Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development

Bruce Wightman, Thomas R. Bürglin, Joseph Gatto, Prema Arasu, and Gary Ruvkun

Department of Molecular Biology, Massachusetts General Hospital and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02114 USA

The heterochronic gene lin-14 controls the temporal sequence of developmental events in the Caenorhabditis elegans postembryonic cell lineage. It encodes a nuclear protein that normally is present in most somatic cells of late embryos and L1 larvae but is absent at later stages. Two lin-14 gain-of-function mutations delete 3'-untranslated sequences causing an inappropriately high level of the lin-14 nuclear protein late in development. These mutations identify a negative regulatory element that controls the formation of the lin-14 protein temporal gradient. The 21-kb lin-14 gene is differentially spliced to generate three lin-14 transcripts that encode protein products with variable amino-terminal regions and a constant carboxy-terminal region. The sequence of the gene revealed no protein sequence similarity to any proteins in various data bases.

[Key Words: C. elegans; lin-14; heterochronic gene; gain-of-function mutation; 3' UTR]

Received June 3, 1991; revised version accepted July 22, 1991.

Regulatory genes have been shown to control the spatial and temporal configuration of the many cell types generated during development in Caenorhabditis elegans, Drosophila, Xenopus, and presumably all metazoans. Many of these genes are expressed in striking spatial or temporal patterns that are consistent with the developmental defects caused by mutations in the corresponding genes (Costa et al. 1988; Ingham 1988; Ruiz i Altaba and Melton 1989; Ruvkun and Giusto 1989). A major step in the dissection of the molecular basis of pattern formation is to discern how these genes become activated in spatial and temporal patterns.

Temporal pattern formation genes have been identified by heterochronic mutations in C. elegans, which perturb the time at which many cell lineages and cell types are generated during postembryonic development [Ambros and Horvitz 1984, 1987; Ambros 1989]. The lin-14 heterochronic gene plays a central role in this temporal regulation of the cell lineage. Loss-of-function [lf] lin-14 alleles cause the precocious execution of cell lineages normally observed in descendent cells one or two larval stages later. Gain-of-function [gf] lin-14 alleles affect the same cell lineages but cause the opposite transient in cell fate: Early cell lineages are normal, but later cells reiterate the early cell lineages normally associated with their ancestor cells. For example, in the development of the lateral hypodermis of lin-14(lf) mutants, the blast cell T skips its characteristic larval stage 1 (L1) sublineage and, instead, expresses a sublineage normally associated with its granddaughter blast cell T.ap. In lin-14(gf) mutants, the blast cell T expresses its normal L1-specific sublineage, but during L2 its granddaughter T.ap reiterates this L1-specific sublineage normally associated with blast cell T (Ambros and Horvitz 1984). The lin-14(gf) mutations have been shown by genetic criteria to cause more lin-14 gene activity than normal, whereas the lin-14(lf) mutations are due to loss of lin-14 gene activity (Ambros and Horvitz 1987). These data suggest that during normal development a relatively high lin-14 gene activity during early larval stages, for example, in cell T, is reduced later in development, for example, in cell T.ap, to form a temporal developmental switch.

The product of the lin-14 gene is a nuclear protein, the abundance of which forms a steep temporal gradient during C. elegans development: It is present at high levels in most but not all nuclei of embryos and L1 animals but is undetectable at later stages [Ruvkun and Giusto 1989]. For example, in the case of the T lineage, during normal development the lin-14 protein is present at high levels.
in the T-cell nucleus but is not observable in the T.ap-cell nucleus. In lin-14(gf) mutants, this temporal gradient is disrupted: The lin-14 protein is observed at all stages of development (Ruvkun and Giusto 1989), for example, in both cells T and T.ap (as well as T.apap, etc.), and these cells reiterate L1-specific cell lineages. In some lin-14 null mutants, no lin-14 protein is detectable, and cells express L2-specific lineages during the L1 stage (Ambros and Horvitz 1987; Ruvkun and Giusto 1989). Thus, the normally sharp decrease in the cell nucleus. In lin-14(gf) mutants, this temporal gradi-ent is disrupted: The lin-14 protein is observed at all stages of development. The lin-14 protein is detectable, and cells express L2-specific lineages during the L1 stage.

Results
lin-14 gene structure and DNA sequence
To determine the wild-type lin-14 DNA sequence and changes in selected lin-14 mutants, we isolated cDNA clones from mixed stage and embryonic wild-type [N2] C. elegans and from a mutant strain [MT1842 = lin-14(n536n838)], which contains a lin-14(gf) mutation [n536] and a cis-acting lin-14(lf) mutation [n838]. In addition to the cDNAs, the corresponding genomic regions from wild type were sequenced to determine the locations of introns and exons (Figs. 1 and 2). The sequences of these cDNAs and the corresponding genomic regions revealed that the lin-14 gene contains 13 exons with intron sizes varying from 44 bases to 12 kb. Differential splicing of exons 1–3 or exon 4 to the common exons 5–13 was observed in the cDNAs (Figs. 1 and 2). Exon 4 is located within the 12-kb intron (between exons 3 and 5) of the other transcripts and is separated by a 101-bp intron from exon 5. Polymerase chain reaction (PCR) with primers specific for exons 1 and 5 revealed a third transcript, which contains exons 1 and 3 spliced to exon 5 but lacks exon 2 (data not shown). The presence of all three transcripts was verified by PCR and RNase protection by using primers and probes from exons 1–5 (see Materials and methods). We designate the exon 4-containing transcript lin-14A, the exon 1,2,3-containing transcript lin-14B1, and the exon 1- and 3-containing transcript lin-14B2. The total lengths of the lin-14B1, lin-14B2, and lin-14A transcripts were inferred from cDNAs and genomic DNA sequence to be 3272, 3197, and 3338 bases, respectively, without the poly[A] tail, all of which are similar to the 3.5-kb lin-14 mRNA species detected with probes specific to either transcript (Ruvkun et al. 1989). PCR with primers specific for exons 1, 4, and 5 showed that no lin-14 transcripts containing exons 1 and 4 are produced, whereas, as also shown by the existing cDNA clones, lin-14 transcripts containing exons 1 and 5 and exons 4 and 5 are produced (data not shown). In addition, no consensus splice acceptor sequences are located within 140 bases of the 5’ end of exon 4. Although the 5’ ends of these transcripts have not been determined, these data suggest that the lin-14A transcript and the lin-14B1/B2 transcripts may have distinct 5’ ends and may use distinct promoters.

