublished observations). One other gene in the EDR, dpr-1 (formerly end-2), which encodes a nuclear receptor type transcription factor, can also rescue the endoderm differentiation defects of EDR mutants. However, RNAi of three genes (end-1, dpr-1, and end-3) does not produce significant enhancement compared with end-1,3(RNAi) alone, and further studies indicate that dpr-1 performs a later role in endoderm development (E. Newman-Smith and J.H.R., unpublished observations).

Other Maternal Pathways Provide Sequential Permissive States for SKN-1 and MED-1/2 Action, Respectively

High levels of SKN-1 or MED-1/2 are sufficient to drive expression of their downstream targets (med-1/2 and end-1/3, respectively) in ectopic lineages, resulting in conversion of non-EMS blastomeres into mesendodermal precursors (Maduro et al., 2001; our unpublished results). However, in the context of normal EMS development, other factors are required apparently to set the permissive state for target gene activation (Fig. 4). In embryos maternally depleted for the CCCH zinc finger protein POS-1, although SKN-1 is expressed at normal levels in EMS, a med-1 reporter gene fails to be activated and E and MS adopt C-like fates, similar to med-1/2(–) embryos (Tabara et al., 1999; Maduro et al., 2001). Similarly, embryos lacking the maternal spn-4 gene function show the same defect in E and MS specification (Gomes et al., 2001); however, accumulation of SKN-1 protein and expression of a med-1 reporter are unaffected in these mutants. Rather, end-1 expression is abolished in spn-4 mutants (Gomes et al., 2001), suggesting
that SPN-4 functions at the next tier in the E and MS gene regulatory hierarchy (Fig. 4). Both pos-1 and spn-4 mutant embryos display other phenotypes, consistent with other roles in the early embryo (Tabara et al., 1999; Gomes et al., 2001). POS-1 is primarily cytoplasmic and SPN-4 contains an RNA recognition motif, suggesting that the wild-type role of these proteins in med and end gene activation, respectively, is indirect and may involve regulation of RNA metabolism or expression. In constructing a coherent picture for how endoderm fate is assigned to E, it will be important to delineate the mechanisms by which POS-1 and SPN-4 establish permissive states that allow for the sequential activation of med-1/2 and end-1/3, respectively.

Elaboration of Intestinal Fate: The end-1/3 Target Genes and GATAS Galore

As end-1 and -3 are the earliest known genes expressed specifically in the E lineage, they likely function at the top of a regulatory cascade that ultimately activates and maintains expression of terminal differentiation genes in the mature intestine. Consistent with this notion, overexpression of either end-1 or -3 in pregastrulation stage embryos is sufficient to promote endodermal differentiation from the descendants of nonendodermal precursors and to repress the differentiation of the cell types (i.e., mesoderm and ectoderm derivatives) normally made by those progenitors (Zhu et al., 1997; our unpublished observations). However, while end-1/3 expression appears to be temporally restricted to the early E lineage (no later than the E8 stage), many terminal genes are not activated until much later in development. There must be other regulators, therefore, that function downstream of end-1/3 to initiate and maintain the program of intestinal differentiation.

The elt-2 gene, which encodes another GATA-type transcription factor, appears to be one such regulator of differentiation activated by END-1/3. ELT-2 was identified by its ability to bind to a pair of tandem GATA sites in the promoter of ges-1, a gene encoding a gut-specific esterase involved in digestion (Fukushige et al., 1998; Kennedy et al., 1993). Unlike the preceding chain of regulators, skn-1, med-1/2, or end-1/3, whose expression is detectable for only a few cell generations (Bowerman et al., 1993; Maduro...
et al., 2001; Zhu et al., 1997; our unpublished observations), expression of elt-2 begins in the immediate descendants of E and continues throughout the life of the animal (Fukushige et al., 1998). The elt-2 promoter itself contains GATA binding sites (Fukushige et al., 1998), suggesting that expression of elt-2 could be initiated by END-1/3. Indeed, ectopically expressed END-1 and -3 promote widespread expression of elt-2 (Zhu et al., 1997; our unpublished results) and GFP-tagged END-3 associates in vivo with the elt-2 promoter in the E daughter nuclei during early interphase (our unpublished data).

