Cell, Vol. 75, 843–854, December 3, 1993, Copyright ©1993 by Cell Press

The C. elegans Heterochronic Gene *lin-4* Encodes Small RNAs with Antisense Complementarity to *lin-14*

Rosalind C. Lee,*† Rhonda L. Feinbaum,*‡ and Victor Ambros†
Harvard University
Department of Cellular and Developmental Biology, Cambridge, Massachusetts 02138

Summary

***lin-4* is essential for the normal temporal control of diverse postembryonic developmental events in C. elegans. *lin-4* acts by negatively regulating the level of LIN-14 protein, creating a temporal decrease in LIN-14 protein starting in the first larval stage (L1). We have cloned the C. elegans *lin-4* locus by chromosomal walking and transformation rescue. We used the C. elegans clone to isolate the gene from three other Caenorhabditis species; all four Caenorhabditis clones functionally rescue the *lin-4* null allele of C. elegans. Comparison of the *lin-4* genomic sequence from these four species and site-directed mutagenesis of potential open reading frames indicated that *lin-4* does not encode a protein. Two small *lin-4* transcripts of approximately 22 and 61 nt were identified in C. elegans and found to contain sequences complementary to a repeated sequence element in the 3′ untranslated region (UTR) of *lin-14* mRNA, suggesting that *lin-4* regulates *lin-14* translation via an antisense RNA-RNA interaction.**

Introduction

A genetic pathway of heterochronic genes in Caenorhabditis elegans acts to specify the temporal fates of cells during larval development, thereby controlling the timing and sequence of events in diverse postembryonic cell lineages (Ambros and Horvitz, 1984; Ambros and Horvitz, 1987; Ambros, 1989). Mutations in the heterochronic genes can cause either precocious development, in which normally late developmental programs are expressed at early larval stages, or retarded development, in which normally early developmental programs are reiterated at later stages (Chalfie et at., 1981; Ambros and Horvitz, 1984). It is likely that the expression of stage-specific developmental programs by particular cells requires the accurate temporal regulation of the products of the heterochronic genes.

lin-4 acts early in C. elegans larval development to affect the timing of developmental events at essentially all larval stages and in diverse cell types (Chalfie et al., 1981; Ambros and Horvitz, 1987). Animals carrying a *lin-4* loss-of-function (*lf*) mutation, *lin-4(e912)*, display reiterations of early fates at inappropriately late developmental stages; cell lineage patterns normally specific for the L1 are reiterated at later stages, and the animals execute extra larval molts (Chalfie et al., 1981). The consequences of these heterochronic developmental patterns include the absence of adult structures (such as adult cuticle and the vulva) and the prevention of egg laying.

lin-14 null *(0)* mutations cause a phenotype opposite to that of *lin-4(lf)* and are completely epistatic to *lin-4(lf)*, which is consistent with *lin-4* acting as a negative regulator of *lin-14* (Ambros and Horvitz, 1987; Ambros, 1989). *lin-14(0)* mutants skip the expression of L1-specific events and precociously execute programs normally specific for the L2, L3, L4, and adult stages. *lin-14* gain-of-function *(gf)* mutations, which cause inappropriately high *lin-14* activity at late stages of development, result in a retarded phenotype virtually identical to that of *lin-4(lf)* (Ambros and Horvitz, 1987). These observations indicate that in wild-type development a high level of *lin-14* activity in the early L1 stage specifies L1-specific programs, and lower levels of *lin-14* activity in the late L1 specify later stage-specific programs. Thus, the normal developmental progression from the execution of L1 programs to later programs depends critically on the *lin-4*-dependent decrease in *lin-14* activity.

The temporal decrease in *lin-14* activity reflects a decrease in the level of LIN-14 protein. LIN-14 protein is normally abundant in the nuclei of late-stage embryos and younger L1 larvae and then is barely detectable by the L2 (Ruvkun and Giusto, 1989). *lin-14* transcripts are constant throughout development, indicating that *lin-14* is negatively regulated posttranscriptionally (Wightman et al., 1993 [this issue of *Cell*]). In *lin-4* mutant animals, as in *lin-14(gf)* mutants, the level of LIN-14 protein remains abnormally high late in development (Arasu et al., 1991). Mapping of the *lin-14(gf)* mutations to the 3′UTR of *lin-14* mRNA (Wightman et al., 1991), and gene fusion experiments (Wightman et al., 1993), define the *lin-14* 3′UTR as a necessary and sufficient cis-negative regulatory element of LIN-14 protein level. The temporal decrease in LIN-14 protein levels requires both *lin-4* in trans and *lin-14* 3′UTR sequences in cis. These observations suggest that the *lin-4* gene product or products act via the *lin-14* 3′UTR to inhibit, directly or indirectly, the translation of *lin-14* mRNA. To determine the mechanism by which *lin-4* developmentally regulates the level of LIN-14 protein, we cloned the *lin-4* locus by chromosomal walking and transformation rescue. Our analysis of the *lin-4* genomic sequence indicates that *lin-4* does not encode a protein product. We have identified two small *lin-4* transcripts of approximately 22 and 61 nt. These *lin-4* RNAs are complementary to a repeated sequence in the 3′UTR of *lin-14* (Wightman et al., 1991, 1993), supporting a model in which these *lin-4* RNAs could regulate *lin-14* translation by an antisense mechanism.

*The first two authors contributed equally to this work.
†Present address: Dartmouth College Department of Biology, Hanover, New Hampshire 03755.
‡Present address: Massachusetts General Hospital Department of Molecular Biology, Boston, Massachusetts 02114.

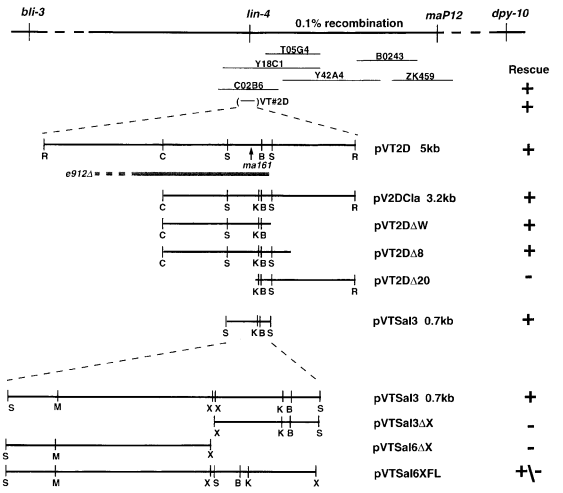

Figure 1. Partial Genetic and Physical Map of the *lin-4* Region of LGII

Recombination distance between *lin-4* and *map12* is based on RFLP mapping as described in Experimental Procedures. The *lin-4–dpy-10* and *lin-4–bli-3* distances are approximately 0.5% and 2% recombination, respectively (Wood et al., 1988). YAC (Y prefix) and cosmid (T, B, Z, and C prefixes) clones are from the C. elegans physical mapping project (Coulson et al., 1986, 1988) and were obtained from A. Coulson. Genomic clone pVT2D was identified by hybridization to Y18C1, as described in Experimental Procedures and in the text. Derivatives of pVT2D were constructed by treatment with restriction enzymes or ExoIII nuclease (see Experimental Procedures). Dotted lines indicate a change in scale. Plus: the cloned DNA rescued the mutant phenotype of *lin-14*(*e912*), as described in Experimental Procedures. Minus: no rescue was observed in any of at least three transgenic lines. Nonrescuing lines were confirmed to contain the plasmid DNA by PCR. Plus/minus: weak rescue was observed, i.e., the appearance of vulval morphogenesis and lateral alae on the adult cuticle was observed, but without complete rescue of egg-laying defects. The *e912* lesion deletes most of pVT2D sequences. One breakpoint was determined to lie between two closely positioned BamHI and SalI sites at one end of pVT2D by Southern blot analysis (Figure 2). The *e912* deletion also extends to sequences corresponding to two other clones, pVT6G and pVT1C (data not shown). Presumably, pVT6G and pVT1C lie to the left of pVT2D, as drawn here. The orientation of pVT2D with respect to the genetic and physical maps has not been established.

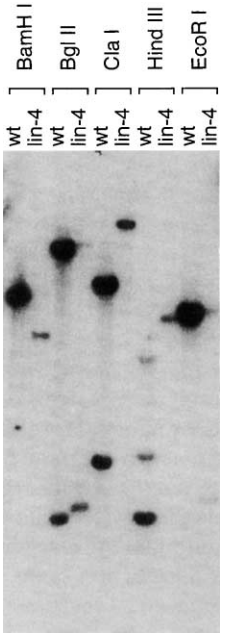

Figure 2. Southern Blot of Genomic DNAs from Wild-Type and *lin-4(e912)* Animals Digested with Various Restriction Endonucleases and Hybridized with pVT2D Probe

Similar amounts of wild-type and *lin-4(e912)* DNA were loaded. For each enzyme, the hybridization in *e912* DNA is to a band of weaker intensity and of altered size (smaller for each enzyme except ClaI) than that in wild-type DNA, consistent with a deletion in *e912* DNA of sequences corresponding to the pVT2D probe.

Results

Identification of C. elegans lin-4 Genomic DNA

lin-4 is located on the C. elegans genetic map in a region of LGII between *bli-3* and *dpy-10* (Wood et al., 1988) (Figure 1). We mapped the positions of restriction fragment length polymorphisms (RFLPs) in the *bli-3-dpy-10* interval with respect to *lin-4*, as described in Experimental Procedures. Among the RFLPs that we analyzed, *maP12*, detected by cosmid ZK459, mapped closest to *lin-4*, and to the right (Figure 1). We used cosmid and yeast artificial chromosome (YAC) clones to the left of ZK459 as probes to southern blots of wild-type and *lin-4(e912)* DNA. The YAC clone Y18C1 detected a restriction fragment alteration associated with *lin-4(e912)* DNA. This lesion, which results in the absence of a 5 kb EcoRI hybridizing band, was not detected by an overlapping YAC clone, Y42A4.