lin-14 proteins
The longest open reading frame in the lin-14B1 transcript starts with an ATG in exon 2 at position 3193 of
Gain-of-function mutations in \textit{lin-14}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure2.png}
\caption{Figure 2. [See following page for legend.]}\label{fig:2}
\end{figure}

\begin{table}[h]
\centering
\begin{tabular}{|c|c|}
\hline
Mutation & Amino Acid Change \\
\hline
--- & --- \\
\hline
\end{tabular}
\caption{Gain-of-function mutations in \textit{lin-14}}\label{tab:mutations}
\end{table}
Because of this clustering of cDNA clone 5' ends, we tentatively designate the 5'-most exon of cDNAs 518, 538, and e31 as exon number 1. After nucleotide 480, there is an intron gap in the sequence of -2.3 kb, on the basis of restriction mapping. We therefore begin numbering again at 2701, where a genomic EcoRI fragment begins. Similarly, throughout the 5' end of the gene, wherever there determined. The 5'-most eDNA isolated [538 from these introns match the consensus for are more extended and conserved than in other higher eukaryotes (Emmons 1988), and all intron splice acceptor and donor sites in region in the 3' UTR, were represented in at least one cDNA clone. The consensus sites for transcripts, was detected (Krause and Hirsch 1987)...

Figure 2. The DNA sequence of the lin-14 gene. Nucleotides are numbered from the 5'-most genomic sequence that we have determined. The 5'-most cDNA isolated [538 from lin-14(n536n838)] starts just to the left of the 5.0-kb BglII fragment [position 259]. Two other independent cDNAs, e31 [from WT] and 518 [from lin-14(n536n838)], have their 5' ends within 29 bases of that position. Because of this clustering of cDNA clone 5' ends, we tentatively designate the 5'-most exon of cDNAs 518, 538, and e31 as exon number 1. After nucleotide 480, there is an intron gap in the sequence of -2.3 kb, on the basis of restriction mapping. We therefore begin numbering again at 2701, where a genomic EcoRI fragment begins. Similarly, throughout the 5' end of the gene, wherever there is an intron gap, numbering begins at a position estimated from restriction mapping. All exons are marked and, except for a 95-base region in the 3' UTR, were represented in at least one cDNA clone. The consensus sites for C. elegans splice donor and acceptor sites are more extended and conserved than in other higher eukaryotes (Emmons 1988), and all intron splice acceptor and donor sites in these introns match the consensus for C. elegans splice sites well. No trans-spliced leader, like those observed in other lin-14 starts just to the left of the 5.0-kb BglII fragment [position 259]. Two other independent cDNAs, e31 [from WT] and 518 [from lin-14(n536n838)], have their 5' ends within 29 bases of that position. Because of this clustering of cDNA clone 5' ends, we tentatively designate the 5'-most exon of cDNAs 518, 538, and e31 as exon number 1. After nucleotide 480, there is an intron gap in the sequence of -2.3 kb, on the basis of restriction mapping. We therefore begin numbering again at 2701, where a genomic EcoRI fragment begins. Similarly, throughout the 5' end of the gene, wherever there is an intron gap, numbering begins at a position estimated from restriction mapping. All exons are marked and, except for a 95-base region in the 3' UTR, were represented in at least one cDNA clone. The consensus sites for C. elegans splice donor and acceptor sites are more extended and conserved than in other higher eukaryotes (Emmons 1988), and all intron splice acceptor and donor sites in these introns match the consensus for C. elegans splice sites well. No trans-spliced leader, like those observed in other C. elegans transcripts, was detected (Krause and Hirsch 1987). The initiator methionine for the lin-14 open reading frames on the lin-14A, lin-14B1, and lin-14B2 transcripts are tentative. The initiator reading frame upstream of the lin-14A initiator methionine region is shown (Legend continued on following page)
the genomic DNA sequence and ends at position 19,919, yielding a protein of 539 amino acids (Fig. 3) (59,091 daltons). Two codons upstream of the first AUG is a stop codon, which confines the reading frame. Splicing of exon 1–3 in the lin-14B2 transcript generates a stop codon in-frame at the splice junction. The first AUG downstream of this junction yields a smaller protein of 515 amino acids (56,718 daltons), which could also be produced by the lin-14B1 transcript. In the lin-14A transcript, exon 4 contains an open reading frame that splice to the common reading frame in exon 5 (Fig. 3). The reading frame is open from the 5′ end of exon 4 and can be extended >236 bases farther upstream on the genomic DNA sequence to a stop codon at position 15,788. The reading frame is open from the 5′ end of exon 4 and can be extended >236 bases farther upstream on the genomic DNA sequence to a stop codon at position 15,788. The first methionine codon downstream of the stop codon at 15,788 is found at position 16,148, suggesting that this may be the amino terminus of a 537-amino-acid lin-14A protein (59,305 daltons). This protein has a 63-amino-acid amino-terminal domain that is distinct from the 65-amino-acid amino-terminal domain of the putative lin-14B1 protein, and the 38-amino-acid amino-terminal domain of lin-14B2. No sequence similarity between the different amino termini of the lin-14 proteins could be detected. All of the putative initiation codons compare poorly with the C. elegans consensus start codons (C. Fields, pers. comm.). The predicted lin-14A and lin-14B1 proteins are so similar in molecular mass that they should not be resolvable on SDS gels. In fact, only one migrating species at ~67 kD is detected on immunoblots with anti-lin-14 antibodies (Fig. 4). Because the apparent molecular mass of this species is so different from those predicted by the DNA sequence, we could not assess which transcripts encode the proteins detected.