Following activation of elt-2 transcription by the ENDS, ELT-2 apparently maintains expression of its own structural gene. Ectopic elt-2 itself can promote elt-2 expression and widespread intestinal differentiation (Fukushige et al., 1998) and GFP-tagged ELT-2 binds to the elt-2 promoter in vivo (Fukushige et al., 1999). Thus, expression of elt-2 is apparently initiated by END-1/3; however, its continued expression is maintained thereafter by positive autoregulation. An elt-2 knockout mutation results in a progressive loss of gut integrity, suggesting that, unlike the regulators that function upstream, elt-2 expression is continuously required to maintain the differentiated state (Fukushige et al., 1998).

Like the MED and END GATA factors, ELT-2 also shares a function with another factor. Although the GATA sites in the ges-1 promoter are required for expression in the intestine (Aamodt et al., 1991), the elt-2 knockout mutation does not block ges-1 expression, which suggested that another regulator acts in parallel with elt-2 (Fukushige et al., 1998). The ELT-7 GATA factor fulfilled this prediction (K. Strohmaier and J.H.R., unpublished observations). elt-7 is similar to elt-2 in its expression, ability to convert nonendodermal precursors into gut progenitor cells, and capacity to bind to its own gene. Depletion of elt-7 activity by RNAi in an elt-2(−) strain results in a synergistic phenotype: though ges-1 and many other gut-specific genes are still expressed in either single mutant, their expression is abolished when both gene functions are absent and an underdeveloped intestine is formed. Thus, like med-1/2 and end-1/3, genetic redundancy operates at the next level of the endoderm gene regulatory network.

Finally, as if six GATA factors (MEDs, ENDS, and ELT-2/7) were not enough for endoderm development in this simple creature, a tiny (72-amino-acid) GATA factor encoded by an elt-2-adjacent gene, elt-4, is also expressed in the developing intestine (T. Fukushige and J. McGhee, personal communication). The ELT-4 zinc finger is nearly identical to that of ELT-2, suggesting that they may share some function. However, the later expression of ELT-4 and its inability to activate gut differentiation ectopically suggest that it may not be functionally redundant with ELT-2/7.

The transition from END-1/3 to ELT-2/7 (and possibly ELT-4) marks a conceptual shift from specification to differentiation of the endoderm. The ENDS provide the trigger for activation of the endoderm developmental program, and repress other differentiation programs; they then hand off the job to ELT-2/7, which, by an autoregulatory loop, appear to provide a “lock-down” system for maintenance of the differentiated state.

In addition to ELT-2/7, there is evidence for additional target genes in the early E lineage acting immediately downstream of end-1/3. One hallmark of E specification is a delay in the cell cycle times of the E daughter cells, Ea and Ep, until such time as they have ingressed into the interior of the embryo, marking the onset of gastrulation (Sulston et al., 1983). In the absence of end-1/3, the E daughter cells display a more rapid cell cycle time and divide on the ventral surface of the embryo (Zhu et al., 1997; our unpublished results). As ELT-2/7 expression is detected after the E daughters have already ingressed, an additional END-1/3-dependent zygotic function, the identity of which is unknown, must exist to delay the Ea and Ep cell cycles.

Somewhat paradoxically, at least two maternal genes appear to contribute to the gastrulation function of end-1/3. Mutants lacking the function of GAD-1, a WD motif-containing protein, or EM-B-5, a putative chromatin structure regulator similar to yeast SPT6, demonstrate a failure of the E daughter cells to gastrulate properly, although endoderm is specified correctly in most embryos (Knight and Wood, 1998; Nishiwaki et al., 1993). The effect on gastrulation may be indirect, as gad-1 and emb-5 mutant embryos are pleiotropic for other developmental defects. Instead, the gastrulation defect may be attributable to their effect on the expression levels of end-1/3, as skn-1 and med-1/2 mutants that still produce endoderm often do not gastrulate correctly and there is evidence that the gad-1 and emb-5 mutations result in diminished expression levels of the end genes (E. W. Witze, M.F.M., I. Mengarelli, and J.H.R., unpublished observations). The study of mutants defective in gastrulation, therefore, may reveal new insights into the transcriptional activation of end-1/3.

