Molecular Characterization of Animal microRNAs: Sequence, Expression, and Stability.

by

Nelson C. Lau

B.S. Biochemistry and Molecular Biology
University at Albany, SUNY, 1999

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL
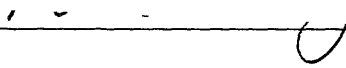FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
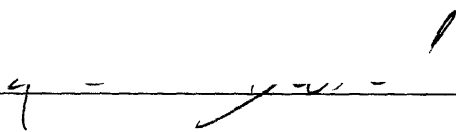
SEPTEMBER 2004

© 2004 Nelson C. Lau. All Rights Reserved.

The author hereby grants to MIT permission to reproduce and to
distribute publicly paper and electronic copies of this thesis
document in whole or in part.
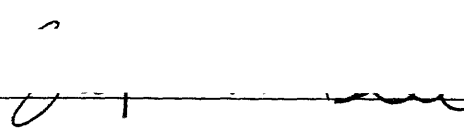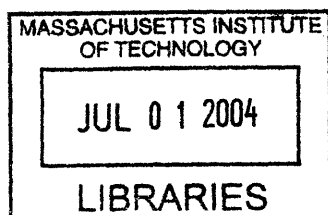

Signature of Author:

_____
Department of Biology
(June 30, 2004)

Certified by:

_____
David P. Bartel
Professor of Biology
Thesis Supervisor

Accepted by:

_____
Stephen P. Bell
Professor of Biology
Chair, Biology Graduate Committee

1

Molecular Characterization of Animal microRNAs: Sequence, Expression, and Stability.

by

Nelson C. Lau

ABSTRACT

Multicellular organisms possess natural gene-regulatory pathways that employ small RNAs to negatively regulate gene expression. In nematodes, the small temporal RNAs (stRNAs), *lin-4* and *let-7*, negatively regulate genes important in specifying developmental timing. A gene-silencing pathway present in plants, fungi and animals called RNA interference, involves the conversion of long double-stranded RNA into short interfering RNAs, which can serve to negatively regulate endogenous genes or suppress the replication of viruses and transposons.

To investigate how wide a role small RNAs play in regulating gene expression in animals, we developed a RNA cloning procedure and first applied it to the cloning of small RNAs from the nematode, *Caenorhabditis elegans*. In addition to cloning *lin-4* and *let-7* sequences, our study revealed a large number of conserved and highly expressed small RNAs with features reminiscent of stRNAs. Because not all of these small RNAs were expressed in temporal fashion, we and others have referred to this novel class of tiny RNAs as microRNAs.

We completed an extensive census of microRNA (miRNA) genes in *C. elegans* by cloning and bioinformatics searches to lay the groundwork for future functional studies. Our census marked the detection of nearly 90 *C. elegans* miRNAs, estimated an upper-bound of about 120 miRNA genes in *C. elegans*, and detailed the conservation and clustering of miRNA sequences. We also determined the high molecular abundance of several miRNAs in *C. elegans* and Hela cells.

In an effort to understand the reason for the high molecular abundance of miRNAs, we constructed an inducible miRNA-expressing cell line to measure the stability of animal miRNAs. Time course measurements suggested a long (>24 hours) half-life for two well-conserved miRNAs. These cells lines may also be useful for other functional studies, such as validation of putative mRNA target genes.

Acknowledgements

Table of Contents

Introduction

microRNAs: Nuggets of RNA Reveal a Goldmine of Biology

Although microRNAs (miRNAs) remained hidden in the "genomic dirt" of intergenic regions and introns for more than 25 years after the first mutants were isolated [1, 2], they now represent a treasure trove of new biology. Not only are these small RNAs found in many branches of multi-cellular organisms, they also exert an extraordinary level of control on gene expression, and have refreshed our appreciation of RNA's functional diversity in eukaryotes. While the importance of miRNAs in development is undisputed, establishing the integral roles of each miRNA remains a challenge.

Progress in miRNA research has been rapid and dramatic. The catalog of miRNAs in diverse organisms has swelled in recent years (Figure 1). Before 2001, only two miRNA genes were known, but quickly the discovery of other miRNAs surged as cloning and computational methods homed in on small RNA genes. Identification and prediction of target mRNAs regulated by miRNAs have recently seen similar increases. I will survey how genetic, biochemical, and genomic approaches have uncovered this rich repository of gene riboregulators, and review our expanding knowledge about the biogenesis and function of miRNAs.

A Tale of Worm Mutants

The current flood of miRNA knowledge began with research performed in Victor Ambros's lab and Gary Ruvkun's lab who were interested in the problem of heterochrony – the proper ordering and timing of developmental processes. What drove their research was a set of *Caenorhabditis elegans* mutants with various defects in the manner by which the cells should divide and differentiate according to their programmed lineage fates [3]. Of these lineage-defective, or *lin* mutants, two became a focus of study in determining how heterochronic genes

interact – *lin-4* and *lin-14*. While the *lin-4* loss-of-function (LOF) mutant contained somatic

cells that reiterated early larval divisions and whose maturity was retarded, the *lin-14* LOF

mutants instead displayed cells that precociously differentiate [4]. Cloning and characterization

of *lin-14* by the Ruvkun lab indicated that proper division of those certain somatic cells

depended on a downregulation of LIN-14 during development [5, 6]. The intriguing connection

between *lin-4* and *lin-14* was first hinted by three observations: (1) *lin-4* LOF mutants showed

similar phenotypes to the *lin-14* gain-of-function (GOF) mutants which failed to downregulate

LIN-14; (2) accumulation of LIN-14 was also seen in the *lin-4* mutant; and (3) *lin-14* LOF

mutants in a *lin-4* LOF background partially suppressed the *lin-4* heterochronic defects [7-10].

After the *lin-14* gene was cloned, the gene was found to encode a protein that appeared

localized to the nucleus and was temporally regulated [5, 6]. Characterization of the *lin-14* GOF

mutants revealed lesions in the 3' untranslated region (UTR) of the *lin-14* mRNA which

suggested sequence elements that might be acted upon by the *lin-4* gene to negatively regulate

LIN-14 levels [7-9, 11]. Although one might have expected that *lin-4* could be an RNA-binding

protein, such notions were dispelled when the Ambros lab cloned a ~700 nt genomic fragment

capable of rescuing the *lin-4* mutant, and which could not encode a significant protein product

[12]. So, the Ambros lab turned their attention to the transcript encoded by the *lin-4* locus, and

found that *lin-4* actually encoded two small non-coding RNAs, a ~60 nt RNA capable of folding

into a foldback structure, and a ~22 nt RNA thought to derive from the longer RNA (Figure 2).

Developmental Northern blots indicated that *lin-4* RNA expression was temporally regulated,

because it was detected only after the late larval stage 1 (L1), which coincided with a drop in

LIN-14 levels [8, 12]. Thanks to active communication between the Ambros and Ruvkun labs,

an illuminating molecular connection was proposed for *lin-4* and *lin-14* – the sequence of *lin-4*

could base-pair to sites in *lin-14* that matched exactly the locations of the deregulating genetic lesions. Although the base pairing interactions contained bulges and mismatches, both the *lin-4* RNA and the binding sites of the *lin-14* 3'UTR were conserved in the related nematode, *C. briggsae*. Furthermore, this interaction caused LIN-14 levels to be repressed while RNA levels remained unchanged, indicating a mechanism of translational repression. Together, these observations put forth an unprecedented model of gene regulation in nematodes.

The *lin-4* – *lin-14* story was compelling yet very peculiar, because neither *lin*-4 homologs nor analogous regulatory examples had been seen in other animals. Although another heterochronic gene, *lin-28*, was later discovered to be a second target of regulation by *lin*-4, this example was also isolated to the study of *C. elegans* [13]. Not until the year 2000 did additional studies in heterochronic genes by the Ruvkun lab broaden the appeal of RNAs like *lin-4*. While the previous studies examined early larval development in worms, the Ruvkun lab reinvestigated heterochronic genes specifying the later larval stages. They employed a screen that isolated mutants defective in the proper developmental timing of adult nematode structures, such as longitudinal cuticle structures, and pin-pointed developmental roles for genes like *let-7, lin-41, lin-42 and daf-12* [14]. The previous insights of the *lin-4* RNA and its ability to repress target heterochronic genes evidently proved valuable in guiding the subsequent cloning and functional assignment of *let-7*. In a case of déjà vu, the Ruvkun lab showed that *let-7* was also a non-coding ~22-nt RNA that could negatively regulate the translation of its target mRNA, *lin-41*, by base-pairing to conserved sites in the *lin-41* 3' UTR (Figure 2) [14]. A larger ~60-nt precursor RNA capable of forming a foldback structure was also detected for *let-7*, and it was temporally expressed only after the late L3 stage [14].

8

The complex gene orchestration needed to properly time worm development was now clarified by the key genetic switches of *lin-4* and *let-7* (Figure 2). In this pathway, the L1 to L2 transition and the L3 to L4 transition could be triggered by the expression of the *lin-4* and *let-7* RNAs, respectively. This resulted in reductions in LIN-14 and LIN-41 levels, and genes affected downstream of *lin-14* and *lin-41* could subsequently adjust for cells to decide their proper developmental fate. The importance of this elegant regulatory pathway is evident from the conservation of these genes and the mode of regulation amongst other nematodes, and defects in pathway components result in serious phenotypic abnormalities, such as bursting vulva, sterility, and lethality [10, 14].

Nematodes are interesting and peculiar creatures for a number of reasons, but could they also be the only animal with small regulatory RNAs? Previous database searches had failed to reveal homologs for *lin-4*, however, the Ruvkun lab's query on dipteran and mammalian genomes suggested the presence of homologs for *let-7*. To test their hypothesis, the Ruvkun lab and a team of collaborators assembled and assayed an impressive array of RNA samples from various phyla in the animal kingdom. Some samples lacked any detectable *let-7* RNA signal, but surprisingly all bilateral animals were confirmed for *let-7* expression. Furthermore, *let-7* was detected only in late stages of flies; and in human tissues, its signal was absent from bone marrow, which consists of mostly of undifferentiated cells [15]. These results lead the Ruvkun lab to propose that *let-7* temporal regulation and developmental timing function might also be broadly conserved amongst animals. By coining the term 'small temporal RNAs' (stRNAs) for *lin-4* and *let-7*, the Ruvkun lab anticipated other small endogenous RNAs were out there, waiting to be found.

Panning Transcriptomes for Small RNAs

In addition to the breakthrough in detecting *let-7* across animals, two other lines of research pointed to a universe of small endogenous RNAs waiting to be uncovered: RNA interference (RNAi) and RNA genomics (RNomics). RNAi was born out of early research on post-transcriptional gene silencing (PTGS), whereby a natural process conserved from plants to animals seemed to be silencing the expression of endogenous and artificial genes at the RNA level (for a review see [16]). Seminal work by Fire and Mello indicated that PTGS might be triggered by long double-stranded RNA (dsRNA), and with RNA playing a central role, the new term RNA interference (RNAi) was coined for the silencing phenomenon [17, 18]. Soon after, experiments by Hamilton and Baulcombe suggested that small RNA species between 20-25 nt were being generated in the course of PTGS [17, 18], while work in *Drosophila* extracts showed that these small RNAs derived from the dsRNA [19]. The enzyme thought to initiate RNAi was an RNase III called Dicer, which dices long dsRNA into small ~22 nt RNAs [20]. These short RNAs generated by Dicer were called small interfering RNAs (siRNAs) because they specify the interference of expression of a target mRNA by base pairing to the mRNA and guiding the cleavage and degradation of that mRNA [21]. The siRNAs exert their function by associating into an assembly of proteins to form the RNA-Induced Silencing Complex, or the RISC, which contains an endoribonuclease [22]. Suspecting more than coincidence in the similar lengths of siRNAs and stRNAs, several labs showed that Dicer generated mature stRNAs, because a mutant worm lacking Dicer accumulated stRNA foldback precursors and exhibited pleiotropic defects similar to *lin-4* and *let-7* defects [23-26]. These observations lead to the hypothesis that additional Dicer substrates might be expressed endogenously in plants and animals.

Meanwhile, RNomics approaches being applied to eukaryotes and prokaryotes indicated a plethora of small (<300 nt) RNAs that could be detected at high levels [27, 28]. Computational screens and immunoprecipitation experiments indicated that eukaryotes like yeast and even archaeabacteria contained many novel non-coding RNAs that represented new classes of small nucleolar RNAs (snoRNAs), known to have roles in modifying and regulating ribosomal RNAs through base-pairing interactions [29, 30]. Likewise, a functional genomics effort uncovered a host of small non-coding RNAs in prokaryotes, which were later shown to have roles in regulating the expression of other bacterial genes, either through antisense mechanisms or association with protein factors to affect translation or transcription termination [28, 31]. An unbiased purification and cloning of small RNAs from mouse brain also identified many highly-expressed non-coding RNAs that appeared to correspond to additional snoRNAs [32]. Work in RNAi and stRNAs hinted at additional endogenous small RNAs, but previous investigation in mouse brain only cloned RNAs >50 nt long. Could cloning of even smaller RNAs from other organisms be informative or would it be doomed to sifting through many small degradation intermediates of longer RNAs?

The first sign that RNA cloning would be informative came from endeavors by the Tuschl lab to directionally clone siRNAs produced from exogenous dsRNA added to *Drosophila* embryo lysates [21]. Not only did the Tuschl lab recover siRNAs, but they also obtained clones that differed from degradation products of structural RNAs (i.e. tRNA and rRNA) or mRNAs. Some of these alternative clones instead matched retrotransposon sequences, suggesting that RNAi, known to play a role in suppressing transposon hopping, was indeed generating small RNAs of ~22nt from endogenous double-stranded transcripts from mobile elements [21]. The Tuschl lab continued to clone small RNAs from *Drosophila* as well as HeLa cells, while the

Ambros lab and Bartel lab (Chapter 1 of this thesis) directionally cloned small RNAs from

C.elegans. Although our cloning techniques all differed, all three labs converged on the

discovery of a preponderance of endogenous small ~22nt RNAs in animals [33-35]. Two small

RNAs were discovered by all three groups, while the Tuschl and Bartel groups managed to clone

let-7 sequences, indicating the efforts had at least identified an important positive control.

Developmental Northern blots confirmed the expression of small ~22-nt long RNAs in worms,

flies and human cells, but unlike the small temporal RNA *let-7*, several of the small RNAs,

including the two commonly discovered by all three groups, seemed to be expressed throughout

worm and fly development. So, a new classification was created – the term "microRNAs"

(miRNAs) was coined for this extensive class of ~22 nt noncoding RNAs that appeared to be

processed from short endogenous stem-loop precursors resembling those of *lin-4* and *let-7*.

Thanks to the nearly-complete worm, fly and human genomes, mapping and

characterizing these miRNA clones could be easily accomplished, and several features could be

defined for this class of small RNA genes. First, a segment of nearby genomic sequence and the

miRNA sequence could be folded by RNA-folding algorithms into foldback structures

reminiscent of the foldback structures of the *lin-4* and *let-7* precursors (Figure 3). Indeed,

Northern blots could confirm the presence of several miRNA precursors [33-35], and the *dcr-1*

mutant defective in Dicer seemed to accumulate the miRNA precursor [35]. Second, discrete

single loci could be matched to the great majority of the miRNAs, as opposed to repeat elements

like transposons seen earlier for siRNAs cloned by the Tuschl lab. Third, the miRNA loci often

sit far away from annotated genes, often in intergenic expanses, or at closest within the introns of

protein-coding genes. Those miRNA genes in isolated intergenic regions might represent

independent transcription units, while many miRNAs within introns are likely expressed from

the host gene. Some miRNAs even appeared to cluster together, with only a few nt separating each stem loop, suggesting multiple miRNAs could be transcribed from operon-like units (Figure 3C) [33, 34]. Fourth, the miRNA genes were surprisingly conserved in related species, and moderately conserved amongst animals – the vast majority of C.elegans miRNAs had detectable homologs in C.briggsae, and miR-1, miR-2, and miR-34 shared the distinction with let-7 RNA as being nearly perfectly conserved between worms, flies and humans (Figure 3A). Fifth, the mature miRNA seemed to predominantly derive from just one side of the foldback precursor, either from the 5' or 3' side of the foldback, although one small RNA clone had been obtained from the other side of miR-56 (that less abundant RNA was named miR-56*) [34]. Such characteristics described above later become important in defining whether future candidate small RNAs could be assigned as miRNAs.

These initial miRNA cloning efforts did not nearly approach saturation, and left much room for further miRNA gene discovery. Large scale small RNA cloning projects were eventually completed for the invertebrates, C. elegans and D. melanogaster [36-38]. Meanwhile, cloning efforts were significantly expanded in vertebrates, with one focus directed at specific tissues in mice [39, 40], while another effort examined miRNAs in early development stages of zebrafish [41]. Multiple groups cloned many miRNAs from various mammalian samples, ranging from immortal cell lines, to mouse embryonic stem cells, to cancer tissue samples, and to bone marrow [39, 42-45]. Some mammalian miRNAs have been identified from purified protein complexes. For example, an RNA cloning study of a ~15S ribonucleoprotein (RNP) complex containing Gemin3, Gemin4 and eIF2C2 yielded a surprising number of novel miRNAs and implicated an miRNP in which miRNAs exert their biological function [46]. Following earlier observations that lin-4 fractionated with polysomes on sucrose gradients [47], polysomes

from mammalian neurons were isolated and subjected to small RNA cloning, further enlarging

the list of animal miRNAs identified by molecular means [48]. Recently, a few miRNAs have

even been found in a mammalian virus [49]. Five new miRNAs from Epstein-Barr virus (EBV)

were discovered by RNA cloning, and Northern blots not only confirmed their expression but

also indicated that cell lines in different latency stages of EBV infection can exhibit differential

expression of the viral-specific miRNAs [49].

Animals are clearly rich repositories of miRNAs, but could other multi-cellular

organisms harbor such genes? The RNAi phenomenon had been clearly observed in plants, fungi

and even protozoans, and Dicer homologs were also evident in *Arabidopsis thaliana*,

Neurospora crassa, and *Schizosaccaromyces pombe*, suggesting that small RNAs might reach

deep into ancient eukaryotic lineages [20]. Naturally, RNA cloning efforts were extended to

plants, fungi and trypanosomes, and many endogenous small RNAs could be identified as unique

from degradation products of structural and messenger RNAs [50-56]. However, only

endogenous siRNAs derived from retrotransposons or genomic repeat elements could be

discerned from RNA cloning in *T.brucei* and *S.pombe*, respectively. Nevertheless, miRNAs

could be identified in plants, although a few features are distinct from those commonly observed

in animal miRNAs. The plant miRNAs were similar to animal miRNAs in that (1) they were

highly expressed 20-22nt single-stranded RNAs often detectable on Northern blots, (2) they were

well conserved among different plant species (although not conserved with any animal

miRNAs), (3) they were derived from one strand of a hairpin precursor structure, and (4) their

processing was dependent on an RNase III enzyme, DCL1, the plant Dicer homolog [51, 55].

However, plant miRNAs deviate from animals in that the precursors of plant miRNAs have not

been detected on Northern blots, and sometimes required much longer stretches of genomic

sequence (>150 bp vs. 80 bp in animal miRNAs) to form a foldback structure that could be a suitable DCL1 substrate [51] (Figure 3B). Additionally, some plant miRNAs match multiple loci, where some matches appear to represent the original miRNA gene, but other matches corresponded to the antisense sequence of a protein coding gene, hinting at antisense recognition of mRNA targets by plant miRNAs [50, 51]. This feature would later bear upon fruitful efforts to find additional plant miRNA targets [57].

Mining Genome Databases for More miRNAs

Conservation of RNAi from animals to protozoa indicates an ancient role for small RNAs carrying out gene silencing processes, with some researchers postulating that RNAi serves as an innate immune response against RNA viruses, transposons, or other aberrant gene expression [58, 59]. However, the existence of miRNAs seems restricted to plants and animals, suggesting that miRNAs might have emerged with the advent of multi-cellular life. Given the lack of sequence conservation between any plant and animal miRNA sequence, the miRNA genes might have emerged independently after the split between animal and plant lineages. However, within each respective lineage, miRNA sequences have been quite well conserved, suggesting that miRNA sequences are retained due to important, conserved target interactions. This conservation of miRNA sequence and striking conservation in precursor structure within the respective animal and plant lineages provided a means to computationally mine genomes for new miRNA candidates. Bioinformatic methods could test our understanding of miRNA features, and might readily identify new miRNA genes in a fashion unbiased by miRNA molecular abundance or cloning efficiency. Although a handful of new candidate miRNAs were identified by standard

BLAST searches for homologs of cloned miRNAs, several groups have employed more

sophisticated algorithms to identify broader sets of novel candidate animal miRNAs.

The two most sensitive algorithms available for finding miRNA genes are MirSeeker,

which was applied against fly genomes, and MirScan, which has been applied against nematode

and vertebrate genomes. In MirSeeker, aligned foldback structures between *D.melanogaster* and

D.pseudoobscura were scored against the pattern of sequence divergence being higher in the

loops rather than in the arms of the stem loop to identify positive candidates. MirSeeker could

capture 75% of the previously cloned and conserved *Drosophila* miRNAs and allowed for the

validation of 48 additional miRNAs [60]. MirScan, on the other hand, used a list of aligned stem

loops loosely conserved between two or more genomes, and assigns scores based on structural

and sequence similarities between the candidate and a set of previously known *C.elegans*

miRNAs stemloops. When applied to foldback sequences conserved between *C.elegans* and

C.briggsae, MirScan captured 93% of previously cloned *C.elegans* miRNAs that possessed

C.briggsae homologs, and suggested 35 new miRNAs that were validated by cloning, Northern

blots or PCR amplification [36]. MirScan has also been applied to finding new human miRNAs

that are conserved in mouse and fish, yielding 188 vertebrate miRNAs that stand out in a natural

peak from other candidate foldback structures [41]. The strengths of MirSeeker and MirScan are

their sensitivity and specificity, which enables firm estimates on the upper bounds of miRNA

genes – about ~1% the number of the protein coding genes in invertebrate genomes (120 genes

in worms, 110 genes in flies) [36, 60]. About the same percentage of the human genes appear to

be miRNAs, although for firm estimates on the total miRNA genes, the verdict is still out and

awaits the sequencing of additional vertebrate genomes and the development of a second

generation of miRNA gene finding programs [61, 62].

Other computational efforts have been applied to predicting *C.elegans* miRNAs [38, 63].

Although these efforts had modest abilities in pinpointing previously cloned miRNAs, some new

miRNAs were identified, but many of the remaining candidates await verification.

Computational searches for miRNA genes are not limited to animals. A newly-developed

algorithm, called MirCheck, looks for foldback structures in *Arabidopsis* and *Oryza* genomes,

and then scans them for conserved 21-mer sequences that could represent a novel miRNA [64].

It also checks for homologous matches between conserved 21-mer sequences and the antisense

of annotated genes which could represent potential novel miRNA targets. In total, MirCheck has

uncovered 23 new *Arabidopsis* miRNA loci represented within 7 new families [64].

With so many labs prospecting for possible miRNAs, a set of guidelines and a sequence

repository were eventually established to aid researchers in evaluating candidate sequences and

to facilitate proper nomenclature of miRNAs. Some of the critical criteria for classifying

miRNAs have included specific, detectable expression, phylogenetic conservation, derivation

from a foldback precursor, and processing by an RNase III enzyme [65]. Additionally, a

distinction in origins was spelled out for miRNAs versus siRNAs – miRNAs arise generally from

endogenous gene units encoding transcripts with short fold-back structures, while groups of

siRNAs derive from long dsRNA precursors from viral genes, transposons, and other aberrant

genes. These criteria became vital in differentiating miRNAs from other small RNAs that may

represent endogenous siRNAs or other less characterized small RNAs like tncRNAs [38, 53].

The vast majority of miRNA genes have been named numerically by the order in which they

were discovered, with highly similar miRNAs sometimes sharing the same number followed

with a letter or number suffix. A clearinghouse for miRNA sequences has been established at

the Rfam databases of RNA families to catalog a non-redundant list of existing miRNAs and to assign unique names to newly discovered miRNAs [66].

How Do Cells Make miRNAs?

In the eukaryotic cell, transcription is often a highly regulated process, and generally, transcribed RNA species must be processed into their mature, functional state. So when miRNAs appeared on the scene, it was natural to wonder what transcriptional and maturation processes would be involved for this new class of RNAs. Developmental and tissue-panel Northern blots as well as RNA cloning had illuminated the temporal expression and spatial specificity seen for many miRNAs, which argued for active transcriptional control of these small RNA genes [33-35, 39, 40]. As of yet, though, no definitive promoters or biochemical determinations have established the RNA polymerase responsible for transcribing miRNAs, although a repressive DNA element and a protein factor has been implicated in the temporal control of *let-7* expression in *C.elegans* [67]. Two other studies also indicate the temporal regulation of *let-7* expression in *Drosophila* may be mediated by an ecdysone hormonal signal which is known to affect the transcription of many genes [68, 69]. It is possible that miRNAs might be transcribed by RNA Pol III, which generally transcribes shorter RNAs like tRNAs and some spliceosomal RNAs, because the ~60-70 nt size of the foldback precursors is similar to tRNA lengths. However, there is mounting evidence to indicate that miRNA precursors themselves derive from longer transcripts that would be predominantly transcribed by RNA Pol II. This evidence includes observations that miRNA sequences could be mapped to large cDNAs in EST databases [33, 70], multiple miRNAs can cluster within a region akin to an operon [33,

34], miRNAs mapped to introns often reside on the transcribed strand of the host gene [36, 37, 60], and RT-PCR can detect long transcripts encompassing the fold-back precursor [37, 71].

Experimental proof that miRNAs could derive from longer transcripts was demonstrated with mammalian cell extracts that showed primary transcripts containing a single miRNA or a cluster of miRNAs could be processed into ~70 nt RNAs (putative foldback precursors), for which this product was then competent for further processing into ~22 nt RNAs (presumably the mature miRNA) [71]. These long primary transcripts were dubbed pri-miRNAs, which appear to be processed in the nucleus [71]. The factor responsible for pri-miRNA processing in animals is Drosha, an evolutionarily-conserved, nuclear-localized RNAse III, which trims away the flanking sequences from the precursor miRNA (pre-miRNA) and leaves behind a 3' two-nt overhang, a signature of RNase III enzymes (Figure 4, Step A2) [72]. Drosha cleavage can define the 5' end of mature miRNAs that derive from the 5' arm of the pre-miRNA, suggesting that Dicer need only cleave the portion of the pre-miRNA proximal to the loop.

How does the pre-miRNA that is generated in the nucleus by Drosha flow into the cytoplasm for further processing by Dicer? Although passive diffusion might be an option, several labs suspected an active process for pre-miRNA transport, given that other non-coding RNAs like tRNAs are actively exported to the cytoplasm by protein machinery [73]. Three labs independently demonstrated that Exportin-5 (Exp5), a RanGTP-dependent transport protein, was functionally different from other nuclear export receptors, in that its primary role was to transport pre-miRNAs across the nuclear membrane (Figure 4, Step A3) [74-76]. Genetic evidence further supporting Exp5 as the primary nuclear transporter of miRNAs was seen in an *Arabidopsis* mutant defective in the HASTY gene, a homolog of Exp5, which displays

pleiotropic developmental defects similar to other plant mutants where miRNA production is compromised (Figure 4, Step P4) [77].

Although plants and animals may share similarity in protein factors for miRNA nuclear export, there is a notable difference in pre-miRNA formation. Animals require a nuclear and a cytoplasmic RNase III to cleave the different ends of the pre-miRNA, while in plants, only DICER-LIKE-1 (DCL1), one of the four Dicer-like RNase III enzymes in plants, has a demonstrated role in processing miRNAs [51, 55, 78]. Given that DCL1 is primarily nuclear localized in plants, it is likely that DCL1 supplies both cleavage events on a pre-miRNA, perhaps so efficiently that fold-back precursors do not get a chance to accumulate (Figure 4, Step P2 and P3) [51, 61]. Alternatively, animal pre-miRNA might tend to accumulate because RNase III cleavage events are partitioned, and because transport of pre-miRNAs across the nucleus is relatively slow [79]. Since animals utilize two different RNase III enzymes to cut the different ends of pre-miRNAs, this may explain how the boundaries of the mature miRNA are defined, but the exact sequence and/or structural determinants specifying the register of Drosha or Dicer cleavage is still unclear [72, 80]. Perhaps equally puzzling are the recent reports indicating that while worms and humans have only one identifiable Dicer enzyme, flies have two Dicers; *Drosophila* DCR-1 processes miRNA, while DCR-2 forms siRNAs [81, 82]. Despite these unresolved questions, our clearer understanding of miRNA biogenesis has allowed researchers to use either Pol II or Pol III promoters to robustly express miRNAs at will now in many cells [44, 79, 80]; and this technique in turn will lead to better expression systems for small hairpin RNAs used in functional RNAi genomic studies [83].