Data bank searches with the amino acid sequences of both lin-14 proteins revealed no significant sequence similarity to any proteins. Because the lin-14 protein is nuclear, we also examined the sequence closely for motifs identified in other nuclear proteins (Ginsberg et al. 1984; Bell et al. 1988; Evans and Hollenberg 1988; Murre et al. 1989; Scott et al. 1989) and found no similarity. A 20-amino-acid region (positions 417–436) common to both lin-14 proteins could potentially form an amphipathic α-helix with two basic regions separated by an acidic region on the hydrophilic face (Fig. 3).

The lin-14 proteins also contain a high proportion of proline, glutamate, serine, and threonine, or PEST residues, as has been observed in various unstable proteins (Rogers et al. 1986; Nash et al. 1988). Given that the protein appears to be quickly degraded late in L1 (Ruvkun and Giusto 1989), these sequences could mediate that instability.

lin-14[n355] and lin-14[n536] gain-of-function mutations are located in the 3′ UTR

Both lin-14(gf) mutations were shown previously to map within the 3.8-kb EcoRI fragment at the 3′ end of the gene (Ruvkun et al. 1989). We isolated genomic DNA segments from this region from both lin-14(gf) mutants n355 and n536 to pinpoint the location of each mutation by DNA sequencing. The n355 mutation is an insertion or inversion of at least 10 kb of unknown DNA sequences at position 20,172, 256 bp 3′ to the termination codon of the lin-14 protein-coding region common to all lin-14 transcripts (Figs. 1 and 2). No other change was found between this breakpoint and the end of the lin-14 protein-coding region. This breakpoint separates the distal 1.35 kb of the 3′ UTR and the normal polyadenylation signal site from the proximal part of the lin-14 transcripts.

The other lin-14(gf) mutation, n536, is a 607-bp deletion starting at position 20,216, 300 bp downstream from the 3′ end of the lin-14 open reading frame (Fig. 2). In addition, a C → T transition at position 20,167, 7 bases upstream of the n355 breakpoint, was detected. We did not determine which of these two changes causes the lin-14 mutant phenotype. The region deleted by the n536 mutation overlaps the region that is rearranged in the n355 mutation.

To verify that these mutations affected the transcript as well, we examined the lin-14 message detected on Northern blots of RNA from each mutant. As shown previously, instead of the normal 3.5-kb lin-14 RNA, n355 mutants produce two species of 3.0 and 2.2 kb, and

Figure 2. (Continued)
in italics. A gap of 95 bp separates the end of the e31 cDNA from the next cDNAs farther downstream (549 and 505; both isolated from wild type). These cDNAs also were generated from the lin-14 transcript; the 549 cDNA hybridizes on Northern blots to the same size mRNAs as the e31 cDNA: 3.5 kb in wild-type and 3.0 kb in lin-14(n536) mutants. Because no good splice acceptor or donor sites are present in the 95-base interval, we can assume that this region is transcribed without interruption. The 570 cDNA also hybridizes to the 3.5-kb lin-14 mRNA and extends farther downstream, ending in a 27-base-long poly[A] tail. A good polyadenylation signal site (AATAAA) is present 11 bp upstream from the poly[A] tail. The lin-14B1 transcript that we assemble from these cDNA clones is 3272 bases long without the poly[A] tail. The sequence of the lin-14B2 transcript was determined by the sequencing of a clone derived from PCR of N2 cDNA. The lin-14A transcript was identified by the 557 cDNA clone. This transcript is 3338 bases long. The locations of the lin-14(n536) and lin-14(n355) mutations are shown above the wild-type sequence. The location of the n383 mutation was ascertained from three cDNA clones isolated from this mutant and compared to the wild-type DNA sequence in this region from one cDNA and genomic clones, as well as the sequence of this region from the genomic clone from the n536 mutant. DNA sequence analysis of the fragment inserted into the lin-14 3′ UTR in the lin-14(n355) mutant revealed a consensus C. elegans splice donor site 7 bases into the rearrangement. This donor site may be spliced to an acceptor site located ~800 bases 3′ in the inserted DNA to generate the 2.2-kb transcript, whereas the 3.0-kb transcript may be not spliced at this point. The inverted repeat in the 3′ UTR is shown below the sequence with arrows for each stem-loop. Within the main loop are two smaller stem-loops, also shown with inverted arrows. Restriction sites for EcoRI and BglII are shown in bold. Italicized boldface type is used to demark the polyadenylation site. The sequence has been entered into GenBank/EMBL; accession numbers are x60230, x60231, x60232, and x60233.
Wightman et al.