Emergent Themes in Endoderm Specification

Our discussion has accounted for the core regulators known to direct endoderm specification and differentiation in C. elegans. These findings present a picture of the regulatory pathway for endoderm development that follows a familiar theme in development: cell fates are specified by stepwise restriction of developmental potential, followed by activation of instructive signals that direct terminal differentiation. We can explain the existence of the regulators in the C. elegans endoderm regulatory network based on two emergent principles as shown in Fig. 4: first, progression through successive tiers is punctuated by a cell division. Second, the regulators at each level not only activate the genes at the next tier in the cascade, but also carry out additional functions not encompassed by the next regulatory tier.

A direct correlation exists between cell division and successive levels of regulation downstream of SKN-1 (Fig. 4). The onset of med-1/2 expression is detectable in EMS in
four-cell-stage embryos; expression peaks in the E and MS nuclei (Maduro et al., 2001; our unpublished observation). The end-1 and -3 transcripts are first detectable in the E cell; expression peaks in the E daughters, Ea and Ep (Zhu et al., 1997; our unpublished observation). Finally, elt-2 and -7 expression is first detected toward the end of the Ea/Ep cell cycles (Fukushige et al., 1998; K. Strohmaier and J.H.R., unpublished observation).

Why should progression through these three levels in the gene regulatory cascade be correlated with the cell cycle? One possibility is that activation of genes at each successive tier, and establishment of a new transcriptional state, requires a complete cell cycle. This might be the case if DNA replication is a prerequisite for formation of a complex competent for transcription initiation, i.e., to allow for remodeling of chromatin and/or interaction of regulatory factors with their targets. In fact, the notion that sequential steps in the regulatory cascade are punctuated by one cell cycle was presaged many years ago when it was found that a round of DNA synthesis in the first cell cycle following birth of the E cell is essential for expression of at least one ELT-2 target, ges-1 (Edgar and McGhee, 1988; Fukushige et al., 1998). Alternatively, the correlation between transitions in the regulatory hierarchy and the occurrence of cell division may be coincidental and might simply reflect the fact that the rapid pace of cell division and sequential gene activation occur over similar time frames in the early C. elegans embryo, as can be addressed by experiments analogous to those described earlier (Edgar and McGhee, 1988).

Though we do not know whether cell division between tiers is an essential element in the deployment of the gene regulatory network, the existence of each tier is easily justified by considering the regulatory function at each step (Fig. 4). Initially, maternal factors set embryonic polarity and establish broad domains of specification, resulting in the segregation of functional SKN-1 activity to a single blastomere, EMS. This is achieved by differential translation of maternal skn-1 mRNA in P, descendants and abrogation of its activity in the non-EMS lineage in which it is translated. SKN-1 then defines EMS fate broadly by activating expression of both the next tier of regulators, MED-1/2, and at least one other target gene that encodes the ligand for induction of secondary mesoderm in the AB lineage. The combined activities of MED-1/2 and Wnt/ MAPK signaling through POP-1 lead to activation of end-1/3 specifically in E. SKN-1 is an active transcription factor in EMS; if SKN-1 were to bypass med-1/2 and activate end-1/3 directly, end-1/3 would be expressed in EMS, before POP-1 repression could occur. Hence, the intermediate regulators MED-1/2 provide a means of delaying end-1/3 activation until the POP-1 repression/activation switch can function to distinguish the EMS daughters. Conversely, if SKN-1 activity was delayed until after the division of EMS, there might be insufficient time to activate expression of the signaling ligand that induces secondary mesoderm during a narrow window of time in the AB lineage. Hence, the need to integrate other SKN-1 functions and Wnt/elt-2/7 expression occurs over similar time frames in the early C. elegans embryo, as can be addressed by experiments analogous to those described earlier (Edgar and McGhee, 1988).