We cloned the wild-type sequences corresponding to the *e912*-affected 5 kb band based on its differential hybridization to Y18C1 and Y42A4, as described in Experimental Procedures. Three classes of clones were identified, each corresponding to a unique EcoRI insert: pVT2D (containing a 5 kb insert), pVT1C (a distinct 5 kb insert), and pVT6G (a 3.5 kb insert). Probes from pVT2D (Figure 2), pVT1C, pVT6G, and an overlapping cosmid clone, C02B6 (data not shown), detected restriction fragment aberrations on southern blots of *lin-4(e912)* DNA, indicating that the *e912* lesion must extend over several kb of DNA. We have not characterized the *e912* lesion in detail, except to determine that sequences corresponding to both pVT2D and pVT1C are at least partially deleted, and that the pVT6G insert fragment seems to be rearranged and possibly duplicated (data not shown; see legend to Figure 1).

Functional *lin-4* sequences were localized to pVT2D by transformation rescue. pVT2D, pVT2DCla (a 3.2 kb subclone of pVT2D), and C02B6 (a cosmid clone that overlaps pVT2D) rescued the mutant phenotype of *lin-4(e912)* animals in transgenic lines generated by microinjection of the cloned DNA (Figure 1; see also Experimental Procedures). These results indicate that the *lin-4* gene lies within the 3.2 kb EcoRI/ClaI insert fragment of pVT2DCla. The *e912* lesion is a deletion of most of the sequences of pVT2D, including about half of the sequences in pVT2DCla (Figure 1). Subclones and ExoIII deletion derivatives of PVT2DCla (see Experimental Procedures) were tested for rescue of *lin-4(e912)* defects in transgenic animals. A subclone (pVTSal3) containing a 693 bp SalI fragment (see

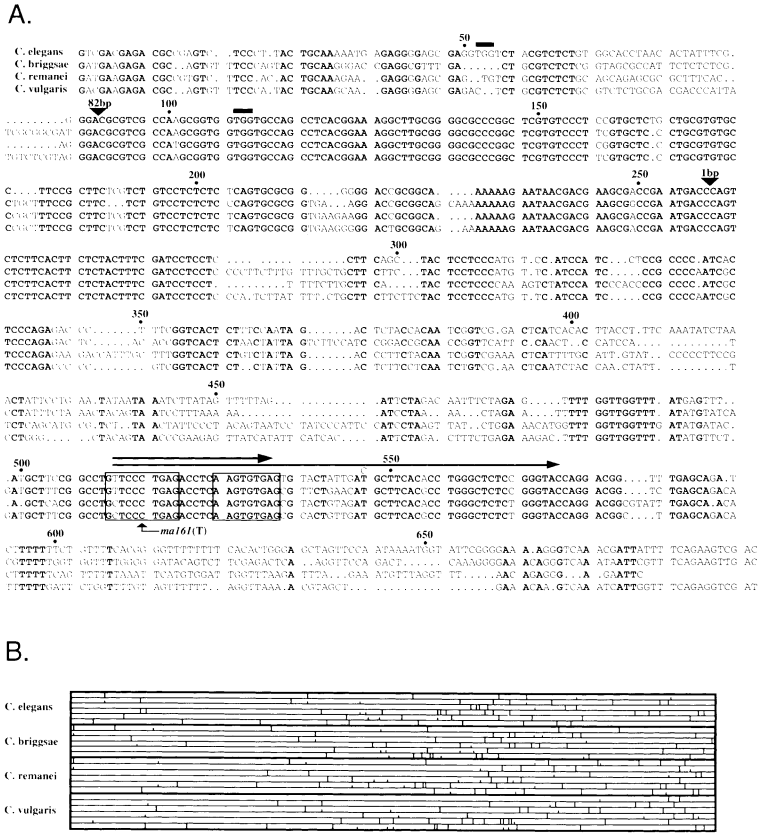

Figure 3. Sequence Analysis of *lin-4*

(A) Comparison of *lin-4* sequences from four Caenorhabditis species: C. elegans, C. briggsae, C. remanei, and C. vulgaris.Genomic *lin-4* clones were identified and prepared as described in Experimental Procedures. DNA sequence was determined for both strands of each species; only one strand is shown here. Numbering refers to the 693 bp SalI insert of the C. elegans rescuing clone pVTSal3 (see Figure 1). The right end of the C. remanei sequences shown here is the EcoRI site at the end of the C. remanei genomic DNA insert in pVTreA1. The right ends of the C. briggsae and C. vulgaris sequences are defined by sequence alignment with the end of the C. elegans pVTSal3 clone. Dots indicate absence of nucleotides. Sequences conserved in all four species are shown in bold. Note the conserved TTTCAG beginning at nucleotide 682, which is likely the 3′ splice acceptor for an exon of the host gene within which *lin-4* resides. The eight nucleotides that follow are the only coding sequences of the host gene on the pVTSal3 rescuing fragment. Sequences complementary to a repeated sequence in the 3′UTR of *lin-14* mRNA (Wightman et al., 1991, 1993) are boxed, *lin-4*S and *lin-4*L transcripts are indicated according to the results of nuclease protection, primer extension, and Northern blot experiments (see Figures 4, 5, 6, and 7). The position of the ethylmethane sulfonate–induced C-to-T change associated with *ma161* is shown by an arrow below the line. The positions of mutations that were induced in vitro and retain *lin-4* function (see Results and Experimental Procedures) are shown above the C. elegans line. In separate experiments, insertions of 81 and 1 bp were introduced at positions 90 and 259, respectively. Triplets that were converted to stop codons are shown by bars at positions 52 and 109. A T-to-C mutation introduced at position 547 is indicated by a C above the C. elegans sequence.

(B) Predicted open reading frames (in all six frames) for the wild-type *lin-4* sequences of four Caenorhabditis species. Short bars are predicted AUG codons, and tall bars are predicted stop codons. The top three lines and bottom three lines for each species correspond to the three rightward and three leftward reading frames, respectively.

below) also fully rescues *lin-4(e912)* (Figure 1). Less than a hundred base pairs (bp) of the corresponding genomic SalI fragment remain in *lin-4(e912)* DNA, strongly suggesting that *e912* completely eliminates *lin-4* gene activity (Figure 1). The fact that the 693 bp SalI fragment rescues in either orientation in the vector (i.e., pVTSal6 or pVTSal3) indicates that rescue probably does not depend on sequences within the vector and that the complete *lin-4* gene is included in these 693 bp.

Phylogenetic Comparison of *lin-4* Genomic Sequence

C. briggsae and C. remanei *lin-4* clones were identified by hybridization to a C. elegans *lin-4* probe, and a C. vulgaris *lin-4* clone was constructed using polymerase chain reaction (PCR)–amplified genomic DNA as described in Experimental Procedures, The C. briggsae, C. remanei, and C. vulgaris clones were tested for rescue of the mutant phenotypes of *lin-4(e912)* and were found to be fully functional for rescue in C. elegans. It is likely that the clones from C. briggsae, C. remanei, and C. vulgaris correspond to the *lin-4* locus of each species, because they all function in C. elegans, and because *lin-4* seems to be a single copy gene in *C. elegans* (Figure 2).

The C. elegans *lin-4* genomic clone pVT2DCla and the C. briggsae, C. remanei, and C. vulgaris *lin-4* clones were sequenced as described in Experimental Procedures. Among the four Caenorhabditis species, two main blocks of phylogenetically conserved DNA sequence were observed within the *lin-4* rescuing (pVTSal) region, one from approximately base pair 85 to base pair 350 of pVTSal3, and another from about base pair 500 to base pair 600 (Figure 3A). The sequence conservation consists of stretches of complete identity, interspersed with short stretches of divergence, and insertions and deletions of 1 or more bp. Outside the sequences corresponding to the 693 bp pVTSal3 insert, the sequences of these four species are less well conserved (data not shown). The fact that the *lin-4* clones from four Caenorhabditis species function in C. elegans indicates that they encode similar functionally related gene products.

lin-4 Is Unlikely to Encode a Protein

A probe from pVT2DCla was used to screen a cDNA library. The clones that were isolated and sequenced all contained DNA corresponding to genomic sequences that were almost entirely outside the 693 bp SalI (pVTSal) rescuing region. Comparison of these cDNA sequences with

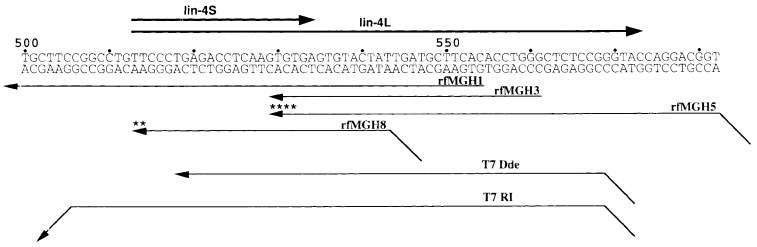

Figure 4. A Segment of the C. elegans *lin-4* Sequence Showing Probes Used in Analysis of *lin-4* Transcripts

*lin-4*S and *lin-4*L transcripts, derived from the experiments described in Figures 5, 6, and 7, are shown above the sequence. DNA oligonucleotide probes (rf perfix) and RNA probes generated by T7 RNA polymerase transcription from restriction enzyme–cut pMspI plasmid are shown as arrows with their 3′ ends at the tip of the arrowhead. rfMGH5 and rfMGH8 have nonhybridizing 5′ tails of 18 and 11 nt, respectively, to aid in the discrimination of undigested probe from protected hybrids. Preparation and radioactive labeling of probes are described in Experimental Procedures. rfMGH1 and rfMGH3 were labeled at the 5′ end using polynucleotide kinase (Ausubel, 1989). Asterisk, 3′ end nucleotides labeled as described in Experimental Procedures.

pVT2D sequences indicated that the corresponding mRNA is spliced from a primary transcript at conventional splice donor and acceptor sites within pVT2D that, for the most part, flank the pVTSal rescuing region (see legend to Figure 3). Therefore, it appears that the *lin-4* gene lies within an intron of another gene. The normal function of the host gene is unknown but is apparently unrelated to *lin-4* function; pVTSal fully rescues the *lin-4(e912)* mutant phenotype in spite of the fact that the *e912* lesion deletes large regions of the host gene outside of pVTSal (data not shown).