Figure 3. Protein sequence of lin-14. The three possible lin-14 proteins predicted from the lin-14 DNA sequence. The lin-14A-specific amino-terminal region is shown above the lin-14B1- and lin-14B2-specific protein. The crossover from the carboxyl terminus of the lin-14A-specific region to the amino terminus of the common region shows how the lin-14A protein sequence continues in the lin-14 common carboxy-terminal region. The star above position 25 designates the first methionine present in the predicted lin-14B2 protein, as splicing of exon 1-3 generates an in-frame stop codon just 5' to this methionine codon. Both lin-14B1 and lin-14B2 products could use this methionine as a translation initiator. Basic residues are shown in boldface type. Regions containing a high proportion of PEST residues are underlined; the region containing the n-14B2-specific protein. The crossover lin-14(n355) n536 predicted from the end of the lin-14 protein-coding region and -69 bases upstream of the n355 breakpoint [Fig. 6]. This structure would be expected to have a free energy of -38.3 kcal/mole. Randomization of the 107-base region that comprises this stem-loop generated structures of a mean free energy of -15 kcal/mole ± 2.5 (n = 53).

Figure 4. Immunoblot with anti-lin-14 and anti-myosin antibodies. Immunoblots with antibodies raised against the lin-14A protein [Ruvkun and Giusto 1989] revealed proteins of 67 kD in L1 animals from both wild-type and the two gain-of-function mutants n536 and n355. These proteins are not detected in the lin-14(n536n838) mutant [Fig. 7e; Ruvkun and Giusto 1989]. Immunoblot analyses of the lin-14 proteins from wild-type and both the gain-of-function mutants support this prediction: No difference in the size of the lin-14 proteins was observed in these mutants [Fig. 4].

In the case of the n355 mutation, either the deleted 1.35 kb of lin-14 3' UTR or the additional foreign sequences fused to the 3' end of the lin-14 message could cause the inappropriate production of lin-14 protein after the L1 stage. However, given that the n536 deletion causes a similar gain-of-function phenotype, these data suggest that the 3' UTR of lin-14 contains a negative regulatory element that mediates the down-regulation of lin-14 protein levels after mid-L1. Both lin-14(gf) mutations delete this element, so that in these mutants the level of the lin-14 protein is higher than normal after L1.

Because an element in the 3' UTR might be expected to act on the message, the sequence of the lin-14 3' UTR was examined for structures of possible significance in the RNAs. Computer analysis of the lin-14 RNA sequence identified one significant secondary structure in the 3' UTR. A stem-loop structure with 21 of 22 matching pairs in the main stem and two side stems of 7 bp is located from 19,980 to 20,084, ~59 bases downstream from the end of the lin-14 protein-coding region and ~69 bases upstream of the n355 breakpoint [Fig. 6]. The location of both lin-14(gf) mutations in the 3' UTR suggests that they do not affect the lin-14 protein. Immunoblot analyses of the lin-14 proteins from wild-type and both the gain-of-function mutants support this prediction: No difference in the size of the lin-14 proteins was observed in these mutants [Fig. 4].

In the case of the n355 mutation, either the deleted 1.35 kb of lin-14 3' UTR or the additional foreign sequences fused to the 3' end of the lin-14 message could cause the inappropriate production of lin-14 protein after the L1 stage. However, given that the n536 deletion causes a similar gain-of-function phenotype, these data suggest that the 3' UTR of lin-14 contains a negative regulatory element that mediates the down-regulation of lin-14 protein levels after mid-L1. Both lin-14(gf) mutations delete this element, so that in these mutants the level of the lin-14 protein is higher than normal after L1.
Figure 5. Northern blots of wild-type and gain-of-function mutants. Poly(A)+ RNA from N2 (lane 1), lin-14(n355) (lane 2), and lin-14(n536); lin-4(e912) (lane 3) animals, probed with cDNA 549 [A], a wild-type clone that spans the location of the n355 and n536 breakpoints, and with cDNA 570 [B], which maps to the right of the breakpoints (see Fig. 2). The 2.2-kb lin-14(n355) mRNA hybridizes to the probe to the left of the breakpoint but not to the probe to the right of the breakpoint. The 3.0-kb band in n355 is only faintly apparent in this blot. The 3.0-kb lin-14(n536) mRNA hybridizes to both probes. The lin-4(e912) mutation has no effect on lin-14 mRNA size or level (data not shown). The lanes shown are all from one Northern blot probed sequentially with the two probes.