Expression of Terminal Differentiation Genes

The end point of endoderm development is, by definition, the expression of gene products that function in the differentiated intestine. Several such genes have been identified through genetic and biochemical approaches, and in a few cases, their expression and promoter requirements have been studied (Table 2). As mentioned earlier, the esterase-encoding ges-1 gene is expressed when the developing gut consists of only four cells (Aamodt et al., 1991; Kennedy et al., 1993). This early expression of ges-1 appears anomalous among genes expressed exclusively in the intestine. For example, the acid phosphatase-encoding pho-1 gene is expressed beginning late in embryogenesis (Beh et al., 1991; T.
Fukushige and J. McGhee, personal communication), while the cpr-1 gene, which encodes a cysteine protease, is first expressed after hatching (Britton et al., 1998). Mutation of two GATA consensus sites in the cpr-1 promoter abolishes expression, implicating the GATA endoderm network in cpr-1 regulation (Britton et al., 1998); if ELT-2/7 or -4 are responsible for activating its expression, there is presumably an additional event that delays activation until larval development has begun. A similar situation exists with the metallothionein genes mtl-1 and mtl-2, which are activated exclusively in the intestine by exposure to cadmium or heat stress (Freedman et al., 1993). Expression of mtl-1 and mtl-2 reporters requires ELT-2 activity and GATA binding sites in the mtl promoters (Moilanen et al., 1999).

Molecular approaches, such as in situ hybridization of expressed sequence tags (ESTs) from cDNA libraries and “promoter trapping” experiments, have also identified intestine-specific genes (e.g., Tabara et al., 1996; Hope, 1991). More recently, transcriptional profiling using microarrays can, in a single experiment, reveal the relative expression of an entire suite of genes at a particular developmental stage or under a specific condition (e.g., Reinke et al., 2000). Experiments using this approach have relied primarily on mutant backgrounds, developmental staging, or environmental manipulation to generate differences between populations of mRNAs. Kim et al. (2001) pooled the results of 553 such microarray experiments and grouped transcripts that are coregulated under a variety of experimental conditions. The data are represented as a three-dimensional “terrain map” in which groups of similarly regulated mRNAs are clustered, with gene density of each cluster represented by the height, thereby creating “mountains” of correlated gene expression patterns (Fig. 5A). The 8th largest cluster represents 803 transcripts, some of which are known to be specifically enriched in the intestine, and includes genes encoding digestive enzymes such as esterase E, acid phosphatase, intermediate filament (MH33 antigen), aspartic protease, cysteine protease, and aminopeptidase P.

---

**TABLE 2**

Examples of Terminal Differentiation Genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product</th>
<th>Expression</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>ges-1</td>
<td>Esterase E</td>
<td>E to adult</td>
<td>Kennedy et al., 1993</td>
</tr>
<tr>
<td>pho-1</td>
<td>Acid phosphatase</td>
<td>Most intestinal cells, strongest from late embryogenesis</td>
<td>Beh et al., 1991; T. Fukushige and J. McGhee, personal communication</td>
</tr>
<tr>
<td>F10C1.7</td>
<td>Intermediate filament</td>
<td>Early embryo to adult</td>
<td>T. Fukushige and J. McGhee, personal communication</td>
</tr>
<tr>
<td>asp-1</td>
<td>Aspartic protease</td>
<td>Late embryonic and early larval stages</td>
<td>Tcherepanova et al., 2000</td>
</tr>
<tr>
<td>cpr-1</td>
<td>Cysteine protease</td>
<td>Postembryonic gut</td>
<td>Britton et al., 1998</td>
</tr>
<tr>
<td>W03G9.4</td>
<td>Aminopeptidase P</td>
<td>Larvae and adults</td>
<td>Laurent et al., 2001</td>
</tr>
<tr>
<td>vha-6</td>
<td>V-ATPase</td>
<td>Larvae and adults</td>
<td>Oka et al., 2001</td>
</tr>
<tr>
<td>mtl-2</td>
<td>Metallothionein</td>
<td>Inducible in larvae and adults</td>
<td>Freedman et al., 1999</td>
</tr>
<tr>
<td>vit-2</td>
<td>Vitellogenin</td>
<td>Adult intestine</td>
<td>Spieth et al., 1988</td>
</tr>
</tbody>
</table>