RISC and miRNAs, Some Assembly (and Unwinding) Required

In both plants and animals, Dicer processes long dsRNA into siRNAs, for which both

sense and antisense small RNAs are generated. This presented a quandary as to why only one

strand of a pre-miRNA accumulates as the dominant miRNA sequence. Despite the presence of

bulges in the pre-miRNA stem, several miRNA* clones obtained from extensive miRNA cloning

studies could be matched to the dominant miRNA to form duplexes with two nt 3' overhangs, a

signature of siRNA duplexes, suggesting that Dicer processes pre-miRNAs like a perfectly

complementary duplex [36, 51]. This conundrum of asymmetric miRNA accumulation was

finally explained in recent work by two groups that demonstrated a single strand of an siRNA

duplex could often be more effective at initiating RNAi than the opposite strand of the duplex

[84, 85]. These groups noticed that the 5' end of the more effective siRNA strongly correlated

with the duplex end with the lower relative thermodynamic base-pairing energy [84, 85]. When

this observation was extended to duplexes of miRNAs and miRNA*'s from flies, the elegant rule

proposed above predicted the correct accumulation of the miRNA strand in the majority of cases.

For miRNA duplexes where the unwinding energy of either duplex end was similar, it was

shown that the miRNA* also accumulates appreciably [37, 84]. So, a model was proposed

where a RNA helicase samples and selects the weaker duplex end for unwinding and loading of

one siRNA or one miRNA strand over the other into the RISC or miRNP, respectively, and it is

presumed the non-incorporated strand is degraded quickly (Figure 4, Step 5) [84, 85]. In the

natural role for RNAi in silencing RNA viruses and transposons, multiple siRNAs are generated,

and it is hard to imagine why the RNAi response would depend on strand selectivity, therefore

the siRNA/miRNA helicase with strand bias might have evolved to ensure more efficient and

specific assembly of miRNAs, where maximizing the levels and specificity of miRNAs is important for miRNA function.

What are the steady-state levels of miRNAs in the cell? Quantitative Northern analysis has provided some estimates of miRNA levels in nematode and Hela cells, and the most abundant miRNAs in *C.elegans*, miR-2 and miR-58, accumulate to ~50 thousand copies per average cell, while *lin-4* and *let-7* RNAs are present in 3-5 thousand copies per average cell [36]; the measurement of *let-7* and other miRNAs in worms could be higher since *let-7* has been postulated to have cell-specific expression [67]. In Hela cells, *let-7* is expressed at about one thousand copies per cell, but other tissue-specific miRNAs assayed in human tissues may exceed 100 thousand copies per cell [79]. The molecular abundance of miRNAs are much higher than most mRNAs, and even rivals the abundant levels of "house-keeping" RNAs like U6 RNA of the spliceosome (~50 – 100 thousand copies per cell). Thus, miRNAs and their associated proteins represent one of the most abundant RNPs in the cell. Why do animal miRNAs accumulate to such high levels? Only miniscule amounts of siRNA are needed to prompt gene silencing, because the mode of RNA cleavage by RISC is catalytic [17, 86]. However, miRNAs like *lin-4* and *let-7* do not appear to mediate significant mRNA turnover, and bind to target sites with imperfect complementarity, suggesting that high cellular concentrations might be important for target interaction or function in repressing translation, which is supported by experiments that use siRNAs to mimic and examine the miRNA target interaction [87, 88]. Other models have further proposed miRNAs might have many mRNA targets, so high miRNA levels might be needed to effectively suppress multiple target mRNAs [89].

Comparing and contrasting siRNAs and miRNAs has revealed insights not only in the biogenesis of small regulatory RNAs, but also in their maturation. While origins of miRNAs and

siRNAs are clearly different, their biogenesis undoubtedly intersects because Dicer and an

unidentified helicase handle both siRNAs and miRNAs. Several reports have extended this

theme implicating that the miRNP shares the same biological activity and core components as

the RISC, namely the ability to cleave a target mRNA. Around the time plant miRNAs and

predicted plant mRNA targets were reported by the Bartel lab [51, 57], the Carrington lab

detected and mapped the *in vivo* sites of cleavage of several Scarecrow-Like (*SCL*) mRNAs,

which correspond to a conserved site antisense to miR171 [90]. The cleavage sites on the *SCL*

mRNAs were primarily at the 10th nt opposite from the 5' end of the miRNA, which is a pattern

characteristic of RNA cleavage by siRNA-programmed RISC [21, 90, 91]. Meanwhile, the

Zamore lab demonstrated that an artificial mRNA containing a perfect complementary site for

let-7 could be efficiently cleaved by endogenous *Drosophila* and human *let-7* [92].

Furthermore, they showed that immunoprecipitations of the miRNP (containing eiF2C2, Gemin3

and Gemin4) contained cleavage activity for the *let-7* perfect complementary site. This notion

of common protein complexes for miRNAs and siRNAs has been reinforced by detection of

miRNAs in the RISC [93-96]. Given the evidence that the miRNP is probably the same as

RISC, only the term RISC will be discussed henceforth.

Although some RISC factors have been determined in Hela and *C.elegans* extracts, the

majority of factors have been examined in *Drosophila*. The list of protein factors known to be in

Drosophila RISC include Ago2, VIG, Tudor-SN, Aubergine, FXR, Dcr-2, and R2D2 [93-98],

and several additional bands on protein gels from RISC purifications remain unidentified. One

particularly elusive but highly sought-after RISC factor is the putative Slicer enzyme which

mediates the endonucleolytic cleavage on the mRNA. Recent reports argue that Slicer is not the

factor Tudor-SN, a micrococcal nuclease homolog, because metal co-factor requirements and the

presence of a phosphate on the 5' end of the 3' cleavage product are inconsistent with the activity of a micrococcal nuclease [99, 100]. An additional complexity of dissecting RISC is that various isoforms probably exist in organisms, reflected in part by the various different sizes of RISC purified from different labs [22, 46, 98, 101, 102]. Purification of siRNA-containing RNPs from trypanosomes and biochemical fractionation of miRNAs from worms and mammalian cells strongly suggest that RISC is associated with ribosomes and polyribosomes in the course of its action to cleave mRNA or when mediating translation repression [47, 48, 54, 103]. Thus, factors involved in the translation machinery, and even membrane-associated factors, like eIF2C1 (also known as GERp95), will likely contribute to diversity of RISC isoforms [102]. Tissue-specific isoforms of RISC might also exist, given that a putative helicase, *Armitage*, is preferentially expressed in *Drosophila* ovaries and is important for RISC assembly [97, 104].

Even though a clearer model of miRNA biogenesis and maturation is emerging (Figure 4), several steps in the miRNA pathway remain unresolved. For example, the actions of the numerous Paz and Piwi Domain (PPD) proteins are still mysterious [105]. This class of proteins includes the founding members, Piwi, Argonaute, and Zwille/Pinhead, for which the PAZ protein domain is named after, and is a deeply conserved but enigmatic class with important roles in eukaryotic development and stem cell regulation [106]. PPD proteins are known to act in the RNAi and miRNA pathway, but the roles of these proteins are sometimes exclusive. For example, in *C.elegans* the depletion of *alg-1* and *alg-2* by dsRNA treatment causes severe developmental defects due to perturbations of miRNA function in RISC, however, the RNA silencing machinery using siRNAs appears unaffected in the *alg-1/alg-2* depleted worms [23]. Conversely, mutant worms defective in *rde-1* are unable to perform RNAi, yet have no effect on production or function of miRNAs [23, 107]. In plants, different Argonaute (Ago) genes also

seem to have different roles – Ago1 is necessary for proper accumulation and function of miRNAs and ~21-nt siRNAs [108-110], while Ago4 utilizes ~24-nt siRNAs to stimulate DNA and histone methylation [111]. Where the PPD proteins act and how they fit into the miRNA pathway is unclear, but one PPD protein, Ago-2 in *Drosophila*, appears to be at the center of small RNA pathways, because it interacts with Dicer-1 and Dicer-2, is associated with both siRNAs and miRNAs, and copurifies with RISC activity [93-96, 98].

Another important but unresolved question in miRNA biogenesis and maturation are the various protein factors that contribute to miRNA accumulation or stability. In plants, a dsRNA binding protein, HYL1, is required for miRNA accumulation but is not necessary for RNAi [112, 113], and yet the animal homologs, R2D2 in flies and RDE-4 in worms do not affect miRNA levels but instead are thought to mediate the transfer and loading of siRNAs from Dicer to the RISC [81, 114]. The mechanism of action for these proteins is unclear, and animal homologs specific for miRNAs has not been identified, but perhaps these proteins add an additional level of diversity in RISC assembly pathways. An additional factor in plants, HEN1, appears to affect miRNA steady-state levels, thus eliciting developmental defects, but has only been characterized genetically and not extensively enough to establish its role in the pathway [55, 115]. Plant viral proteins like p19, p21 and HC-Pro suppress RNAi (and thus allow the virus to evade the antiviral defense and replicate), but can also perturb miRNA function and levels [116-118]; and recent evidence that p19 and p21 can bind miRNA duplexes suggest they act downstream of Dicer processing [119, 120]. A fourth plant viral protein, p69, also appears to upregulate miRNA accumulation and function, but only partially blocks the RNAi pathway [121]. Determining where these viral proteins act might reveal additional steps in the miRNA pathway. Termination of the miRNA pathway, namely the degradation of miRNAs, also has not been defined, although

the recent isolation of a siRNA-specific RNase, ERI-1, and studies of miRNA stability in mammalian cells could lend some insight to this subject [79, 122].

The Search for miRNA Function (and Targets)

A key to understanding the function of each miRNA relies on knowing which target mRNAs can be regulated by a miRNA, and this, in turn, can depend on understanding the biochemical mechanism of target interaction and the regulatory effect of miRNA binding. If there is near perfect complementarity between the miRNA and the mRNA, mRNA cleavage can ensue, and RNA levels may diminish (Figure 5A) [90, 92, 123, 124]. In plants, many target mRNAs can be predicted to base-pair extensively to miRNAs, suggesting that cleavage is a dominant mechanism of action [57, 64]. Cleavage of these plant targets by miRNAs has been verified experimentally and *in vivo,* and interestingly, the majority of these targets are transcription factors [57, 64, 70, 90, 113, 117, 123, 125, 126]. This lends to the hypothesis that plant cells use miRNAs to downregulate transcription factor messages, and thus alter the transcriptome towards states that promote cell differentiation [57, 61]. Very few mRNAs outside of plants, however, are known to be targeted by this cleavage mechanism because extensive complementarity with an miRNA is rarely detected; two notable exceptions being *HOXB8* cleavage by miR-196 and BALF5 cleavage by miR-BART2 [49, 124].

For genetically identified miRNA targets in animals, a bulged interaction with mismatches between the miRNA and the mRNA is more commonly seen, which severely impairs mRNA cleavage, and instead inhibits productive translation of the mRNA (Figure 2 and Figure 5B) [8, 12-14, 127-131]. This mode of translation inhibition does not appear to be restricted to miRNAs, because transfection of siRNAs that base-pair imperfectly to an mRNA

site (thus mimicking a miRNA interaction) can also downregulate protein expression [87, 132, 133]. The biochemical mechanism of RISC-mediated translational repression is still quite obscure: a few hints are that repression may occur after translation initiation and that RISC associates with polyribosomes [47, 48, 103, 134]. There is also experimental data indicating that binding of the 5' portion of the miRNA sequence to the mRNA is a critical determinant for translational repression, but a strong interaction at the 3' portion of the miRNA may compensate for weaker interactions at the 5' end of the miRNA [133]. Another recent study also suggests that the size and configuration of the internal mismatch bulge of the miRNA binding site might be important for target recognition [135].

Even before the sequence determinants of miRNA-mRNA interaction were examined experimentally, the importance of the 5' portion of the miRNA for target recognition had already been recognized. Sequence gazing and testing of computationally predicted targets indicated that conserved sites in the 3' UTR of *Drosophila* genes known to be downregulated post-transcriptionally during development showed uncanny base-pairing with the 5' portion of a subset of fly miRNAs [136, 137]. Amongst families of highly related miRNAs, the conservation has almost always been observed in the 5' portions of the miRNA sequence, supporting the hypothesis that miRNA families together may be regulating a common class of mRNA targets [36-38, 61, 130, 136]. This interaction between the first 2-8 bases of the miRNA sequence and the target site has been referred to as the 'seed' interaction [138], such that RISC might be envisioned to scan target sites with this 7-bp recognition sequence in the miRNA, and after "seeding" the initial target recognition, additional interactions within the 3' end of the miRNA stabilizes RISC for gene-silencing action (Figure 5B). Predictions that enforce the requirement for a conserved 'seed' interaction exhibit substantial specificity [138], and the 'seed' interaction

might also explain off-target silencing of genes from siRNA transfection experiments, because many of these off-targets profiled from array experiments share sequence elements that base pair to the 5' portion of the siRNA [139].

While a number of computational efforts have predicted many targets for animal miRNAs [135, 137, 138, 140, 141], assessing the validity of these targets necessitates some caution until a conclusive assay confirms miRNA-dependent downregulation. Some of this caution stems from whether the methods require conservation of the 3' UTR sites amongst different animal genomes, and whether the correct methodology was chosen for shuffling control sequences to match dinucleotide biases in animal genomes [61, 138]. Three studies have employed different reporter techniques to lend experimental support to these target predictions. Stark and colleagues used a transgenic fly sensor that fuses entire 3' UTR sequences from predicted fly targets to a fluorescent protein construct; this reporter construct is co-injected with a miRNA-expressing construct, and cell clones in the imaginal discs diminish in fluorescence when the expressed miRNA downregulates sensor expression [128, 137]. Because the algorithm by Lewis and colleagues predicts with most confidence targets that contain at least two miRNA binding sites, only the segment of the 3' UTR that is flanked by two sites was fused to a luciferase reporter, which was then assayed in mammalian cells expressing or transfected with miRNA sequences [138]. Finally, Kiriakidou and colleagues report a variant of the cell-based luciferase assay, whereby only a single site, or a miRNA Response Element (MRE), is incorporated into the 3' UTR of the luciferase construct [135]. Six *Drosophila* targets exhibited regulation by two fly miRNAs (the total number of targets tested is not clear) [137], 11 out of 15 mammalian targets with two miRNA sites also indicated miRNA-dependent downregulation [138], and 10 mammalian targets with single MREs that conform to the "proper" interior bulge

interaction appear to be downregulated by an miRNA, although 11 other target MREs that display significant base-pairing to miRNAs but do not conform to "binding rules" fail to be downregulated [135].

The list of animal miRNA targets that are supported by these reporter assays (see Table 1, entries marked with a †) is still very small compared to the number of targets predicted, but this list can highlight a potential breadth of function amongst predicted animal targets and inform the range of regulation imparted by miRNAs. While the majority of predicted plant miRNA targets serve as transcription factors, the predicted animal targets are not concentrated in a particular ontology class, and have functions ranging from transcription regulation to signal transduction to basic metabolism [135, 137, 138, 140, 141]. For those targets whose luciferase reporter expression could be reduced by an miRNA, the level of regulation ranged between 2–5 fold in repression for a single MRE [135], while two-site 3' UTR segments exhibited a range of 1.6–21 fold repression (the average repression is 7.6 fold) [138]. Multiple synthetic miRNA imperfect complementary sites (>4x) inserted into reporter gene UTRs can also cause ~4 fold repression of the reporter by an miRNA in mammalian cells [80, 132, 142]. These levels of reporter gene silencing in mammalian cells are comparable to downregulation of LIN-14 by *lin-4* RNA in the nematode: a reporter gene with the LIN-14 3'UTR (which contains six *lin-4* complementary sites) can be repressed 20 fold relative to a control 3' UTR reporter gene [8, 143]. Interestingly, the temporal regulation of LIN-14 protein in wild-type worms is a 10 fold decrease, but in mutants lacking *lin-4* RNA or *lin-4* binding sites in *lin-14*, LIN-14 is still temporally downregulated modestly, hinting that additional factors are in play to downregulate LIN-14 in the worm [8]. Since *let-7* sites in the *lin-14* 3' UTR have been proposed [14], and translation repression appears cooperative in combinations of multiple bulged interaction sites to siRNAs

mimicking miRNA action [87, 133], it is easy to imagine multiple miRNAs acting combinatorially upon a given mRNA's 3' UTR (Figure 5B).

The number and breadth of predicted mammalian targets, the range in repression levels observed in nature and for reporter gene assays, and the notion of combinatorial regulation by multiple miRNAs on single mRNAs has prompted the micromanager hypothesis for interpreting the roles of miRNAs in metazoan gene regulation [89]. One aspect of this complex hypothesis is that targets might be "tuned" to various levels of protein expression by miRNAs, depending on the mechanism of repression and the number of miRNAs acting combinatorially upon a target. Target regulation by miRNAs could be envisioned to yield different gradients of gene expression, such that an RNA gradient could be initiated by a miRNA that cleaves its target (Figure 5A), or a target could be tuned by multiple miRNAs to generate a gradual gradient (Figure 5B). In *Drosophila* embryogenesis, spatial and temporal regulation of mRNAs and proteins is vital; translational control of mRNAs is a common theme; and gradients of protein and mRNA are set between the axes of the embryo [144]. Recent evidence suggesting that miRNAs might regulate *oskar*, a gene that is important for proper fly embryogenesis [104], could hint that many other genes known to be expressed in a temporal or spatial gradient could be regulated post-transcriptionally by miRNAs (Figure 5B). Since the *C.elegans* homolog of *hunchback*, *hbl-1*, appears to be regulated by miRNAs [130, 131], it is tempting to speculate that miRNAs might also influence the establishment of the *hunchback* protein gradient in the *Drosophila* embryo. While the micromanager hypothesis is thought-provoking, it remains to be tested experimentally.

Despite advances in target prediction and testing of targets, verifying the biological validity of these targets lags behind the pace of gene finding efforts and progress in

understanding miRNA biogenesis. Table 1 catalogs the extent of miRNAs with experimentally tested target mRNAs, though one should keep in mind that targets assessed by reporter gene assays can only suggest but not confirm the miRNA's biological role. Of the 100 miRNAs in *C.elegans*, only 3 miRNAs in *C.elegans* have ascribed biological functions – *lin-4*, *let-7*, and *lsy-6*. A similar dearth of functional assignments for miRNAs persists in *Drosophila* and mammals. Only 3 of the 78 known *Drosophila* miRNAs have verified biological function (*bantam,mir-7,* and *mir-14*), and amongst the >200 known mammalian miRNAs, only 3 miRNAs (*mir-196, mir-181,* and *mir*-BART2) have *in vivo* evidence indicating the miRNA's function (a report on miR-23's regulation of mammalian *Hes-1* was retracted [145, 146]). Clearly, it is still early days in determining the roles of miRNAs in metazoans.

In plants, however, the proportion of functionally assigned miRNAs is significantly higher. Of the 92 known *Arabidopsis* miRNAs representing 22 families, all but a few have assignable functions as negative regulators of transcription factor genes and other genes involved in development. The greater success of miRNA functional assignment in plants can be attributed to: (1) the conservation of the extensive complementarity of the miRNA target site, which significantly improves the signal-to-noise ratio and confidence in predictions [57, 64, 147]; (2) the existence of GOF mutants in target genes where the lesion maps to the miRNA target site [57, 148, 149]; (3) cleavage products from targets can be detected *in vivo* by Northern blots [90, 117, 126, 150]; and (4) the target cleavage sites mapped *in vivo* by 5' rapid amplification of cDNA ends (RACE) fits the specificity of Slicer cleavage – 10 bases downstream from the 5' end of the miRNA [64, 70, 90, 113, 117, 126, 147, 150]. While the majority of predicted plant targets are transcription factors, newly identified targets have widened the ontology to include laccases, superoxide dismutases, and ATP sulfurylases [64]. The attributes of plant targets would

suggest mRNA cleavage is the dominant mode of regulation by plant miRNAs, however miR-172 instead inhibits productive translation of its target, *AP2,* leaving open the possibility that bulged interactions commonly seen with animal miRNAs might also operate in plants [70, 125]. If so, the current lists of plant targets might also be far from complete, but expanding the list of plant targets will have to await the refinement of prediction algorithms for animal miRNAs.

miRNA Function In Vivo: A Return to Mutants

Compared to functional genomics efforts on miRNAs, the classical genetics performed in this "modern" age have actually been the most informative on the function of a select few animal and plant miRNAs. Indeed, it was classical genetic methods that revealed the developmental roles of *lin-4* and *let-7*, and more recent genetics have added *hbl-1/lin-57*, another heterochronic gene, as a second target of *let-7* [130, 131] . The power of genetics recently uncovered a novel miRNA in *C.elegans*, *lsy-6*, that had not been previously reported by cloning and computational efforts [129]. Two sensory neurons in *C.elegans,* although symmetric in their location, are functionally distinct because they express different olfactory receptors [151]. A genetic screen of worms defective in this functional asymmetry identified genes that operate in a network to specify selective expression of the olfactory receptors [152]. While chipping away at this regulatory network, Johnston and Hobert proposed that *lsy-6* (a lateral symmetry mutant, pronounced "lousy") could negatively regulate *cog-1*, which in turn negatively regulates other genes downstream [129]. Through transgenic rescue experiments, they whittled the *lsy-6* locus down to a miRNA-expressing fragment, and the sequence they guessed for the mature *lsy-6* RNA could base-pair to a conserved element in the *cog-1* 3' UTR, an interaction that was subsequently confirmed genetically. Although *lsy-6* is the third example of nematode miRNAs

regulating genes post-transcriptionally, the demonstration of this miRNA in specifying neuronal asymmetry impressively expands the influence of miRNAs beyond heterochronic development.

MicroRNA mutants in *Drosophila* were finally exposed when the *bantam* locus was discovered to encode a miRNA. The mutant was isolated in an enhancer-P element screen and was initially characterized as a regulator of growth and cell size, because deletion of *bantam* reduces the size of larvae and causes pupa lethality [153]. The lack of any protein-coding genes in the vicinity of mapped *bantam* mutations led Brennecke and colleagues to explore the possibility of a miRNA gene, and determination of *bantam*'s sequence prompted a computational search that identified *hid*, an apoptosis gene, as a candidate target for *bantam* regulation [128]. The Cohen lab's fluorescent protein sensor, discussed earlier, indicated widespread *in vivo* expression of *bantam*, and also demonstrated that the *hid* 3' UTR could be directly controlled by *bantam*. Reduction of endogenous Hid by *bantam* in genetic crosses further proved the interaction between the two genes and suggests that *bantam*'s role could be an oncogene that promotes cell growth by repressing apoptosis [128]. The revelation that P-element mutants at miRNA genes can elicit a phenotype is prompting many fly labs to reinvestigate their stocks, and a second example to emerge from this revelation is *mir-14*. Although direct targets for miR-14 have not been confirmed, *mir-14* was found to genetically repress cell death caused by other apoptotic genes [154]. Interestingly, the deletion of *mir-14* does not cause lethality like the *bantam* deletion, but instead causes an accumulation of fat droplets in the fly head [154].

Sometimes, the function of animal miRNAs has been revealed simply from ectopic expression by transgenes. Over-expression of *mir-7* in flies causes notched wings with vein defects and proliferation of bristles, which are similar to defects seen in other *Drosophila* mutants of the *Notch* signaling pathway [137]. Three mammalian miRNAs, *mir-181*, *mir-142s*

and *mir-223*, are upregulated in hematopoeitic cells that begin to display markers of differentiation compared to undifferentiated progenitor cells [44]. By overexpressing these three miRNAs separately in progenitor cells *in vitro*, Chen and colleagues observed the population of cells shifting towards expressing a B-cell marker with a *mir-181* transgene, or a shift towards expressing a T-cell marker with *mir-142s* and *mir-223* transgenes. When progenitor cells overexpressing *mir-181* were transplanted into lethally irradiated mice, a preference for B-cell formation was also observed, confirming *mir-181*'s role in modulating B-cell differentiation [44]. Targets for *mir-181*, *mir-142s,* and *mir-223* remain to be identified or confirmed, but *in vivo* targets have been confirmed for two other miRNAs, *mir-196* in mice and *mir*-BART2 from EBV. Using 5' RACE, cleavage sites were mapped for miR-196–directed cleavage of *Hox-B8* mRNA, an important master regulatory gene in development [124], and for *mir*-BART2-directed cleavage of EBV-BALF5, the viral DNA polymerase [49]. Mapping of these cleavage targets is certainly compelling, but the functional consequence of perturbing these miRNAs or the site of regulation in the targets remains to be examined.

Long before miRNAs were even known in plants, DCL1 mutants had been identified multiple times in screens for defects in female reproductive development or floral patterning (for review see [155]); while mutants in the Ago1 and Pinhead/Zwille genes had been known to affect various aspects of general plant architecture [156-158]. Now that plant miRNA targets have been identified, studies of these targets highlight specific defects previously seen amongst the pleiotropic defects of the DCL, Ago, and Zwille mutants. For example, defects in proper leaf patterning could be explained by the recent re-evaluation of the PHABULOSA (*PHB*) and PHAVOLUTA (*PHV*) mutants, which are dominant GOF alleles in homeodomain zinc-finger (HD-ZIP) transcription factors. The initial hypothesis for how these alleles distorted leaf

structures from their normal polarized patterns presumed that point mutations in the START

domain of these proteins were somehow constitutively activating the protein through a single

amino-acid change [148], but computational predictions suggested instead that the point

mutations were disrupting an interaction with MIR165 and MIR166 [57]. *In vitro* analysis in

wheat-germ lysates of RNA targets bearing wild type or mutant *PHV* sequences supported the

hypothesis of loss of miR-165/166 regulation [123]. Lesions that change the miRNA binding site

but not the protein code of REVOLUTA, another HD-ZIP protein, appeared to exert similar

defects as *PHV*, thus increasing the support for MIR165/166 function in leaf patterning [149].

Recently, miR-165 has been suggested to accumulate asymmetrically in plant structures [110],

and a maize gene, *rolled leaf1* is also a likely target of regulation by miR-165 [159]. The depth

of functional conservation of MIR166 has also been suggested to extend even into the most

ancient of plants [147].

The first plant mutant in a miRNA locus was actually discovered based on a

misclassification of a serrated leaf morphology defect in the *jawD* mutant. Reexamination of

jawD pointed to crinkly and unevenly shaped leaves akin to defects from snapdragon mutants in

the *CINCINNATA* gene, a member of the TCP transcription factor family [150]. When

microarray analysis of *jawD* indicated some but not all TCP mRNAs were being downregulated,

Palatnik and colleagues deduced that the downregulated genes were being affected by a very

small sequence element shared by both TCP mRNAs and the *jawD* locus. This element turned

out to be a new miRNA, mir-JAW, which was highly expressed in leaves and was overexpressed

in the *jawD* mutant. Additional evidence was presented for TCP mRNAs being cleaved by miR-

JAW, and for the rescue of *jawD* defects by overexpressing TCP. The highly similar miR-159

was also shown to regulate the transcription factor MYB33, which also affects leaf morphology

[150]. When plants were transformed with TCP and MYB transgenes resistant to miRNA regulation, growth was stunted and cotyledon leaves sometimes fused, which harks back to defects similarly observed in DCL mutants [150, 155].

The second plant miRNA mutant to be reported was a line that overexpressed miR172a-2 (this mutant is also called *eat-D*) [70]. The *MIR172a-2* mutant precociously flowered with deformities, and resembled LOF alleles of *APETALA2 (AP2)*, a floral homeotic gene. A perfect binding site for miR172 was found in the *AP2* transcript, and was conserved amongst *AP2* homologs like *toe1*, an activation tagged mutant that flowered late instead of early. *AP2* and *TOE1* cleavage products and decreased levels of *AP2* protein in the mutant overexpressing miR172 confirmed the model that miR172 was silencing *AP2* family genes to trigger proper floral development by removing the floral suppression activity of *AP2* genes [70]. An independent reverse genetics approach with MIR172–resistant *AP2* transgenes also supported this model [125]. Perhaps the most intriguing result to come from both studies was that *AP2* RNA levels were unaffected by the overexpression of miR172, indicating that despite the extensive complementarity between miR172 and *AP2*, inhibition of productive translation was the dominant mechanism for gene silencing instead of mRNA cleavage [70, 125].

Evoking the philosophical overtones of the Ying and Yang, MIR162 and MIR168 can negatively regulate DCL1 and AGO1, respectively, which in turn would reduce miR162 and miR168 levels [57, 109, 113, 126, 160]. The consequences for elevated DCL1 levels is not known, but elevated AGO1 levels clearly cause reduced fertility and pleiotropic defects overlapping with null mutants of *dc1*, *hen1*, and *hyl1* [109]. Fully deconvolving the feedback loops of MIR162 and MIR168 regulation will be complicated given that their targets also modulate the levels of other plant miRNAs. The importance of these feedback loops can be

appreciated in light of *jawD* and the *MIR172a2* mutant, where elevated miRNA levels can impose serious developmental defects. Alternatively, the consequences of *jawD* and the *MIR172a2* mutant are the result of ectopic expression of miRNAs in cells that would normally not contain these miRNAs. Although evidence for negative feedback regulation in the animal miRNA pathway has not yet been observed, it remains a viable possibility since overexpression of *bantam* and *mir-14* is detrimental to fly development.