Discussion

A negative regulatory element in the lin-14 3' UTR is necessary to generate the lin-14 temporal switch

We have demonstrated that the normal down-regulation of the level of lin-14 protein, which occurs during late L1 and causes cells to switch from L1-specific fates to L2-specific fates, depends on a regulatory element in the 3' UTR of the lin-14 gene. Two lin-14(gf) mutations inappropriately accumulate lin-14 protein in nuclei at the L2-L4 and adult stages (Ruvkun and Giusto 1989). Analysis of these mutations revealed that both are lesions in the 3' UTR of the lin-14 gene, 270 and 300 bases downstream from the 3' end of the open reading frame (Fig. 2). These mutations delete sequences from the 3' UTR of the lin-14 mRNA and, in turn, cause an inappropriately high level of lin-14 protein at post-L1 stages (Figs. 5 and 7). There is no change in the lin-14 protein-coding region in these mutants, and Western blots showed no change in the size of the lin-14 proteins (Fig. 4). Because the lin-14(gf) mutations are not in protein-coding regions, they do not act by altering the activity or stability of the lin-14 protein. The location of the regulatory element in the 3' UTR suggests that the element controls a post-transcriptional process such as transcript stability, nuclear export, or translation.

Post-transcriptional regulation has been demonstrated in the iron-regulated transferrin receptor and ferritin genes in mammals [Casey et al. 1988]. In these cases, related stem-loop structures in the 3' UTR of the transferrin receptor mRNA regulate transcript stability, and in the 5' UTR of the ferritin gene mRNA, control translation [Casey et al. 1988]. Mutations in these stem-loop structures disrupt the regulation [Rouault et al. 1988]. In the case of the ferritin 5' UTR stem-loop structure, proteins have been identified that bind to those elements [Leibold and Munro 1988]. In the mouse protamine gene, 3' UTR sequences have been shown to mediate negative translational control [Braun et al. 1989]. In yeast, genetic evidence has been obtained that proteins bound to the 3' end of mRNAs can influence their translation: Suppressors of mutations in poly(A)−binding protein have been isolated in a ribosomal protein gene [Sachs and Davis 1989]. In the human immunodeficiency virus (HIV), the rev trans-activator protein has been shown to interact
Figure 7. In situ immunofluorescent detection of lin-14 nuclear protein in wild-type, lin-14(gf), and lin-14(lf) mutant animals. Specimens were fixed and incubated with affinity-purified anti-lin-14 antibodies followed by a secondary fluorescein isothiocyanate-labeled (FITC) goat anti-rabbit antibody and 4,6-diamido-2-phenylindole (DAPI)-stained to visualize all nuclei. Immunofluorescent photographs were taken with FITC [F] filters to show lin-14 protein staining or DAPI [D] filters to show all nuclei. Representative nuclei of a cell type are identified by white lines: (G) Gonad; (H) hypoderm; (I) intestine; (M) body wall muscle; (N) nerve ring; (V) ventral cord neurons. Photomicrographs are generally lateral views, in one of two planes. If a hypodermal cell is pointed out, the plane of focus is lateral and large nuclei are all hypodermal [H] and small nuclei are neuronal [N]. If intestinal [I] and ventral cord neurons [V] are labeled, the plane of focus is medial and all large nuclei are intestinal and the line of small nuclei ventrally are the ventral nerve cord and small dorsal nuclei are muscle. The head of the animal is identified in most photomicrographs by the intensely staining nerve ring nuclei (N). (a) (Top left) FITC; wild-type early L1 stage, lateral plane, showing bright lin-14 protein staining in all hypodermal nuclei; (top middle) FITC of same L1 animal with plane of focus at the midline showing intense lin-14 staining in the large intestinal nuclei and the smaller nuclei of the ventral cord neurons toward the interior (ventral) and the dorsally located body wall muscle nuclei on the outer section of this animal. The nerve ring also shows intense lin-14 nuclear staining. (Top right) DAPI of same L1 animal with plane of focus at the midline as in the top middle, (Bottom left) FITC; wild-type L2 stage animal focused at the midline showing no detectable lin-14 staining; (bottom right) DAPI of same animal showing ventral cord neuronal nuclei interspersed with Pn.a-derived neural nuclei and Pn.p ventral hypodermal nuclei that do not accumulate lin-14 protein. (b) (Top left) FITC, ventral view of lin-14(n536gf) mutant, early L1 stage, showing lin-14 protein in the ventral cord neurons and in muscle nuclei. Both gain-of-function mutants show inappropriate accumulation of lin-14 protein in L2 and later stages, although the n536 allele and lin-14 protein staining is consistently about half as intense in lin-14(n536gf) mutant animals after the L1 stage. (c) (Top) FITC, lin-14(n536n838lf) mutant L1 and L2 stages (L1 is the smaller animal above), showing weak lin-14 protein staining in neuronal and hypodermal nuclei of the L1 stage animal and no lin-14 protein in the L2 stage animal, unlike the lin-14(n536gf) mutants. (Bottom) DAPI. (d) (Top left) FITC, lin-14(n355n534lf) mutant L1 stage, which is a partial revertant of the n355(gf) mutant, showing little or no lin-14 protein in most nuclei in the two animals shown. (Bottom) DAPI. (Top right) FITC, lin-14(n355n534) L3 stage, displaying faint lin-14 nuclear protein staining in some neurons of the anterior nerve ring. All other cell lineages of the later developmental stages show no detectable lin-14 protein [data not shown]. (e) (Top left) FITC, lateral view of lin-14(n536n540lf) mutant L1 stage, showing no lin-14 protein in any nuclei. (Bottom) DAPI. (Top right) FITC, lin-14(n536n540) L1 immunostained with anti-myosin antibodies, showing body wall myosin (Miller et al. 1986) to prove that these animals were fixed and permeabilized properly.
with an RNA structure in the viral transcript (Zapp and Green 1989). This system may mediate export of the mRNA from the nucleus because mutations in the rev gene disrupt export of the HIV unspliced transcript from the nucleus (Emerman et al. 1988). Sequences in the 3' UTR of fem-3 have been shown to mediate its polyadenylation (Ahringer and Kimble 1991). These experiments reveal that sequences in various regions of the transcript can regulate polyadenylation, export of a transcript from the nucleus, the half-life of the transcript, or translation of that transcript.