---

**FIG. 5.** Distribution of endoderm genes in the C. elegans gene expression map. (A) The endoderm-specifying regulators skn-1, med-1/2, end-1/3, and elt-2/7 do not cluster in this view of the entire expression terrain map (Kim et al., 2001). (B) An enlargement of the “gut mountain” (outlined by a box in A) reveals that many terminal differentiation genes (ges-1, srg-12, mtl-2, and cpr-1) show coordinate control. For both panels, Vxlnsight was used to produce terrain maps and locate individual genes, as described in Kim et al. (2001).
as proteases, carboxylesterases, and lipases (Fig. 5B). Of 8 genes known to be expressed in differentiated intestine, 5 were found in this cluster (Kim et al., 2001). At least 15 of the 803 genes can be immediately classified as transcription factors by sequence homology, 9 of which are nuclear receptors. However, there are also some unexpected groupings of endoderm regulators: elt-2 and elt-7, while clustered, are found in Mount 1, a group that includes muscle and neuronal genes, and while END-1 and -3 are in separate mountains, SKN-1 and END-1 map very closely in the same mountain. Hence, correlative transcriptional profiling is not likely to identify all genes whose expression is endoderm-specific. In addition, it is apparent that not all genes in the “gut mountain” are specific for the endoderm; for example, the egl-5 Hox gene, which controls cell fates in the posterior of the worm, maps to the gut mountain even though its expression is not gut-specific (Kenyon et al., 1997). Future microarray experiments that enrich for intestine-specific mRNAs should identify the complete set of endoderm-expressed genes and will clarify the relevance of correlated profiling analyses to the dissection of gene regulatory networks.

**Other Transcriptional Regulators Expressed in the Intestine**

Many other regulators have been identified that show expression in the E lineage; most of these are also expressed in other lineages as well. As the number is rapidly expanding and the role of these factors in endoderm development is unknown, we will not list them, but will describe a few examples here.

The best studied of these is PHA-4, an HNF3/forkhead-related transcription factor required for pharynx differentiation (Horner et al., 1998; Kalb et al., 1998). Although PHA-4 is expressed at high levels in the pharynx and rectum (where it programs organ-specific differentiation), it is also expressed at low levels in the intestine. Wide-spread overexpression of elt-2 drives ectopic pha-4 expression, and ELT-2 binds the pha-4 promoter in vitro, consistent with direct regulation by ELT-2 (Kalb et al., 1998).

Intriguingly, maternal genes that function in endoderm specification also have zygotic functions in the gut. Both skn-1 and pop-1 are zygotically expressed in the intestine (Bowerman et al., 1993; Lin et al., 1998). SKN-1 is apparently required to activate genes in response to heat or oxidative stress (K. Blackwell, personal communication), while the LIT-1/POP-1 pathway provides anteroposterior patterning information, similar to its role in other lineages (Schroeder and McGhee, 1998; Hermann et al., 2000; Lin et al., 1998).

An anticipated role for regulators expressed downstream of elt-2/7 is the control of genes that function in intestinal morphogenesis. Indeed, the intestines of embryos lacking the function of both genes generally do not contain a lumen or other features of a differentiated gut (K. Strohmaier and J.H.R., unpublished observation). During morphogenesis of the developing gut tube, three pairs of cells in the anterior portion of the gut tube rotate relative to the other cells.
is required specifically. Asymmetry is correlated with the occurrence of intestinal endoderm, and lay the groundwork for studies that correlate transcription factor activity to morphogenesis.

Several other transcription factors have been identified that are specifically expressed in the developing intestine; however, none is expressed as early as either the ENDs or ELT-2/7, suggesting that all may be subservient to these core regulators in the gene regulatory network. While the function of these transcription factors is obscure at present, the large number of endoderm-expressed genes will make it possible to dissect the regulatory interactions through which they act. Moreover, these factors are likely to reveal conserved elements in the metazoan endoderm gene regulatory network (see below).