An inevitable goal for plant and animal miRNA labs would be to learn the phenotypic consequence of perturbing many more miRNAs *in vivo*. Hunting for mutants is challenging, since no simple set of phenotypes can be pinned upon miRNAs as a class, many miRNA sequences are coded by multiple loci, and highly similar miRNAs in gene families can also be functionally redundant. Rather than be at the mercy of serendipitous breakthroughs like *lin-4, let-7,* and other miRNA mutants, however, labs have initiated efforts to knock-out miRNA genes in animals. However, recent reports suggest a short cut over genetic mutants might be achieved with 2'-O-methyl RNA (2OMe) oligos directed antisense to the miRNA sequence. Not only can 2OMe oligos interrupt RISC activity in *Drosophila* lysate and in tissue culture [161, 162], but it can also phenocopy heterochronic defects when directed against *let-7* in worms [161]. Although 2OMe technology cannot rival the rigor of a genetic lesion, it could greatly inform geneticists on what defects to focus on when hunting for miRNA mutants, just as how genomic RNAi has revolutionized functional genomics.

Summary

The field of miRNAs began from the study of two mutant worms, *lin-4* and *let-7*, that revealed an elegant form of gene regulation by the action of small, endogenous RNAs. Since the

unveiling of numerous miRNAs in animals and plants, several examples illustrating the biological importance of miRNA regulation have emerged. Although much of this introduction expounds on work external of this thesis, the cloning of miRNAs has helped to catalyze these subsequent advances in miRNA biology

The stRNA worm mutants and the observation that siRNA function in many eukaryotes provided us the hunch that Dicer products might exist in plants and animals for the purpose of regulating endogenous genes. The first chapter of this thesis reports our initial foray into cloning and characterizing small endogenous RNAs from *C.elegans*. I devised a RNA cloning strategy that exploits the molecular features of Dicer products (namely a small RNA with a 5' phosphate and a 3' hydroxyl) to efficiently build cDNA libraries with minimal RNA breakdown products. We examined the sequence characteristics, genomic locations, phylogenetic distribution, and expression patterns of these RNA clones, and concluded that they represent an abundant class of endogenously expressed, tiny RNAs with probable regulatory roles. Together with labs that simultaneously published similar discoveries, we named this new class of RNAs the microRNAs.

To expand our census of the miRNAs in *C.elegans*, I describe in the second chapter a lab collaborative effort to identify the majority of remaining miRNA genes through large scale RNA cloning and computational methods. A list of almost 90 miRNA genes were confirmed in *C.elegans* and several groups of miRNAs were shown to have interesting developmental expression patterns as well as significant conservation to mammalian miRNAs. We also determined the molecular abundance of miRNAs in *C.elegans* and Hela cells.

The high molecular abundance of miRNAs beckoned the analysis of miRNA stability. In the third chapter, Hela cell lines with inducible expression of a miRNA were constructed. By monitoring miRNA levels either in the decay phase of halted transcription or at half-maximal

steady state from induction, we hoped to determine the half-life of two model miRNAs, *mir-1* and *mir-124*. Although some features of these inducible lines remain to be characterized and improved, a lower estimate of miRNA stability is presented. The study suggests a long half-life for miRNAs in animals, which could partially explain the high steady-state levels.

Together, these studies characterize many of the molecular characteristics of animal miRNAs, as well as catalog most of the miRNA genes in the nematode, *C.elegans*. Determination of miRNA gene features has been and will continue to be critical for future efforts to find miRNAs in other genomes. The expression data, abundance levels, and stability measurements provided in this thesis will also prove valuable in functional studies that begin to address the fundamental roles of miRNA activity in organisms.

Acknowledgements

REFERENCES

1. Meneely, P.M., and Herman, R.K. (1979). Lethals, steriles and deficiencies in a region of the X chromosome of Caenorhabditis elegans. Genetics *92*, 99-115.

2. Horvitz, H.R., and Sulston, J.E. (1980). Isolation and genetic characterization of cell-lineage mutants of the nematode Caenorhabditis elegans. Genetics *96*, 435-454.

3. Chalfie, M., Horvitz, H.R., and Sulston, J.E. (1981). Mutations that lead to reiterations in the cell lineages of C. elegans. Cell *24*, p59-69.

4. Ambros, V., and Horvitz, H.R. (1984). Heterochronic mutants of the nematode Caenorhabditis elegans. Science *226*, 409-416.

5. 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.

6. 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.

7. Wightman, B., Burglin, T.R., 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.

8. 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*, 855-862.

9. 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.

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

11. Wickens, M. (1990). In the beginning is the end: regulation of poly(A) addition and removal during early development. Trends Biochem Sci *15*, 320-324.

12. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843-854.

13.   Moss, E.G., Lee, R.C., and Ambros, V. (1997). The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. Cell *88*, 637-646.

14.   Reinhart, B.J., Slack, F.J., Basson, M., Bettinger, J.C., Pasquinelli, A.E., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature *403*, 901-906.

15.   Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M., Maller, B., Srinivasan, A., Fishman, M., Hayward, D., Ball, E., Degnan, B., Muller, P., Spring, J., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation across animal phylogeny of the sequence and temporal regulation of the 21 nucleotide *let-7* heterochronic regulatory RNA. Nature *408*, 86-89.

16.   Hutvágner, G., and Zamore, P.D. (2002). RNAi: nature abhors a double-strand. Curr Opin Genet Dev *12*, 225-232.

17.   Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature *391*, 806-811.

18.   Hamilton, A.J., and Baulcombe, D.C. (1999). A novel species of small antisense RNA in posttranscriptional gene silencing. Science *286*, 950-952.

19.   Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

20.   Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. (2001). Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature *409*, 295-296.

21.   Elbashir, S.M., Leneckel, W., and Tuschl, T. (2001). RNA interference is mediated by 21- and 22- nucleotide RNAs. Genes Dev. *15*, 188-200.

22.   Hammond, S.C., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates posttranscriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

23.   Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference

regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.

24. Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. Science *293*, 834-838.

25. Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654-2659.

26. Knight, S.W., and Bass, B.L. (2001). A role for the RNase III enzyme DCR-1 in RNA interference and germ line development in C. elegans. Science *293*, 2269.

27. Eddy, S.R. (2001). Non-coding RNA genes and the modern RNA world. Nat Rev Genet *2*, 919-929.

28. Gottesman, S. (2002). Stealth regulation: biological circuits with small RNA switches. Genes Dev *16*, 2829-2842.

29. Lowe, T.M., and Eddy, S.R. (1999). A computational screen for methylation guide snoRNAs in yeast. Science *283*, 1168-1171.

30. Omer, A.D., Lowe, T.M., Russell, A.G., Ebhardt, H., Eddy, S.R., and Dennis, P.P. (2000). Homologs of small nucleolar RNAs in Archaea. Science *288*, 517-522.

31. Wassarman, K.M., Repoila, F., Rosenow, C., Storz, G., and Gottesman, S. (2001). Identification of novel small RNAs using comparative genomics and microarrays. Genes Dev *15*, 1637-1651.

32. Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P., and Brosius, J. (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J *20*, 2943-2953.

33. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

34. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science *294*, 858-862.

35. Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in Caenorhabditis elegans. Science *294*, 862-864.

36. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003a). The microRNAs of *Caenorhabditis elegans*. Genes Dev *17*, 991-1008.

37. Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

38. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003b). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.

39. Lagos-Quintana, M., Rauhut, R., Meyer, J., Borkhardt, A., and Tuschl, T. (2003). New microRNAs from mouse and human. RNA *9*, 175-179.

40. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Current Biology *12*, 735-739.

41. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003b). Vertebrate microRNA genes. Science *299*, 1540.

42. Dostie, J., Mourelatos, Z., Yang, M., Sharma, A., and Dreyfuss, G. (2003). Numerous microRNPs in neuronal cells containing novel microRNAs. RNA *9*, 631-632.

43. Houbaviy, H.B., Murray, M.F., and Sharp, P.A. (2003). Embryonic stem cell-specific microRNAs. Dev Cell *5*, 351-358.

44. Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83-86 Published online December 84, 2003 2010.1126/science.1091903.

45. Michael, M.Z., SM, O.C., van Holst Pellekaan, N.G., Young, G.P., and James, R.J. (2003). Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res *1*, 882-891.

46. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev *16*, 720-728.

47. Olsen, P.H., and Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. Developmental Biology *216*, 671-680.

48. Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. (2004). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci U S A *101*, 360-365.

49. Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of Virus-Encoded MicroRNAs. Science *304*, 734-736.

50. Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell *14*, 1605-1619.

51. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

52. Catalanotto, C., Azzalin, G., Macino, G., and Cogoni, C. (2002). Involvement of small RNAs and role of the qde genes in the gene silencing pathway in Neurospora. Genes Dev *16*, 790-795.

53. Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

54. Djikeng, A., Shi, H., Tschudi, C., and Ullu, E. (2001). RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA *7*, 1522-1530.

55. Park, W., Li, J., Song, R., Messing, J., and Chen, X. (2002). CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. Cur. Biol. *12*, 1484-1495.

56. Wang, J.F., Zhou, H., Chen, Y.Q., Luo, Q.J., and Qu, L.H. (2004). Identification of 20 microRNAs from Oryza sativa. Nucleic Acids Res *32*, 1688-1695.

57. Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

58. Vance, V., and Vaucheret, H. (2001). RNA silencing in plants - defense and counterdefense. Science *292*, 2277-2280.

59. Plasterk, R.H. (2002). RNA silencing: the genome's immune system. Science *296*, 1263-1265.

60. Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. (2003). Computational identification of Drosophila microRNA genes. Genome Biol *4:R42*, 1-20.

61.   Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

62.   Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P., and Burge, C.B. (2004). Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. RNA, in press.

63.   Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G., and Kim, J. (2003). Computational and experimental identification of C. elegans microRNAs. Mol Cell *11*, 1253-1263.

64.   Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets. Mol Cell *In press*.

65.   Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Marshall, M., Matzke, M., Ruvkun, G., and Tuschl, T. (2003a). A uniform system for microRNA annotation. RNA *9*, 277-279.

66.   Griffiths-Jones, S. (2004). The microRNA Registry. Nucleic Acids Res *32*, D109-111.

67.   Johnson, S.M., Lin, S.Y., and Slack, F.J. (2003). The time of appearance of the C. elegans let-7 microRNA is transcriptionally controlled utilizing a temporal regulatory element in its promoter. Dev Biol *259*, 364-379.

68.   Bashirullah, A., Pasquinelli, A.E., Kiger, A.A., Perrimon, N., Ruvkun, G., and Thummel, C.S. (2003). Coordinate regulation of small temporal RNAs at the onset of Drosophila metamorphosis. Dev Biol *259*, 1-8.

69.   Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M., and Ambros, V. (2003). Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity. Dev Biol *259*, 9-18.

70.   Aukerman, M.J., and Sakai, H. (2003). Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. Plant Cell *10*, 10.

71.   Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J *21*, 4663-4670.

72.   Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

73. Gorlich, D., and Kutay, U. (1999). Transport between the cell nucleus and the cytoplasm. Annu Rev Cell Dev Biol *15*, 607-660.

74. Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA *10*, 185-191.

75. Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98 Published online on November 20, 2003,2010.1126/science.1090599.

76. Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. *17*, 3011-3016.

77. Bollman, K.M., Aukerman, M.J., Park, M.Y., Hunter, C., Berardini, T.Z., and Poethig, R.S. (2003). HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development *130*, 1493-1504.

78. Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and Functional Diversification of Small RNA Pathways in Plants. PLoS Biol *2*, E104.

79. Lau, N.C., and Bartel, D.P. (unpublished data).

80. Zeng, Y., and Cullen, B.R. (2003). Sequence requirements for micro RNA processing and function in human cells. RNA *9*, 112-123.

81. Liu, Q., Rand, T.A., Kalidas, S., Du, F., Kim, H.E., Smith, D.P., and Wang, X. (2003). R2D2, a bridge between the initiation and effector steps of the Drosophila RNAi pathway. Science *301*, 1921-1925.

82. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. Cell *117*, 69-81.

83. Dykxhoorn, D.M., Novina, C.D., and Sharp, P.A. (2003). Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol *4*, 457-467.

84. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

85. Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

86. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. (1999). Targeted mRNA degradation by double-stranded RNA *in vitro*. Genes Dev. *13*, 3191-3197.

87. Doench, J.G., Peterson, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev. *17*, 438-442.

88. Saxena, S., Jonsson, Z.O., and Dutta, A. (2003). Small RNAs with imperfect match to endogenous mRNA repress translation. Implications for off-target activity of small inhibitory RNA in mammalian cells. J Biol Chem *278*, 44312-44319.

89. Bartel, D.P., and Chen, C.-Z. (2004). Micromanagers of Gene Expression: The Potentially Widespread Influence of Metazoan MicroRNAs. Nat Rev Genet *5*, 396-400.

90. Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

91. Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W., and Tuschl, T. (2001). Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. EMBO J *20*, 6877-6888.

92. Hutvágner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

93. Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev *16*, 2491-2496.

94. Hammond, S.M., Boettcher, S., Caudy, A.C., Kobayashi, R., and Hannon, G.J. (2001). Argonaute2, a link between genetic and biochemical analyses of RNAi. Science *293*, 1146-1150.

95. Ishizuka, A., Siomi, M.C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev *16*, 2497-2508.

96. Caudy, A.A., Ketting, R.F., Hammond, S.M., Denli, A.M., Bathoorn, A.M., Tops, B.B., Silva, J.M., Myers, M.M., Hannon, G.J., and Plasterk, R.H. (2003). A micrococcal nuclease homologue in RNAi effector complexes. Nature *425*, 411-414.

97. Tomari, Y., Du, T., Haley, B., Schwarz, D.S., Bennett, R., Cook, H.A., Koppetsch, B.S., Theurkauf, W.E., and Zamore, P.D. (2004). RISC assembly defects in the Drosophila RNAi mutant armitage. Cell *116*, 831-841.

98.   Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., and Sontheimer, E.J. (2004). A
        Dicer-2-Dependent 80S Complex Cleaves Targeted mRNAs during RNAi in Drosophila.
        Cell *117*, 83-94.

99.   Schwarz, D.S., Tomari, Y., and Zamore, P.D. (2004). The RNA-Induced Silencing
        Complex Is a Mg(2+)-Dependent Endonuclease. Curr Biol *14*, 787-791.

100.   Martinez, J., and Tuschl, T. (2004). RISC is a 5' phosphomonoester-producing RNA
        endonuclease. Genes Dev.

101.   Nykänen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small
        interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.

102.   Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-
        stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

103.   Nelson, P.T., Hatzigeorgiou, A.G., and Mourelatos, Z. (2004). miRNP:mRNA
        association in polyribosomes in a human neuronal cell line. RNA *10*, 387-394.

104.   Cook, H.A., Koppetsch, B.S., Wu, J., and Theurkauf, W.E. (2004). The Drosophila SDE3
        homolog armitage is required for oskar mRNA silencing and embryonic axis
        specification. Cell *116*, 817-829.

105.   Schwarz, D.S., and Zamore, P.D. (2002). Why do miRNAs live in the miRNP? Genes
        Dev *16*, 1025-1031.

106.   Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family:
        tentacles that reach into RNAi, developmental control, stem cell maintenance, and
        tumorigenesis. Genes Dev *16*, 2733-2742.

107.   Tabara, H., Sarkissian, M., Kelly, W.G., Fleenor, J., Grishok, A., Timmons, L., Fire, A.,
        and Mello, C.C. (1999). The *rde-1* gene, RNA interference, and transposon silencing in
        C. elegans. Cell *99*, 123-132.

108.   Morel, J.B., Godon, C., Mourrain, P., Béclin, C., Boutet, S., Feuerbach, F., Proux, F., and
        Vaucheret, H. (2002). Fertile hypomorphic *ARGONAUTE*(ago1) mutants impaired in
        post-transcriptional gene silencing and virus resistance. Plant Cell *14*, 629-639.

109.   Vaucheret, H., Vazquez, F., Crata, P., and Bartel, D.P. (2004). The action of
        ARGONAUTE1 in the miRNA pathway and its regulation by the miRNA pathway are
        crucial for plant development. Genes Dev., 1201404.

110. Kidner, C.A., and Martienssen, R.A. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature *428*, 81-84.

111. Zilberman, D., Cao, X., and Jacobsen, S.E. (2003). ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. Science *299*, 716-719.

112. Han, M.H., Goud, S., Song, L., and Fedoroff, N. (2004). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci U S A *101*, 1093-1098.

113. Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol *14*, 346-351.

114. Tabara, H., Yigit, E., Siomi, H., and Mello, C.C. (2002). The dsRNA binding protein RDE-4 interacts with RDE-1, DCR-1, and a DExH-box helicase to direct RNAi in C. elegans. Cell *109*, 861-871.

115. Boutet, S., Vazquez, F., Liu, J., Beclin, C., Fagard, M., Gratias, A., Morel, J.B., Crete, P., Chen, X., and Vaucheret, H. (2003). Arabidopsis HEN1: a genetic link between endogenous miRNA controlling development and siRNA controlling transgene silencing and virus resistance. Curr Biol *13*, 843-848.

116. Mallory, A.C., Reinhart, B.J., Bartel, D.P., Vance, V.B., and Bowman, L.H. (2002). A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and microRNAs in tobacco. PNAS.

117. Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. Dev Cell *4*, 205-217.

118. Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A.J. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol *132*, 1382-1390.

119. Dunoyer, P., Lecellier, C.H., Parizotto, E.A., Himber, C., and Voinnet, O. (2004). Probing the MicroRNA and Small Interfering RNA Pathways with Virus-Encoded Suppressors of RNA Silencing. Plant Cell.

120. Chapman, E.J., Prokhnevsky, A.I., Gopinath, K., Dolja, V.V., and Carrington, J.C. (2004). Viral RNA silencing suppressors inhibit the microRNA pathway at an intermediate step. Genes Dev., 1201204.

121. Chen, J., Li, W.X., Xie, D., Peng, J.R., and Ding, S.W. (2004). Viral Virulence Protein Suppresses RNA Silencing-Mediated Defense but Upregulates the Role of MicroRNA in Host Gene Expression. Plant Cell.

122. Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature 427, 645-649.

123. Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev 17, 49-63.

124. Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science 304, 594-596.

125. Chen, X. (2003). A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. Science, Published online September 11 2003; 2010.1126/science.1088060.

126. Xie, Z., Kasschau, K.D., and Carrington, J.C. (2003). Negative feedback regulation of Dicer-like1 in Arabidopsis by microRNA-guided mRNA degradation. Curr Biol 13, 784-789.

127. Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The lin-41 RBCC gene acts in the C. elegans heterochronic pathway between the let-7 regulatory RNA and the LIN-29 transcription factor. Mol Cell 5, 659-669.

128. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell 113, 25-36.

129. Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature 426, 845-849.

130. Abrahante, J.E., Daul, A.L., Li, M., Volk, M.L., Tennessen, J.M., Miller, E.A., and Rougvie, A.E. (2003). The Caenorhabditis elegans hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. Dev Cell 4, 625-637.

131. Lin, S.Y., Johnson, S.M., Abraham, M., Vella, M.C., Pasquinelli, A., Gamberi, C., Gottlieb, E., and Slack, F.J. (2003). The C elegans hunchback homolog, hbl-1, controls temporal patterning and is a probable microRNA target. Dev Cell 4, 639-650.

132. Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A 100, 9779-9784.

133. Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev 18, 504-511.

134. Seggerson, K., Tang, L., and Moss, E.G. (2002). Two genetic circuits repress the Caenorhabditis elegans heterochronic gene lin-28 after translation initiation. Dev Biol 243, 215-225.

135. Kiriakidou, M., Nelson, P.T., Kouranov, A., Fitziev, P., Bouyioukos, C., Mourelatos, Z., and Hatzigeorgiou, A. (2004). A combined computational-experimental approach predicts human microRNA targets. Genes Dev., 1184704.

136. Lai, E.C. (2002). MicroRNAs are complementary to 3'UTR motifs that mediate negative post-transcriptional regulation. Nature Genetics 30, 363-364.

137. Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila microRNA targets. PLOS Biol. 1, E60.

138. Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell 115, 787-798.

139. Jackson, A.L., Bartz, S.R., Schelter, J., Kobayashi, S.V., Burchard, J., Mao, M., Li, B., Cavet, G., and Linsley, P.S. (2003). Expression profiling reveals off-target gene regulation by RNAi. Nat Biotechnol 21, 635-637.

140. Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biology 5, R1.

141. Rajewsky, N., and Socci, N.D. (2004). Computational identification of microRNA targets. Dev Biol 267, 529-535.

142. Zeng, Y., Wagner, E.J., and Cullen, B.R. (2002). Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. Mol Cell 9, 1327-1333.

143. Ha, I., Wightman, B., and Ruvkun, G. (1996). A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. Genes Dev *10*, 3041-3050.

144. Kuersten, S., and Goodwin, E.B. (2003). The power of the 3' UTR: translational control and development. Nat Rev Genet *4*, 626-637.

145. Kawasaki, H., and Taira, K. (2003). Retraction: Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. Nature *426*, 100.

146. Kawasaki, H., and Taira, K. (2003). Hes1 is a target of microRNA-23 during retinoic-acid-induced neuronal differentiation of NT2 cells. Nature *423*, 838-842.

147. Floyd, S.K., and Bowman, J.L. (2004). Gene regulation: ancient microRNA target sequences in plants. Nature *428*, 485-486.

148. McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature *411*, 709-713.

149. Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. Curr Biol *13*, 1768-1774.

150. Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature *425*, 257-263.

151. Hobert, O., Johnston, R.J., Jr., and Chang, S. (2002). Left-right asymmetry in the nervous system: the Caenorhabditis elegans model. Nat Rev Neurosci *3*, 629-640.

152. Chang, S., Johnston, R.J., Jr., and Hobert, O. (2003). A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of C. elegans. Genes Dev *17*, 2123-2137.

153. Hipfner, D.R., Weigmann, K., and Cohen, S.M. (2002). The bantam gene regulates Drosophila growth. Genetics *161*, p1527-1537.

154. Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

155. Schauer, S.E., Jacobsen, S.E., Meinke, D.W., and Ray, A. (2002). *DICER-LIKE1*: blind men and elephants in *Arabidopsis* development. Trends Plant Sci. *7*, 487-491.

156. Lynn, K., Fernandez, A., Aida, M., Sedbrook, J., Tasaka, M., Masson, P., and Barton, M.K. (1999). The PINHEAD/ZWILLE gene acts pleiotropically in Arabidopsis development and has overlapping functions with the ARGONAUTE1 gene. Development *126*, 469-481.

157. Bohmert, K., Camus, I., Bellini, C., Bouchez, D., Caboche, M., and Benning, C. (1998). AGO1 defines a novel locus of *Arabidopsis* controlling leaf development. EMBO J. *17*, 170-180.

158. Moussian, B., Schoof, H., Haecker, A., Jurgens, G., and Laux, T. (1998). Role of the ZWILLE gene in the regulation of central shoot meristem cell fate during Arabidopsis embryogenesis. EMBO J. *17*, 1799-1809.

159. Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature *428*, 84-88.

160. Bartel, B., and Bartel, D.P. (2003). MicroRNAs: At the Root of Plant Development? Plant Physiol *132*, 709-717.

161. Hutvagner, G., Simard, M.J., Mello, C.C., and Zamore, P.D. (2004). Sequence-specific inhibition of small RNA function. PLoS Biol *2*, E98.

162. Meister, G., Landthaler, M., Dorsett, Y., and Tuschl, T. (2004). Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA *10*, 544-550.

163. Moss, E.G., and Tang, L. (2003). Conservation of the heterochronic regulator Lin-28, its developmental expression and microRNA complementary sites. Dev Biol *258*, 432-442.

Table 1. MicroRNAs with experimental data informing of their biological function. (Adapted from [61] and updated with recently identified miRNAs/target genes)

miRNA	Target Gene(s)	Biological Role of miRNA/Target Gene(s) ‡	Refs
Nematodes			
lin-4 RNA*	*Ce lin-14* §; *Ce lin-28*§	Timing of early larval developmental transitions	[8, 12, 13]
let-7 RNA*	*Ce lin-41* §;*Ce hbl-1*§	Timing of late larval developmental transitions	[14, 127]
lsy-6 RNA*	*Ce cog-1* §	Left/right asymmetry of chemoreceptor expression	[129]
Insects			
bantam RNA*	*Dm hid* §	Apoptosis and growth control during development	[128]
miR-2 family	*Dm grim*†; *Dm reaper*†; *Dm sickle*†	Promotes apoptosis	[136, 137]
miR-7*	*Dm HLHm3*†; *Dm hairy*†; *Dm m4*†	Interprets Notch-mediated decisions in neuronal development	[136, 137]
miR-14*	unknown	Promotes apoptosis and regulates fat metabolism	[137, 154]
Mammals			
miR-1	*Hs G6PD*†; *Hs BDNF*†	Oxidative Stress Response; Growth factor and neuronal development	[138]
miR-19a	*Hs PTEN*†	Tumor suppression	[138]
miR-23a	*Hs SDF-1*†; *Hs BRN-3b*†	Growth and localization of hematopoietic progenitor cells; neuronal development	[138]
miR-26a	*Hs SMAD-1*†	Regulates TGF-dependent gene expression	[138]
miR-34	*Hs Delta1*†; *Hs Notch1*†	Activates Notch during cell-fate decision; modulate cell-fate decisions during development	[138]
miR-101	*Hs ENX-1*†; *Hs N-MYC*†	Proliferation of hematopoeitic cells and other gene regulation; cell differentiation and proliferation	[138]
miR-130	*Hs MCSF*†	Mononuclear phagocytic lineage regulation	[138]
miR-142s	unknown	Hematopoietic differentiation	[44]
miR-181*	unknown	Hematopoietic differentiation	[44]
miR-223	unknown	Hematopoietic differentiation	[44]
miR-196	*Mm HOXB8*§; *Mm HOXC8*†; *Mm HOXD8*†; *Mm HOXA7*†	Master regulation of early body pattering and development	[124]
let-7 RNA family	*Hs LIN-28*†; *Hs SMC1L1*†	Function in mammals unknown, but see above homlogs; linked to maintenance of chromosome structure.	[135, 163]
miR-15	*Hs DMTF1*†	Transcriptional regulation in response to cyclin D	[135]
miR-16	*Hs CGI-38*†	Putative signal transduction role in the brain	[135]
miR-23	*Hs FLJ21308*†	Unknown	[135]
miR-24	*Hs MAPK14*†	Signal transduction	[135]
miR-103-1	*Hs FBXW1B*†	Regulation of ubiquitination by F-box proteins	[135]
miR-141	*Hs CLOCK*†	Regulation of circadian rhythms	[135]
miR-145	*Hs FLJ13158*†	Unknown	[135]
miR-199b	*Hs LAMC2*†	Basement membrane component	[135]
Plants			
miR156/157	*At SPL2* family txn factors§	Floral meristem identity	[50, 51, 57, 90, 113]
miR159	*At MYB33* family txn factors §	Leaf development	[55, 57, 150]
miR-JAW*	*At TCP4* family txn factors §	Leaf development and embryonic patterning	[150]
miR160	*At ARF10* and *ARF17* family txn factors§	Auxin response and development	[57, 117]
miR167	*At ARF8* and *ARF6* txn factors§	Auxin response and development	[55, 57, 117]
miR161	*At PPR* gene At1g06580	Unknown	[113]
miR162	*At DCL1*§	miRNA biogenesis	[126, 160]
miR168*	*At* AGO1§	miRNA biogenesis and RISC function	[57, 109, 113]
miR164	*At CUC1* and *CUC2* family txn factors§	Shoot apical meristem formation and organ separation	[57, 117]
miR165/166	*At REV* family txn factors §; *Zm RLD1* family txn factors §	Axial meristem initiation and leaf polarity specification	[57, 110, 123, 149, 159]

Table 1. (continued)

miRNA	Target Gene(s)	Biological Role of miRNA/Target Gene(s) ‡	Refs
Plants (continued)			
miR169	*At CBF-HAP2*	Unknown	[57]
miR170/171	*At SCL-6-III, -IV* family txn factors§	Root radial patterning	[50, 51, 57, 90]
miR172*	*At AP2* family txn factors §	Flower development and timing transition to flowering	[55, 70, 117, 125]
miR393	*At TIR1* family F-box proteins§; At3g23690§	Regulation of ubiquitination; unknown transcription factor	[64]
miR394	At1g27340§	Regulation of ubqiquitination by F-box protein	[64]
miR395	*At APS* family ATP sulfuryases§	Adaptation to soil sulfur content	[64]
miR396	*At GRL* family txn factors§	Growth regulation	[64]
miR397	At2g29130§; At2g38080§; At5g60020§	Oxidation reactions in metabolism	[64]
miR398	*At CSD1*§, *CSD2*§; At3g15640§	Free radical metabolism; Electron transport	[64]
Viruses			
miR-BART2	*EBV* BALF5§	Regulation of viral DNA replication	[49]

Abbreviations: *Caenorhabditis elegans, Ce*; *Drosophila melanogaster, Dm*; human, *Hs*; mouse, *Mm*; *Arabidopsis thaliana, At*; *Zea maize, Zm*; Epstein-Barr virus, *EBV*; transcription factors, txn factors.