These precedents suggest that during late L1, the negative regulatory element in the lin-14 3' UTR may normally down-regulate the stability of the lin-14 mRNA, its polyadenylation, its export from the nucleus, or its translation. The secondary structure predicted in the lin-14 3' UTR (Fig. 6) could mediate this regulation, and the lin-14(gf) mutations might delete sequences that are necessary for the proper function or formation of this RNA structure. However, it is not clear that this structure is actually involved in the process of down-regulation because neither gain-of-function mutant actually deletes the stem–loop itself.

We suggest that at the mid-L1 stage, a newly activated or synthesized regulatory factor binds to the 3' UTR of the lin-14 mRNA and down-regulates its processing or translation to protein. A candidate for binding to the 3' UTR of the lin-14 mRNA is the lin-4 gene product. The heterochronic lineage defects of a lin-4 mutant are the same as those of lin-14(gf) mutants (Chalfie et al. 1981; Ambros and Horvitz 1984), and lin-14(lf) mutants are epistatic to lin-4 mutants (Ambros 1989), suggesting that lin-4 negatively regulates lin-14. As in lin-14(gf) mutants, the level of the lin-14 protein is inappropriately high in late larval stages in a lin-4 mutant, suggesting that lin-4 negatively regulates the expression of lin-14 (Arasu et al., this issue). Although this lin-4-mediated regulation of lin-14 need not be direct, analysis of whether the lin-4 product binds to the lin-14 3' UTR will test whether lin-4 regulates lin-14 in this manner.

The lin-14 gene encodes multiple protein products

The lin-14 DNA sequence revealed 13 exons in the lin-14 gene, with introns ranging in size from ~12 kb to 44 bp. The lin-14B1 and lin-14B2 unspliced transcripts are ~15 kb larger than the lin-14A unspliced transcript. The transcripts encode proteins with distinct amino-terminal domains and a common carboxy-terminal region. This production of multiple related protein products is similar to that observed for the Ultrabithorax (Ubx), E74, and E75 genes from Drosophila (Beachy et al. 1985; Burtis et al. 1990; Segraves and Hogness 1990). Two of the Ubx proteins are expressed in distinct but overlapping cell types and specify distinct cell fates (Mann and Hogness 1990). Similarly, the multiple lin-14 protein products could be expressed in particular sets of cells or at particular developmental times to specify distinct cell types.

A combination of genetic and molecular evidence sug-

V

Gains of it function mutations in lin-14

lin-14 nuclear proteins are not similar to any known class of regulatory protein

Although the nuclear localization of lin-14 proteins suggested that we might find sequence similarities to other known nuclear proteins, we have found no homology to the classes of transcription factors and RNA-splicing proteins sequenced so far. Analysis of the lin-14 protein sequences did, however, reveal provocative potential
structures. The common region of the lin-14 proteins contains a potential amphipathic helix with two basic patches and an acidic patch of amino acid residues on the helical face opposite the hydrophobic face (Fig. 3). A computer search with just this region revealed no homologs in the data base. Amphipathic helices with basic patches have been observed in the myc family of transcription factors (Murre et al. 1989). The predicted amphipathic helix is located within the protein sequence which is slightly reminiscent of zinc finger proteins but distinct from those observed so far in this gene family (Evans and Hollenberg 1988).

The absence of known proteins showing sequence similarity to the lin-14 proteins suggests that they represent the first example of this regulatory gene class. Because the lin-14 proteins are localized to the nucleus, they may function to regulate gene expression. The potential basic amphipathic helix hints that the protein could bind to DNA or RNA. The lin-14 proteins could activate L1 stage lineage-specific genes and/or repress L2 stage and later lineage-specific genes. Genetic epistasis experiments suggest that the downstream gene lin-29 may be negatively regulated by the lin-14 proteins (Ambros 1989).

**Generation of the lin-14 temporal gradient**

During wild-type development, the lin-14 protein levels decrease by a factor of 25 from mid-L1 to late L1, a period of 6 hr [B. Wightman and G. Ruvkun, unpubl.]. The lin-14 protein levels decrease at the same rate in both dividing and nondividing cells (Ruvkun and Giusto 1989), suggesting that, at least in nondividing cells, breakdown of the nuclear membrane is not necessary for the degradation of the lin-14 protein and that nuclear proteases must control this process. In addition, these observations suggest that in the nondividing cells, cell-cycle-regulated proteases, like those that control cyclin protein levels [Murray and Kirschner 1989], probably do not regulate lin-14 protein levels. Protease[s] that degrade the lin-14 protein could be induced late in the L1 stage or could be present constitutively, with their effects only visible after the synthesis of new lin-14 protein is abolished in late L1. The presence of PEST sequences in the lin-14 protein and our observation that addition of a PEST amino acid by the n838 mutation decreases the level of lin-14 protein supports the notion that the rate of lin-14 protein degradation is relevant to the formation of the lin-14 protein gradient.