Redundant GATA Factor “Pairs” in Germ Layer Specification

Of 11 GATA factors encoded in the C. elegans genome, 5 (EN D-1/3, ELT-2/7, and perhaps ELT-4) function specifically in endodermal development, 2 specify mesoderm (MED-1/2), and the remaining 4 (ELT-1, -3, -5, -6) function specifically in the ectoderm (Zhu et al., 1997; Fukushige et al., 1998; Maduro et al., 2001; Gilleard and McGhee, 2001; Koh and Rothman, 2001; T. Fukushige and J. McGhee, personal communication; K. Strohmaier and J.H.R., unpublished results). It is intriguing that 8 of these (MED-1/2; END-1/3; ELT-2/7; ELT-5/6) constitute genetically redundant pairs. Of the remaining 3, only elt-1 is convincingly not redundant with another gene (Page et al., 1997); deletions of elt-3 or elt-4 do not result in any recognizable phenotype (Gilleard and McGhee, 2001; J. McGhee, personal communication). Curiously, only ELT-1 contains 2 GATA-type zinc fingers, a characteristic feature of the vertebrate GATA factors (Lowry and Atchley, 2000). The carboxyl zinc finger of vertebrate GATA factors is involved in binding to a cognate GATA site, and shares the most homology with the carboxyl finger of ELT-1 and the single fingers of the remaining C. elegans GATA factors (Lowry and Atchley, 2000; reviewed in Newton et al., 2001). However, new data suggest that the amino finger, which has previously been described as being important for interactions with coregulators, can also interact with DNA (Newton et al., 2001). One possibility, then, is that the single-finger GATA factor pairs achieve a higher promoter-binding specificity through heterodimeric interactions, perhaps necessitating the existence of pairs of partially redundant, single-finger GATA factors.

Endoderm Regulatory Networks in Other Nematodes

The elucidation of the regulatory network underlying C. elegans endoderm development provides a strong base from which to explore the mechanisms that drive evolutionary change within such networks. Comparisons of noncoding regions between C. elegans and C. briggsae have pointed to regions important for regulation (e.g., Kennedy et al., 1993; Gower et al., 2001; MacMorris et al., 1994). Moreover, at least some components of endoderm specification appear to be conserved in C. briggsae. At the top of the cascade, C. briggsae contains a skn-1 homolog (Kent and Zahler, 2000). A C. briggsae end-1 homolog, which can rescue endoderm formation in C. elegans EDR-deficient mutants, more closely resembles C. elegans end-1 than C. elegans end-3, indicating that the divergence of end-1 and end-3 occurred before the briggase/elegans split (J. Kasmir, J. Zhu, M. F. Madura, and J. H. Rothman, unpublished observations). The C. briggsae ges-1 gene is expressed in a similar manner when transgenes are introduced into either C. briggsae or C. elegans, implying conservation of the upstream regulatory network (Kennedy et al., 1993). While the 5' flanking sequences are not well conserved, several short regions show high conservation, including a 17-mer sequence that is completely conserved between the two species (Kennedy et al., 1993). Intriguingly, deletion of a conserved promoter element, containing two GATA sites, in either species produces an expression component in the pharynx and tail, suggesting that a similar mechanism blocks ges-1 activation outside the intestine (Egan et al., 1995). As expected, such studies have identified more similarities than differences and one suspects that the C. elegans and C. briggsae endoderm regulatory networks will be largely identical. As such, C. briggsae may be more useful for identifying putative regulatory regions by sequence conservation, as a “proving ground” for models advanced by work in C. elegans, and a tool for identifying mechanisms that drive microevolutionary changes in the architecture of cis-acting regulatory elements. In some cases, the genetic redundancy within the core endoderm regulatory network means that such short time-scale changes can also be inferred from a single species. For example, although the med genes are nearly identical, the divergences are focused in noncoding regions and exclude known sites for interaction with transcriptional regulators, including SKN-1 and GATA factors.

Analyses in more deeply divergent nematodes have provided insights into developmental strategies that have been adopted in evolution to restrict endoderm to the appropriate blastomere during early embryogenesis. Early embryogenesis in C. elegans is highly mosaic, particularly with regard to generation of endoderm from the E cell. Even in the
marine nematode Enoplus brevis, which shows highly indeterminate development, E has the cell-intrinsic capacity to make intestine (Voronov and Panchin, 1998). However, unlike C. elegans, a related and morphologically similar nematode, Acroboloides nanus, shows highly regulative development, which extends even to specification of the endoderm precursor (Wiegner and Schierenberg, 1998, 1999): endoderm fate can be assigned to any early blastomere cultured in isolation. Thus, unlike C. elegans, repressive intercellular interactions prevent the acquisition of endoderm fate in A. nanus, in which all early blastomeres are potentiated to make endoderm. With the core C. elegans endoderm specification components in hand, it will be of interest to ask at what level the regulatory mechanisms that restrict endoderm to the E cell lineage diverge between A. nanus and C. elegans.