‡ semicolons indicate the separate roles of each listed target, in respective order, otherwise a single role applies to all target genes listed; * phenotype in an organism with either a mutant miRNA or from transgene overexpression; § genetically or *in vivo* verified targets; † predicted targets supported by 3' UTR-reporter assays. Plant miRNAs listed include all members within the gene family.

Figure Legends

Figure 1. Rapid and Recent Progress in the miRNA Field. (A) miRNA genes, and (B) miRNAs targets discovered in the last decade. *Counts in 2004 are as of May 30, 2004.

Figure 2. Orchestration of Molecular Events in the *C.elegans* Heterochronic Pathway by *lin-4* and *let-7*. A developmental time line from embryo (E) through the larval stages 1-4 (L1-L4) to adult (A) is detailed from top to bottom. Towards the L1 to L2 transition, the precursor RNA for *lin-4* is expressed. The red colored bases in the precursor structure represent the single-stranded, mature *lin-4* RNA that is formed by Dicer processing. The mature *lin-4* RNA base-pairs to the 3' UTR (straight portion of the wavy transcript) of target mRNAs to inhibit productive formation of LIN-14 and LIN-28. A similar mechanism for *let-7* occurs at the L3 to L4 transition. A complex flux of gene product levels, reflected in the bar graphs on the right, imparts the necessary cues for proper development.
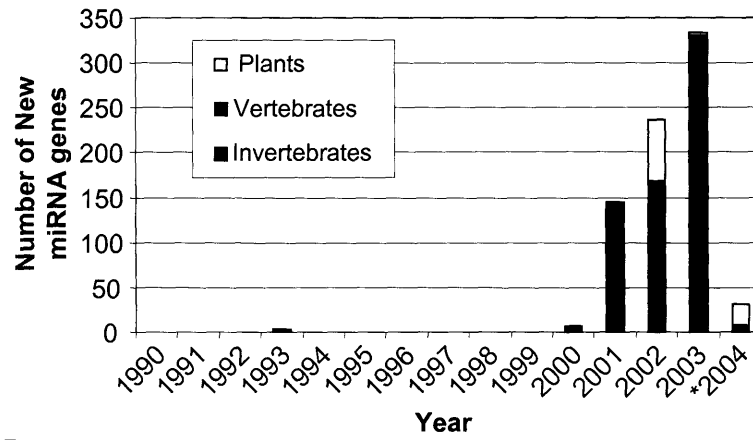
Figure 3. Sequences and Secondary Structures of Representative miRNAs and pre-miRNAs. (A) Conservation of sequence and secondary structure of *mir-1* RNAs in animals, and (B) *MIR164* in plants. (C) The *mir-35 –41* cluster in *C.elegans*. The mature miRNA sequence is colored in red. Abbreviations: *C. elegans, Ce; C. briggsae, Cb; D. melanogaster, Dm; H. sapiens, Hs; A. thaliana, At; O. sativa, Os.*

Figure 4. MicroRNA biogenesis in Animals and Plants. Figure is adapted and modified from [61], and details on the pathway are elaborated in the text. Steps A1-A6 describe animal specific processing events; steps P1-P4 illustrate plant-specific processing events, and Step 5 is a maturation step functionally shared in both animals and plants. Red strand denotes the predominant miRNA, blue strand is the miRNA*, hashed strands represent degraded miRNA; and monophosphates (P) marks the 5' end of each miRNA strand. Brackets signify possible transient plant pre-miRNA intermediates that have not yet been detected.

Figure 5. Modes of miRNA-Target Regulation. This figure uses adaptations from [89]. (A) miRNA targets can be regulated to form a sharp gradient of expression, possibly by catalytic miRNA cleavage of the mRNA. (B) Gradual gradients of protein expression can be established by multiple miRNAs that act via moderate translational repression, where each successive miRNA contributes to increasing levels of gene silencing.

Figure 1.

A

B

Figure 2.

Figure 3.

A

Ce-mir-1   Cb-mir-1   Dm-mir-1   Hs-mir-1b

B

At-   At-   Os-   Os-
MIR164a   MIR164b   MIR164a   MIR164b

C

59
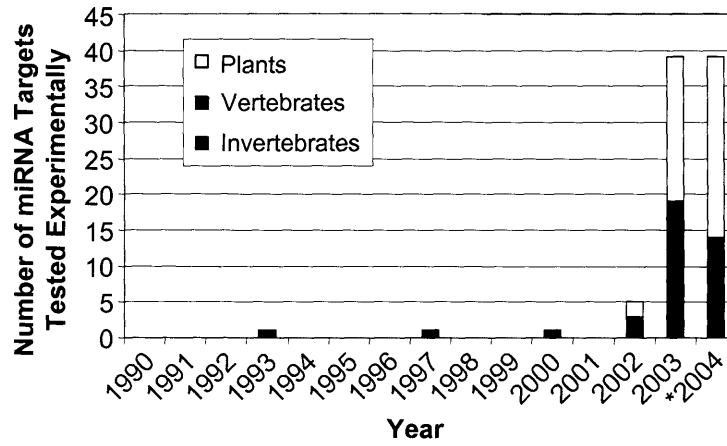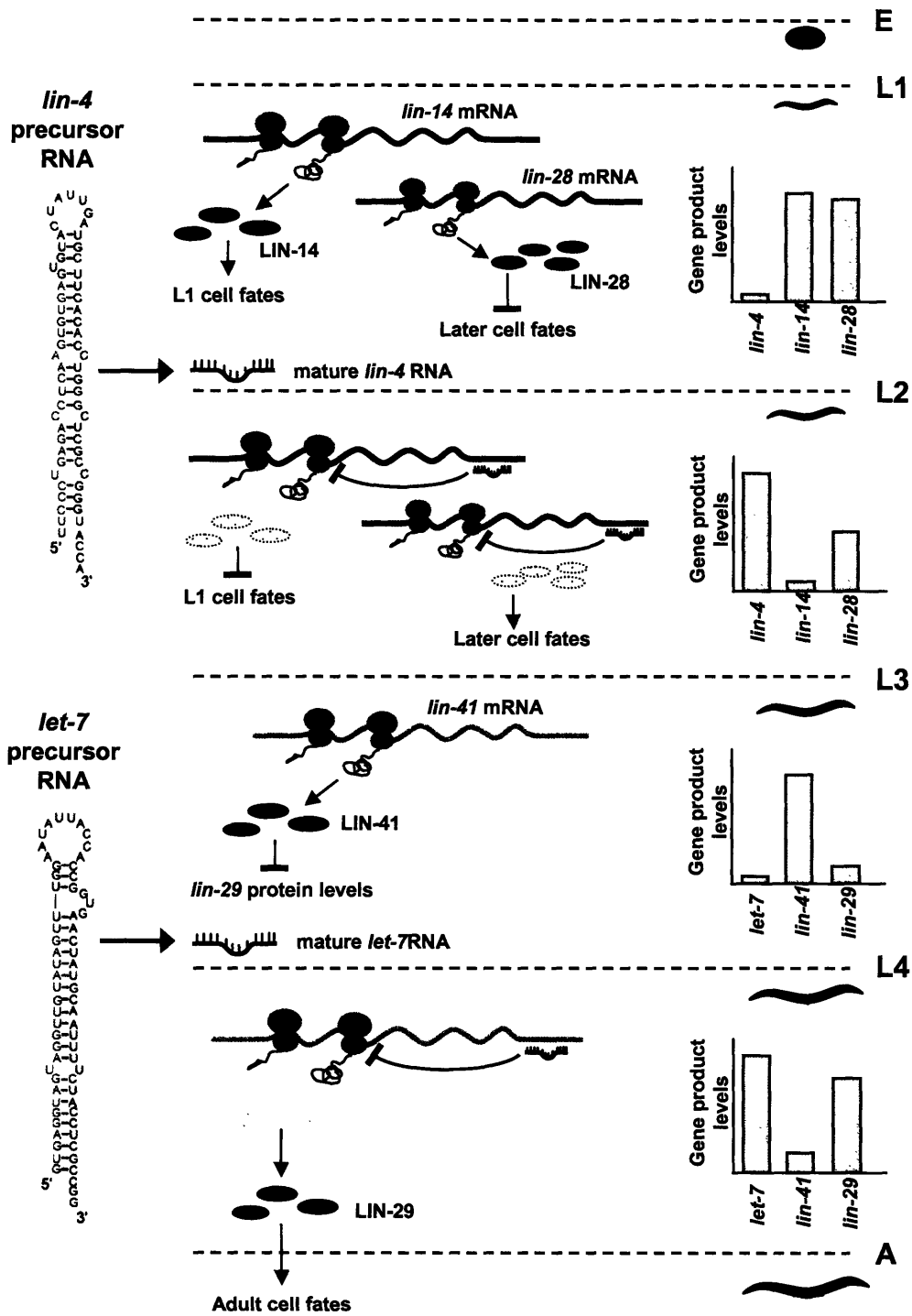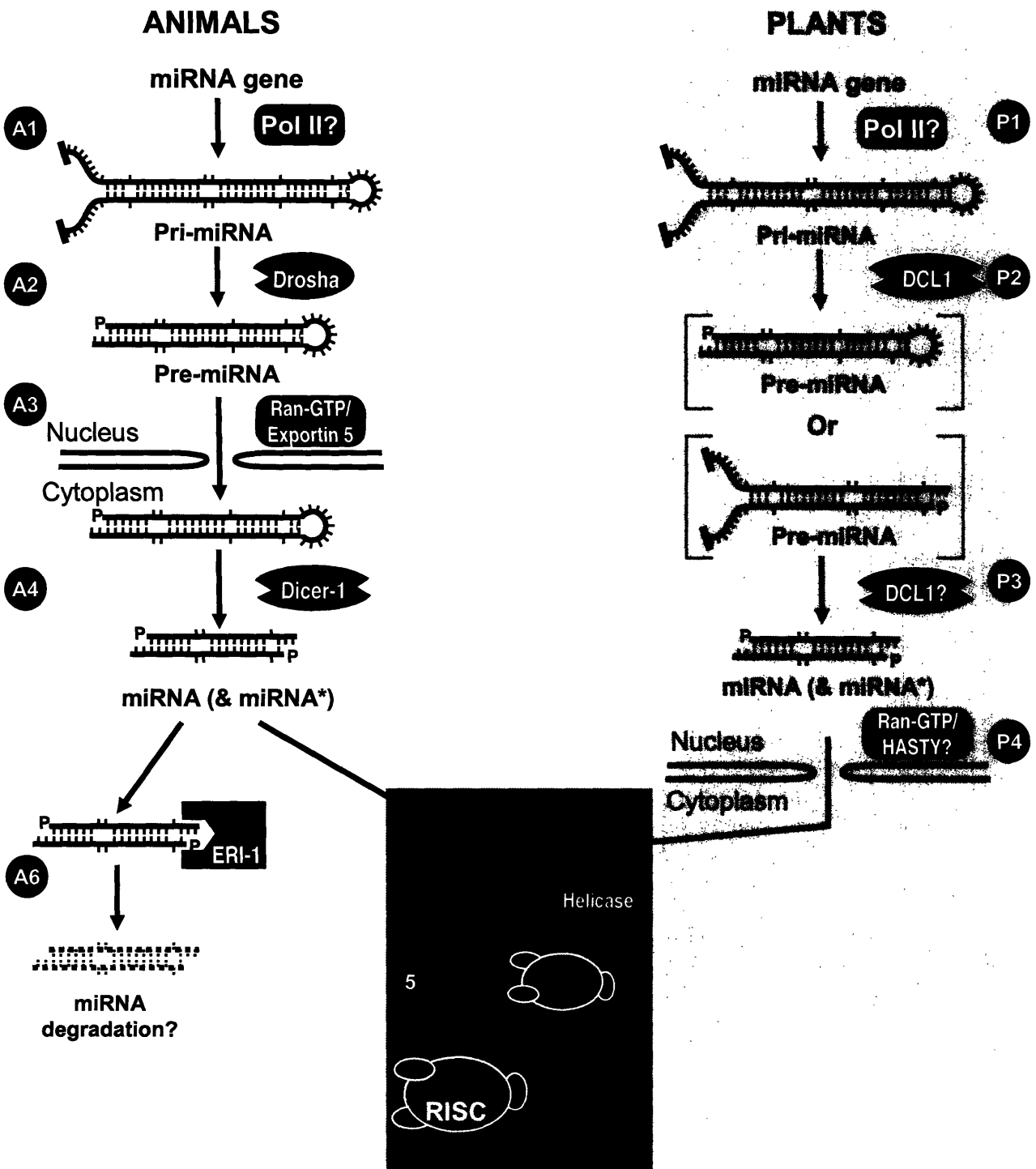
Figure 4.

Figure 5.

An Abundant Class of Tiny RNAs with Probable

Regulatory Roles in *Caenorhabditis elegans*

The work presented in this chapter was a collaborative effort between me, Lee Lim, and Earl Weinstein. Specifically, Lee Lim performed various bioinformatics analyses, including determinations of miRNA sequence compositions, phylogenetic analyses of miRNA homologs in *C.briggsae*, and automated predictions of miRNA precursor secondary structures. Earl Weinstein assisted in automating the analysis of clone sequences. I performed all of the other experiments.

Two small temporal RNAs (stRNAs), *lin-4* and *let-7*, control developmental timing in *Caenorhabditis elegans*. We find that these two regulatory RNAs are members of a large class of 21–24-nucleotide non-coding RNAs, called microRNAs (miRNAs). We report on 55 novel miRNAs in *C. elegans*. The miRNAs have diverse expression patterns during development: a *let-7* paralog is temporally co-expressed with *let-7*; miRNAs encoded in a single genomic cluster are co-expressed during embryogenesis; still other miRNAs are expressed constitutively throughout development. Potential orthologs of several novel miRNA genes were identified in *Drosophila* and human genomes. The abundance of these tiny RNAs, their expression patterns, and their evolutionary conservation imply that, as a class, miRNAs have broad regulatory functions in animals.

Two types of short RNAs, both about 21–25 nt in length, serve as guide RNAs to direct posttranscriptional regulatory machinery to specific mRNA targets. Small temporal RNAs (stRNAs) control developmental timing in *Caenorhabditis elegans* (*1-3*). They pair to sites within the 3´-untranslated region (3´ UTR) of target mRNAs, causing translational repression of these mRNAs, thereby triggering the transition to the next developmental stage (*1-5*). Small interfering RNAs (siRNAs), which direct mRNA cleavage during RNA interference (RNAi) and related processes, are the other type of short regulatory RNAs (*6-12*). Both stRNAs and siRNAs are generated by processes requiring Dicer, a multidomain protein with tandem RNAse III domains (*13-15*). Dicer

cleaves within the double-stranded portion of precursor molecules to yield the 21–25 nt guide RNAs.

lin-4 and *let-7* have been the only two stRNAs identified, and so the extent to which this type of small non-coding RNA normally regulates eukaryotic gene expression is only beginning to be understood (*1-5*). RNAi-related processes protect against viruses or mobile genetic elements, yet these processes are known to normally regulate only one other mRNA, that of *Drosophila Stellate* (*16-20*). To investigate whether RNAs resembling stRNAs or siRNAs might be playing a more general role in gene regulation, we isolated and cloned endogenous *C. elegans* RNAs that have the expected features of Dicer products. Tuschl and colleagues showed that such a strategy is feasible when they fortuitously cloned endogenous *Drosophila* RNAs while cloning siRNAs processed from exogenous dsRNA in an embryo lysate (*12*). Furthermore, other efforts focusing on longer RNAs have recently uncovered many novel non-coding RNAs (*21, 22*).

Dicer products, such as stRNAs and siRNAs, can be distinguished from most other oligonucleotides that might be present in *C. elegans* by three criteria: a length of about 22 nt, a 5′-terminal monophosphate, and a 3′-terminal hydroxyl group (*12, 13, 15*). Accordingly, a procedure was developed for isolating and cloning *C. elegans* RNAs with these features (*23*). Of the clones sequenced, 330 matched *C. elegans* genomic sequence, including 10 representing *lin-4* RNA and one representing *let-7* RNA. Another 182 corresponded to *E. coli* genomic sequence. *E. coli* RNA clones were expected because the worms were cultured with *E. coli* as the primary food source.

Three hundred of the 330 *C. elegans* clones have the potential to pair with nearby genomic sequence to form fold-back structures resembling those thought to be needed for

Dicer processing of *lin-4* and *let-7* stRNAs (Fig. 1) (*24*). These 300 clones with

predicted fold-backs represent 54 unique sequences: *lin-4*, *let-7*, and 52 other RNAs

(Table 1). Thus, *lin-4* and *let-7* RNAs appear to be members of a larger class of non-

coding RNAs that are about 20–24 nt in length and processed from fold-back structures.

We and the two other groups with concurrent reports refer to this class of tiny RNAs as

microRNAs, abbreviated miRNAs, with individual miRNAs and their genes designated

miR-# and *mir-#*, respectively (*25, 26*).

We propose that most of the miRNAs are expressed from independent

transcription units, previously unidentified because they do not contain an open reading

frame (ORF) or other features required by current gene-recognition algorithms. No

miRNAs matched a transcript validated by an annotated *C. elegans* expressed-sequence

tag (EST), and most were at least 1 kb from the nearest annotated sequences (Table 1).

Even the miRNA genes near predicted coding regions or within predicted introns are

probably expressed separately from the annotated genes: If most miRNAs were

expressed from the same primary transcript as the predicted protein, their orientation

would be predominantly the same as the predicted mRNA, but no such bias in orientation

was observed (Table 1). Likewise, other types of RNA genes located within *C. elegans*

intronic regions are usually expressed from independent transcription units (*27*).

Whereas both *lin-4* and *let-7* RNAs reside on the 5′ arm of their fold-back

structures (*1, 3*), only about a quarter of the other miRNAs lie on the 5′ arm of their

proposed fold-back structures, as exemplified by miR-84 (Table 1; Fig. 1A). All the

others reside on the 3′ arm, as exemplified by miR-1 (Table 1; Fig. 1B). This implies

that the stable product of Dicer processing can reside on either arm of the precursor and

that features of the miRNA or its precursor, other than the loop connecting the two arms, must determine which side of the fold-back contains the stable product.

When compared to the RNA fragments cloned from *E. coli*, the miRNAs had unique length and sequence features (Fig. 2). The *E. coli* fragments had a broad length distribution, ranging from 15 to 29, which reflects the size-selection limits imposed during the cloning procedure (*23*). In contrast, the miRNAs had a much tighter length distribution, centering on 21–24 nt, coincident with the known specificity of Dicer processing (Fig 2A). The miRNA sequence composition preferences were most striking at the 5′ end, where there was a strong preference for U and against G at the first position and then a deficiency of Us at positions 2–4 (Fig. 2B). miRNAs were also generally deficient in C, except at position 4. These composition preferences were not present in the clones representing *E. coli* RNA fragments.

The expression of 20 cloned miRNAs was examined, and all but two (miR-41 and miR-68) were readily detected on Northern blots (Fig. 3). For these 18 miRNAs with detectable expression, the dominant form was the mature 20–24 nt fragment(s), though for most, a longer species was also detected at the mobility expected for the fold-back precursor RNA. Fold-back precursors for *lin-4* and *let-7* have also been observed, particularly at the stage in development when the stRNA is first expressed (*1, 14, 15*).

Because the miRNAs resemble stRNAs, their temporal expression was examined. RNA from wild-type embryos, the four larval stages (L1–L4), and young adults was probed. RNA from *glp-4 (bn2)* young adults, which are severely depleted in germ cells (*28*), was also probed because miRNAs might have critical functions in the germ line, as suggested by the finding that worms deficient in Dicer have germ line defects and are

sterile (*14, 29*). Many miRNAs have intriguing expression patterns during development

(Fig. 3). For example, the expression of miR-84, an miRNA with 77% sequence identity

to *let-7* RNA, was found to be indistinguishable from that of *let-7* (Fig. 3). Thus, it is

tempting to speculate that miR-84 is an stRNA that works in concert with *let-7* RNA to

control the larval-to-adult transition, an idea supported by the identification of plausible

binding sites for miR-84 in the 3' UTRs of appropriate heterochronic genes (*30*).

Nearly all of the miRNAs appear to have orthologs in other species, as would be

expected if they had evolutionarily conserved regulatory roles. About 85% percent of the

novel miRNAs had recognizable homologs in the available *C. briggsae* genomic

sequence, which at the time of our analysis included about 90% of the *C. briggsae*

genome (Table 1). Over 40% of the miRNAs appeared to be identical in C. briggsae, as

is seen *lin-4* and *let-7* RNAs, (*1, 3*). Those miRNAs not absolutely conserved between *C.*

briggsae and *C. elegans* might still have important functions, but might have more

readily co-varied with their target sites because, for instance, they might have fewer

target sites. It is noteworthy that when the sequence of the miRNA differs from that of its

homologs, there is usually a compensatory change in the other arm of the fold-back to

maintain pairing, providing phylogenetic evidence for the existence and importance of

the fold-back secondary structures. *let-7*, but not *lin-4*, has discernable homologs in more

distantly related organisms, including *Drosophila* and human (*31*). At least seven other

miRNA genes (*mir-1*, *mir-2*, *mir-34*, *mir-60*, *mir-72*, *mir-79*, and *mir-84*) appear to be

conserved in *Drosophila*, and most of these (*mir-1*, *mir-34*, *mir-60*, *mir-72*, and *mir-84*)

appear to be also conserved in human (*24*). The most highly conserved novel miRNA,

miR-1, is expressed throughout *C. elegans* development (Fig. 3) and so is unlikely to control developmental timing but might instead control tissue-specific events.

The distribution of miRNA genes within the *C. elegans* genome is not random (Table 1). For example, clones for six miRNA paralogs clustered within an 800-bp fragment of chromosome II (Table 1). Computer folding readily identified the fold-back structures for the six cloned miRNAs of this cluster, and predicted the existence of a seventh paralog, miR-39 (Fig. 1D). Northern analysis confirmed the presence and expression of miR-39 (Fig. 3). The homologous cluster in *C. briggsae* appears to have eight related miRNAs. Some of the miRNAs in the *C. elegans* cluster are more similar to each other than to those of the *C. briggsae* cluster, and vice versa, indicating that the size of the cluster has been quite dynamic over a short evolutionary interval, with expansion and perhaps also contraction since the divergence of these two species.

Northern analysis of the miRNAs of the *mir-35–41* cluster showed that these miRNAs are highly expressed in the embryo and in young adults (with eggs), but not at other developmental stages (Fig. 3). For the six detectable miRNAs of this cluster, longer species with mobilities expected for the respective fold-back RNAs also appear to be expressed in the germ line, as indicated by the observation that L4 animals, which have developing gonads but not embryos, express these longer RNAs, whereas germ line-deficient adults do not (Fig. 3) (*30*).

The close proximity of the miRNA genes within the *mir-35–41* cluster (Fig. 1D) suggests that they are all transcribed and processed from a single precursor RNA, an idea supported by the coordinate expression of these genes (Fig. 3). This operon-like organization and expression brings to mind several potential models for miRNA action.

For example, each miRNA of the operon might target a different member of a gene family for translational repression. At the other extreme, they all might converge on the same target, just as *lin-4* and *let-7* RNAs potentially converge on the 3′ UTR of *lin-14* (*3*).

Another four clusters were identified among the sequenced miRNA clones (Table 1). Whereas the clones from one cluster were not homologous to clones from other clusters, the clones within each cluster were usually related to each other, as seen with the *mir-35–41* cluster. The last miRNA of the *mir-42–44* cluster is also represented by a second gene, *mir-45*, which is not part of the cluster. This second gene appears to enable more constitutive expression of the miRNA (miR-44/45) as compared to the first two genes of the *mir-42–44* cluster, which are expressed predominantly in the embryo (Fig. 3).

Dicer processing of stRNAs differs from that of siRNAs in its asymmetry: RNA from only one arm of the fold-back precursor accumulates, while the remainder of the precursor quickly degrades (*15*). This asymmetry extends to nearly all the miRNAs. For the 35 miRNAs yielding more than one clone, in only one case, miR-56, were RNAs cloned from both arms of a hairpin (Fig. 1C, Table 1). The functional miRNA appears to be miR-56 and not miR-56*, as indicated by analysis of sequence conservation between *C. elegans* and *C. briggsae* orthologs, analogy to the other constituents of the *mir-54–56* cluster, and Northern blots detecting RNA from only the 3′ arm of the fold-back (*30*).

We were surprised to find that few, if any, of the cloned RNAs had the features of siRNAs. No *C. elegans* clones matched the antisense of annotated coding regions. Of the 30 *C. elegans* clones not classified as miRNAs, 15 matched fragments of known RNA

genes, such as tRNA and ribosomal RNA. Of the remaining 15 clones, the best candidate for a natural siRNA is GGAAAACGGGUUGAAAGGGA. It was the only *C. elegans* clone perfectly complementary to an annotated EST, hybridizing to the 3′ UTR of gene ZK418.9, a possible RNA-binding protein. Even if this and a few other clones do represent authentic siRNAs, they would still be greatly outnumbered by the 300 clones representing 54 different miRNAs. Our cloning protocol is not expected to preferentially exclude siRNAs; it was similar to the protocol that efficiently cloned exogenous siRNAs from *Drosophila* extracts (*12*). Instead, we propose that the preponderance of miRNAs among our clones indicates that in healthy, growing cultures of *C. elegans*, regulation by miRNAs normally plays a more dominant role than does regulation by siRNAs.

Irrespective of the relative importance of miRNAs and siRNAs in the normal regulation of endogenous genes, our results show that small RNA genes, of the type exemplified by *lin-4* and *let-7*, are more abundant in *C. elegans* than previously appreciated. Results from a parallel effort that directly cloned small RNAs from *Drosophila* and HeLa cells demonstrates that the same is true in other animals (*25*), a conclusion further supported by the orthologs to the *C. elegans* miRNAs that we identified through database searching. Many of the miRNAs that we identified are represented by only a single clone (Table 1), suggesting that our sequencing has not reached saturation and that there are over a hundred miRNA genes in *C. elegans*.

We presume that there is a reason for the expression and evolutionary conservation of these small non-coding RNAs. Our favored hypothesis is that these novel miRNAs, together with *lin-4* and *let-7* RNAs, constitute an important and abundant class of riboregulators, pairing to specific sites within mRNAs to direct the

posttranscriptional regulation of these genes (*32*). The abundance and diverse expression

patterns of miRNA genes implies that they function in a variety of regulatory pathways,

in addition to their known role in the temporal control of developmental events.

References and Notes

1. R. C. Lee, R. L. Feinbaum, V. Ambros, *Cell* **75**, 843 (1993).

2. B. Wightman, I. Ha, G. Ruvkun, *Cell* **75**, 855 (1993).

3. B. J. Reinhart, et al., *Nature* **403**, 901 (2000).

4. E. G. Moss, R. C. Lee, V. Ambros, *Cell* **88**, 637 (1997).

5. F. J. Slack, et al., *Mol Cell* **5**, 659 (2000).

6. A. J. Hamilton, D. C. Baulcombe, *Science* **286**, 950 (1999).

7. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **404**, 293 (2000).

8. P. D. Zamore, T. Tuschl, P. A. Sharp, D. P. Bartel, *Cell* **101**, 25 (2000).

9. S. Parrish, J. Fleenor, S. Xu, C. Mello, A. Fire, *Mol. Cell* **6**, 1077 (2000).

10. D. Yang, H. Lu, J. W. Erickson, *Curr Biol* **10**, 1191 (2000).

11. C. Cogoni, G. Macino, *Curr Opin Genet Dev* **10**, 638 (2000).

12. S. M. Elbashir, W. Lendeckel, T. Tuschl, *Genes Dev* **15**, 188 (2001).

13. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).

14. A. Grishok, et al., *Cell* **106**, 23 (2001).

15. G. Hutvágner, et al., *Science* **293**, 834 (2001).

16. F. G. Ratcliff, S. A. MacFarlane, D. C. Baulcombe, *Plant Cell* **11**, 1207 (1999).

17. R. F. Ketting, T. H. Haverkamp, H. G. van Luenen, R. H. Plasterk, *Cell* **99**, 133 (1999).

18. H. Tabara, et al., *Cell* **99**, 123 (1999).

19. S. Malinsky, A. Bucheton, I. Busseau, *Genetics* **156**, 1147 (2000).

20. A. A. Aravin, et al., *Cur. Biol.* **11**, 1017 (2001).

21. A. Huttenhofer, et al., *Embo J* **20**, 2943 (2001).

22. K. M. Wassarman, F. Repoila, C. Rosenow, G. Storz, S. Gottesman, *Genes Dev* **15**, 1637 (2001).

23. Short endogenous *C. elegans* RNAs were cloned using a protocol inspired by Elbashir et al. (*12*), but modified to make it specific for RNAs with 5´-terminal phosphate and 3´-terminal hydroxyl groups. In our protocol (*24*), gel-purified 18–26 nt RNA from mixed-stage worms was ligated to a pre-adenylylated 3´-adaptor oligonucleotide in a reaction using T4 RNA ligase but without ATP. Ligated RNA was gel-purified, then ligated to a 5´-adaptor oligonucleotide in a standard T4 RNA ligase reaction. Products from the second ligation were gel-purified, then reverse transcribed and amplified by using the primers corresponding to the adaptor sequences. To achieve ligation specificity for RNA with a 5´-terminal phosphate and 3´-terminal hydroxyl, phosphatase and phosphorylase treatments, useful for preventing circularization of Dicer products (*12*), were not included our protocol. Instead, circularization was avoided by using the pre-adenylylated 3´-adaptor oligonucleotide and omitting ATP during the first ligation reaction.