To determine whether the *lin-4* rescuing region might encode a protein from a rare mRNA not represented in the cDNA library that we screened, we compared the potential protein coding sequences of the *lin-4* rescuing region from the four species. These four *lin-4* genes contain no conserved protein sequence that begins and ends with conventional start and stop codons or that can be predicted to be assembled using conventional C. elegans splice donor and acceptor sites (Wood et al., 1988) (Figure 3B). One relatively long predicted open reading frame of 143 amino acids in the region from base pair 1 to base pair 429 of the C. elegans sequence was tested by inserting 82 bp at an MluI site (base pair 90) and by inserting one base at a Tth111I site (base pair 259) (Figure 3A). Similarly, three mutations were introduced by oligonucteotide-mediated in vitro mutagenesis (see Experimental Procedures) that each disrupt this and other putative open reading frames in the C. elegans clone: nonsense codons were created at base pairs 53 and 111, and a conserved ATG at base pair 546 was converted to ACG (Figure 3A). All of these mutant constructs fully rescued *lin-4(e912)*. These observations are inconsistent with a LIN-4 protein product.

Two Small *lin-4* Transcripts

The fact that *lin-4* negatively regulates *lin-14* activity in trans strongly implies a *lin-4* gene product. Northern blot and RNAase protection analysis using strand-specific probes (Figure 4) that covered the 693 bp *lin-4* rescuing sequence detected two small transcripts (Figure 5). Both transcripts mapped to the same region by Northern blots and RNAase protection experiments (see below) and are transcribed in the same orientation (See Figures 3 and 4). The larger transcript, *lin-4*L, appears to be 61 nt, and the

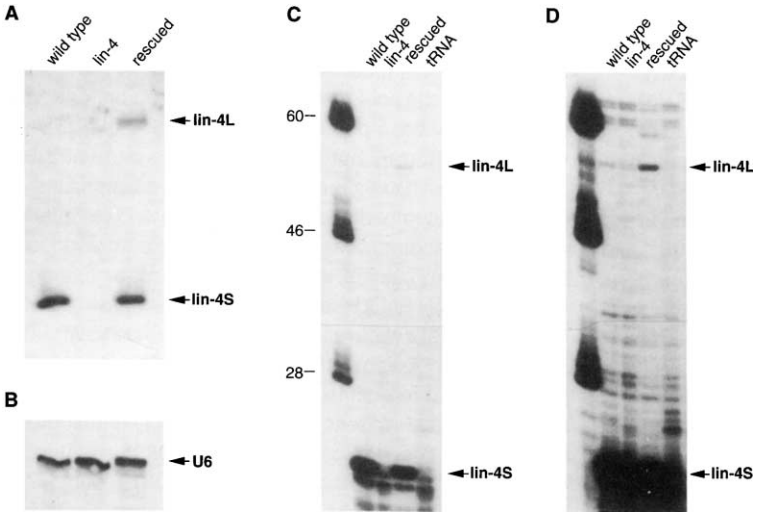

Figure 5. Identification of Two Small *lin-4* Transcripts

(A) Northern blot of total RNA from wild-type, *lin-4(e912)* (strain VT371) and *lin-4(e912)* rescued by transformation with pVTSal6 (strain VT510), probed with radiolabeled *lin-4* RNA probe. Although not shown here, similar northern blots have been done with DNA oligonucleotides loaded as size markers.

(B) Northern blot shown in (A), stripped and reprobed with an oligonucleotide complementary to U6 snRNA, to control for RNA loading.

(C) RNAase protection analysis of total RNA from wild-type N2, *lin-4(e912)*, and a *lin-4* line rescued by pVTSal6 (strain VT510). The *lin-4* RNA probe used here and in (A) was generated as a runoff transcript from pMspI digested with EcoRI (T7R1, Figure 4). This *lin-4* probe covers nucleotides 506–568 and does not include the 3′ end of *lin-4*L; hence, *lin-4*L does not completely protect the probe in this assay. Probe fragments protected by *lin-4*S and *lin-4*L are indicated by the arrows. RNA size markers generated by runoff transcription of pBS are shown at the left. The gel was exposed to film for 16 hr with an intensifying screen.

(D) Same gel shown in (C), exposed to film for 96 hr with intensifying screen to better visualize the product protected by *lin-4*L. For both the Northern blots and RNAase protection experiments, the estimated error in the lengths of RNA molecules is ±2 nt.

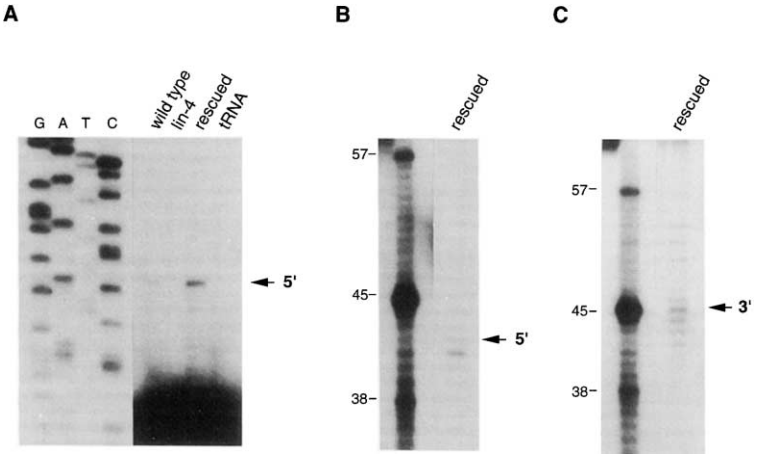

Figure 6. Mapping 5′ and 3′ Ends of *lin-4*L

(A) Primer extension analysis of total RNA from wild-type N2, *lin-4(e912)*, and the rescued line VT510. rfMGH3 (see Figure 4) was used as a primer. An Mp18 sequencing ladder provides size markers. The estimated error in the size of the extension product (arrow), determined by comparison with the sequencing ladder as well as to labeled oligonucleotide markers (data not shown), is ± 1 nt. The single extended product of 49 nt corresponds to *lin-4*L transcription initiating at the T located at position 513 of the C. elegans DNA sequence (Figures 3 and 4).

(B) S1 mapping of the 5′ end of *lin-4*L. The probe used was rfMGH1 (Figure 4). S1 analysis of total RNA from VT510 is shown, but an identical pattern of S1 products was also detected in wild-type N2 RNA. The size of 5′ end–labeled oligonucleotide markers is indicated to the left. Three S1 digestion products of 40, 41, and 42 nt were obtained. The largest product (arrow) corresponded to the same 5′ end as that identified by primer extension. The estimated error in the size of the S1-protected RNAs is ± 1 nt.

(C) S1 mapping of the 3′ end of *lin-4*L. The rfMGH5 probe (Figure 4) was labeled at the 3′ end as described in Experimental Procedures. S1 analysis of total RNA from VT510 is shown, but an identical pattern of S1 products was also detected in N2 RNA. The two most abundant species were 44 and 45 nt long; the 45 nt product (arrow) corresponds to a 3′ end of *lin-4*L at the C located at base pair 573 of pVTSal3 (Figures 3 and 4).

smaller, *lin-4*S, is approximately 22 nt in length. *lin-4*L and *lin-4*S were present in total RNA isolated from mixed-stage populations of wild-type N2 and *lin-4(e9l2)* transgenic animals rescued by plasmid pVTSal6 (strain VT510) but were not detected in RNA from *lin-4* mutant animals (see Figures 5, 6, and 7). *lin-4L* is barely detectable on Northern blots and in RNAase protection assays of RNA from mixed-stage N2 animals (Figure 5). *lin-4*L levels appear to be elevated somewhat, compared with those of the wild type, in RNA isolated from VT510 animals (Figure 5) or VT509 animals (a distinct rescued line; data not shown). This may be due to overexpression of *lin-4*L from the rescuing arrays. In contrast with *lin-4*L, *lin-4*S appears to be very abundant and is easily detected in total RNA from both mixed-stage N2 and rescued *lin-4(e912)* strains.

Primer extension and S1 analysis of total RNA from mixed-stage N2, *lin-4*-rescued animals, or both was used to determine the 5′ end of *lin-4*L (Figures 6A and 6B). Primer extension experiments gave a single extended product of 49 nt (Figure 6A), indicating that *lin-4*L transcription initiates at the T at position 513 of the C. elegans DNA sequence (see Figures 3 and 4). The oligonucleotide primer used to map the 5′ end of *lin-4*L (rfMGH-3, see Figure 4) should not have detected *lin-4*S transcripts, owing to the limited complementarity (6 bp) between the primer and *lin-4*S. S1 analysis confirmed the 5′ end of *lin-4*L determined by the above primer extension experiment. Three S1 digestion products of 40, 41, and 42 nt were obtained (Figure 6B). The largest product corresponded to the same 5′ end as that identified by primer extension. The smaller bands presumably are due to breathing of the DNA:RNA hybrid during the S1 digestion.