Other heterochronic genes could control developmental timing of the C. elegans cell lineage by participating in the generation of the lin-14 temporal gradient at any of these levels, by participating in the reception of this signal, or by interacting with the lin-14 protein to specify distinct cellular responses to this general temporal control signal. In Arasu et al. (this issue), we show that the two other heterochronic genes, lin-28 and lin-4, act antagonistically to set the level of the lin-14 protein during postembryonic development. We also show that the heterochronic gene lin-29 does not affect the level of the lin-14 protein and thus is a downstream gene. Therefore, the lin-28 and lin-4 products are candidates to interact directly with the lin-14 mRNA or protein, and the lin-29 gene is a potential downstream gene to which the lin-14 nuclear protein could bind to regulate its expression.

**Materials and methods**

**Cloning procedures**

cDNA was synthesized from poly[A]+ RNA according to Ausubel et al. (1987) and cloned into λgt10 with either XbaI–EcoRI linkers or NotI polylinkers. Some libraries were obtained from Stuart Kim (Stanford University, Stanford, CA). λgt10 libraries were screened according to standard methods by using either the 32P-labeled 5-kb BglII restriction fragment or the 32P-labeled 3.8-kb EcoRI restriction fragment as probe. cDNAs and genomic EcoRI and BglII restriction fragments from cosmid EEG4 (see Ruvkun et al. 1989) were subcloned into Bluescribe + [pBS +]. Genomic DNA from the lin-14(n355) mutant strain was prepared as described in Ruvkun et al. (1989), partially digested with Sau3A, and cloned into pBl8 cosmid vector. Cosmid-bearing colonies were screened by using a 32P-labeled 1.9-kb EcoRI restriction fragment. Genomic DNA from the lin-14(n536) mutant strain was prepared the same way, completely digested with HindIII, and size-selected on an agarose gel, fragments of ~6 kb were cloned into pBS +. The resulting library was screened with a 32P-labeled 3.8-kb EcoRI fragment.

**PCR studies**

cDNA was prepared from N2 (wild-type) poly[A]+ RNA and used as a template for PCR. Paired oligonucleotides corresponding to sequences in exons 1 and 2, exons 1 and 4, exons 1 and 5, exons 4 and 5 and a eDNA clone that included exons 1-3 and 5, were analyzed on urea-polyacrylamide gels.

**Northern and RNase protection analysis**

Northern blots were standard 1.2% agarose–formaldehyde gels transferred to nitrocellulose and probed with 32P-labeled probes as indicated (Ausubel et al. 1987). For RNase protection assays, total RNA obtained from N2 was hybridized to antisense RNA probes produced in vitro by standard methods [Ausubel et al. 1987]. Probes were produced from a genomic subclone spanning exons 4 and 5 and a cDNA clone that included exons 1–3 and 5, both in BS +. Hybridizations were performed at 42°C and digested with RNase A and RNase T1, and the resulting fragments were analyzed on urea–polyacrylamide gels.

**Sequencing**

Sequencing was performed on either single-stranded or double-stranded templates according to the chain-termination method with 35S-labeled dATP [Amersham and New England Nuclear] and Sequenase [U.S. Biochemical]. In brief, cDNAs in pBS + were sequenced from both ends with forward and reverse primers. Sequence was then extended by using synthesized oligonucleotides [Applied Biosystems] as primers until all sequence was double stranded. Genomic subclones were sequenced from the ends by using forward and reverse primers. Sequence was extended with newly synthesized primers or previously synthesized primers derived from the cDNAs. Genomic sequence of
parts of some introns is only single stranded. The n355 genomic clone was directly sequenced on cosmid DNA, from just before the end of the open reading frame to beyond the breakpoint, by using primers derived from the cDNA sequence. cDNA clones obtained from mutants containing the n355 mutation were sequenced, demonstrating no other divergence from wild-type sequence from exon 10 to the 3’ end of the coding sequence [data not shown]. The n536 genomic clone was sequenced in the region flanking the n838 point mutation and from 50 bp 5’ of the n355 breakpoint to 120 bp 3’ of the n536 breakpoint.

Sequences were entered on Apple Macintosh systems with a talking sequence editor written by T.R. Bürglin. Sequences were aligned and assembled with a multiple sequence editor generously provided by Will Gilbert, and sequences were analyzed by David Miller (Duke) and used at a dilution of 1:200. Fluorescent nuclei were visualized with a Zeiss Universal microscope as reported previously [Ruvkun and Giusto 1989].

Acknowledgments

We thank Victor Ambros and Mike Finney for helpful comments on the manuscript, Candy Lee for some of the subclones, Susan Pong for some of the sequencing, and Mike Finney for some of the cDNA. T.R.B. was supported by EMBO fellowship ALTF179-1988. This work was supported by a grant from Hoechst AG. This paper is dedicated to the memory of Jones. The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby identified as belonging solely to the author for whose separate account the publication costs are payable.

Gain-of-function mutations in lfn-14

n355 is lethal, a Drosophila sex-determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell 55: 1037–1046.

Braun, R.E., J.J. Peschon, R.R. Behringer, R.L. Brinster, and R.D. Palmiter. 1989. Proteamine 3’-untranslated sequences regulate temporal translational control and subcellular localization of growth hormone in spermatids of transgenic mice. Genetics & Dev. 3: 793–807.