24. Supplemental material describing methods and predicted fold-back secondary structures for the miRNAs of Table 1 and some of their homologs in other species is available at *Science* Online at www.sciencemag.org/cgi/content/full/294/5543/858/DC1

25. M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, *Science* **294**, 853 (2001).

26. R. C. Lee, V. Ambros, *Science* **294**, 862 (2001).

27. *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998).

28. M. J. Beanan, S. Strome, *Development* **116**, 755 (1992).

29. S. W. Knight, B. L. Bass, *Science* **293**, 2269 (2001).

30. N. C. Lau, L. P. Lee, E. Weinstein, D. P. Bartel, data not shown

31. A. E. Pasquinelli, et al., *Nature* **408**, 86 (2000).

32. This begs the question as to why more riboregulators have not been found previously. Perhaps they had not been identified biochemically because of a predisposition towards searching for protein rather than RNA factors. They could be identified genetically—this was how *lin-4* and *let-7* were discovered (*1-3*)—however, when compared to mutations in protein-coding genes, point substitutions in these short RNA genes would be less likely and perhaps less disruptive of function. Furthermore, mutations that map to presumed intergenic regions with no associated RNA transcript detectable on a standard RNA blot might be put aside in favor of other mutants.

33. Wormbase is available on the Web at www.wormbase.org.

34. Sequencing traces representing 2.5–3-fold average coverage of the *C. briggsae* genome were obtained at www.ncbi.nlm.nih.gov/Traces

35. I. L. Hofacker, et al., *Monatshefte f. Chemie* **125**, 167 (1994).

36. We thank C. Ceol for guidance in culturing and staging *C. elegans*, R. Horvitz for the use of equipment and facilities, T. Tuschl, G. Ruvkun, B. Reinhart, A. Pasquinelli, A. Ensminger, and C. Mello for helpful discussions, and P. Zamore, T. Orr-Weaver, and M. Lawrence for comments on this manuscript, and T. Tuschl, V. Ambros, G. Ruvkin, R. Horvitz, P. Sharp, and J. Hodgkin for discussions on nomenclature.

Table 1. miRNAs cloned from *C. elegans*. 300 RNA clones represented 54 different miRNAs. Also included are miR-39, miR-65, and miR-69, three miRNAs predicted based on homology and/or proximity to cloned miRNAs. miR-39 and miR-69 have been validated by Northern analysis (Fig. 3), whereas miR-65 is not sufficiently divergent to be readily distinguished by Northern analysis. All *C. elegans* sequence analysis relied on WormBase, release WS45 (*33*). Some miRNAs were represented by clones of different lengths, due to heterogeneity at the miRNA 3′-terminus. The observed lengths are indicated, as is the sequence of the most abundant length. Comparison to *C. briggsae* shotgun sequencing traces revealed miRNA orthologs with 100% sequence identity (+++) and potential orthologs with >90% (++) and >75% (+) sequence identity (*24, 34*). Five miRNA genomic clusters are indicated with square brackets. Naming of miRNAs was coordinated with the Tuschl and Ambros groups (*25, 26*).

miRNA gene	Number of clones	miRNA Sequence	Length	*C. briggsae* homology	Fold-back arm	Chromosome and distance to nearest gene
lin-4	10	UCCCUGAGAC CUCAAGUGUG A	21	+++	5'	II	
let-7	1	UGAGGUAGUA GGUUGUAUAG UU	22	+++	5'	X	
mir-1	9	UGGAAUGUAA AGAAGUAUGU A	21	+++	3'	I 3.7 kb from start of T09B4.3, antisense
mir-2	24	UAUCACAGCC AGCUUUGAUG UGC	22–23	+++	3'	I 0.6 kb from start of M04C9.6b
mir-34	3	AGGCAGUGUG GUUAGCUGGU UG	22	+++	5'	X 2.1 kb from end of Y41G9A.4, antisense
mir-35	9	UCACCGGGUG GAAACUAGCA GU	22	+	3'	II 1.3 kb from end of F54D5.12, antisense
mir-36	1	UCACCGGGUG AAAAUUCGCA UG	22	+	3'	II 1.2 kb from end of F54D5.12, antisense
mir-37	2	UCACCGGGUG AACACUUGCA GU	22	++	3'	II 1.1 kb from end of F54D5.12, antisense
mir-38	1	UCACCGGGAG AAAAACUGGA GU	22	+	3'	II 1.0 kb from end of F54D5.12, antisense
mir-39	0	UCACCGGGUG UAAAUCAGCU UG	predicted	++	3'	II 0.8 kb from end of F54D5.12, antisense
mir-40	2	UCACCGGGUG UACAUCAGCU AA	22	+	3'	II 0.7 kb from end of F54D5.12, antisense
mir-41	2	UCACCGGGUG AAAAAUCACC UA	22	+	3'	II 0.6 kb from end of F54D5.12, antisense
mir-42	1	CACCGGGUUA ACAUCUACAG	20	+++	3'	II 1.2 kb from end of ZK930.2, antisense
mir-43	1	UAUCACAGUU UACUUGCUGU CGC	23	+++	3'	II 1.1 kb from end of ZK930.2, antisense
mir-44	3	UGACUAGAGA CACAUUCAGC U	21	+++	3'	II 1.0 kb from end of ZK930.2, antisense
mir-45				+++	3'	II 0.7 kb from end of K12D12.1, antisense
mir-46	2	UGUCAUGGAG UCGCUCUCUU CA	22	+++	3'	III 3.0 kb from end of ZK525.1, antisense
mir-47	6	UGUCAUGGAG GCGCUCUCUU CA	22	+++	3'	X 1.8 kb from end of K02B9.2, antisense
mir-48	11	UGAGGUAGGC UCAGUAGAUG CGA	22–24	+++	5'	V 6.1 kb from start of Y49A3A.4
mir-49	1	AAGCACCACG AGAAGCUGCA GA	22	+++	3'	X 2.7kb from end of F19C6.1, antisense
mir-50	2	UGAUAUGUCU GGUAUUCUUG GGUU	24	++	5'	I in intron of Y71G12B.11a
mir-51	6	UACCCGUAGC UCCUAUCCAU GUU	23	++	5'	IV 0.4 kb from end of F36H1.6, antisense
mir-52	47	CACCCGUACA UAUGUUUCCG UGCU	22–25	+++	5'	IV 4.6 kb from end of Y37A1B.6, antisense
mir-53	2	CACCCGUACA UUUGUUUCCG UGCU	24	–	5'	IV 1.9 kb from end of F36H1.6, antisense
mir-54	2	UACCCGUAAU CUUCAUAAUC CGAG	24	+	3'	X 5.5 kb from end of F09A5.2, antisense
mir-55	5	UACCCGUAUA AGUUUCUGCU GAG	23	+	3'	X 5.3 kb from end of F09A5.2, antisense
mir-56	5	UACCCGUAAU GUUUCCGCUG AG	22	+	3'	X 5.2 kb from end of F09A5.2, antisense
	2	UGGCGGAUCC AUUUUGGGUU GUA	23	+	5'	X 5.2 kb from end of F09A5.2, antisense
mir-57	9	UACCCUGUAG AUCGAGCUGU GUGU	24	+++	5'	II 0.9 kb from start of AF187012-1.T09A5
mir-58	31	UGAGAUCGUU CAGUACGGCA AU	21–22	+++	3'	IV in intron of Y67D8A.1
mir-59	1	UCGAAUCGUU UAUCAGGAUG AUG	23	+	3'	IV 1.8 kb from start of B0035.1a, antisense
mir-60	1	UAUUAUGCAC AUUUUCUAGU UCA	23	++	3'	II 1.5 kb from end of C32D5.5
mir-61	1	UGACUAGAAC CGUUACUCAU C	21	+	3'	V 0.4 kb from end of F55A11.3, antisense
mir-62	1	UGAUAUGUAA UCUAGCUUAC AG	22	+++	3'	X in intron of T07C5.1
mir-63	1	UAUGACACUG AAGCGAGUUG GAAA	24	–	3'	X 1.7 kb from start of C16H3.2, antisense
mir-64	2	UAUGACACUG AAGCGUUACC GAA	23	–	5'	III 0.25 kb from start of Y48G9A.1
mir-65	0	UAUGACACUG AAGCGUAACC GAA	predicted	+	5'	III 0.10 kb from start of Y48G9A.1
mir-66	10	CAUGACACUG AUUAGGGAUG UGA	23–24	–	5'	III in coding sequence of Y48G9A.1
mir-67	2	UCACAACCUC CUAGAAAGAG UAGA	24	+++	3'	III 4.4 kb from end of EGAP1.1
mir-68	1	UCGAAGACUC AAAAGUGUAG A	21	–	3'	IV 3.3 kb from start of Y51H4A.22
mir-69	0	UCGAAAAUUA AAAAGUGUAG A	predicted	–	3'	IV 0.6 kb from start of Y41D4B.21, antisense
mir-70	1	UAAUACGUCG UUGGUGUUUC CAU	23	+	3'	V in intron of T10H9.5
mir-71	5	UGAAAGACAU GGGUAGUGA	19, 20, 22	+++	5'	I 7.8 kb from start of M04C9.6b
mir-72	9	AGGCAAGAUG UUGGCAUAGC	20, 21, 23	–	3'	II 0.21 kb from end of F53G2.4, antisense
mir-73	2	UGGCAAGAUG UAGGCAGUUC AGU	23	++	3'	X 2.9 kb from start of T24D8.6, antisense
mir-74	7	UGGCAAGAAA UGGCAGUCUA CA	22	++	3'	X 3.2 kb from start of T24D8.6, antisense
mir-75	2	UUAAAGCUAC CAACCGGCUU CA	22	++	3'	X 3.5 kb from start of F47G3.3
mir-76	1	UUCGUUGUUG AUGAAGCCUU GA	22	++	3'	III 3.0 kb from start of C44B11.3, antisense
mir-77	1	UUCAUCAgGC CAUAGCUGUC CA	22	+++	3'	II 1.5 kb from start of T21B4.9, antisense
mir-78	2	UGGAGGCCUG GUUGUUUGUG C	21	–	3'	IV 2.0 kb from start of Y40H7A.3, antisense
mir-79	1	AUAAAGCUAG GUUACCAAAG CU	22	+++	3'	I 2.3 kb from end of C12C8.2
mir-80	25	UGAGAUCAUU AGUUGAAAGC CGA	23	+++	3'	III 4.7 kb from end of F44E2.2, antisense
mir-81	7	UGAGAUCAUC GUGAAAGCUA GU	22	+++	3'	X in intron of T07D1.2, antisense
mir-82	6	UGAGAUCAUC GUGAAAGCCA GU	22	+++	3'	X 0.11 kb from start of T07D1.2
mir-83	1	UAGCACCAUA UAAAUUCAGu AA	22	++	3'	IV 5.0 kb from start of C06A6.2
mir-84	3	UGAGGUAGUA UGUAAUAUUG UA	22, 24	+	5'	X 0.8 kb from end of B0395.1, antisense
mir-85	1	UACAAAGUAU UUGAAAAGUC GUGC	24	++	3'	II in intron of F49E12.8, antisense
mir-86	6	UAAGUGAAUG CUUUGCCACA GUC	23	+++	5'	III in intron of Y56A3A.7

Figure Legends

Fig. 1. Fold-back secondary structures involving miRNAs (red) and their flanking sequences (black), as predicted computationally using RNAfold (*35*). (**A**) miR-84, an miRNA with similarity to *let-7* RNA. (**B**) miR-1, an miRNA highly conserved in evolution. (**C**) miR-56 and miR-56*, the only two miRNAs cloned from both sides of the same fold-back. (**D**) The *mir-35–41* cluster.

Fig. 2. Unique sequence features of the miRNAs. (**A**) Length distribution of the clones representing *E. coli* RNA fragments (white bars) and *C. elegans* miRNAs (black bars). (**B**) Sequence composition of the unique clones representing *C. elegans* miRNAs and *E. coli* RNA fragments. To avoid over-representation from groups of related miRNAs in this analysis, each set of paralogs was represented by its consensus sequence.

Fig. 3. Expression of novel miRNAs and *let-7* RNA during *C. elegans* development. Northern blots probe total RNA from mixed-stage worms (Mixed), worms staged as indicated, and *glp-4 (bn2)* adult worms (*24*). Specificity controls ruled out cross-hybridization among probes for miRNAs from the *mir-35–41* cluster (*24*). Other blots indicate that, miR-46/47, miR-56, miR-64/65, miR-66, and miR-80 are expressed constitutively throughout development (*30*).

Figure 1.

Figure 2.

A

B

C.elegans miRNAs

5' end   9-15 nt   3' end
          middle portion

E.coli RNA fragments

5' end   5-19 nt   3' end
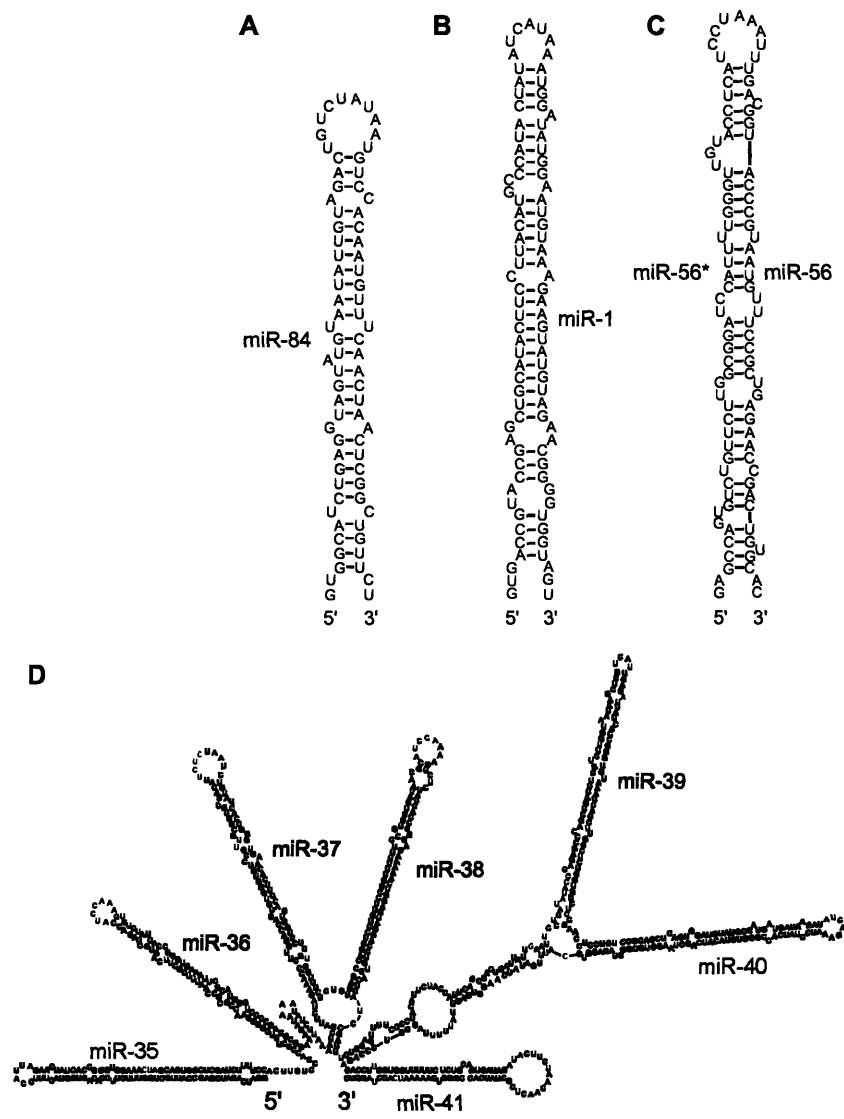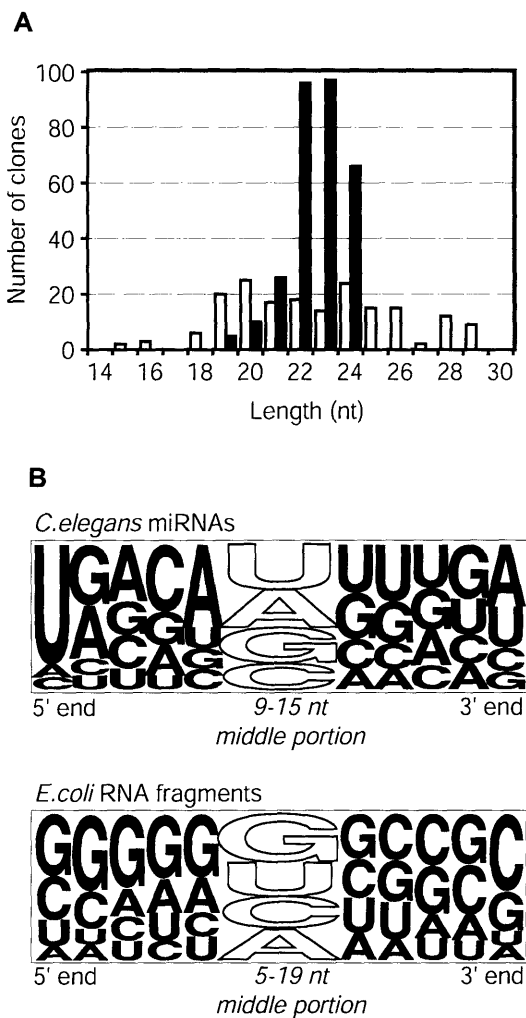   *middle portion*

Figure 3.

The MicroRNAs of *Caenorhabditis elegans*

The work presented in this chapter was a collaborative effort amongst several members in the Bartel and Burge lab. Specifically, Lee Lim executed the computational identification of miRNA genes, performed the phylogenetic analysis of miRNA homologs in *C.briggsae*, and completed the Northern blots testing the presence of candidate miRNA genes. Earl Weinstein carried out the large scale cloning experiments, analyzed the sequence data, and assisted in the Northern analysis of new miRNAs. Aliaa Abdelhakim improved the clone sequence analysis and assisted with Northern analysis. Soraya Yekta developed and applied the PCR method for validating rare miRNAs from cDNA libraries that I had generated. Matthew Rhoades determined miRNA abundance in HeLa cells by performing quantitative Northerns.

My contribution to this effort has been completing the developmental Northern analysis of the majority of *C.elegans* miRNAs (Figure 4), and determining miRNA abundance in nematode cells by quantitative Northerns (Figure 5). I cultured all the various nematode samples and extracted nematode RNAs used throughout the work, and I assisted in RNA cloning procedures and sequence analysis.

MicroRNAs (miRNAs) are an abundant class of tiny RNAs thought to regulate the expression of protein-coding genes in plants and animals. Here we describe a computational procedure to identify miRNA genes conserved in more than one genome. Applying this program, known as MiRscan, together with molecular identification and validation methods, we have identified most of the miRNA genes in the nematode *Caenorhabditis elegans*. The total number of validated miRNA genes stands at 88 with no more than 35 genes remaining to be detected or validated. These 88 miRNA genes represent 48 gene families; 46 of these families (comprising 86 of the 88 genes) are conserved in *C. briggsae*, and 22 families are conserved in humans. More than a third of the worm miRNAs, including newly identified members of the *lin-4* and *let-7* gene families, are differentially expressed during larval development, suggesting a role for these miRNAs in mediating larval developmental transitions. Most are present at very high steady-state levels—over 1000 molecules per cell, with some exceeding 50,000 molecules per cell. Our census of the worm miRNAs and their expression patterns helps define this class of noncoding RNAs, lays the groundwork for functional studies, and provides the tools for more comprehensive analyses of miRNA genes in other species.

Introduction

Noncoding RNAs (ncRNAs) of about 22 nucleotides (nt) in length are increasingly recognized as playing important roles in regulating gene expression in animals, plants, and fungi. The first such tiny regulatory RNA to be identified was the *lin-4* RNA, which controls the timing of *C. elegans* larval development (Lee et al. 1993; Wightman et al. 1993). This 21-nt RNA pairs to sites within the 3´ untranslated region (UTR) of target mRNAs, specifying the translational repression of these mRNAs and triggering the transition to the next developmental stage (Lee et al. 1993; Wightman et al. 1993; Ha et al. 1996; Moss et al. 1997; Olsen and Ambros 1999). A second tiny riboregulator, *let-7* RNA, is expressed later in development, and appears to act in a similar manner to trigger the transition to late-larval and adult stages (Reinhart et al. 2000; Slack et al. 2000). The *lin-4* and *let-7*

RNAs are sometimes called small temporal RNAs (stRNAs) because of their important roles in regulating the timing of larval development (Pasquinelli et al. 2000). The *lin-4* and *let-7* stRNAs are now recognized as the founding members of a large class of ~22-nt noncoding RNAs termed microRNAs (miRNAs), which resemble stRNAs but do not necessarily control developmental timing (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001).

Understanding the biogenesis and function of miRNAs has been greatly facilitated by analogy and contrast to another class of tiny ncRNAs known as small interfering RNAs (siRNAs), first identified because of their roles in mediating RNA interference (RNAi) in animals and posttranscriptional gene silencing in plants (Hamilton and Baulcombe 1999; Hammond et al. 2000; Parrish et al. 2000; Zamore et al. 2000; Elbashir et al. 2001a; Klahre et al. 2002). During RNAi, long double-stranded RNA (either a bimolecular duplex or an extended hairpin) is processed by Dicer, an RNAse III enzyme, into many siRNAs that serve as guide RNAs to specify the destruction of the corresponding mRNA (Hammond et al. 2000; Zamore et al. 2000; Bernstein et al. 2001; Elbashir et al. 2001a). Although these siRNAs are initially short double-stranded species with 5′ phosphates and 2-nt 3′ overhangs characteristic of RNAse III cleavage products, they eventually become incorporated as single-stranded RNAs into a ribonucleoprotein complex, known as the RNA-induced silencing complex (RISC) (Hammond et al. 2000; Elbashir et al. 2001a; Elbashir et al. 2001b; Nykäken et al. 2001; Martinez et al. 2002; Schwarz et al. 2002). The RISC identifies target messages based on perfect (or nearly perfect) antisense complementarity between the siRNA and the mRNA, and then the endonuclease of the RISC cleaves the mRNA at a site near the middle of the siRNA complementarity (Elbashir et al. 2001a; Elbashir et al. 2001b). Similar pathways have been proposed for gene silencing in plants and fungi, with siRNAs targeting mRNA for cleavage during posttranscriptional gene silencing and heterochromatic siRNAs targeting chromatin for histone methylation, triggering heterochromatin formation and consequent transcriptional gene silencing (Hamilton and Baulcombe 1999; Vance and Vaucheret 2001; Hall et

al. 2002; Hamilton et al. 2002; Pickford et al. 2002; Reinhart and Bartel 2002; Volpe et al. 2002; Zilberman et al. 2003).

MicroRNAs have many chemical and functional similarities to the siRNAs. Like siRNAs they are processed by Dicer, and so they are the same length and possess the same 5′-phosphate and 3′-hydroxyl termini as siRNAs (Grishok et al. 2001; Hutvágner et al. 2001; Ketting et al. 2001; Lau et al. 2001; Park et al. 2002; Reinhart et al. 2002). They are also incorporated within a ribonucleoprotein complex, known as the miRNP, which is similar, if not identical to the RISC (Caudy et al. 2002; Hutvágner and Zamore 2002; Ishizuka et al. 2002; Martinez et al. 2002; Mourelatos et al. 2002). In fact, many plant miRNAs match their predicted mRNA targets with near-perfect antisense complementarity, as if they were functioning as siRNAs within a RISC complex (Rhoades et al. 2002), and the plant miR171 and miR165/166 have been shown to specify cleavage of their mRNA targets (Llave et al. 2002b; Tang et al. 2003). The *C. elegans* and *Drosophila* miRNAs do not have as pronounced a tendency to pair with their targets with near perfect complementarity (Rhoades et al. 2002). Nonetheless, some might still direct cleavage of their targets, as suggested by the observation that miRNAs and siRNAs with 3–4 mismatches with their targets can still direct cleavage in plant and animal lysates (Tang et al. 2003). Furthermore, the *let-7* miRNA is present within a complex that can cleave an artificial RNA target when such a target is perfectly complementary to the miRNA (Hutvágner and Zamore 2002). The known biological targets of *lin-4* and *let-7* RNAs have several mismatches within the central region of the miRNA complementary sites, perhaps explaining why in these particular cases the miRNAs specify translational repression rather than mRNA cleavage during *C. elegans* larval development (Lee et al. 1993; Wightman et al. 1993; Ha et al. 1996; Moss et al. 1997; Olsen and Ambros 1999; Reinhart et al. 2000; Slack et al. 2000; Hutvágner and Zamore 2002).

Regulatory targets for most animal miRNAs have not yet been identified. Prediction of plant miRNA targets has led to the proposal that many plant miRNAs function to clear from differentiating cells mRNAs encoding key transcription factors, thereby facilitating plant

development and organogenesis (Rhoades et al. 2002). Confident computational prediction of animal miRNA targets has relied on experimental evidence to first narrow the number of candidate mRNAs (Lai 2002). Nonetheless, as seen for the plant miRNAs, the sequences of the animal miRNAs are generally highly conserved in evolution. For example, 91 of the 107 miRNAs cloned from mammals are detected in the pufferfish (*Fugu rubripes*) genome, implying that they have important functions preserved during vertebrate evolution (Lim et al. 2003).

The first step in a systematic approach to identifying the biological roles of miRNAs is to find the miRNA genes themselves. Because gene-prediction programs had not been developed to identify miRNAs in genomic sequence, miRNA gene identification has been primarily achieved by cloning the small RNAs from size-fractionated RNA samples, sometimes specifically enriching in miRNAs by first immunoprecipitating the miRNP complex, or by using a cloning protocol specific for the 5′ phosphate and 3′ hydroxyl found on Dicer products (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001; Lagos-Quintana et al. 2002; Llave et al. 2002a; Mourelatos et al. 2002; Park et al. 2002; Reinhart et al. 2002; Lagos-Quintana et al. 2003). Once small RNAs have been cloned, the challenge is to differentiate the authentic miRNAs from other RNAs present in the cell, particularly from endogenous siRNAs. Because both miRNAs and siRNAs are Dicer products and both can act to specify mRNA cleavage, miRNAs cannot be differentiated based on their chemical composition or their functional properties. However, miRNAs can be distinguished from siRNAs based on their biogenesis and evolutionary conservation: i) they are 20- to 24-nt RNAs that derive from endogenous transcripts that can form local RNA hairpin structures; ii) these hairpins are processed such that a single miRNA molecule ultimately accumulates from one arm of each hairpin precursor molecule; iii) the sequences of the mature miRNAs and their hairpin precursors are usually evolutionarily conserved; and iv) the miRNA genomic loci are distinct from and usually distant from those of other types of recognized genes, although a few are found within predicted introns but not necessarily in the same orientation as the introns. Endogenous siRNAs differ in that i) they derive from extended dsRNA, ii) each dsRNA precursor gives rise to numerous

different siRNAs iii) they generally display less sequence conservation, and iv) they often perfectly correspond to the sequences of known or predicted mRNAs, transposons, or regions of heterochromatic DNA (Aravin et al. 2001; Djikeng et al. 2001; Elbashir et al. 2001a; Lau et al. 2001; Llave et al. 2002a; Mochizuki et al. 2002; Reinhart and Bartel 2002; Reinhart et al. 2002). Regarding this fourth criterion, miRNAs can also perfectly correspond to sequences of their mRNA targets, but when they do, they still derive from loci distinct from those of their mRNA targets (Llave et al. 2002a; Llave et al. 2002b; Reinhart et al. 2002). Because miRNAs are primarily distinguished based on their biogenesis and evolutionary conservation, the current norms for identification and validation of miRNA genes include experimental evidence for endogenous expression of the miRNA, coupled with evidence of a hairpin precursor, preferably one that is evolutionarily conserved (Ambros et al. 2003).

Some miRNAs might be difficult to isolate by cloning, due to their low abundance or to biases in cloning procedures. Thus, computational identification of miRNAs from genomic sequences would provide a valuable complement to cloning. Recent advances have been made in the computational identification of ncRNA genes through comparative genomics, and complex algorithms have been developed to identify ncRNAs in general (Argaman et al. 2001; Rivas et al. 2001; Wassarman et al. 2001), as well as specific ncRNA families such as tRNAs and snoRNAs (Lowe and Eddy 1997; Lowe and Eddy 1999).