The 3′ end of *lin-4*L was also mapped by S1 analysis. A set of S1 digestion products ranging from 42 to 46 nt was obtained (Figure 6C). The two most abundant species were 44 and 45 nt long; the 45 nt product corresponded to termination of the transcript at the C located at base pair 573 of pVTSal3 (see Figures 3 and 4). The multiple products obtained in this experiment may be in part due to probe heterogeneity caused by incomplete labeling of the four terminal bases of the probe (see Figure 4; also Experimental Procedures) or by breathing of the RNA:DNA hybrid during the S1 digestion, or both. It is also possible that there is natural variation in the 3′ end of *lin-4*L.

*lin-4*S appears to have the same 5′ end as *lin-4*L. RNAase protection experiments using nested probes (Figure 7A) demonstrated that *lin-4*S spans a DdeI site that is located 5 nt downstream of the 5′ end of *lin-4*L (at T 513). The length of the protected product corresponding to *lin-4*S is approximately 5 nt shorter in the case of the DdeI probe compared with the EcoRI probe (Figure 7A), indicating that *lin-4*S starts at the same position as *lin-4*L.

The 3′ end of *lin-4*S was mapped by S1 analysis. An oligonucleotide probe (rfMGH8) labeled at the two 3′ terminal A residues that are complementary to the proposed 5′ end of both *lin-4* transcripts (see Figure 4) was hybridized to total N2 RNA and then digested with S1 nuclease at different temperatures. At 15°C, the S1 digestion produced protected products that ranged from 19–22 nt, while at 37°C, protected products were predominantly 17–21 nt long (see Figure 7B). These mapping data agree with the estimated size for *lin-4*S of 20 ± 2 nt from Northern blots and RNAase protection experiments (see Figure 5). This suggests that *lin-4*S is identical in sequence to the first 22 nt of *lin-4*L (see Figures 3 and 4).

lin-4 Mutations Affect *lin-4*L and *lin-4*S

As mentioned previously, neither *lin-4* transcript was detectable in the single previously isolated *lin-4* mutant, *e912*. This was not unexpected, since the *e912* lesion is a dele-

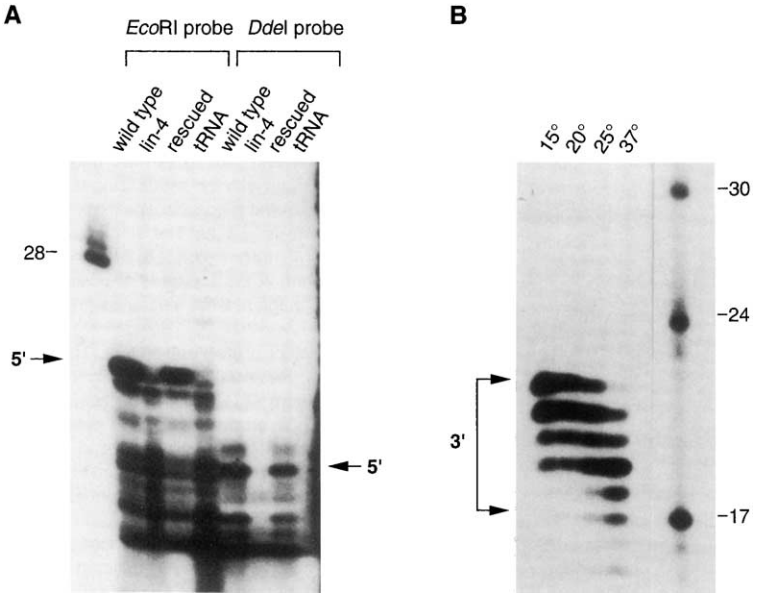

Figure 7. Mapping 5′ and 3′ Ends of *lin-4*S

(A) RNAase protection of total RNA from wild-type N2, *lin-4*(*e912*), and VT510, using nested antisense probes to map the 5′ end of *lin-4*S. Probes were runoff transcripts from pMspI that had been previously digested with either EcoRI, which cuts in the polylinker (T7RI, Figure 4), or DdeI, which cuts within *lin-4* sequences (T7Dde, Figure 4). The size of one RNA marker is indicated to the left. Arrows indicate the probe fragments protected by *lin-4*S. The length of the protected product is approximately 5 nt longer in the case of the EcoRI probe compared with the DdeI probe, indicating that *lin-4*S starts 5 nt upstream of the end of the DdeI probe.

(B) S1 analysis of total RNA from wild-type N2 using 3′ end–labeled rfMGH8 as a probe. The temperature at which the S1 digestion was performed is indicated above each lane. S1 digestion was for 1 hr. The size of 5′ end-labeled oligonucleotide markers is indicated to the right.

tion or rearrangement that removes at least 5 kb in the *lin-4* region, including the entire *lin-4*S- and *lin-4*L-transcribed sequences. To test further the functional significance of the *lin-4* transcribed sequences, we used a noncomplementation screen (see Experimental Procedures) to isolate a novel *lin-4* mutation and then identified the corresponding molecular lesion. Over 20,000 mutagenized chromosomes were screened, and a single novel *lin-4* allele, *ma161*, was identified. DNA from *ma161* animals was amplified by PCR and sequenced. The only sequence alteration in the 693 bp *lin-4* region of *ma161* DNA was a C to T transition at base pair 517 (see Figure 3). This point mutation would presumably alter nucleotide 5 in *lin-4*L and *lin-4*S.

lin-4 Transcripts Are Complementary to the 3′UTR of *lin-14* mRNA

The *lin-4* transcribed sequence was combined in tandem to the sequence of the *lin-14* 3′UTR (Wightman et al., 1991), and this sequence was searched for the formation of *lin-4*:*lin-14* hybrid RNA structures, using the STEMLOOP program of the GCG sequence analysis package (Devereux et al., 1984), as described in Experimental Procedures. Two short blocks of *lin-4* sequence were identified (Figure 3A) that are complementary to an element repeated seven times in the 3′UTR of *lin-14* (Figure 8; Wightman et al., 1991, 1993). The first block of complementary sequence, 5′-GUUCCCUGAG-3′, is (with the exception of the first G) at the very 5′ end of both *lin-4*L and *lin-4*S. The second block of sequence, 5′AAGUGUGAG3′, would lie internal to *lin-4*L and at the 3′ end of *lin-4*S, with the 3′G perhaps missing from *lin-4*S (see Figures 3, 4, and 8).

Discussion

The *lin-4* Locus

The C. elegans heterochronic gene *lin-4* was identified by chromosomal walking from linked polymorphisms, char-

Figure 8. *lin-4* Transcripts and Complementarity between *lin-4* and *lin-14*

(A) Sequences for *lin-4*L and *lin-4*S RNAs, and a proposed secondary structure for *lin-4*L, predicted by the MULFOLD program (see Experimental Procedures). A secondary structure for *lin-4*S is not shown, owing to the uncertainty of the precise 3′ and 5′ nucleotides of *lin-4*S, which significantly affect structural predictions (see text). Sequences complementary to the *lin-14* 3′UTR (Wightman et al., 1991, 1993) are bold. Nucleotides that differ from the C. elegans sequence in one of the other three species (Figure 3) are indicated by the alternative nucleotide in smaller type. Also in smaller type and indicated by an asterisk is a U-to-C mutation introduced in vitro that retains *lin-4* function (see Results and Figure 3A). The position of the *ma161* C-to-U loss-of-function mutation is shown by an arrow.

(B) Complementarity between *lin-4* and seven copies of a repeated element in the 3′UTR of *lin-14* RNA (Wightman et al., 1993). Dots indicate absence of a nucleotide; dashes indicate one or more noncomplementary nucleotides. Only *lin-4*:*lin-14* complementarity that is conserved between C. elegans and C. briggsae (Wightman et al., 1993) is represented. The *lin-4*-complementary elements in *lin-14* vary somewhat among themselves and contain internal blocks of noncomplementary nucleotides of variable length and sequence. Note the five conserved noncomplementary nucleotides (5′ACCUC) of *lin-4* RNA. RNA sequences are deduced from DNA sequence and transcript mapping, and not from direct sequence analysis of RNA.

acterization of the molecular lesion associated with a *lin-4* allele, *e912*, and transformation rescue. *lin-4(e912)* is most likely a *lin-4* null allele, because its lesion removes most of the sequences corresponding to the 693 bp clone that rescues *lin-4* function. This observation, together with the fact that *lin-14* null alleles are completely epistatic to *lin-4(e912)* (Ambros, 1989), suggests that the sole function of *lin-4* is to downregulate *lin-14*.

The C. elegans *lin-4* gene is located within the intron of a gene of unknown function. Although *lin-4* is transcribed in the same orientation as that gene (R. C. L., R. L. F., and V.A., unpublished data), it is unlikely that the expression or function of *lin-4* depends on the host gene. This is because genomic clones that contain essentially only intron sequences of the host gene rescue *lin-4(e912)* extrachromosomally. These rescued animals appear fully wild type, despite the fact that *e912* deletes a substantial portion of the host gene in addition to the *lin-4* rescuing region (R. C. L., R. L. F., and V. A., unpublished data), indicating that the activity of the host gene is probably not required for *lin-4* function or for other aspects of normal development. Although our data indicate that the 693 bp SalI rescuing fragment contains the entire *lin-4* locus, all our rescue experiments were performed using high copy extrachromosomal arrays (Mello et al., 1991). Since we have not tested the 693 bp SalI clones in single copy, we cannot rule out the possibility that some additional sequences may be required for full *lin-4* activity. Generally, rescue of *lin-4* retarded defects by high-copy plasmid arrays was complete, and we did not consistently observe additional developmental defects that might be associated with overexpression of *lin-4* from the arrays.