Conservation of an Ancestral Endoderm Regulatory Networks

All metazoan phyla, with the exception of the sponges, contain an endoderm. As this germ layer was likely invented once during animal evolution, one supposes that at least some elements of the gene regulatory network regulating formation of the endoderm are pervasive across phylogeny. The scattered evidence available is consistent with this postulate. In particular, GATA transcription factors play a key role in endoderm development in many other systems. A striking parallel is seen with the Drosophila SERPENT GATA factor (Rehorn et al., 1996): mutants lacking this gene fail to generate a midgut, the endodermal portion of the digestive tract, and cells that would become endodermal are apparently converted to ectoderm (Reuter, 1994). As in C. elegans, a second GATA factor, dGATAc, apparently acts downstream of this endoderm-specifying GATA factor (Lin et al., 1995b). The GATA family of transcription factors in vertebrates consists of two groups of genes defined by sequence homology and expression pattern. GATA1, -2, and -3 function in hematopoiesis (reviewed in Orkin and Zon, 1997), while GATA4, -5, and -6 function primarily in cardiac and endoderm development (reviewed in Charron and Nemer, 1999). In the zebrafish, faust/GATA5 is required for the formation of the gut tube and other endodermal organs (Reiter et al., 1999, 2001) and overexpression of GATA5 in Xenopus can respecify cells of mesodermal or endodermal origin towards an endoderm fate (Weber et al., 2000). In addition, the C. elegans END-1 GATA factor is capable of activating endoderm development in Xenopus animal caps, which would otherwise produce exclusively ectoderm (Shoichet et al., 2000).

The expression of a number of other transcription factors in the C. elegans endoderm is also consistent with substantial conservation of the endoderm gene regulatory network. For example, members of the odd-skipped (K. Strohmaier and J.H.R., unpublished observations), HNF4 (K. Koh and J.H.R., unpublished observations), Sox (Hanna-Rose and Han, 1999), and HNF3 (Horner et al., 1998; Kalb et al., 1998) gene families are expressed in the C. elegans endoderm and in the digestive systems or developing endoderm of other animals ranging from Drosophila to vertebrates (Zhong et al., 1993; Morrisey et al., 1998; Stainier, 2002). It remains to be seen whether this similarity in expression is superficial or instead reflects genuine conservation of an endoderm regulatory network.

Finally, based on the similarities between C. elegans EM5 specification and mesendoderm development in vertebrates, it has been proposed that the conjunction of endoderm and a subset of mesoderm as a mesendoderm layer is of ancient origin (Rodaway and Patient, 2001). Thus, while the inputs into the regulatory gene cascade are different, there is good evidence to suggest that they impinge upon a common, conserved pathway.

In comparing the known pathways for endoderm development in diverse animals, a general theme emerges: the regulatory inputs that initiate the program for endoderm are dramatically divergent, yet these divergent inputs may activate a well-conserved gene regulatory network that functions in all metazoans. For example, though endoderm differentiation in both Drosophila and C. elegans appears to be regulated by a cascade of GATA factors and other similar regulators, including HNF-4-like and Fkh-like factors, the early events that first select cells for the endoderm pathway are entirely distinct, involving the HKB Sp1/egr-like gap gene product (Bronner et al., 1994) in the former and SKN-1 and Wnt/MAPK signaling in the latter (Bowerman et al., 1993; Rocheleau et al., 1997, 1999; Thorpe et al., 1997).