Here we describe a computational procedure to identify miRNA genes. Using this procedure, together with extensive sequencing of clones (3423 miRNA clones were sequenced), we have detected 30 additional miRNA genes, including previously unrecognized *lin-4* and *let-7* homologs. Extrapolation of the computational analysis indicates that miRNA gene identification in *C. elegans* is now approaching saturation, and that no more than 120 miRNA genes are present in this species. We also identify those genes with intriguing expression patterns during larval development and conditions of nutrient stress, and we show that most miRNAs are expressed at very high levels, with some present in as many copies per cell as the highly abundant U6 snRNA.

This extensive census of worm miRNAs and their expression patterns establishes the general

properties of this gene class and provides resources and tools for studies of miRNA function in

nematodes and other organisms.

Results

Computational prediction of *C. elegans* miRNA genes

We developed a computational tool to specifically identify miRNAs that are conserved in two

genomes and have the features characteristic of known miRNAs. To identify miRNAs in

nematodes, the *C. elegans* genome was first scanned for hairpin structures whose sequences were

conserved in *C. briggsae*. About 36,000 hairpins were found that satisfied minimum requirements

for hairpin structure and sequence conservation. This procedure cast a sufficiently wide net to

capture 50 of the 53 miRNAs previously reported to be conserved in the two species (Lau et al.

2001; Lee and Ambros 2001). These 50 published miRNA genes served as a training set for the

development of a program called MiRscan, which was then used to assign scores to each of the

36,000 hairpins, evaluating them based on their similarity to the training set with respect to the

following features: base pairing of the miRNA portion of the fold-back, base pairing of the rest of

the fold-back, stringent sequence conservation in the 5´ half of the miRNA, slightly less stringent

sequence conservation in the 3´ half of the miRNA, sequence biases in the first five bases of the

miRNA (especially a U at the first position), a tendency towards having symmetric rather than

asymmetric internal loops and bulges in the miRNA region, and the presence of 2–9 consensus base

pairs between the miRNA and the terminal loop region, with a preference for 4-6 bp (Fig. 1A).

The distribution of MiRscan scores for the ~36,000 hairpins illustrated the ability of

MiRscan to discern the 50 miRNA genes of the training set, which fell mostly in the high-scoring

tail of the distribution (Fig. 2). Of the features evaluated by MiRscan, base pairing potential and

sequence conservation played primary roles in distinguishing known miRNAs (Fig. 1B). Some of

the other conserved hairpins also scored highly; 35 had scores exceeding 13.9, the median score of

the 58 known miRNAs (Fig. 2B). These 35 hairpins were carried forward as the top miRNA candidates predicted by MiRscan.

Molecular identification of miRNA genes

Our initial cloning and sequencing of small RNAs from mixed-stage *C. elegans* had identified 300 clones that represented 54 unique miRNA sequences (Lau et al. 2001). For the present study, this approach for identifying miRNAs was scaled-up about 10 fold. In an effort to identify miRNAs not normally expressed in mixed-stage logarithmically growing hermaphrodite worms, RNA was also cloned from populations of *him-8* worms, starved L1, and dauer worms. The *him-8* population was about 40% males, whereas the normal (N2) population was nearly all hermaphrodites (Broverman and Meneely 1994). Starved L1 and dauer worms are arrested in development at larval stages L1 and L3 respectively, with dauer worms having undergone morphological changes that enhance survival following desiccation or other harsh conditions.

As before, some clones matched *E. coli*, the food source of the worms, others corresponded to fragments of annotated *C. elegans* RNAs. Nevertheless, 3423 clones were classified as miRNA clones (Table 1). Most of these represented the 58 miRNA genes previously identified in *C. elegans* (Lau et al. 2001; Lee and Ambros 2001). For example, *lin-4* was represented by 125 clones, *let-7* by 17 clones, and *mir-52* by 404 clones (Table 1). The remaining miRNA clones represented 23 newly identified miRNA loci.

In total, 80 loci were represented by cloned miRNAs (Table 1). Of these, 77 had the classical features of *C. elegans* miRNA genes, in that they had the potential to encode stereotypic hairpin precursor molecules with the 20- to 25-nt cloned RNAs properly positioned within an arm of the hairpin so as to be excised during Dicer processing, and their expression was manifested as a detectable Northern signal in the 20- to 25-nt range. Three other loci, *mir-41*, *mir-249*, and *mir-229*, were also included. The *mir-41* and *mir-249* RNAs were not detected on Northern blots but were still classified as miRNAs because these RNAs and their predicted hairpin precursors appear to be conserved in *C. briggsae*.

The *mir-229* locus was also classified as a miRNA gene, even though it appears to derive from an unusual fold-back precursor. Its precursor appears to be larger than normal, possibly because of an extra 35-nt stem-loop protruding from the 3′ arm of the precursor stem-loop (Supplementary Figure 1). Nonetheless, miR-229 was detectable as a ~25- to 26-nt species on Northern blots, and accumulation of its presumed precursor increased in the *dcr-1* mutant, suggesting that Dicer processes this precursor despite the unusual predicted secondary structure (Supplementary Figure 1). Furthermore, *mir-229* is only 400 bp upstream of a previously recognized miRNA gene cluster including *mir-64*, *mir-65*, and *mir-66*. miR-229 also has significant sequence identity with the miRNAs of this cluster. We provisionally classified *mir-229* as a miRNA and a member of this *C. elegans* cluster. Greater confidence would be warranted if its unusual precursor structure were conserved in another species. A weakly homologous cluster of two potential miRNAs was found in *C. briggsae*, but neither of the predicted *C. briggsae* homologs appeared to have an unusual precursor resembling that of miR-229.

Validation of computationally predicted miRNAs

Of the 23 newly cloned miRNAs, 20 received MiRscan scores, and these scores are indicated in yellow in Figure 2B. The other three were not scored because orthologous sequences in C. briggsae were not identified. A Mann-Whitney test showed that the distribution of scores for these recently cloned miRNAs was not significantly different from that of the previously cloned miRNAs. Because the recently cloned miRNAs were not known during the development of MiRscan, their high scores gave added assurance that MiRscan was not over-fitting its training set. Ten of the 23 newly cloned miRNAs were among the set of 35 high-scoring miRNA gene candidates and served to validate these ten candidates.

The remaining 25 candidate miRNAs that had not been cloned were tested by Northern blots. RNA from *dcr-1* worms was included on the blots to enhance detection of precursor hairpins. Dicer-dependent processing of ~70-nt precursors was detected for six candidates, and ~22-nt miRNAs were detected for miR-250, miR-251, and miR-252 (as shown for miR-250 and

miR-255, Fig. 2C). Despite prolonged exposure times and enrichment for small RNA by size

fractionation, the Northern signals were generally weak, perhaps explaining why these miRNAs

were missed in the current set of 3423 sequenced miRNA clones.

To investigate whether these miRNAs eventually would have been identified after further

cloning and sequencing of our cDNA library of small RNA sequences, a PCR assay was used to

detect the presence of these miRNAs in the library. Using a primer specific to the 3´ segment of the

predicted miRNA, together with a second primer corresponding to the adapter sequence attached to

the 5´ terminus of all the small RNAs, the 5´ segment of the miRNA was amplified, cloned, and

sequenced. This procedure validated five of the six predicted miRNAs for which at least a

precursor could be detected on Northerns, including two of the candidates (miR-253 and miR-254)

for which a mature ~22-nt RNA was not detected on Northern blots. In addition, it identified the 5´

terminus of these five miRNAs, which is difficult to achieve with confidence when using only

bioinformatics and hybridization.

Combining the cloning and expression data, 16 of the 35 computationally identified

candidates were validated (ten from cloning, five from Northerns plus the PCR assay, and one from

Northerns only, which validated the precursor but did not identify the mature miRNA). Of the

remaining 19 candidates, four could be readily classified as false positives. They appear to be non-

annotated larger ncRNA genes, in that probes designed to hybridize to these candidates hybridized

instead to high-molecular weight species that remained constant in the samples from *dcr-1* worms.

The remaining 15 new candidates with high MiRscan scores but without any Northern signal might

also be false positives, or they might be authentic miRNAs that are expressed at low levels or in

only very specific cell types or circumstances. Considering the extreme case in which all the non-

validated candidates are false positives, the minimum specificity of MiRscan for the *C. elegans/C.

briggsae* analysis can be calculated as (29 + 16)/(29 + 35), or 0.70, at a sensitivity level that detects

half of the 58 previously known miRNAs. A summary of the miRNA genes newly identified by

validating computational candidates (16 genes) or by cloning alone (13 genes) is shown in Table 2;

and predicted stem-loop precursors are shown in the supplemental on-line information. Table 2 also includes one additional gene, *mir-239b*, which was identified based on its homology to *mir-239a* and its MiRscan score of 13.6.

Evolutionary conservation of miRNAs

The 88 *C. elegans* miRNA genes identified to this point were grouped into 48 families, each comprising 1–8 genes (data not shown). Within families, sequence identity either spanned the length of the miRNAs, or it was predominantly at their 5´ terminus. All but two of these families extended to the miRNAs of *C. briggsae*. The two families without recognizable *C. briggsae* orthologs each comprised a single miRNA (miR-78 and miR-243). Thus, nearly all (>97%) of the *C. elegans* miRNAs identified had apparent homologs in *C. briggsae*, and all but 6 six of these *C. elegans* miRNAs (miR-72, miR-63, miR-64, miR-66, miR-229, and miR-247) had retained at least 75% sequence identity to a *C. briggsae* ortholog. Of the 48 *C. elegans* miRNA families, 22 also had representatives among the known human miRNA genes (Fig. 3). In that these 22 families included 33 *C. elegans* genes, it appears that at least a third (33/88) of the *C. elegans* genes have homologs in humans and other vertebrates.

Developmental expression of miRNAs

The expression of 62 miRNAs during larval development was examined and compiled together with previously reported expression profiles (Lau et al. 2001) to yield a comprehensive data set for the 88 *C. elegans* miRNAs (Fig. 4). RNA from wild-type embryos, the four larval stages (L1 through L4), and young adults was probed, as was RNA from *glp-4* (*bn2*) young adults, which are severely depleted in germ cells (Beanan and Strome 1992). Nearly two thirds of the miRNAs appeared to have constitutive expression during larval development (Fig. 4A). These miRNAs might still have differential expression during embryogenesis, or they might have tissue-specific expression, as has been observed for miRNAs of larger organisms in which tissues and organs can

be more readily dissected and examined (Lee and Ambros 2001; Lagos-Quintana et al. 2002; Llave et al. 2002a; Park et al. 2002; Reinhart et al. 2002).

Over a third of the miRNAs had expression patterns that changed during larval development (Fig. 4B and C), and there were examples of miRNA expression initiating at each of the four larval stages (Fig. 4B). Expression profiles for miR-48 and miR-241 (which are within 2 kb of each other in the *C. elegans* genome) were similar to those previously reported for *let-7* RNA and miR-84 (Reinhart et al. 2000; Lau et al. 2001) (Fig. 4B). In fact, these four miRNAs appear to be paralogs, with all four miRNAs sharing the same first eight residues (Fig. 3). Another newly identified miRNA, miR-237, is a paralog of the other canonical stRNA, *lin-4* RNA (Fig. 3), although miR-237 exhibited an expression pattern distinct from *lin-4* RNA (Fig. 4E). The existence of these paralogs, as well as other families of miRNAs with expression initiating at the different stages of larval development, supports the idea that *lin-4* and *let-7* miRNAs are not the only stRNAs with important roles in the *C. elegans* heterochronic pathway.

Expression usually remained constant once it initiated, as has been seen for *lin-4* and *let-7* miRNA expression (Fig. 4A and B). Exceptions to this trend included the miRNAs of the *mir-35–mir-41* cluster, which were expressed transiently during embryogenesis (Lau et al. 2001), miR-247, which was expressed transiently in larval stage 3 (and dauer), and miR-248, which was most highly expressed in dauer (Fig. 4C and D). miR-234 was expressed in all stages, but expression was highest in both L1 worms (which had been starved shortly before harvest to synchronize the worm developmental staging) and dauer worms, suggesting that this miRNA might be induced as a consequence of nutrient stress.

Molecular Abundance of miRNAs

The very high cloning frequency of certain miRNAs (e.g. miR-52, represented by more than 400 clones) raised the question as to the molecular abundance of these and other miRNA species. In addition, there was the question of whether the actual molecular abundance of miRNAs in nematodes was proportionally reflected in the numbers of clones sequenced. To address these

questions, quantitative Northerns were used to examine the molecular abundance of 12

representative miRNAs, picked so as to span the range of frequently and rarely cloned sequences

and differing 3′ and 5′ terminal residues (Fig. 5).

To determine the molecular abundance of these 12 miRNAs in the adult worm soma, the

hybridization signals for RNA from a known number of *glp-4* young adult worms were compared

to standard curves from chemically synthesized miRNAs (Hutvágner and Zamore 2002) (Fig. 5).

Accounting for RNA extraction yields and dividing the number of miRNA molecules per worm by

the total number of cells in the worms, yielded averages of up to 50,000 molecules per cell, with the

most abundant miRNAs as plentiful as the U6 snRNA of the spliceosome (Fig. 5C). These are

much higher numbers than those for the typical worm mRNAs, estimated to average about 100

molecules per cell for the 5000 most highly expressed genes in the cell. (This estimate was

calculated based on our yield of 20 pg total RNA per worm cell, assuming that the 5000 most

highly expressed genes have mRNAs averaging 2 Kb in length and represent 3 percent of the total

RNA in an adult worm; it was consistent with estimates based on hybridization kinetics of mRNAs

from mouse tissues (Hastie and Bishop 1976)). Perhaps high concentrations of miRNAs are needed

to saturate the relevant complementary sites within the target mRNAs, which might be recognized

with low affinity because of the non-canonical pairs or bulges that appear to be characteristic of the

animal miRNA–target interactions.

Because these numbers represent molecular abundance averaged over all the cells of the

worm, including cells that might not be expressing the miRNA, there are likely to be some cells that

express even more molecules of the miRNA. To examine the abundance in a single cell type, HeLa

RNA was probed for representative human miRNAs, yielding a similar range of molecular

abundance (Fig. 5C). The high number of miRNA molecules in human cells only increases the

mystery as to why miRNAs had gone undetected for so long, which raises the question of whether

other classes of highly expressed ncRNAs might yet remain to be discovered. A recent large-scale

analysis of full-length cDNAs from mouse indicates the possible existence of hundreds or thousands of expressed ncRNAs in vertebrates (Okazaki et al. 2002).

To address the extent to which the actual molecular abundance of miRNAs in nematodes is proportionally reflected in the numbers of clones sequenced, the abundance of the miRNA within the mixed-stage RNA preparation was compared to the number of clones generated from that preparation (Fig. 5D). The strong positive correlation observed between the molecular abundance and the number of times the miRNAs were cloned indicated that systematic biases in the cloning procedure were not major. At most, these miRNAs were over- or under-represented 5-fold in the sequenced set relative to their actual abundance as measured by quantitative Northerns. We cannot rule out the possibility that certain miRNAs not yet cloned might be refractory to our cloning procedure, e.g. because of a propensity to form secondary structures that preclude adaptor ligation reactions. Nonetheless, on the whole, the cloning frequencies can be used to approximate the molecular abundance of the miRNAs, and we have no reason to suspect that the set of miRNAs identified by cloning differs in any substantive way, other than an overall higher steady-state expression level, from the complete set of *C. elegans* miRNAs.

Other endogenous ~22-nt RNAs of *C. elegans*

Of the 4078 *C. elegans* clones, a large majority represented authentic miRNAs (3423 clones, Table 1). The next most abundant class represented degradation fragments of larger ncRNAs, such as tRNA and rRNA (447 clones) and introns (18 clones). The remaining clones represented potential Dicer products that were not classified as miRNAs. Some corresponded to sense (18 clones) or antisense (23 clones) fragments of known or predicted mRNAs and might represent endogenous siRNAs. Others (143 clones) corresponded to regions of the genome not thought to be transcribed; these might represent another type of endogenous siRNAs, known as heterochromatic siRNAs (Reinhart and Bartel 2002). The possible roles of the potential siRNAs and heterochromatic siRNAs in regulating gene expression are still under investigation. The remaining clones were

difficult to classify because they matched more than one locus, and their loci were of different types (6 clones).

A fourth class of potential Dicer products (38 clones, representing 14 loci) corresponded to miRNA precursors but derived from the opposite arm of the hairpin than the more abundantly expressed miRNA, as has been reported previously for miR-56 in *C. elegans*, miR156d and miR169 in plants, and several vertebrate miRNAs (Lau et al. 2001; Lagos-Quintana et al. 2002; Mourelatos et al. 2002; Reinhart et al. 2002; Lagos-Quintana et al. 2003). Our current data adds another 13 examples of this phenomenon (Fig. 6). In all of our cases, the ~22-nt RNA from one arm of the fold-back was cloned much more frequently than that from the other and was far more readily detected on Northern blots. We designated the less frequently cloned RNA as the miRNA-star (miRNA*) fragment (Lau et al. 2001).

Discussion

We have developed a computational procedure for identifying miRNA genes conserved in two genomes. Using this procedure, together with extensive sequencing of clones from libraries of small RNAs, we have now identified 87 miRNA genes in *C. elegans* (Tables 1 and 2). Together with *mir-88* (Lee and Ambros 2001; Lagos-Quintana et al. 2002; Llave et al. 2002a; Park et al. 2002; Reinhart et al. 2002), which we have not yet cloned nor found computationally, the number of validated *C. elegans* genes stands at 88. More than a third of these genes have human homologs (Fig. 3), and a similar fraction, including previously unrecognized *lin-4* and *let-7* paralogs, are differentially expressed during larval development (Fig. 4). Most miRNAs accumulated to very high steady-state levels, with some at least as plentiful as the U6 snRNA (Fig. 5). Below, we discuss some implications of these results with regard to some of the defining features of miRNA genes in animals, the processing of miRNA precursors, and the number of miRNA genes remaining to be identified.

MiRscan accuracy and the defining features of miRNAs

As calculated in the Results section, the specificity of MiRscan was ≥ 0.70 at a sensitivity that detects half the previously known *C. elegans* miRNAs, when starting from an assembled *C. elegans* genome and *C. briggsae* shotgun reads. This accuracy was sufficient to identify new genes and obtain an upper bound on the total number of miRNA genes in the worm genome (described later). However, it was not sufficient to reliably identify all the conserved miRNA genes in *C. elegans*. The accuracy of MiRscan appears to be at least as high as that of general methods to identify ncRNA genes in bacteria (Argaman et al. 2001; Rivas et al. 2001; Wassarman et al. 2001), but is lower than that of algorithms designed to identify protein-coding genes or specialized programs that predict tRNAs and snoRNAs (Lowe and Eddy 1997; Burge and Karlin 1998; Lowe and Eddy 1999). The relative difficulty in identifying miRNAs can be explained by the low information content inherent in their small size and lack of strong primary sequence motifs. The performance of MiRscan will improve with a more complete and assembled *C. briggsae* genome. We anticipate that using only those sequences conserved in a syntenic alignment of the two genomes would capture fewer of the background sequences, enabling the authentic miRNAs to be more readily distinguished from the false positives.

Improvement would also come from bringing in a third nematode genome, particularly a genome more divergent than those of *C. elegans* and *C. briggsae*. The advantage of such an additional genome is illustrated by our application of MiRscan to the identification of vertebrate miRNAs using three genomes. The version of MiRscan described here, which had been trained on the set of 50 miRNAs conserved in worms, was applied to the assembled human genome, shotgun reads of the mouse genome, and the assembled pufferfish (*Fugu*) genome (Lim et al. 2003). This analysis had a specificity of ≥ 0.71 at a sensitivity that detected three quarters of the previously known vertebrate miRNAs. The accuracy of the vertebrate analysis was therefore substantially improved over that of the *C. elegans/C. briggsae* analysis, even though the vertebrate genomes are 4–30 times larger than those of *C. elegans* and *C. briggsae*, and are expected to have a correspondingly higher number of background sequences. This improved performance can be

attributed to using three genomes, and also to the evolutionary distance between the mammalian and fish genomes, which are distant enough to reduce the number of fortuitously high scoring sequences, yet close enough to retain most of the known miRNAs.

Other improvements in the computational identification of miRNAs will come with the definition of additional sequence and structural features that specify which sequences are transcribed, processed into miRNAs, and loaded into the miRNP. With the exception of sequence conservation, the features that MiRscan currently uses to identify miRNAs (Fig. 1A) are among those that the cell also uses to specify the biogenesis of miRNAs and miRNPs. The utility of these parameters for MiRscan (Fig. 1B) is a function of both the degree to which these features are correctly modeled (or have already been utilized to restrict the number of miRNA candidates; see Fig. 1B legend) and their relative importance *in vivo*. Clearly, much of what defines a miRNA *in vivo* remains to be determined. Sequence elements currently unavailable for MiRscan include transcriptional promoter and termination signals. Additional sequence and structural features important for processing of the primary transcript and the hairpin precursors also remain to be identified (Lee et al. 2002).

MicroRNA biogenesis

The presence of miRNA* species, observed now for 14 of the *C. elegans* miRNAs ((Lau et al. 2001); Fig 6), provides evidence for the idea that Dicer processing of miRNA precursors resembles that of siRNA precursors (Hutvágner and Zamore 2002; Reinhart et al. 2002). We suspect that, with more extensive sequencing of clones, miRNA* sequences will be found for a majority of the miRNA precursors, a notion supported by the identification of additional miRNA* sequences using our PCR assay (data not shown). As observed for both *MIR156d* and *MIR169* in plants (Reinhart et al. 2002), the miRNA:miRNA* segments are typically presented within the predicted precursor, paired to each other with 2-nt 3´ overhangs (Fig. 6)—a structure analogous to that of a classical siRNA duplex. This is precisely the structure that would be expected if both the miRNA and the miRNA* were excised from the same precursor molecule, and the miRNA* fragments were

transient side-products of productive Dicer processing. An alternative model for miRNA

biogenesis and miRNA* formation, which we do not favor but cannot rule out, is that the Dicer

complex normally excises a ~22-nt RNA from only one side of a miRNA precursor but it

sometimes binds the precursors in the wrong orientation and excises the wrong side. In an extreme

version of the favored model, the production of the miRNA* would be required for miRNA

processing and miRNP assembly; in a less extreme version, miRNA* production would be an

optional, off-pathway phenomenon. The idea that ~22-nt RNAs might be generally excised from

both sides of the same precursor stem-loop brings up the question of why the miRNAs and

miRNA*s are present at such differing levels. With the exception of miR-34* (sequenced 17 times)

none of the miRNA*s is represented by more than 3 sequenced clones. Perhaps the miRNAs are

stabilized relative to their miRNA* fragments because they preferentially enter the miRNP/RISC

complex. Alternatively, both the miRNA and the miRNA* might enter the complex, but the

miRNA might be stabilized by interactions with its targets.

Five of the newly identified miRNAs are within annotated introns, all five in the same

orientation as the predicted mRNAs. When considered together with the previously identified

miRNAs found within annotated introns (Lau et al. 2001), 10 of 12 known *C. elegans* miRNAs

predicted to be in introns are in the same orientation as the predicted mRNAs. This bias in

orientation, also reported recently for mammalian miRNAs (Lagos-Quintana et al. 2003), suggests

that some of these miRNAs are not transcribed from their own promoters but instead derive from

the excised pre-mRNA introns (as are many snoRNAs), and it is easy to imagine regulatory

scenarios in which the coordinate expression of a miRNA with an mRNA would be desirable.

The number of miRNA genes in *C. elegans* and other animals

In addition to providing a set of candidate miRNAs, MiRscan scoring provides a means to estimate

the total number of miRNA genes in *C. elegans*. A total of 64 loci have greater than the median

score of the 58 initially reported *C. elegans* miRNAs (Fig. 2B). Note that this set of 58 miRNAs

includes not only the 50 conserved miRNAs of the training set but also all eight of the previously

reported miRNAs that were not in our set of 36,000 potential stem-loops, usually because they lacked easily recognizable *C. briggsae* orthologs. Thus, the estimate calculated below takes into account the poorly conserved miRNAs without MiRscan scores. Four of the 64 high-scoring loci are known to be false positives. Thus, the upper bound on the number of miRNA genes in *C. elegans* would be $2 \times (64 - 4)$, or 120. This upper bound of ~120 genes remained stable when extrapolating from points other than the median, ranging from the top 25th to 55th percentiles. For this estimate, we made the assumption that the set of all *C. elegans* miRNAs has a distribution of MiRscan scores similar to the distribution of initially reported miRNAs. Such an assumption might be called into question, particularly when considering that the initially reported miRNAs served as a training set for the development of MiRscan (even though the scores of the training-set loci have been jackknifed to prevent over-fitting). However, this assumption is supported by two observations. First, the set of newly cloned miRNAs did indeed have a distribution of scores indistinguishable from that of the training set of previously reported miRNAs (Fig. 2B). Second, there is no correlation between the number of times that a miRNA has been cloned and its MiRscan score (Fig. 7). The absence of a correlation between cloning frequency and MiRscan score lessens our concern that miRNAs that are difficult to clone, including those still not present in our set of 3423 sequenced clones, might represent a population of miRNAs that are refractory to computational analysis as well.

This estimate of 120 genes is an upper bound and would decrease if additional high-scoring candidates were shown to be false positives. The extreme scenario, in which all are false positives, places the lower bound of miRNA genes near the number of validated genes, adding perhaps another five genes to account for the low-scoring counterparts of the five computational candidates validated only by Northerns and PCR, yielding a lower bound on the number of *C. elegans* miRNAs of ~93.

Our count of 105 ± 15 miRNA genes in *C. elegans* might underestimate the true count if there are miRNAs with unusual fold-back precursors that were cloned but dismissed as endogenous

siRNAs or degradation fragments. To investigate this possibility, we examined the expression of each small RNA that was cloned more than once but did not appear to derive from a canonical miRNA precursor as predicted by RNAfold. Because most (72/88) of the authentic miRNAs identified to date were represented by multiple clones (Table 1), this analysis should uncover most of the miRNAs coming from non-conventional precursors. This broader analysis detected only a single additional miRNA, miR-229. All of the other sequences that we cloned more than once were minor degradation fragments or processing byproducts of larger ncRNAs (e.g. the 5′ leader sequence of a tRNA). Thus, the number of miRNAs that derive from non-conventional precursors is not sufficient to significantly influence the miRNA gene count.

The estimated number of miRNA genes represents between 0.5 and 1 percent of the genes identified in the *C. elegans* genome, a fraction similar to that seen for other very large gene families with presumed regulatory roles, such as those encoding nuclear hormone receptors (270 predicted genes), C2H2 Zinc-finger proteins (157 predicted genes), and homeodomain proteins (93 predicted genes) (Chervitz et al. 1998; C.elegans Sequencing Consortium 1998). Extending our analysis to vertebrate genomes revealed that 230 ± 30 of the human genes are miRNAs, also nearly 1 percent of the genes in the genome (Lim et al. 2003). The miRNA genes are also among the most abundant of the ncRNA gene families in humans, comparable in number to the genes encoding rRNAs (~650-900 genes), tRNAs (~500 genes), snRNAs (~100 genes), and snoRNAs (~100-200 genes) (Lander et al. 2001). For rRNAs, tRNAs and snRNAs, the hundreds of gene copies in the human genome represent only relatively few distinct genes, probably fewer than 100 distinct genes for all three classes combined. For the miRNAs and snoRNAs, there are many more distinct genes, and each is present in only one or a few copies.

Unlike the other large ncRNA gene families and many of the transcription-factor gene families, there is no indication that miRNAs are present in single-celled organisms such as yeast. A pilot attempt to clone miRNAs from *Schizosaccharomyces pombe* did not detect any miRNAs (Reinhart and Bartel 2002), and there is no evidence that the proteins (such as Dicer) needed for

102

miRNA accumulation in plants and animals are present in *Saccharomyces cerevisiae*. Given the known roles of miRNAs in *C. elegans* development (Lee et al. 1993; Wightman et al. 1993; Reinhart et al. 2000) and the very probable roles of miRNAs in plant development (Rhoades et al. 2002), it is tempting to speculate that the substantial expansion of miRNA genes in animals (and the apparent loss of miRNA genes in yeast) is related to their importance in specifying cell differentiation and developmental patterning, and that the extra layer of gene regulation afforded by miRNAs was crucial for the emergence of multicellular body plans. The identification of most of the worm miRNAs and the quantitation of the number of genes remaining to be found are important steps towards understanding the evolution of this intriguing class of genes and placing them within the gene regulatory circuitry of these and other animals.