lin-4 Is Unlikely to Encode a Protein

Two lines of evidence strongly suggest that none of the *lin-4* genomic sequence encodes protein. First, strategically located frameshift and nonsense mutations, as well as an ATG to ACG mutation, were introduced into several putative open reading frames of the C. elegans genomic clone without affecting rescuing activity. Second, comparison of the C. elegans, C. briggsae, C. remanei, and C. vulgaris *lin-4* genomic sequences (which are functionally interchangeable in C. elegans) revealed a high level of nucleic acid sequence conservation but no conserved open reading frames. Regions of conserved sequence do not show a degeneracy at the third base positions of putative codons that would be expected for protein coding sequences, and they contain deletions or insertions that would cause frameshifts. There are no splice donor or acceptor sites that would compensate for these frameshifts. Thus, if a conserved protein product were encoded by the *lin-4* locus, its synthesis would need to be directed by unorthodox translation start and stop signals. The 693 bp C. elegans fragment rescues in either orientation, arguing against the generation of a *lin-4* protein from the rescuing fragment using translational signals from the vector.

lin-4 Encodes Small RNAs

Northern blots and RNAase protection experiments with *lin-4* genomic probes detected the *lin-4*S and *lin-4*L small RNAs in total RNA of wild-type and rescued *lin-4(e912)* animals. Several observations argue that either *lin-4*L RNA or *lin-4*S RNA or both are the functional *lin-4* gene products. First, the *ma161* mutation, which results in loss of *lin-4* function, affects a conserved C of the *lin-4*L and *lin-4*S transcripts. Second, the DNA sequence encoding *lin-4*L and *lin-4*S is virtually identical in four nematode species (Figure 3A). Finally, the sequence complementarity between the *lin-4* transcripts and the *lin-14* mRNA (Wightman et al., 1991, 1993) suggests a direct base-pairing interaction between *lin-4* and *lin-14* RNAs and lends credence to the idea that *lin-4*S, *lin-4*L, or both may be the active *lin-4* gene product. Regions of very high sequence conservation were identified outside the *lin-4*L- and *lin-4*S-transcribed regions of the *lin-4*-rescuing clones. One of these is located between approximately 140 and 420 bp upstream of the transcribed sequences. This region is essential for *lin-4* function, but its orientation is not critical, because when it is inverted with respect to the transcribed region, partial rescuing activity is observed (Figure 1). These upstream conserved regions could contain regulatory sequences or encode additional small RNAs of an abundance too low to be detected in our experiments. No other *lin-4* products besides *lin-4*L and *lin-4*S were consistently detected in our experiments.

*lin-4*L and *lin-4*S may be transcribed from the same promoter, given that their 5′ ends appear to be identical. *lin-4*S could be generated by a posttranscriptional processing of *lin-4*L, or *lin-4*L may be produced by read-through of a *lin-4*S termination site. It is possible that either *lin-4*L or *lin-4*S is a nonfunctional byproduct of the biogenesis of the other, or that both are derived by the processing of a longer precursor transcript. *lin-4*L and *lin-4*S seem to be synthesized in approximately normal levels in a strain lacking the *lin-14* 3′UTR (R. L. F., unpublished data), suggesting that the generation of the two products likely occurs independently of an interaction with the *lin-14* mRNA. The fact that the lengths determined for *lin-4*S and *lin-4*L by RNAase protection and S1 mapping agree with the lengths estimated by Northern blot analysis indicates that the RNAs are not heavily modified posttranscriptionally (for example by significant splicing, polyadenylation, or covalent linkage to protein). However, these experiments would not necessarily detect slight modification or processing of the RNAs, such as modified nucleotides or cap structures. Further experiments are required to develop a complete description of *lin-4* RNA biogenesis and structure.

Antisense Complementarity to the *lin-14* 3′UTR

Sequence comparison of the *lin-4* transcripts with the 3′UTR of *lin-14* revealed that a sequence element repeated seven times in the *lin-14* 3′UTR is complementary to the *lin-4* transcripts. The significance of this complementarity to *lin-4* function is supported by phylogenetic and genetic observations. First, the seven repeated elements are conserved between the C. elegans and C. briggsae *lin-14* sequences (Wightman et al., 1993), consistent with a conserved *lin-4:lin-14* RNA base pairing and the observed conservation of *lin-4* function between these species. Second, the *lin-4(ma161)* loss-of-function mutation is located within a block of *lin-14*-complementary sequences, and so could affect *lin-4* activity by altering or destabilizing a *lin-4:*

lin-14 RNA hybrid. Third, the seven *lin-4*-complementary elements occur in the *lin-14* region deleted by *lin-14(gf)* alleles (Wightman et al. 1991). These observations strongly support the hypothesis that *lin-4* downregulates LIN-14 protein levels through a direct RNA–RNA interaction with the *lin-14* 3′UTR.

Although the longest stretch of contiguous sequence complementarity between *lin-4* and *lin-14* is 10 nt (Figure 8; Wightman et al., 1993), there is precedence for specific RNA–RNA interactions by such short regions of complementarity. For example, interactions among spliceosomal RNAs and between spliceosomal RNAs and the pre-mRNAs involve similarly short duplexes (Datta and Weiner, 1991; Madhani and Guthrie, 1992; Wassarman and Steitz, 1992; reviewed by Green, 1991; Guthrie, 1991). Since both *lin-4*S and *lin-4*L span the *ma161* lesion, and both include sequences complementary to the 3′UTR of *lin-14*, either could in principle interact with *lin-14* RNA. Although our 5′ and 3′ end-mapping experiments indicate that *lin-4*S might not contain all the *lin-14*-complementary sequences (Figure 8), the ambiguity in these data (± 1–2 nt at each end) leaves open the possibility that *lin-4*S contains the entire *lin-14*-complementary sequence. Since *lin-4*S is significantly more abundant than *lin-4*L, it seems likely that *lin-4*S plays the major role in base pairing with *lin-14* RNA, especially if the interaction is stoichiometric. Furthermore, the predicted secondary structure of *lin-4*L (Figure 8) would sequester within a stem most of the bases that are complementary to *lin-14*, perhaps rendering *lin-4*L inactive for base-pairing with *lin-14* mRNA. We cannot evaluate how the secondary structure of *lin-4*S might affect its function, since the precise 5′ and 3′ ends of *lin-4*S are uncertain, and this variability would significantly alter any proposed secondary structure (data not shown; see also legend to Figure 8).

A definitive proof of the proposed antisense pairing between *lin-4* and *lin-14* RNAs will require the construction of compensatory base-pairing mutations in *lin-14* and *lin-4*. Furthermore, determination of whether the interaction involves a stable complex or a transient (perhaps catalytic) interaction will require careful measurement of *lin-4* and *lin-14* RNA levels and direct tests for a complex in vivo.

Temporal Regulation of *lin-14* by *lin-4*

If *lin-4* RNA inhibits *lin-14* translation by an antisense interaction with the *lin-14* 3′UTR, then the temporal decrease of LIN-14 protein during C. elegans development is probably generated by a stage-specific increase in *lin-4* gene expression. According to this view, *lin-4* is inactive during late embryogenesis and the early part of the L1, permitting synthesis of a high level of LIN-14 protein, which programs the execution of L1 fates. Later in the L1 stage, *lin-4* gene expression is activated, and the *lin-4L* transcripts, *lin-4S* transcripts, or both bind to the complementary sequence elements in the 3′UTR of *lin-14* mRNA. This RNA–RNA interaction results, through some unknown mechanism, in a decrease in LIN-14 protein between the L1 and L2 stages. The resulting reduced level of LIN-14 protein specifies the expression of L2-specific fates in the L2, and L3-specific fates in the L3 (Ambros and Horvitz, 1987). To downregulate *lin-14* globally, *lin-4* would have to be activated stage-specifically in at least those cells that express *lin-14*. A developmental activation of *lin-4* might involve transcriptional induction or posttranscriptional modification of *lin-4* RNA, or both. Little is known about what developmental signal(s) activate *lin-4* gene expression, but it is likely that *lin-4* is not activated until after a food signal initiates postembryonic development, since LIN-14 protein does not decrease until after feeding (Arasu et al., 1991). *lin-4* may be required only in the L1 to modify *lin-14* mRNA into an untranslatable state, or also during later stages to inhibit LIN-14 protein synthesis continuously. Features of this model need to be tested by examining in detail the temporal profile of *lin-4* transcript levels and by artificially manipulating the level of active *lin-4* RNA during development.

Antisense Regulation via 3′UTRs

There are a number of known examples of natural antisense regulatory mechanisms that affect the stability or translatability of target RNAs (Kimelman and Kirschner, 1989; Hildebrandt and Nellen, 1992; reviewed by Simons, 1988; Eguchi et al., 1991). *lin-4* RNA probably does not control the stability of *lin-14* mRNA because the steady state levels of *lin-14* mRNA remain relatively constant during development and are not appreciably altered in a *lin-4* mutant background (Wightman et al., 1993). *lin-4* could inhibit LIN-14 protein synthesis indirectly, for example by modifying *lin-14* mRNA, or by localizing it to a subcellular compartment (such as the nucleus) where it is inaccessible to the ribosomes. Alternatively, *lin-4* RNA may bind to *lin-14* mRNA in the cytoplasm and inhibit its translation by interacting directly with components of the translational machinery. Distinguishing among these broad classes of models will require determination of the subcellular location of *lin-4* and *lin-14* RNAs.

Previously known natural antisense mechanisms that affect translation involve interaction of an antisense RNA with the 5′UTR of the mRNA and appear to affect ribosome binding (Liao et al., 1987; Simons, 1988; Kittle et al., 1989). By contrast, if *lin-4* does directly inhibit translation, then the proposed interaction between *lin-4* RNA and the 3′ end of *lin-14* might represent a novel kind of antisense translational control mechanism. Structural features of the *lin-4–lin-14* complex might interfere with a critical interaction between the 3′ and 5′ ends of the *lin-14* mRNA (Jackson and Standart, 1990; Gallie, 1991; Sachs and Deardorff, 1992). A regulation of poly(A) length might be involved or an interaction with translation factors, such as poly(A) binding protein, that act via the 3′ end (Sachs and Davis, 1989; Jackson and Standart, 1990). *lin-4* RNA may act in conjunction with proteins, but to date there is no genetic evidence for genes other than *lin-4* that regulate *lin-14* via its 3′UTR. It is noteworthy that *lin-4* RNA sequences that are not complementary to *lin-14*, and hence are looped out of the predicted *lin-4:lin-14* hybrid, are completely conserved (Figure 8B). This loop may bind cellular proteins involved in the control of *lin-14* mRNA. Proteins that interact with *lin-4* RNA might also have functions es-

sential for cell viability, and so it may not be surprising that the genes encoding them have apparently not been previously identified by a visible heterochronic phenotype.