If, as seems likely, there exists a conserved endoderm gene regulatory network across metazoans, in which not only specific components, but also the logic of cross-regulatory interactions and genetic circuitry are common, then it will be of great value to reveal the common parameters of the network. Elucidating the entire endoderm regulatory networks in disparate species will not only reveal limits on the degree of conservation and the rules for diversification of gene networks, but may also provide clues as to the steps by which germ layers were invented during the transition from a germ layer-less ancestral metazoan form (perhaps a sponge-like organism) to a diploblast, containing two germ layers. The plethora of transcription factors present in the endoderm, not only in, for example, sea urchins (e.g., Yuh et al., 1998), in which the endoderm consists of many cells, but also in C. elegans, which contains one of the simplest endoderm layers of any animal may prove revealing as to the evolutionary steps leading to assembly of the network.

Why does the C. elegans endoderm, consisting of only 20 nearly identical cells at hatching, express so many transcription factors? Would it not be sufficient to maintain a simple, core regulatory cascade that activates the complete set of intestine-specific differentiation genes? A simplified model that could account for this apparent plethora of transcription factors in the endoderm proposes that the gene regulatory network might have been assembled during evolution from preexisting gene batteries present in a
primordial unicellular eukaryote (Fig. 6). As the endoderm in all animals gives rise to an intestine, one set of differentiation genes expressed in the endoderm of all animals would be those involved in digestion. In unicellular eukaryotes, there exist many independently regulated gene batteries, the members of which act in common metabolic processes (e.g., nitrogen or phosphate metabolism; Marzluf, 1997; Lenburg and O’Shea, 1996). These batteries are coupled together under the control of a single battery regulator, which itself is responsive to varying environmental conditions (e.g., GATA factors in nitrogen metabolism).

Each battery regulator acts through recognition sites specific for the regulatory domains of all members of that particular gene battery. It would be possible, in principle, to create an organ in a multicellular creature that is constitutively dedicated to digestion by tying together the gene batteries under the control of a single, organ-specific regulator. Each of the individual gene batteries could become subordinated to the organ-specific regulator (e.g., an endoderm-promoting GATA factor) by a single step in which a binding site for this regulator (e.g., a GATA site) is created in the regulatory domain of the battery regulator. Although of relatively minor importance to the overall function of the organ, each subordinate regulator would then carry with it a set of subservient genes with their cognate recognition sites. These intermediaries, which link the global, organ-specific regulator to the gene batteries, would be maintained over long evolutionary periods, since a prohibitively large number of changes would be required to alter the recognition sites of the entire set of genes in each battery to one appropriate for the organ-specific regulator. If this model is correct, the gene networks in present-day metazoans might well be palimpsests that reveal underlying unicellular gene batteries subsumed within the larger regulatory networks of organs or germ layers. This proposal will be tested once the complete architecture of the gene regulatory network for endoderm is elaborated and individual circuits are compared to those in present-day unicellular eukaryotes.

Conclusion and Future Prospects

From the analysis of the core pathway for endoderm development in C. elegans, several principles have emerged that seem to guide the construction of the general transcriptional regulatory network for endoderm. It remains to be determined whether these principles actually reflect underlying mechanisms by which gene regulatory events are rapidly and sequentially deployed. As pathways for the specification of other early blastomeres and their descendant lineages in C. elegans are elucidated, it will become evident, for example, whether a cell division-linked gene cascade typifies such pathways in general.

With the identity of the main regulators in hand, the identification of ~800 genes that appear to be coregulated with gut-specific genes based on correlated transcriptional profiling and genome-wide methods for rapidly assessing gene function (Gonczy et al., 2000; Fraser et al., 2000), the C. elegans endoderm is likely to be among the first examples in which a transcriptional regulatory network for germ layer differentiation will be completely elucidated. The goal of such a task is to learn how such networks are assembled, why they are so complex, what features are common to all regulatory networks, how such networks might have evolved from simpler circuits, and to what extent the details of the network are conserved or subject to evolutionary plasticity.

ACKNOWLEDGMENTS

We are grateful to K. Blackwell, T. Fukushima, and J. McGehee for communicating unpublished results and to members of the Rothman laboratory for comments on the manuscript. M.F.M. was supported by a fellowship from the Natural Sciences and Engineering Research Council of Canada. This work was supported in part by grants from the NIH (HD37487) and the March of Dimes (to J.H.R.).

REFERENCES