Materials and Methods

Computational identification of stem-loops

Potential miRNA stem-loops were located by sliding a 110-nt window along both strands of the *C.*

elegans genome (Wormbase Release 45, www.wormbase.org) and folding the window with the

secondary structure-prediction program RNAfold (Hofacker et al. 1994) to identify predicted stem-

loop structures with a minimum of 25 base pairs and a folding free energy of at least 25 kcal/mol

($\Delta G^{\circ}_{folding} \leq -25$ kcal/mol). Sequences that matched repetitive elements were discarded, as were those

with skewed base compositions not observed in known miRNA stem-loops and those which

overlapped with annotated coding regions. Stem-loops that had fewer base pairs than overlapping

stem-loops were also culled. *C. briggsae* sequences with at least loose sequence similarity to the

remaining *C. elegans* sequences were identified among *C. briggsae* shotgun sequencing reads

(November 2001 download from www.ncbi.nlm.nih.gov/Traces) using WU-BLAST with default

parameters and a non-stringent cutoff of E < 1.8 (Gish, W., blast.wustl.edu). These *C. briggsae*

sequences were folded with RNAfold to ensure that they met the minimal requirements for a hairpin

structure as described above. This procedure yielded about 40,000 pairs of potential miRNA

hairpins. For each pair of potential miRNA hairpins, a consensus *C. elegans/briggsae* structure was

generated using the alidot and pfrali utilities from the Vienna RNA package (Hofacker et al. 1998;

Hofacker and Stadler 1999). To create RNA consensus structures, alidot and pfrali combine a Clustal

alignment (Thompson et al. 1994) of a pair of sequences with either the minimum free energy

structures of these sequences (alidot) derived using the Zuker algorithm (Zuker 1994) or the base

pairing probability matrices of these sequences (pfrali) derived using the McCaskill algorithm

(McCaskill 1990).

MiRscan

Of the ~40,000 pairs of hairpins, 35,697 had the minimal conservation and base pairing needed to

receive a MiRscan score. Among this set were 50 of the 53 previously published miRNAs that were

reported to be conserved between *C. elegans* and *C. briggsae* (Lau et al. 2001; Lee and Ambros

2001). [miR-53 is included as a previously reported conserved miRNA because it is nearly identical

to miR-52, which has a highly conserved *C. briggsae* ortholog (Lau et al. 2001; Lee and Ambros

2001). The three conserved genes missing from the ~36,000 pairs of hairpins were *mir-56*, *mir-75*,

and *mir-88*. The reverse complements of *mir-75* and *mir-88* were later observed among the ~36,000

hairpins and given scores (Table 1).] The MiRscan program was developed to discriminate these 50

known miRNA hairpins from background sequences in the set of ~36,000 hairpins. For a given 21-nt

miRNA candidate, MiRscan makes use of the seven features derived from the consensus hairpin

structure illustrated in Figure 1A: x_1, "miRNA base pairing", the sum of the base-pairing probabilities

for pairs involving the 21-nt candidate miRNA; x_2, "extension of base pairing", the sum of the base-

pairing probabilities of the pairs predicted to lie outside the 21-nt candidate miRNA but within the

same helix; x_3, "5′ conservation", the number of bases conserved between *C. elegans* and *C.*

briggsae within the first 10 bases of the miRNA candidate; x_4, "3′ conservation", the number of

conserved bases within the last 11 bases of the miRNA candidate; x_5, "bulge symmetry", the number

of bulged or mismatched bases in the candidate miRNA minus the number of bulged or mismatched

bases in the corresponding segment on the other arm of the stem-loop; x_6, "distance from loop", the

number of basepairs between the loop of the stemloop and the closest end of the candidate; x_7, "initial

pentamer", the specific bases at the first 5 positions at the candidate 5′ terminus.

For a given feature i with value x_i, MiRscan assigns a log-odds score $s_i(x_i) = \log_2\left(\frac{f_i(x_i)}{g_i(x_i)}\right)$,

where $f_i(x_i)$ is an estimate of the frequency of feature value x_i in miRNAs derived from the training

set of 50 known miRNAs, and $g_i(x_i)$ is an estimate of the frequency of feature value x_i among the

background set of ~36,000 hairpin pairs. The overall score assigned to a candidate miRNA is simply

the sum of the log-odds scores for the 7 features: $S = \sum_{i=1..7} s_i(x_i)$. To score a given hairpin, MiRscan

slides a 21 nucleotide window representing the candidate miRNA along each arm of the hairpin,

assigns a score to each window, and then assigns the hairpin the score of its highest-scoring window.

In order to be evaluated, a window was required to be 2–9 consensus base pairs away from the terminal loop.

For features x_1, x_3, x_4, x_5, and x_6, f_i and g_i were obtained by smoothing the empirical frequency distributions from the training and background sets, respectively, using the R statistical package (http://lib.stat.cmu.edu/R/CRAN) with a triangular kernel. Because x_1 and x_2 are not independent of each other, the relative contribution of x_2 was decreased by computing f_2 and g_2 separately subject to the conditions $x_1 \geq 9$ and $x_1 < 9$, in order to account for this dependence. For x_7, a weight matrix model (WMM) was generated for the five positions at the miRNA 5′ terminus. The background WMM, g_7, was set equal to the base composition of the background sequence set. The miRNA WMM, f_7, was derived from the position-specific base frequencies of the 50 training set sequences, using standard unit pseudocounts, and normalizing for the contributions of related miRNAs.

Because both strands of the *C. elegans* genome were analyzed, both a hairpin sequence and its reverse complement were sometimes included in the set of ~36,000 stem-loops. For representation in Figure 2, in such cases both sequences were considered as a single locus that received the score of the higher scoring hairpin. Also, to prevent overscoring of the 50 known miRNA loci within the training set, each known miRNA locus was assigned a jackknife score calculated using a training set consisting of the other 49 miRNAs.

RNA cloning and bioinformatic analyses

Small RNAs were cloned as described previously (Lau et al. 2001), using the protocol available on the web (http://web.wi.mit.edu/bartel/pub/). Sequencing was performed by Agencourt Bioscience. Sequences of known *C. elegans* tRNA and rRNA were removed, and the remaining clones were clustered based on the location of their match to the *C. elegans* genome (C.elegans Sequencing Consortium 1998), downloaded from WormBase (www.wormbase.org). Genomic loci not previously reported to encode miRNAs were examined using the RNA-folding program RNAfold (Hofacker et al. 1994). Two sequences were folded for each locus: one included 15 nt upstream and 60 nt

downstream of the most frequently cloned sequence from that locus; the other included 60 nt

upstream and 15 nt downstream. Sequences for which the most stable predicted folding resembled

the stem-loop precursors of previously validated miRNAs were carried forward as candidate miRNA

loci. Sequences without classical stem-loop precursors were also analyzed further (see Discussion),

but only one, miR-229, was classified as a miRNA. The clones classified as representing potential

fragments of mRNAs (20 clones) and potential antisense fragments of mRNAs (23 clones)

corresponded to predicted ORFs (as annotated in GenBank) or probable UTR segments (100 bp

upstream or 200 bp downstream of the predicted ORF).

Northerns

Expression of candidate miRNA loci was examined using Northern blots and radiolabeled DNA

probes (Lau et al. 2001). To maintain hybridization specificity without varying hybridization or

washing conditions, the length of probes for different sequences was adjusted so that the predicted

melting temperatures of the miRNA-probe duplexes did not exceed 60°C (Sugimoto et al. 1995).

Probes not corresponding to the entire miRNA sequence were designed to hybridize to the 3´ region

of the miRNA, which is most divergent among related miRNA sequences.

PCR validation

A PCR assay was performed to detect the sequences of predicted miRNAs within a cDNA library

constructed from 18- to 26-nt RNAs expressed in mixed-stage worms. This library, the same as that

used for cloning (Lau et al. 2001), consisted of PCR-amplified DNA that comprised the 18- to 26-nt

sequences flanked by 3´- and 5´-adaptor sequences. For each miRNA candidate, a primer specific to

the predicted 3´ terminus of the candidate and a primer corresponding to the 5´-adaptor sequence

common to all members of the library (ATCGTAGGCACCTGAAA) were used at concentrations of

1.0 μM and 0.1 μM respectively (100 μl PCR reaction containing 5 μl of a 400-fold dilution of the

PCR reaction previously used to amplify all members of the cDNA library). The specific primer was

added after the initial denaturation incubation had reached 80°C. Following 20 PCR cycles, the

reaction was diluted 20 fold into a fresh PCR reaction for another 20 cycles. PCR products were cloned and sequenced to both identify the 5' terminus of the miRNA and ensure that the amplified product was not a primer-dimer or other amplification artifact. Specific primers for the reactions that successfully detected candidate miRNAs were ACCATGCCAACAGTTG (miR-250), TAAGAGCGGCACCACTAC (miR-251), TACCTGCGGCACTACTAC (miR-252), GTCAGTGTTAGTGAGG (miR-253), TACAGTCGGAAAGATTTG (miR-254), and GTGGAAATCTATGCTTC (miR-254*).

Quantitative Northerns

MicroRNA standards (purchased from Dharmacon) were diluted to appropriate concentrations in the presence of 1.0 μg/μl carrier RNA in the form of either *E. coli* tRNA or HeLa cell total RNA. Northern analysis was performed (Lau et al. 2001), loading 30 μg of RNA per lane, in the format shown for miR-66 (Fig. 5A). Signals were quantitated using phosphorimaging, standard curves (linear through at least three orders of magnitude, including the region of interpolation) were constructed, and absolute amounts of miRNAs per sample were determined, as illustrated for miR-66 (Fig. 5B). The average number of miRNA molecules per *glp-4* adult nematode was calculated using 19 ng as the average amount of total RNA extracted per worm. This number was determined as the average of three independent extraction trials, from known numbers of synchronized, 2-day-old adult *glp-4(bn2)* hermaphrodites, the same frozen worm population used for the quantitative Northern blots. All extractions were performed as described previously (Lau et al. 2001), except during two of the trials a radiolabeled miRNA was spiked into the preparation during worm lysis. At least 90% of this RNA was recovered, indicating near quantitative yield. Having calculated the number of each miRNA per worm, the average number of miRNAs per cell was calculated using 989 as number of cells per worm. The 989 cells per worm is based on the 959 somatic nuclei of the adult hermaphrodites plus the 30 germ nuclei of 2-day-old adult *glp-4(bn2)* animals (Sulston et al. 1983; Beanan and Strome 1992). Total RNA from known numbers of HeLa cells was determined in an analogous fashion.

Acknowledgments

References

Ambros, V., B. Bartel, D.P. Bartel, C.B. Burge, J.C. Carrington, X. Chen, G. Dreyfuss, S. Eddy, S. Griffiths-Jones, M. Matzke, G. Ruvkun, and T. Tuschl. 2003. A uniform system for microRNA annotation. *RNA*: in press.

Aravin, A.A., N.M. Naumova, A.A. Tulin, Y.M. Rozovsky, and V.A. Gvozdev. 2001. Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in D. melanogaster germline. *Curr. Biol.* **11**: 1017-1027.

Argaman, L., R. Hershberg, J. Vogel, G. Bejerano, E.G. Wagner, H. Margalit, and S. Altuvia. 2001. Novel small RNA-encoding genes in the intergenic regions of Escherichia coli. *Curr. Biol.* **11**: 941-950.

Beanan, M.J. and S. Strome. 1992. Characterization of a germ-line proliferation mutation in C. elegans. *Development* **116**: 755-766.

Bernstein, E., A.A. Caudy, S.M. Hammond, and G.J. Hannon. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**: 295-296.

Broverman, S.A. and P.M. Meneely. 1994. Meiotic mutants that cause a polar decrease in recombination on the X chromosome in Caenorhabditis elegans. *Genetics* **136**: 119-127.

Burge, C.B. and S. Karlin. 1998. Finding the genes in genomic DNA. *Curr. Opin. Struct. Biol.* **8**: 346-354.

Caudy, A.A., M. Myers, G.J. Hannon, and S.M. Hammond. 2002. Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes & Dev.* **16**: 2491-2496.

Chervitz, S.A., L. Aravind, G. Sherlock, C.A. Ball, E.V. Koonin, S.S. Dwight, M.A. Harris, K. Dolinski, S. Mohr, T. Smith, S. Weng, J.M. Cherry, and D. Botstein. 1998. Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**: 2022-2028.

C.elegans Sequencing Consortium. 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium. *Science* **282**: 2012-2018.

Djikeng, A., H. Shi, C. Tschudi, and E. Ullu. 2001. RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. *RNA* **7**: 1522-1530.

Elbashir, S.M., W. Lendeckel, and T. Tuschl . 2001a. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes & Dev.* **15**: 188-200.

Elbashir, S.M., J. Martinez, A. Patkaniowska, W. Lendeckel, and T. Tuschl. 2001b. Functional anatomy of siRNAs for mediating efficient RNAi in Drosophila melanogaster embryo lysate. *EMBO J.* **20**: 6877-6888.

Grishok, A., A.E. Pasquinelli, D. Conte, N. Li, S. Parrish, I. Ha, D.L. Baillie, A. Fire, G. Ruvkun, and C.C. Mello. 2001. Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. *Cell* **106**: 23-34.

Ha, I., B. Wightman, and G. Ruvkun. 1996. A bulged lin-4/lin-14 RNA duplex is sufficient for Caenorhabditis elegans lin-14 temporal gradient formation. *Genes & Dev.* **10**: 3041-3050.

Hall, I.M., G.D. Shankaranarayana, K. Noma, N. Ayoub, A. Cohen, and S.I. Grewal. 2002. Establishment and maintenance of a heterochromatin domain. *Science* **297**: 2232-2237.

Hamilton, A., O. Voinnet, L. Chappell, and D. Baulcombe. 2002. Two classes of short interfering RNA in RNA silencing. *EMBO J.* **21**: 4671-4679.

Hamilton, A.J. and D.C. Baulcombe. 1999. A novel species of small antisense RNA in posttranscriptional gene silencing. *Science* **286**: 950-952.

Hammond, S.C., E. Bernstein, D. Beach, and G.J. Hannon. 2000. An RNA-directed nuclease mediates posttranscriptional gene silencing in Drosophila cells. *Nature* **404**: 293-296.

Hastie, N.D. and J.O. Bishop. 1976. The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* **9**: 761-774.

Hofacker, I.L., M. Fekete, C. Flamm, M.A. Huynen, S. Rauscher, P.E. Stolorz, and P.F. Stadler. 1998. Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucleic Acids Res.* **26**: 3825-3836.

Hofacker, I.L., W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. 1994. Fast folding and comparison of RNA secondary structures. *Monatshefte f. Chemie* **125**: 167-188.

Hofacker, I.L. and P.F. Stadler. 1999. Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comput. Chem.* **15**: 401-414.

Hutvágner, G., J. McLachlan, A.E. Pasquinelli, E. Balint, T. Tuschl, and P.D. Zamore. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. *Science* **293**: 834-838.

Hutvágner, G. and P.D. Zamore. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**: 2056-2060.

Ishizuka, A., M.C. Siomi, and H. Siomi. 2002. A *Drosophila* fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes & Dev.* **16**: 2497-2508.

Ketting, R.F., S.E.J. Fischer, E. Bernstein, T. Sijen, G.J. Hannon, and R.H.A. Plasterk. 2001. Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. *Genes & Dev.* **15**: 2654-2659.

Klahre, U., P. Crete, S.A. Leuenberger, V.A. Iglesias, and F. Meins. 2002. High molecular weight RNAs and small interfering RNAs induce systemic posttranscriptional gene silencing in plants. *Proc. Natl. Acad. Sci. U S A* **99**: 11981-11986.

Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science* **294**: 853-858.

Lagos-Quintana, M., R. Rauhut, J. Meyer, A. Borkhardt, and T. Tuschl. 2003. New microRNAs from mouse and human. *RNA* **9**: 175-179.

Lagos-Quintana, M., R. Rauhut, A. Yalcin, J. Meyer, W. Lendeckel, and T. Tuschl. 2002. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**: 735-739.

Lai, E.C. 2002. MicroRNAs are complementary to 3'UTR motifs that mediate negative post-transcriptional regulation. *Nature Genetics* **30**: 363-364.

Lander E.S., L.M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W Fitz Hugh, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.

Lau, N.C., L.P. Lim, E.G. Weinstein, and D.P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858-862.

Lee, R.C. and V. Ambros. 2001. An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* **294**: 862-864.

Lee, R.C., R.L. Feinbaum, and V. Ambros. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**: 843-854.

Lee, Y., K. Jeon, J.T. Lee, S. Kim, and V.N. Kim. 2002. MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* **21**: 4663-4670.

Lim, L.P., M.E. Glasner, S. Yekta, C.B. Burge, and D.P. Bartel. 2003. Vertebrate microRNA genes. *Science*: in press.

Llave, C., K.D. Kasschau, M.A. Rector, and J.C. Carrington. 2002a. Endogenous and silencing-associated small RNAs in plants. *Plant Cell* **14**: 1605-1619.

Llave, C., Z. Xie, K.D. Kasschau, and J.C. Carrington. 2002b. Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. *Science* **297**: 2053-2056.

Lowe, T.M. and S.R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**: 955-964.

Lowe, T.M. and S.R. Eddy. 1999. A computational screen for methylation guide snoRNAs in yeast. *Science* **283**: 1168-1171.

Martinez, J., A. Patkaniowska, H. Urlaub, R. Luhrmann, and T. Tuschl. 2002. Single-Stranded Antisense siRNAs Guide Target RNA Cleavage in RNAi. *Cell* **110**: 563-574.

McCaskill, J.S. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**: 1105-1119.

Mochizuki, K., N.A. Fine, T. Fujisawa, and M.A. Gorovsky. 2002. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in Tetrahymena. *Cell* **110**: 689-699.

Moss, E.G., R.C. Lee, and V. Ambros. 1997. The cold shock domain protein LIN-28 controls developmental timing in C. elegans and is regulated by the lin-4 RNA. *Cell* **88**: 637-646.

Mourelatos, Z., J. Dostie, S. Paushkin, A. Sharma, B. Charroux, L. Abel, J. Rappsilber, M. Mann, and G. Dreyfuss. 2002. miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes & Dev.* **16**: 720-728.

Nykäken, A., B. Haley, and P.D. Zamore. 2001. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**: 309-321.

Okazaki, Y., M. Furuno, T. Kasukawa, J. Adachi, H. Bono, S. Kondo, I. Nikaido, N. Osato, R. Saito, H. Suzuki, et al. 2002. Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* **420**: 563-573.

Olsen, P.H. and V. Ambros. 1999. The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. *Dev. Biol.* **216**: 671-680.

Park, W., J. Li, R. Song, J. Messing, and X. Chen. 2002. CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana. Cur. Biol.* **12**: 1484-1495.

Parrish, S., J. Fleenor, S. Xu, C. Mello, and A. Fire. 2000. Functional anatomy of a dsRNA trigger: differential requirement for the two trigger strands in RNA interference. *Mol. Cell* **6**: 1077-1087.

Pasquinelli, A.E., B.J. Reinhart, F. Slack, M.Q. Martindale, M. Kuroda, B. Maller, A. Srinivasan, M. Fishman, D. Hayward, E. Ball, B. Degnan, P. Muller, J. Spring, J. Finnerty, J. Corbo, M. Levine, P. Leahy, E. Davidson, and G. Ruvkun. 2000. Conservation across animal phylogeny of the sequence and temporal regulation of the 21 nucleotide *let-7* heterochronic regulatory RNA. *Nature* **408**: 86-89.

Pickford, A.S., C. Catalanotto, C. Cogoni, and G. Macino. 2002. Quelling in Neurospora crassa. *Adv. Genet.* **46**: 277-303.

Reinhart, B.J. and D.P. Bartel. 2002. Small RNAs correspond to centromere heterochromatic repeats. *Science* **297**: 1831.

Reinhart, B.J., F.J. Slack, M. Basson, J.C. Bettinger, A.E. Pasquinelli, A.E. Rougvie, H.R. Horvitz, and G. Ruvkun. 2000. The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* **403**: 901-906.

Reinhart, B.J., E.G. Weinstein, M.W. Rhoades, B. Bartel, and D.P. Bartel. 2002. MicroRNAs in plants. *Genes & Dev.* **16**: 1616-1626.

Rhoades, M.W., B.J. Reinhart, L.P. Lim, C.B. Burge, B. Bartel, and D.P. Bartel. 2002. Prediction of plant microRNA targets. *Cell* **110**: 513-520.

Rivas, E., R.J. Klein, T.A. Jones, and S.R. Eddy. 2001. Computational identification of noncoding RNAs in E. coli by comparative genomics. *Curr. Biol.* **11**: 1369-1373.

Schwarz, D.S., G. Hutvágner, B. Haley, and P.D. Zamore. 2002. Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol. Cell* **10**: 537-548.

Slack, F.J., M. Basson, Z. Liu, V. Ambros, H.R. Horvitz, and G. Ruvkun. 2000. The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. *Mol. Cell* **5**: 659-669.

Sugimoto, N., S. Nakano, M. Katoh, A. Matsumura, H. Nakamuta, T. Ohmichi, M. Yoneyama, and M. Sasaki. 1995. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* **34**: p11211-11216.

Sulston, J.E., E. Schierenberg, J.G. White, and J.N. Thomson. 1983. The embryonic cell lineage of the nematode Caenorhabditis elegans. *Dev. Biol.* **100**: 64-119.

Tang, G., B.J. Reinhart, D.P. Bartel, and P.D. Zamore. 2003. A biochemical framework for RNA silencing in plants. *Genes & Dev.* **17**: 49-63.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673-4680.

Vance, V. and H. Vaucheret. 2001. RNA silencing in plants - defense and counterdefense. *Science* **292**: 2277-2280.

Volpe, T., C. Kidner, I. Hall, G. Teng, S. Grewal, and R. Martienssen. 2002. Heterochromatic silencing and histone H3 lysine 9 methylation are regulated by RNA interference. *Science* **297**: 1833-1837.

Wassarman, K.M., F. Repoila, C. Rosenow, G. Storz, and S. Gottesman. 2001. Identification of novel small RNAs using comparative genomics and microarrays. *Genes & Dev.* **15**: 1637-1651.

Wightman, B., I. Ha, and G. Ruvkun. 1993. Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell* **75**: 855-862.

Zamore, P.D., T. Tuschl, P.A. Sharp, and D.P. Bartel. 2000. RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. *Cell* **101**: 25-33.

Zilberman, D., X. Cao, and S.E. Jacobsen. 2003. ARGONAUTE4 Control of Locus-Specific siRNA Accumulation and DNA and Histone Methylation. *Science* **9**: Science online.

Zuker, M. 1994. Prediction of RNA secondary structure by energy minimization. *Methods Mol. Biol.* **25**: 267-294.

Table 1. Cloning frequency and MiRscan scores of *C. elegans* miRNAs. 3423 clones from logarithmically growing mixed-stage worms and worms from the indicated stages or mutant (dauer, starved L1, and *him-8*) represented 79 different miRNAs (and 80 different miRNA genes, because the miR-44/45 miRNA appears to be encoded at two loci). Genes not represented in the set of ~36,000 stem-loops did not receive scores (NS). Note that the previously reported "miR-68" clone is not included. This RNA was not detected on Northern blots and neither it nor its predicted precursor appears to be conserved in another species. Accordingly, it is now classified as an endogenous siRNA. Two other *C. elegans* loci previously thought to encode miRNAs (*mir-69* and *mir-89*) also do not satisfy the current criteria for classification as miRNA genes (Ambros et al. 2003) and were not considered during the course of this study. One previously reported gene, *mir-88*, was not represented in our set of sequenced clones, but is detected on Northern blots as a ~22 nt RNA (V. Ambros, personal communication) and thus satisfies the current criteria for classification as an miRNA gene.

miRNA	MiR-scan Score	Mixed stage	Dauer	Starved L1	him-8	Total
let-7 RNA	13.8	15	0	0	2	17
lin-4 RNA	15.8	48	46	4	27	125
miR-1	14.7	43	17	7	9	76
miR-2	6.2	138	46	20	9	213
miR-34	14.1	13	25	5	9	52
miR-35	14.4	23	0	1	2	26
miR-36	14.6	21	0	1	5	27
miR-37	9.6	8	0	1	2	11
miR-38	8.9	10	0	1	0	11
miR-39	9.5	11	0	0	1	12
miR-40	15.4	12	0	4	2	18
miR-41	12.0	2	0	0	0	2
miR-42	9.5	10	4	3	1	18
miR-43	17.5	8	1	9	0	18
miR-44/45	16.6/17.4	22	3	3	4	32
miR-46	11.3	14	11	9	3	37
miR-47	16.5	19	7	4	5	35
miR-48	12.0	52	1	0	8	61
miR-49	13.1	1	0	1	1	3
miR-50	14.6	10	16	5	1	32
miR-51	12.0	16	5	2	2	25
miR-52	11.6	287	70	18	29	404
miR-53	12.4	20	6	1	4	31
miR-54	9.4	49	40	9	13	111
miR-55	13.8	47	32	16	15	110
miR-56	NS	40	16	9	6	71
miR-57	12.1	31	11	8	3	53
miR-58	17.5	181	51	27	31	290
miR-59	18.5	1	0	0	0	1
miR-60	14.1	20	6	3	7	36
miR-61	13.7	8	5	1	3	17
miR-62	15.1	4	4	6	0	14
miR-63	NS	7	1	0	1	9
miR-64	NS	11	4	8	3	26
miR-65	7.4	22	7	3	2	34
miR-66	NS	68	25	6	7	106
miR-67	16.8	3	0	0	0	3
miR-70	11.6	11	8	3	6	28
miR-71	17.9	53	72	23	22	170
miR-72	NS	49	22	10	9	90
miR-73	11.3	13	7	1	1	22
miR-74	17.9	35	12	6	7	60
miR-75	12.6	14	3	2	2	21
miR-76	14.9	1	2	6	3	12
miR-77	14.2	17	3	0	2	22
miR-78	NS	5	1	1	0	7
miR-79	14.2	14	3	3	3	23
miR-80	17.1	121	27	20	17	185
miR-81	18.8	32	24	6	12	74
miR-82	16.3	36	12	6	11	65
miR-83	15.2	12	12	2	8	34
miR-84	−3.3	12	2	1	4	19
miR-85	17.5	10	0	0	12	22
miR-86	16.3	46	57	30	17	150
miR-87	16.7	1	0	0	0	1
miR-88	−7.9					0
miR-90	14.0	5	37	14	9	65
miR-124	15.7	7	16	7	5	35
miR-228	17.5	1	13	8	3	25
miR-229	NS	2	1	0	0	3
miR-230	16.8	0	0	0	1	1
miR-231	14.1	1	2	0	0	3
miR-232	13.8	4	7	2	1	14
miR-233	16.4	1	8	4	0	13
miR-234	14.3	0	0	1	0	1
miR-235	1.9	5	21	1	8	35
miR-236	16.8	3	6	2	1	12
miR-237	11.9	3	0	0	0	3
miR-238	14.0	0	4	1	0	5
miR-239a	12.7	4	0	0	1	5
miR-239b	13.6					0
miR-240	12.5	0	0	0	1	1
miR-241	14.9	7	0	0	3	10
miR-242	9.9	0	0	1	1	2
miR-243	NS	1	0	1	0	2
miR-244	13.4	0	2	5	0	7
miR-245	13.8	0	1	0	0	1
miR-246	12.8	0	0	0	1	1
miR-247	NS	0	2	0	0	2
miR-248	14.6	0	2	0	0	2
miR-249	13.7	0	2	1	0	3
miR-250	18.4					0
miR-251	15.5					0
miR-252	17.7					0
miR-253	16.9					0
miR-254	15.7					0
miR-255	16.4					0
Total clones		1821	851	363	388	**3423**

Table 2. Newly identified *C. elegans* miRNA genes. For predicted stem-loop precursors, see the supplemental on-line material available at www.genesdev.org. Genes were identified and validated as indicated in the ID method column: MS, candidate gene had high MiRscan score (Table 1); C, miRNA was cloned and sequenced (Table 1); N, expression of the mature miRNA was detectable on Northern blots; D, the miRNA stem-loop precursor was detected on Northern blots and enriched in RNA from *dcr-1* animals, but the mature miRNA was not detected; PCR, targeted PCR amplification and sequencing detected the miRNA in a library of *C. elegans* small RNAs; H, the locus was closely homologous to that of a validated miRNA. For the miRNAs cloned and sequenced, some miRNAs were represented by clones of different lengths, due to heterogeneity at the miRNA 3′ terminus. The observed range in length is indicated, and the sequence of the most abundant length is shown. For the RNAs that have not been cloned, the 5′ terminus was determined by the PCR assay, but the 3′ terminus was not determined. For *mir-250*, *mir-251*, and *mir-252*, the length of the miRNA sequence shown was inferred from the Northern blots; for other miRNAs not cloned, the length was not determined (n.d.). For *mir-254*, the PCR assay detected ~22-nt RNAs from both sides of the fold-back, representing both the miRNA and the miRNA*. Their relative positions within the precursor suggest that the RNA from the 5′ arm is 22 nt and the RNA from the 3′ arm is 23 nt. The RNA from the 3′ arm was chosen as the miRNA because of its similarity to the human miR-19 gene family. The miR-255 gene is known only as the precursor, a conserved stem-loop with Dicer-dependent processing (Fig 2b). Comparison to *C. briggsae* shotgun traces from the *C. briggsae* Sequencing Consortium (obtained from www.ncbi.nlm.nih.gov) revealed miRNA orthologs with 100% sequence identity (+++) and potential orthologs with >90% (++) and >75% (+) sequence identity. To indicate the genomic loci of the genes, the chromosome (Chr.), distance to nearest annotated gene, and the orientation relative to that gene, sense (s) or antisense (as) are specified.