The involvement of 3′UTR sequences in mRNA localization, stability, or translation and the developmental roles of posttranscriptional regulatory mechanisms mediated by 3′UTR sequences are becoming increasingly clear. For example, the mRNAs of mouse protamines (Kwon and Hecht, 1991), Drosophila *bicoid, hunchback* (Wharton and Struhl, 1991), and *nanos* (Gavis and Lehmann, 1992), and nematode *fem-3* (Ahringer and Kimble, 1991; Ahringer et al., 1992) and *tra-2* (Goodwin et al., 1993) contain developmentally significant posttranscriptional regulatory sequences in their 3′UTRs. As yet, no antisense regulatory RNAs have been shown to be involved in the above cases, but the possibility has not been ruled out for any of them. Thus, *lin-14* may not be the only developmental gene that is posttranscriptionally regulated via its 3′ end by a small antisense RNA. Because the regulatory affect of a small antisense RNA such as *lin-4* is directed toward the translation of a specific antisense partner mRNA, translational control of gene expression by small antisense RNAs may be a particularly effective strategy for providing a developmental switch of very high specificity.

Small Regulatory RNAs

lin-4 may represent a class of developmental regulatory genes that encode small antisense RNA products, but such genes might be difficult to identify by standard genetic approaches. If *lin-4*S is active in *lin-14* shut-off, it may be among the smallest functional RNAs so far identified. The mouse and human *Xist* genes are believed to encode active RNAs important in the control of X chromosome inactivation (Brockdorff et al., 1992), and the mouse *H19* gene also appears to encode an untranslated RNA (Brannan et al., 1990). However, the products of both these genes are much larger than *lin-4* RNA; the *Xist* RNAs are 15–17 kb and the *H19* RNAs are 2.6 kb. Small RNA molecules play a critical role in essential cellular processes such as RNA processing (Green, 1991; Guthrie, 1991) or editing (Simpson, 1990). However, even compared with this group, the *lin-4* RNAs are unusually small. At 61 and 22 bases, respectively, *lin-4*L and *lin-4*S are shorter than the 80 base guide RNA molecules from Leishmania (Blum et al., 1990). Given its small size, *lin-4*S may interact directly with relatively few cellular proteins.

The scarcity of *lin-4* alleles is consistent with the relatively small size of the *lin-4* locus. We identified a single allele, *ma161*, after screening a relatively large number (>20,000) of mutagenized genome equivalents. Other small RNA genes within introns have been identified in C. elegans as part of a genomic sequencing project (Sulston et al., 1992) (R. Durbin, personal communication) and in mouse (Leverette et al., 1992) and human (Tycowski et al., 1993). The small target size for mutations that affect such genes without altering the host gene and the lack, in many cases, of distinguishing features in their genomic sequence might contribute to the difficulty of identifying other small RNA genes like *lin-4*.

Experimental Procedures

Nematode Strains

C. elegans strains used in this work were the following: wild-type C. elegans var Bristol (strain N2) (Brenner, 1973); C. elegans var Bergerac; *lin-4(e912)* (strain VT371); *lin-4(e912)vab-9(e1744)/mnC1* (VT499); *bli-3(e768)lin-4(e912)dpy-10(e128);lin-14(n179ts)* (VT373); *lin-4(e912)/mnC1* (VT460); *lin-4(e912)dpy-10(e128)/mnC1; lin-14(n179ts)* (VT441); *lin-4(e912)*; *maEx114(pVTSal3,pRF4)* (VT509); and *lin-4(e912); maEX-115(pVTSal6,pRF4)* (VT510). Other Caenorhabditis species used in this work are the following: C. briggsae, obtained from the Caenorhabditis Genetics Center, C. vulgaris, and C. remanei, obtained from Scott Baird.

Mutagenesis

Noncomplementation screens were performed by treating N2 males with ethylmethane sulfonate (Brenner, 1973), mating them with hermaphrodites of the strain VT441 *lin-4(e912)dpy-10(e128)/mnC1; lin-14(n179ts)*, and screening the non-Dpy F1 for hermaphrodites that displayed a Lin phenotype (Chalfie et al., 1981).

RFLP Mapping

Physical genetic mapping of Bristol/Bergerac RFLPs in the *lin-4* region was performed as described elsewhere (Ruvkun et al., 1989). A strain congenic to Bristol for all chromosomes except the *lin-4* region of LGII was obtained by performing nine successive backcrosses of *lin-4(e912)* to Bergerac *lin-4(+)*, isolating the *lin-4(+)* allele at each backcross. This strain and Bergerac were crossed with *bli-3(e768)lin-4(e912)dpy-10(e128)/+++* males to obtain *bergerac lin-4(+)/bristol bli-3(e768)lin-4(e912)dpy-10(e128)* animals. Progeny from these hermaphrodites were screened for Lin-non-Dpy, Dpy-non-Lin, and Bli-non-Lin recombinants. DNA from these recombinants was analyzed by southern blotting (see below) using as probes clones situated to the left of *dpy-10* on the C. elegans physical map (Coulson et al., 1986, 1988).

Southern Blots

Preparation of genomic DNA, agarose gel electrophoresis, and Southern blotting were performed using standard techniques (Ausubel et al., 1989). Probes were prepared from cosmid or YAC DNAs by random primed synthesis (Feinberg and Vogelstein, 1983). YAC DNA was fractionated by field inversion gel electrophoresis (MJ Research) after spheroplast lysis in the gel (Carle and Olson, and 1985). YAC DNA was recovered from gel slices by electroelution, followed by phenol extraction and ethanol precipitation.

Isolation of Genomic Clones

C. elegans genomic DNA was cut with EcoRI, ligated into λZAP vector (Stratagene), and plated on CES200. Replicas of the library were probed, in the presence of total yeast carrier DNA, with YAC DNA labeled by random primed synthesis. Plaques that hybridized to Y18C1 probe but not to Y42A4 probe were identified, the inserts were excised in vivo as Bluescript plasmid clones, and the plasmids were reanalyzed by restriction enzyme digestion and Southern hybridization with the above YAC probes. Plasmids that retained their differential hybridization pattern were pVT2D, pVT3C, and pVT6G (see Results). A probe from pVT2D was used to identify clones containing *lin-4* sequences in λZAP libraries of genomic DNA from C. briggsae and C. remanei. The C. briggsae *lin-4* sequence was determined for a 1.4 kb EcoRV clone pVTbr12, and the C. remanei *lin-4* sequence was determined for a 2.0 kb EcoRI clone pVTreA1. Short stretches of conserved sequence outside the rescuing region were used as a guide to design primers (rcl17 and rcl18) to amplify *lin-4* sequences from C. vulgaris worms. The 0.9 kb PCR product from C. vulgaris was inserted into Bluescript vector, creating clone pVTvul4c.

DNA Sequencing

Sequencing was performed by the method of Sanger (Sanger et al., 1977) using the Sequenase 2.0 kit (United States Biochemicals). Sequence was analyzed with a Macintosh, using DNA Strider 1.2 (Marck, 1988) or MULFOLD (Jaeger et al., 1989a, 1989b; Zuker, 1989) software, and with a VAX, using the GCG Sequence Analysis Software Package (version 7.1) by Genetics Computer Group, Incorporated

(Devereux et al., 1984). Sequence alignment was generated by the GCG program LINEUP and optimized manually.

Single-stranded *lin-4* sequences (one strand only) were amplified from *ma161* genomic DNA using PCR and a 500-fold excess of primer rf2 to rcl2. The region amplified using these primers included all of the 693 bp SalI fragment and about 100 bp 5′ of this region. Two separate PCR amplifications were performed, and the single-stranded pools were sequenced in parallel to avoid potential sequence alterations introduced by Taq polymerase. Primers used for sequencing were rcl2, rcl7, and rf3.

C. elegans Transformation

Plasmids for injection were prepared by standard alkaline lysis miniprep procedure, followed by polyethylene glycol precipitation (Ausubel et al., 1989). Plasmids were injected at concentrations ranging from 10–30 μg/ml, together with pRF4 (which confers a dominant Rolling phenotype [Mello et al., 1991]) at 100–150 μg/ml in TE (10 mM Tris [pH 8.0], 1 mM EDTA). The cosmid C02B6 was injected at 15 μg/ml with PRF4 at 150 μg/ml. Clones were injected into either VT499 or VT460. For VT499, rescue of the *lin-4* defects was scored by examining F1 Vab-Rollers and determining whether they were Non-Lin. For VT460, rescue was signified by F1 Rolling animals that produced progeny among which none of the Lin animals were Rolling, and all of the non-Lin animals were Rolling. From such rescued F1 animals, transformed lines were established that continued to exhibit rescue. Lines were also established for clones that failed to rescue, and these were analyzed by PCR to establish that they contained the injected DNA. Phenotypes (egg laying, body length, adult alae) were assayed using the dissecting microscope.

Exonuclease III Deletions and In Vitro Mutagenesis

pVT2DCla was double-digested with ApaI–ClaI or PstI–EcoRV and treated with ExonucleaseIII as described (Ausubel et al., 1989). Oligonucleotide-directed mutagenesis of pVT2Dcla or pVTSal3 was performed as described (Kunkel et al., 1987), except that single-stranded template DNA was prepared by superinfection of cells containing the Bluescript plasmid (Stratagene) with M13 K07 helper phage.