Burtis, K.C., C.S. Thummel, C.W. Jones, F.D. Karim, and D.S. Hogness. 1990. The Drosophila 74EF early puff contains E74, a complex ecdysone-inducible gene that encodes two ets-related proteins. Cell 61: 85–99.

Casey, J.L., M.W. Hentze, D.M. Koeller, S.W. Caughman, T.A. Rouault, R.D. Klausner, and J.B. Hartford. 1988. Iron-responsive elements: Regulatory RNA sequences that control mRNA levels and translation. Science 240: 924–928.

Chalfie, M., H.R. Horvitz, and J.E. Sulston. 1981. Mutations that lead to reiterations on the cell lineages of C. elegans. Cell 24: 59–69.

Costa, M., M. Weir, A. Coulson, J.E. Sulston, and C. Kenyon. 1988. Posterior pattern formation in C. elegans involves position-specific expression of a gene containing a homeobox. Cell 55: 747–756.

Devereux, J., P. Haebeli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12: 387–395.

Emerman, M., R. Vazeux, and K. Peden. 1988. The rev gene product of the human immunodeficiency virus affects envelope-specific RNA localization. Cell 57: 1155–1165.

Emmons, S.W. 1988. The genome. In The nematode Caenorhabditis elegans (ed. W.B. Wood), pp. 47–80. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Evans, R.M. and S.M. Hollenberg. 1988. Zinc fingers: Gilt by association. Cell 52: 1–3.

Ginsberg, A.M., B.O. King, and R.G. Roeder. 1984. Xenopus SS gene transcription factor, tllA: Characterization of a cDNA clone and measurement of RNA levels through development. Cell 39: 479–489.

Ingham, P. 1988. The molecular genetics of embryonic pattern formation in Drosophila. Nature 335: 25–34.

Krause, M. and D. Hirsch. 1987. A trans-spliced leader sequence on actin mRNA in C. elegans. Cell 49: 753–761.

Leibold, E.A. and H.N. Munro. 1988. Cytoplasmic protein binds in vitro to a highly conserved sequence in the 5’ untranslated region of ferritin heavy- and light-subunit mRNAs. Proc. Natl. Acad. Sci. 85: 2171–2175.

Mann, R. and D.S. Hogness. 1990. Functional dissection of ultratransborax proteins in D. melanogaster. Cell 60: 597–610.

Miller, D.M., F.E. Stockdale, and J. Karn. 1986. Immunological identification of the genes encoding the four myosin heavy chain isoforms of Caenorhabditis elegans. Proc. Natl. Acad. Sci. 83: 2305–2309.

Moerman, D.G., G.M. Benian, and R.H. Waterston. 1986. Molecular cloning of the muscle gene unc-22 in Caenorhabditis elegans. Tcr transposon tagging. Proc. Natl. Acad. Sci. 83: 2579–2583.

Murray, A.W. and M.W. Kirschner. 1989. Cyclin synthesis drives the early embryonic cell cycle. Nature 329: 275–280.

Murre, C., P.S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56: 777–783.

Nash, R., G. Tokiwa, S. Anand, K. Erickson, and A.B. Futcher. 1988. The WH1+ gene Saccharomyces cerevisiae tethers cell division to cell size and is a cyclin homolog. EMBO J. 7: 4335–4346.
Rogers, S., R. Wells, and M. Rechsteiner. 1986. Amino acid sequences common to rapidly degraded proteins: The PEST hypothesis. Science 234: 364–368.

Rouault, T.A., M.W. Hentze, S.W. Caughman, J.B. Hartford, and R.D. Klausner. 1988. Binding of a cytosolic protein to the iron-responsive element of human ferritin messenger RNA. Science 241: 1207–1210.

Ruiz i Altaba, A. and D.A. Melton. 1989. Involvement of the Xenopus homeobox gene Xhox3 in pattern formation along the anterior-posterior axis. Cell 57: 317–326.

Ruvkun, G. and J. Giusto. 1989. The Caenorhabditis elegans heterochronic gene lin-14 encodes a nuclear protein that forms a temporal developmental switch. Nature 338: 313–319.

Ruvkun, G., V. Ambros, A. Coulson, R. Waterston, J. Sulston, and H.R. Horvitz. 1989. Molecular genetics of the Caenorhabditis elegans heterochronic gene in lin-14. Genetics 121: 501–516.

Sachs, A. and R.W. Davis. 1989. The poly[A] binding protein is required for poly[A] shortening and 60S ribosomal subunit-dependent translation initiation. Cell 58: 857–867.

Scott, M.P., J.W. Tamkun, and G.W. Hartzell. 1989. The structure and function of the homeodomain. Biochim. Biophys. Acta 989: 25–48.

Segraves, W.A. and D.S. Hogness. 1990. The E75 ecdysone-inducible gene responsible for the 75B early puff in Drosophila encodes two new members of the steroid receptor superfamily. Genes & Dev. 4: 204–219.

Zapp, M.L. and M.R. Green. 1989. Sequence-specific RNA binding by the HIV-1 rev protein. Nature 342: 714–716.
Negative regulatory sequences in the lin-14 3'-untranslated region are necessary to generate a temporal switch during Caenorhabditis elegans development.

B Wightman, T R Bürglin, J Gatto, et al.

*Genes Dev.* 1991, 5:
Access the most recent version at doi:10.1101/gad.5.10.1813