Preparation of Nematode RNA

Animals were harvested from nematode growth medium agarose plates, washed several times in M9 buffer, collected by centrifugation, and frozen at −70°C. An equal volume of 2× LETS (200 mM LiCl, 20 mM EDTA, 20 mM Tris [pH 7.8], 2% SDS) and 2 vol of phenol, $CHCl_3$, isoamyl alcohol (IAA) were added to the frozen worm pellet, which was then quickly thawed. Animals were lysed by vortexing with glass beads for several minutes. The aqueous phase was separated by centrifugation and reextracted three times with phenol, $CHCl_3$, IAA. RNA was precipitated from this aqueous mixture by adding LiCl (final concentration 0.2 M) and 3 vol of ethanol and incubating at −20°C overnight. RNA was collected by centrifugation, and contaminating DNA was digested with RQ1 DNAase (Promega, 5–10 U/mg of RNA) for 30 min in 40 mM Tris (pH 7.9), 10 mM NaCl, 6 mM $MgCl_2$, and 0.1 mM $CaCl_2$ at 37°C. RQ1 DNAase was removed by extraction with phenol, $CHCl_3$,. IAA, and RNA was reprecipitated by the addition of NaOAc (final concentration, 0.3 M) and 3 vol of ethanol at −20°C.

Northern Blots

Total RNA, 30 μg per lane, was separated by electrophoresis in TBE (89 mM Tris [pH 8.0], 89 mM sodium borate, and 2 mM EDTA) through an 11% urea–acrylamide gel (0.75mm × 16cm × 16cm). RNA was electrophoretically transferred to Zeta-Probe membrane (Bio-Rad) in 0.5 × TBE at 250 mA overnight at 4°C. RNA was crosslinked to the membrane by UV irradiation (1200 μJ, Stratagene UV Stratalinker), and then the membrane was baked at 80°C for 30 min. Antisense *lin-4* probe was synthesized by runoff transcription from the plasmid pMspI digested with EcoRI using T7 RNA polymerase and standard transcription conditions (Ausubel et al., 1989). pMspI was constructed by insertion of an MspI fragment from bp506 to bp568 of *lin-4* (Figure 3) into the ClaI site of pBluscript SK(−). Hybridization and washing conditions with pMspI antisense probe were according to the Zeta-Probe specifications of the manufacturer for hybridization with RNA probes, except that the hybridization temperature was 42°C and the most stringent wash was 0.2 × SSC, 0.1% SDS at room temperature. Northern blots were stripped by washing at 90°C in 0.1 × SSC, 0.5% SDS and then were reprobed with an anti-U6 oligonucleotide probe, rfMGH12. rfMGH12 was 5′ end-labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase according to standard procedures (Ausubel et al., 1989). Hybridization and wash conditions were as designated for oligonucleotide probes in the Zeta-Probe Blotting Membranes instruction manual. In some cases, DNA oligonucleotides (4 μg of 69mer, 3 μg of 33mer, 7 μg of 19mer) were run as size markers on the gel. This marker lane was cut from the gel prior to blotting and stained in 10 μg/ml ethidium bromide for several hours and visualized by UV illumination.

S1 Nuclease Protection

S1 analysis using oligonucleotide probes was as described (Ausubel et al., 1989). Total RNA (30 μg) was analyzed in each sample. The initial hybridization of probes (10^5 cpm/sample) to RNA was under standard conditions except for the rfMGH8 probe, in which case hybridization was in an aqueous solution (final concentration, 1 M NaCl, 0.17 M HEPES [pH7.5], and 0.3 mM EDTA [pH 8]) at 50°C overnight. S1 digestion reactions contained 300 units of S1 nuclease and were at 37°C unless otherwise noted. S1 digestion products were analyzed on 10–12% urea-acrylamide gel, and the gels were exposed wet to XAR-5 film. Probes were generated as described below. rfMGH1 was 5′ end-labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase under standard reaction conditions. rfMGH5 was labeled at its 3′ end as follows: 300 ng of rfMGH5 and 350 ng of rfMGH6 were resuspended in 4 μl of TEN50 (10 mM Tris [pH7.5], 1 mM EDTA [pH 8.0], 50 mM NaCl), heated to 90°C, and then annealed at room temperature. The annealed mixture was diluted 1:3 in TEN50. Of this annealed oligonucleotide mixture, 1 μl was incubated with 10 μl of [^{32}P]dATP (6000 Ci/mmol), 10 μl of [^{32}P]dCTP (6000 Ci/mmol), 5 U of DNA polymerase I Klenow fragment in 50 mM Tris (pH 7.5), 10 mM $MgCl_2$, 1 mM DTT, and 50 μg/ml BSA at 37°C for 15 minutes. The reaction was stopped by heating to 75°C for 10 minutes, and the oligonucleotides were ethanol precipitated in the presence of 10 μg of carrier tRNA, resuspended in 1X TBE plus 80% formamide, heated to 90°C, and separated by electrophoresis through a 10% urea–acrylamide gel. The band corresponding to 3′ end–labeled rfMGH5 was cut from the gel, and the oligonucleotide was eluted by soaking the gel slice in 0.3 M NaOAC overnight at room temperature. The solution was extracted with phenol, $CHCl_3$, IAA, and 3′ end–labeled rfMGH5 was precipitated by addition of 3 vol of ethanol. 3′ end–labeling of rfMGH8 was similar to that described for rfMGH5, except that rfMGH7 was used in the annealing reaction, which created a duplex with a 2 nt 5′ overhang for the fill-in reaction, and only [^{32}P]dATP was used in the labeling reaction. Oligonucleotide markers were labeled with [γ-^{32}P]ATP and T4 polynucleotide kinase.

RNAase Protection

RNAase protection experiments were according to standard procedures (Ausubel et al., 1989) with the following modifications: 15 μg of total RNA and 5 × 10^5 cpm of probe were used in each reaction. To reduce contaminating salts in the samples prior to loading the gel, several additional ethanol precipitations, at room temperature and with no added salt, were performed. Gels were exposed wet to XAR-5 film. Probes were generated by runoff transcription of 0.5 μg of pMspI (predigested with either EcoRI or DdeI) using T7 RNA polymerase and 60 μCi of [^{32}P]UTP (800 Ci/mmol). RNA markers were generated by runoff transcription of pBS (cut with AccI, SmaI or EcoRI) with T3 RNA polymerase in the presence of [^{32}P]UTP under standard reaction conditions.

Primer Extension

Primer extension analysis was as described (Ausubel et al., 1989), with the following modifications: 30 μg of total RNA and 200 units of MMLV reverse transcriptase were used in each reaction. rfMGH3 was 5′ end–labeled using [γ-^{32}P]ATP and T4 polynucleotide kinase. Reverse transcription was at 42°C. The extended products were separated on a 10% polyacrylamide gel that was exposed wet to XAR-5 film. Size markers were generated by 5′ end–labeling oligonucleotides of specific lengths with T4 polynucleotide kinase and [λ-^{32}P]ATP or by sequencing mp18 using Sequenase, −40 reverse primer, and [^{32}P]-dATP according to standard protocols (Sanger et al., 1977).

Oligonucleotides

rf2 (amplification of *ma161*), 5′-TAGTGGAACTCTATTGT-3′.
rf3 (sequencing *ma161*), 5′-CTAGACAATTTCTAGAG-3′.
rfa (sequencing), 5′-GACGCGTCGCCAAGCGGTGGTG-3′.
rfb (sequencing), 5′-AGTGAAGAGACTGGGTCAT-3′.
rfc (sequencing), 5′-CCTGTTCCCTGAGACCTCAA-3′.
rfd (sequencing), 5′-ACCGTCCTGGTACCCGGAGA-3′.
rcl2 (amplification and sequencing *ma161*), 5′-ACCGCTTCTTCTCCCTA-3′.
rcl3 (sequencing), 5′-GGTAAGTGTGATGAGTC-3′.
rcl4 (amplification of C. vulgaris and sequencing), 5′-CAACAGCTCCATCACTC-3′.
rcl7 (sequencing *ma161*), 5′-GACGAAGCGACCGAATG-3′.
rcl8 (mutagenesis), 5′-GCTGGCACTACCACC-3′.
rcl9 (mutagenesis), 5′-CGTAGACTACCTCGC-3′.
rcl17 (amplification of C. vulgaris and sequencing), 5′-GGCACAGGGGACAATCA-3′.
rcl18 (amplification), 5′-GGAATTC(GT) (AG)TT(GA)TA(GA)TC(AG)AC(TC)TCTG-3′.
rcl21 (sequencing), 5′-AAAACAGCCGCCTT-3′.
rfMGH1 (5′ end S1 mapping of *lin-4*L), 5′-GTGAAGCATCAATAGTACACTCACACTTGAGGTCTCAGGGAACAGGCCGGAAGCATAAACTC-3′.
rfMGH3 (primer extension analysis), 5′-CCCAGGTGTGAAGCATCAATAGTACACTCACAC-3′.
rfMGH5 (3′ end mapping *lin-4*L), 5′-CTCACGAATTCGACTACCGTCCTGGTACCCGGAGACGCCAGGTGTGAAGCATCAATAGTACA CTC-3′.
rfMGH6 (3′ end labeling of rfmgh5), 5′-GTGTGAGTGTACTATTGATGCTTCACACCT-3′.
rfMGH7 (3′ end labeling of rfmgh8), 5′-TTCCCTGAGACCTCAAG-3′.
rfMGH8 (3′ end mapping *lin-4*S), 5′-GACTGCTAAGATAGTACACTCACACTTGAGGTCTCAGGG-3′.
rfMGH12 (anti-U6), 5′-GCAGGGGCCATGCTAATCTTCTCTGTATTGTTCCAATTTTAGTATATGTTCTCCCTATAGTGAGTCGTATTA-3′.
va22 (mutagenesis), 5′-GTCCTGGTACCCGGAGAGCCCAGGTGTGAAGCGTCAATAG-3′.

Acknowledgments

We are grateful to Xianjie Yang for isolating the recombinants used in mapping *lin-4*, to Amy Mills for in vitro mutagenesis, to Scott Baird and David Fitch for providing nematode strains, to Bob Barstead for his cDNA library, to Ann Rougvie for her C. elegans EcoRI genomic library, to Steve Roels for extensive assistance with the computer sequence analysis, and to Bruce Wightman and Gary Ruvkun for the C. briggsae genomic library and for communicating their results prior to publication. The sequence complementarity between *lin-4* and *lin-14* RNAs was noticed simultaneously by the Ambros and Ruvkun labs as a result of a cooperative sharing of unpublished data. This research was supported by research grants to V. A. from the National Institutes of Health and the National Science Foundation (grants GM34028 and 9200604, respectively), and postdoctoral fellowships to R. F. from the Daymon Runyon–Walter Winchell Foundation and the United States Public Health Service (GM15396). We also thank the Massachusetts General Hospital Department of Molecular Biology and Hoechst Aktiengesellschaft for financial support for R. F., and we thank Gary Ruvkun for providing research facilities for R. F. Some nematode strains were provided by the Caenorhabditis Genetics Center, which is funded by the National Institutes of Health National Center for Research Resources.

Received August 6, 1993; revised September 17, 1993.

References

Ahringer, J., and Kimble, J. (1991). Control of the sperm-oocyte switch in *Caenorhabditis elegans* hermaphrodites by the *fem-3* 3′ untranslated region. Nature *349*, 346–348.

Ahringer, J., Rosenquist, T. A., Lawson, D. N., and Kimble, J. (1992). The *Caenorhabditis elegans* sex determining gene *fem-3* is regulated post-transcriptionally. EMBO J. *11*, 2303–2310.

Ambros, V. (1989). A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans. Cell *57*, 49–57.

Ambros, V., and Horvitz, H. R. (1984). Heterochronic mutants of the nematode *Caenorhabditis elegans*. Science *266*, 409–416.

Ambros, V., and Horvitz, H. R. (1987). The *lin-14* locus of *Caenorhabditis elegans* controls the time of expression of specific postembryonic developmental events. Genes Dev. *1*, 398–414.

Arasu, P., Wightman, B., and Ruvkun, G. (1991). Temporal regulation of *lin-14* by the antagonistic action of two other heterochronic genes, *lin-4* and *lin-28*. Genes Dev. *5*, 1825–1833.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1989). Current Protocols in Molecular Biology (New York: John Wiley and Sons).

Blum, B., Bakalara, N., and Simpson, L. (1990). A model for RNA editing in kinetoplastid mitochondria: "guide" RNA molecules transcribed from maxicircle DNA provide the edited information. Cell *60*, 189–198.

Brannan, C. I., Dees, E. C., Ingram, R. S., and Tilghman, S. M. (1990). The product of the *H19* gene may function as an RNA. Mol. Cell. Biol. *10*, 28–36.

Brenner, S. (1973). The genetics of *Caenorhabditis elegans*. Genetics *77*, 79–94.

Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., Swift, S., and Rastan, S. (1992). The product of the mouse *Xist* gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell *71*, 515–526.

Carle, G. F., and Olson, M. V. (1985). An electrophoretic karyotype for yeast. Proc. Natl. Acad. Sci. USA *82*, 3756–3760.

Chalfie, M., Horvitz, H., and Sulston, J. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, 59–69.

Coulson, A., Sulston, J., Brenner, S., and Karn, J. (1986). Toward a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA *83*, 7821–7825.

Coulson, A., Waterston, R., Kiff, J., Sulston, J., and Kohara, Y. (1988). Genome linking with yeast artificial chromosomes. Nature *335*, 184–186.

Datta, B., and Weiner, A. M. (1991). Genetic evidence for base pairing between U2 and U6 snRNA in mammalian mRNA splicing. Nature *352*, 821–824.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. *12*, 387–395.

Eguchi, Y., Itoh, T., and Tomizawa, J. (1991). Antisense RNA. Annu. Rev. Biochem. *60*, 631–652.

Feinberg, A., and Vogelstein, B. (1983). A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. *132*, 6–13.

Gallie, D. R. (1991). The cap and poly(A) tail function synergistically to regulate mRNA translational efficiency. Genes Dev. *5*, 2108–2116.

Gavis, E. R., and Lehmann, R. (1992). Localization of *nanos* RNA controls embryonic polarity. Cell *71*, 301–313.

Goodwin, E. B., Okema, P. G., Evans, T. C., and Kimble, J. (1993). Translational regulation of *tra-2* by its 3′ untranslated region controls sexual identity in C. elegans. Cell *75*, 329–339.

Green, M. R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. *7*, 559–599.

Guthrie, C. (1991). Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. Science *253*, 157–163.

Hildebrandt, M., and Nellen, W. (1992). Differential antisense transcription from the Dictyostelium *EB4* gene locus: implications on antisense-mediated regulation of mRNA stability. Cell *69*, 197–204.

Jackson, R. J., and Standart, N. (1990). Do the poly(A) tail and 3′ untranslated region control mRNA translation? Cell *62*, 15–24.

Jaeger, J. A., Turner, D. H., and Zuker, M. (1989a). Improved predictions of secondary structures for RNA. Proc. Natl. Acad. Sci. USA *86*, 7706–7710.

Jaeger, J. A., Turner, D. H., and Zuker, M. (1989b). Predicting optimal

and suboptimal secondary structure for RNA. Meth. Enzymol. *183*, 281–306.

Kimelman, D., and Kirschner, M. W. (1989). An antisense mRNA directs the covalent modification of the transcript encoding fibroblast growth factor in Xenopus oocytes. Cell *59*, 687–696.

Kittle, J. D., Simons, R. W., Lee, J., and Kleckner, N. (1989). Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5′ end of the target RNA and a loop in the anti-sense RNA. J. Mol. Biol. *210*, 561–572.

Kunkel, T. A., Roberts, J. D., and Kakour, R. A. (1987). Rapid and efficient site-specific mutagenesis without phenotypic selection. Meth. Enzymol. *154*, 367–382.

Kwon, Y. K., and Hecht, N. B. (1991). Cytoplasmic protein binding to highly conserved sequences in the 3′ untranslated region of mouse *protamine 2* mRNA, a translationally regulated transcript of male germ cells. Proc. Natl. Acad. Sci. USA *88*, 3584–3588.

Leverette, R. D., Andrews, M. T., and Maxwell, E. S. (1992). Mouse U14 snRNA is a processed intron of the cognate *hsc70* heat shock pre-messenger RNA. Cell *71*, 1215–1221.

Liao, S.-M., Wu, T.-H., Chiang, C., Susskind, M., and McClure, W. (1987). Control of gene expression in bacteriophage P22 by a small antisense RNA. Genes Dev. *1*, 197–203.

Madhani, H. D., and Guthrie, C. (1992). A novel base-pairing interaction between U2 and U6 snRNAs suggests a mechanism for the catalytic activation of the spliceosome. Cell *71*, 803–817.

Marck, C. (1988). 'DNA Strider': a C program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucl. Acids Res. *16*, 1829–1836.

Mello, C. C., Kramer, J. M., Stinchcomb, D., and Ambros, V. (1991). Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J. *10*, 3959–3970.

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

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

Sachs, A. B., and Davis, R. W. (1989). The poly(A) binding protein is required for poly(A) shortening and 60S ribosomal subunit–dependent translation initiation. Cell *58*, 857–867.

Sachs, A. B., and Deardorff, J. A. (1992). Translation initiation requires the PAB-dependent poly(A) ribonuclease in yeast. Cell *70*, 961–973.

Sanger, F., Nicklen, S., Coulson, A. (1977). DNA sequencing by chain terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Simons, R. W. (1988). Naturally occurring antisense RNA control: a brief review. Gene *72*, 35–44.

Simpson, L. (1990). RNA editing: a novel genetic phenomenon? Science *250*, 512–513.

Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-Mieg, J., Qui, L., Dear, S., Coulson, A., Craxton, M., Durbin, R., Berks, M., Metzstein, M., Hawkins, T., Ainscough, R., and Waterston, R. (1992). The *C. elegans* genome sequence project: a beginning. Nature *356*, 37–41.

Tycowski, K. T., Shu, M.-D., and Steitz, J. A. (1993). A small nucleolar RNA is processed from an intron of the human gene encoding ribosomal protein S3. Genes Dev. *7*, 1176–1190.

Wassarman, D., and Steitz, J. A. (1992). Interactions of small nuclear RNAs with precursor messenger RNA during *in vitro* splicing. Science *257*, 1918–1925.

Wharton, R. P., and Struhl, G. (1991). RNA regulatory elements mediate control of Drosophila body pattern by the posterior morphogen *nanos*. Cell *67*, 955–967.

Wightman, B., Burglin, T., Gatto, J., Arasu, P., and Ruvkun, G. (1991). Negative regulatory sequences in the *lin-14* 3′-untranslated region are necessary to generate a temporal switch during *Caenorhabditis elegans* development. Genes Dev. *5*, 1813–1824.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in C. elegans. Cell *75*, this issue.

Wood, W. B., and the community of *C. elegans* researchers, eds. (1988). The Nematode *Caenorhabditis elegans* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press).

Zuker, M. (1989). On finding all suboptimal foldings of an RNA molecule. Science *244*, 48–52.

GenBank Accession Numbers

The accession numbers for the Caenorhabditis sequences reported in this paper are U01829, U01830, U01837, and U01838.