miRNA gene	ID method	miRNA sequence	miRNA length (nt)	*C. briggsae* homology	Fold-back arm	Chr.	Distance to nearest gene		
mir-124	MS, C, N	UAAGGCACGCGGUGAAUGCCA	21	+++	3'	IV	within intron of	C29E6.2	(s)
mir-228	MS, C, N	AAUGGCACUGCAUGAAUUCACGG	21-24	+++	5'	IV	0.2 kb downstream of	T12E12.5	(as)
mir-229	C, N	AAUGACACUGGUUAUCUUUUCCAUCG	25-27	–	5'	III	0.4 kb upstream of	*mir-64*	(s)
mir-230	MS, C, N	GUAUUAGUUGUGCGACCAGGAGA	23	++	3'	X	0.4 kb downstream of	F13D11.3	(as)
mir-231	MS, C, N	UAAGCUCGUGAUCAACAGGCAGAA	23-24	++	3'	III	10.4 kb upstream of	*lin-39*	(s)
mir-232	C, N	UAAAUGCAUCUUAACUGCGGUGA	23-24	+++	3'	IV	1.1 kb downstream of	F13H10.5	(as)
mir-233	MS, C, N	UUGAGCAAUGCGCAUGUGCGGGA	19-23	+++	3'	X	within intron of	W03G11.4	(s)
mir-234	MS, C, N	UUAUUGCUCGAGAAUACCCUU	21	+++	3'	II	1.5 kb downstream of	Y54G11B.1	(as)
mir-235	C, N	UAUUGCACUCUCCCCGGCCUGA	22	+	3'	I	0.6 kb upstream of	T09B4.7	(s)
mir-236	MS, C, N	UAAUACUGUCAGGUAAUGACGCU	21-25	+++	3'	II	0.3 kb downstream of	C52E12.1	(as)
mir-237	C, N	UCCCUGAGAAUUCUCGAACAGCUU	23-24	+	5'	X	3.4 kb upstream of	F22F1.2	(as)
mir-238	MS, C, N	UUUGUACUCCGAUGCCAUUCAGA	21-23	++	3'	III	2.0 kb upstream of	*mir-80*	(s)
mir-239a	C, N	UUUGUACUACACAUAGGUACUGG	22-23	++	5'	X	6.0 kb upstream of	C34E11.1	(s)
mir-239b	H	UUUGUACUACACAAAAGUACUGG	n.d.	++	5'	X	7.0 kb upstream of	C34E11.1	(s)
mir-240	C, N	UACUGGCCCCCAAAUCUUCGCU	22	++	3'	X	1.7 kb upstream of	C39D10.3	(s)
mir-241	MS, C, N	UGAGGUAGGUGCGAGAAAUGA	21	++	5'	V	1.8 kb upstream of	*mir-48*	(s)
mir-242	C, N	UUGCGUAGGCCUUUGCUUCGA	21	++	5'	IV	0.9 kb downstream of	*nhr-78*	(as)
mir-243	C, N	CGGUACGAUCGCGGCGGGAUAUC	22-23	–	3'	IV	1.0 kb upstream of	R08C7.1	(s)
mir-244	C, N	UCUUUGGUUGUACAAAGUGGUAUG	23-25	+++	5'	I	1.6 kb downstream of	T04D1.2	(as)
mir-245	C, N	AUUGGUCCCCUCCAAGUAGCUC	22	+++	3'	I	1.9 kb downstream of	F55D12.1	(s)
mir-246	C, N	UUACAUGUUUCGGGUAGGAGCU	22	++	3'	IV	0.4 kb downstream of	ZK593.8	(s)
mir-247	C, N	UGACUAGAGCCUAUUCUCUUCUU	22-23	–	3'	X	1.9 kb upstream of	C39E6.2	(as)
mir-248	MS, C, N	UACACGUGCACGGAUAACGCUCA	23	++	3'	X	within intron of	AH9.3	(s)
mir-249	C	UCACAGGACUUUUGAGCGUUGC	22-23	++	3'	X	2.7 kb upstream of	Y41G9A.6	(s)
mir-250	MS, N, PCR	UCACAGUCAACUGUUGGCAUGG	~22	++	3'	V	0.1 kb downstream of	*mir-61*	(s)
mir-251	MS, N, PCR	UUAAGUAGUGGUGCCGCUCUUAUU	~24	+++	5'	X	0.2 kb downstream of	F59F3.4	(as)
mir-252	MS, N, PCR	UAAGUAGUAGUGCCGCAGGUAAC	~23	+++	5'	II	1.8 kb downstream of	VW02B12L.4	(as)
mir-253	MS, D, PCR	CACACCUCACUAACACUGACC	n.d.	++	5'	V	within intron of	F44E7.5	(s)
mir-254	MS, D, PCR	UGCAAAUCUUUCGCGACUGUAGG	n.d.	++	3'	X	within intron of	ZK455.2	(s)
mir-255	MS, D	–	n.d.				1.5 kb upstream of	F08F3.9	(as)

Figure Legends

Figure 1. Criteria used by MiRscan to identify miRNA genes among aligned segments of two genomes. (A) The seven components of the MiRscan score for *mir-232* of *C. elegans/C. briggsae*. These components are annotated in the context of the MiRscan prediction for *mir-232*, with the residues of the predicted miRNA circled in pink, which match the residues of the validated miRNA (Table 2), circled in green. In parenthesis are the scores for each component, which were added together to give the total score of 13.9. MiRscan predictions are visualized within the consensus *C. elegans/C. briggsae* secondary structure, as generated using ClustalW (Thompson et al. 1994) and Alidot (Hofacker and Stadler 1999). Shown is the *C. elegans* sequence with residues colored to indicate conserved sequence and pairing potential. Residues conserved in *C. briggsae* are red; residues that vary while maintaining their predicted paired or unpaired state are blue (with variant residues that maintain pairing also circled in black); residues that maintain neither sequence nor pairing are in gray. (B) Estimated relative importance of each MiRscan criterion. Estimates were based on the relative entropy between the training set of 50 previously identified nematode miRNAs and the background set of ~36,000 potential stem-loops. Because pairing and conservation were used to identify the potential stem-loops, the total contributions of these types of criteria for distinguishing miRNA genes from non-protein-coding genomic sequence were underestimated. Likewise, the total contribution of the distance from the loop was underestimated because only those candidates 9–2 bp from the loop were evaluated.

Figure 2. Computational identification of miRNA genes. (A) The distribution of MiRscan scores for 35,697 *C. elegans* sequences that potentially form stem-loops and have loose conservation in *C. briggsae*. Note that the Y-axis is discontinuous so that the scores of the 50 previously reported miRNA genes that served as the training set for MiRscan can be more readily seen (red). Scores for these 50 genes were jackknifed to prevent inflation of their values because of their presence in the training set. (B) An expanded view of the high-scoring tail of the

distribution. This view captures 49 of the 50 genes of the training set (red). The median score of

the 58 previously reported miRNA loci that satisfy the current criteria for designation as miRNA

genes (Ambros et al. 2003) is 13.9 (green arrow). Note that this median score was the midpoint

between the scores of the 29th and the 30th highest scoring loci of the 50-member training set; i.e.,

it was designated the median score after including the eight previously reported miRNA genes

that were not in the training set because they were lost during the identification of conserved

hairpins, usually because they lacked sufficient *C. briggsae* homology. Scores of genes validated

by cloning are indicated (yellow), as are scores of 6 genes that have not yet been cloned but were

verified by Northern analysis (purple). (C) Examples of miRNA genes identified by MiRscan

with the Northern blots that served to validate them. Stem-loops were annotated as in Fig. 1A,

except the DNA rather than RNA sequence is depicted. The Northern blots show analysis of

RNA from either wild-type (N2) or *dcr-1* worms, isolated using either our standard protocol

(Std.) or with an additional polyethylene glycol precipitation step to enrich for small RNAs

(Enr.). Homozygous worms of the *dcr-1* population have reduced Dicer activity, increasing the

level of miRNA precursors (e.g., miR-250-L and miR-255-L), which facilitated the validation of

miRNA loci, especially those for which the mature miRNA was not detected (e.g., miR-255).

The miR-250 stem-loop shown received a MiRscan score of 14.7. The mir-250 reverse

complement received an even greater score of 18.4, but was not detected by Northern analysis.

Thus, the predicted *mir-250* gene was assigned the score of the higher scoring, though incorrect,

alternative stem-loop (Table 1, Fig. 2B).

Figure 3. Alignments of *C. elegans* and human miRNA sequences that can be grouped together

in families. Human miRNAs (*Hs*) are those identified in human cells (Lagos-Quintana et al.

2001; Mourelatos et al. 2002) or are orthologs of miRNAs identified in other vertebrates (Lagos-

Quintana et al. 2002; Lagos-Quintana et al. 2003; Lim et al. 2003).

Figure 4. Expression of *C. elegans* miRNAs during larval development. Total RNA was analyzed from mixed-stage N2 worms (M), embryos (E), larval stages (L1, L2, L3, L4), adults (A), *glp-4(bn2)* adults (G), N2 dauers (D), mixed-stage *him-8(e1489)* worms (H), and N2 starvation-arrested L1 larvae (sL1). Intense signals are represented as black rectangles, and faint signals are represented as gray rectangles. Of the 87 *C. elegans* miRNAs identified, six could not be detected on developmental Northerns (miR-41, miR-78, miR-249, miR-253, miR-254, and miR-255). (A) miRNAs constitutively expressed throughout nematode development. (B) stRNAs, *lin-4* and *let-7*, and similarly expressed miRNAs, which commence expression during larval development and remain expressed through adulthood. (C) miRNAs with discontinuous developmental expression patterns. (D) Northern analysis of miRNAs with enhanced expression in the dauer stage, as marked by the arrow. To control for loading, the blot used for both miR-234 and miR-248 and the blot used for miR-247 were re-probed for the U6 snRNA (U6). Quantitation with a phosphorimager showed that the lane-to-lane variation in U6 signal was as great as three-fold. Normalizing to the U6 signal, the miR-248 signal was four-fold greater in dauer than in most other stages, except for *glp-4* adults, where it was two-fold greater, while the miR-234 signal was highest in dauer and L1, with a signal in these stages about two-fold greater than the average of the other stages. (E) Northern analysis of the *lin-4* RNA and its paralog, mir-237.

Figure 5. Quantitative analysis of miRNA expression. (A) Northern blot used to quantify the abundance of miR-66. RNA prepared from the wild-type (N2) mixed-stage worms used in cloning and from *glp-4(bn2)* young adult worms were run in duplicate with a concentration course of synthetic miRNA standard. The signal from the standard did not change when total RNA from HeLa cells replaced *E. coli* tRNA as the RNA carrier, showing that the presence of other miRNAs did not influence membrane immobilization of the miRNA or hybridization of the probe. (B) Standard curve from quantitation of miR-66 concentration course. The best fit to the

data is a line represented by the equation $y = 3.3x^{0.96}$ ($R^2 = 0.99$). Interpolation of the average signal in the *glp-4* lanes indicates that the *glp-4* samples contain 240 pg of miR-66 (dashed lines). (C) Molecular abundance of miRNAs and U6 snRNA. Amounts of the indicated RNA species in the *glp-4* samples were determined as shown in panels A and B. The average number of molecules per cell was then calculated considering the number of animals/cells used to prepare the sample and the yield of a radiolabeled miRNA spiked into the preparation at an early stage of RNA preparation. Analogous experiments were performed to determine the amounts of the indicted human miRNAs in HeLa RNA samples. (D) Correlation between miRNA molecular abundance and cloning frequency. The number of molecules in the mixed-stage RNA samples were determined as described for the *glp-4* samples and then plotted as a function of the number of times the miRNAs were cloned from this mixed-stage population (Table 1). The line is best fit to the data and is represented by the equation $y = 0.32x$ ($R^2=0.78$).

Figure 6. miRNA (red) and miRNA* (blue) sequences within the context of their predicted fold-back precursors. The number of sequenced clones is shown in parentheses. Colored residues are those for the most frequently cloned species. There was 3′ heterogeneity among the sequenced clones for some miRNA*s and most miRNAs. Heterogeneity at the 5′ terminus was not seen among the sequenced clones for the miRNA*s and was rare among those for the miRNAs; when it occurred it was not observed for more than one of the many clones representing each miRNA.

Figure 7. Plot illustrating the absence of a correlation between the MiRscan score of a cloned miRNA and the number of times that miRNA was cloned and sequenced. Ten of 80 cloned loci of Table 2 were not scored (left panel) because potential homologs of these genes were not identified among the available *C. briggsae* sequencing reads.

A

B

C

Score: 14.7

(Reverse
complement
score: 18.4)

Std.   Enr.

miR-250-L

miR-250

Score: 16.4

Std.   Enr.

miR-255-L

lin-4 family

Hs miR-125b-1
Hs miR-125b-2
Ce lin-4
Ce miR-237

let-7 family

Hs let-7d
Hs let-7e
Hs let-7a-1
Hs let-7a-2
Hs let-7a-3
Hs let-7a-4
Ce let-7
Hs let-7f-1
Hs let-7f-2
Hs miR-98
Hs let-7g
Hs let-7i
Hs let-7b
Hs let-7c
Hs miR-196-1
Hs miR-196-2
Ce miR-84
Ce miR-48
Ce miR-241

mir-1 family

Hs miR-1b
Hs miR-1d
Ce miR-1
Hs miR-206

mir-9 family

Hs miR-9-1
Hs miR-9-2
Ce miR-244

mir-10 family

Hs miR-100-1
Hs miR-100-2
Hs miR-99b
Ce miR-57
Hs miR-10a
Hs miR-10b
Hs miR-99a
Ce miR-51

mir-19 family

Hs miR-19a
Hs miR-19b-1
Hs miR-19b-2
Ce miR-254

mir-25 family

Hs miR-92-1
Hs miR-92-2
Ce miR-235
Hs miR-25-1
Hs miR-25-2
Hs miR-32

mir-29 family

Hs miR-29b-1
Hs miR-29b-2
Hs miR-29b-3
Hs miR-29c
Hs miR-29a-1
Hs miR-29a-2
Ce miR-83

mir-31 family

Ce miR-72
Hs miR-31
Ce miR-73

mir-34 family

Ce miR-34
Hs miR-34
Hs miR-122a

mir-50 family

Ce miR-62
Ce miR-50
Hs miR-190
Ce miR-90

mir-74 family

Hs miR-185
Ce miR-74

mir-76 family

Ce miR-76
Hs miR-187

mir-79 family

Ce miR-79
Hs miR-131
Ce miR-75

mir-80 family

Ce miR-81
Ce miR-82
Ce miR-80
Hs miR-143

mir-105 family

Hs miR-105-1
Hs miR-105-2
Ce miR-232

mir-124 family

Hs miR-124a-1
Hs miR-124a-2
Hs miR-124a-3
Ce miR-124
Ce miR-228
Hs miR-183

mir-133 family

Hs miR-133a-1
Hs miR-133a-2
Hs miR-133b
Ce miR-245

mir-137 family

Ce miR-234
Hs miR-137

mir-141 family

Ce miR-236
Hs miR-141

mir-193 family

Ce miR-240
Hs miR-193

mir-220 family

Ce miR-253
Hs miR-220

Figure 4. Lim et al.

A

miR-66 standard (pg)

B

C

D

miR-244* (1)

miR-90 (41)

miR-83 (34)

miR-75 (21)

miR-71* (3)

miR-56 (71)

miR-55 (110)

miR-52* (2)

miR-45 (32)

miR-42 (18)

miR-34* (17)

miR-2 (213)

miR-1 (76)

MicroRNA Stability Determined In an Inducible Cell Line

ABSTRACT: MicroRNAs are abundantly expressed, ~22 nt long non-coding RNAs with known and potential regulatory functions. The molecular abundance of microRNAs (miRNAs) in the animal cell is very high: the number of molecules of some individual miRNAs equals the number of molecules of the U6 snRNA of the spliceosome. What accounts for the high molecular abundance of miRNAs: ample transcription or prolonged stability? To provide reagents for addressing this and other questions, we generated stable, inducible cell lines allowing for induction or repression of miRNA expression upon doxycycline (Dox) addition, with induction and repression levels exceeding 10 fold. We measured the decay and induction kinetics of the well-conserved and tissue-specific miRNAs, miR-1 and miR-124, and propose that miRNA half lives are at least 24 hours and could possibly be much longer. The high stability of miRNAs observed in this study suggests that miRNA (and siRNA) clearance is most dependent on dilution of the miRNA, either by cell death or cell division, and that active processes would be required to achieve a more rapid diminution of miRNA function.

Abbreviations: bp, base pair; Dox, doxycycline; miRNA, microRNA; PCR, polymerase chain reaction; Tet, tetracycline; tTA, tetracycline Transcriptional Activator; rtTA, reverse tetracycline Transcriptional Activator,

INTRODUCTION

The details of microRNA biogenesis, function, and mechanism have recently expanded in rapid fashion. The downstream products and intermediates of miRNA gene transcription have been elucidated, and several novel protein factors and their mechanisms of action have recently been identified (for review, see [1]). miRNA biogenesis proceeds in several discreet steps. Although miRNA transcriptional regulation is unclear, evidence suggests that miRNAs are transcribed as long primary transcripts >100-nt long , which subsequently are processed by the nuclear RNAse III, Drosha, into ~60-nt hairpin-like precursor miRNAs (pre-miRNAs) [2-5]. These hairpin precursors are then transported into the cytoplasm by a well-conserved export pathway involving Exportin-5 and Ran-GTP [6-8]. Once in the cytoplasm, pre-miRNAs are recognized and cleaved by a cytoplasmic RNase III, Dicer, to yield ~19-nt long duplexes with additional 2-nt, 3' overhangs [9-12]. Finally, an unidentified helicase discriminates and chooses the energetically weaker end of the miRNA duplex for unwinding, and thus directs the asymmetrical loading of a single, mature miRNA strand into the microRNA ribonucleoprotein (miRNP) complex [13, 14].

The exact function of the miRNA within the miRNP is currently thought to only involve the guiding of the miRNP to target mRNAs by complementarity between the miRNA and the 3' untranslated region of the mRNA, for example with *lin-4* RNA – *lin-14* mRNA and *let-7* RNA – *lin-41* mRNA [15-19]. When the miRNP interacts with the target mRNA, it can prompt either inhibition of productive translation or mRNA cleavage to downregulate protein expression. What is classically observed for downregulation of small temporal RNA targets, *lin-14* and *lin-41* is negligible destabilization of the mRNA [16, 18], but little else is known about the mechanism other than association of the miRNP with polyribosomes, suggesting that the

silencing event occurs after translation initiation [20-22]. The second activity seen recently for miRNP-mRNA interactions is mRNA degradation caused by cleavage of the mRNA within the complementary site, at a position that is 10-base pairs away from the 5' end of the miRNA [23-28]. Although the manner by which the miRNP decides between these two activities is not completely understood, the degree of complementarity between the miRNA and the target mRNA appears to contribute importantly – near-complete complementarity appears necessary for mRNA cleavage activity, while the translational repression activity can tolerate more mismatches and bulges in the interaction [23, 29-31].

The biogenesis and mechanistic functions of miRNAs share many similarities with small interfering RNAs (siRNAs) of the RNA interference (RNAi) pathway. Despite differences in their genomic origins and primary transcript structure, both miRNAs and siRNAs transiently exist as 19-nt duplexes with 2-nt 3' overhangs, as a result of processing from Dicer-like endonucleases, and either miRNA or siRNA duplexes are unwound asymmetrically by perhaps the same helicase [9-14, 32]. The RNA-induced silencing complex (RISC) is known to utilize siRNAs for target mRNA recognition and cleavage, a hallmark of RNAi [33-35]. Biochemical and immunoprecipitation purifications of RISC has demonstrated the association of miRNAs, while reciprocal experiments with immunoprecipitated miRNPs have demonstrated RISC activity, indicating functional redundancy and factors intimately shared between the miRNA pathway and the RNAi pathway [23, 36, 37]. Although evidence points to convergence and intersection between miRNA and siRNA pathways, distinctions are clearly noted in different Dicer endonucleases playing different roles in miRNA or siRNA processing [38-40].

While our understanding of miRNA biogenesis and function has progressed, less is known about miRNA metabolism and stability. Interest in the stability of miRNAs arose from

our determinations of the high molecular abundance miRNAs in animal cells. In the nematode, *Caenorhabditis elegans*, the steady-state number of molecules of particular miRNAs range from ~1,000 to >50,000 , and similar quantities are seen for miRNAs in HeLa cells [12]. Numerous miRNAs have been cloned from nematode and HeLa cells (>80 in nematodes, >40 in HeLa cells)[41-44]. Thus, miRNPs represent one of the more abundant ribonucleoprotein complexes in the cell. Two models could explain this high molecular abundance: (1) miRNAs and miRNPs might be rapidly and constantly synthesized at very high levels to accommodate fast RNA turnover, or (2) miRNPs are very stable, so moderate yet constant miRNA/miRNP synthesis results in high steady-state levels.

To distinguish between these two models, we determined the stability of two well-conserved animal miRNAs in HeLa cells. By placing miRNA expression under inducible regulation, we could examine the kinetics of miRNA decay and induction by monitoring miRNA levels directly after Dox addition, and from our analysis we estimate a lower limit of miRNA stability. We also present data on the minimal rate of pre-miRNA turnover.

RESULTS

Inducible and Repressible Expression of miRNAs in HeLa Cells

Classical methods for studying RNA turnover have utilized chemical agents that globally shut off transcription, but pleiotropic damages to other cellular processes often confound interpretation of the results. A non-biased and contemporary method to determine RNA turnover rates is to place the gene under inducible regulation by an induction agent that exerts little effect on other cellular processes [45]. We established an inducible expression system for miRNAs in a HeLa cell line, choosing HeLa cells as our model system because of its ease in

handling and its well-characterization of endogenous miRNA levels [12, 43]. Two model

miRNAs were chosen for this study, human *mir-1d* and *mir-124a*, because the primary, mature

sequences of these two miRNAs are well-conserved in other animals, and both are not expressed

in HeLa cells [43]. The Tetracycline (Tet) Inducible System was employed to regulate

expression of these miRNAs because the Tet system is well validated for inducible expression in

mammalian cells [46].

To facilitate robust expression of the mature miRNA, we designed expression constructs

using ~500-bp long genomic fragments which have been shown to provide the appropriate

primary transcript (pri-miRNA) needed for Drosha processing and other miRNA processing

steps [2-5]. Expression plasmids were then stably transfected into HeLa cells that already

contained the chimeric transcriptional activator VP16 – Tet repressor or VP16 – reverse Tet

repressor, creating cells that would responsively transcribe or stop transcription of a pri-miRNA

upon addition of Dox, respectively (Figure 1). A miRNA Invader® Assay was employed to

screen and isolate desired clones, and generally 1 out of 24 screened clones exhibited adequate

induction or repression of miRNA expression. Four cell line clones were chosen for further

characterization, and each is named according to the miRNA it expresses and the mode of

regulation (Tet-On cells turn on miRNA expression upon Dox addition, while Tet-Off cells turn

off miRNA expression upon Dox addition): mir-1d^{ON} (line C2), mir-1d^{OFF} (line A1), mir-

124a^{ON} (line A2), and mir-124a^{OFF} (line A1). Initial and subsequent Northern blot analysis of

total RNA from these cells revealed mature miRNAs with the same heterogeneity pattern as

endogenous miR-1d and miR-124 from tissues and other cell lines, suggesting that the

expression constructs are recapitulating endogenous expression of these miRNA genes [20, 22,

47].

An 8-day time course was carried out to evaluate the range of miRNA induction and repression and to assess the consistency of the basal rate of un-induced miRNA expression. Constant amounts of total RNA from time points either treated with Dox or mock-treated were analyzed by Northern analysis and miRNA levels were quantified and normalized against U6 snRNA levels (Figures 2 and 3). Both mir-1d^{ON} and mir-124a^{ON} cells exhibited progressive increases in miRNA expression after Dox addition, although miRNA levels dropped at the 192-hour time point, making it difficult to determine the final steady-state level (Figures 2A and 3A). At and prior to the 192-hour time point the Tet-On cells did reach maximal confluence, which might have affected miRNA transcription and final expression levels. Although there is leaky basal expression of miRNA in both mir-1d^{ON} and mir-124a^{ON} cells during mock treatment, this level remains quite low, and at maximal miRNA expression in the Dox-treated cells there is a 21-fold and 14-fold induction level over the average mock baseline in mir-1d^{ON} and mir-124a^{ON} cells, respectively.

Tet-Off cells expressing miR-1d or miR-124a were cultured for 1 week without Dox to allow for steady-state expression to be reached, and then were treated with Dox to shut off miRNA transcription for the 8-day time course. By comparing the presence of miR-1d and miR-124a precursor signal in the mock-treated cells to the complete lack of pre-miRNA signal in the Dox-treated cells (Figure 2B and 3B), we conclude that miRNA transcription is tightly shut off, with any remaining hairpin precursor is either degraded or driven towards processing into mature miRNA. The basal expression of miR-1d and miR-124a remained high during all mock-treatments, while decay in miRNA signal is quite obvious in cells treated with Dox. Comparing the average baseline miRNA expression of mock-treated cells to miRNA levels in cells after 48 hours of Dox treatment indicate a 6-fold and 12-fold reduction in the mir-1d^{OFF} and mir-124a^{OFF}

cells, respectively. This implied a relatively long half-life for the mature miRNA, and given the doubling rates of the actively dividing HeLa cells, we conclude the apparent decay seen in this time course was predominantly reflecting dilution of the miRNA upon cell divisions (data not shown).

An Upper Limit for the Rate of Precursor Disappearance

To further characterize the responsiveness of the Tet-Inducible System in controlling miRNA expression, we conducted shorter time courses with the Tet-Off cells. While mature miRNA levels remained essentially constant through 12 hours, we observed a sharp decrease in signal for the miR-1d precursor (pre-miR-1d) and the miR-124a precursor (pre-miR-124a) 3 hours after Dox addition (Figure 4A). In fact, disappearance of pre-miR-124a can be observed 40 minutes after Dox addition (Figure 4B), indicating that transport of Dox into the cell and exertion of changes in tTA action on transcription occurs on a fast time scale compared to the time scale for investigating miRNA stability (>12 hours). Other investigators using the Tet-Inducible System in HeLa cells for studying RNA turnover have successfully controlled transcription for determining RNA half-lives under 10 minutes [48]. Although precursor decay by processes unrelated to miRNA biogenesis cannot be ruled out, this short time course does suggest that nucleus-to-cytoplasmic transport and precursor processing by Dicer occurs in a time-frame of under 40 minutes.

Measurement of miRNA Decay

Analyzing the data from the 8-day time courses of miR-1d^{OFF} and miR-124a^{OFF} cells by correcting for the rate of cell division suggested that miRNA stability was significantly longer

than the doubling rate of HeLa cells, which we measured to be 19-21 hours (data not shown). The 8-day time course measurements of miR-124a^{OFF} cells displayed the closest fit to pseudo-first order decay, so we modeled the effect of different cell doubling rates against the measured miR-124a levels. Because the result of these calculations is highly dependent on the doubling rate, we modeled the decay kinetics using a range of cell doubling rates. Even a conservative estimate of a 24 hour cell doubling rate yielded a miRNA half life of greater than 100 hours, supporting the notion that miRNAs are very stable molecules.

A limitation of loading a constant amount of RNA on a standard RNA Northern blot is the ~16 fold miRNA signal dilution after 4 days, resulting in signals barely detectable above the background of the exposure. In an attempt to overcome the signal dilution from cell division, we modified the time course and Northern procedure so that all accumulated total RNA from cell divisions could be loaded in a single lane on the gel (see Materials and Methods). Several independent trials of 4-day time courses using the modified harvesting and Northern procedure were conducted on miR-1d^{OFF} and miR-124a^{OFF} cells, and the direct miRNA signal was quantitated and plotted (Figure 5). The decay rates were determined separately for each trial by fitting single-exponential decay (Figure 5B and 5C and Table 1). The variability in the calculated half-lives between trials reflects the inability to normalize miRNA signal against an internal control, despite attempts to equalize RNA loading and transfer. Nevertheless, all calculated half lives were longer than 24 hours, further supporting the claim that miRNAs are stable.

Slow Kinetics of Steady-State miRNA Induction

Kinetic theory of eukaryotic gene expression states that RNA stability ultimately dictates both the rate of decay and rate of accumulation, so that the half life of RNA decay would also equal the half-maximal time point of steady-state RNA expression after induction of transcription [49, 50]. To see if miRNA stability could be assessed from a system of miRNA induction, we investigated miR-1d and miR-124a expression levels in Tet-On cells constantly cultured at sub-confluent levels and induced with Dox. We observed robust but surprisingly slow accumulation of miRNA levels, normalized to U6 snRNA levels (Figure 6). While two trials indicate a possible steady state at 96 hours (TC-III in Figure 6B and 6E), the other four trials seem to suggest continuing increases in miRNA levels beyond 96 hours. Inspection of early time points also suggests a ~24 hour lag in miRNA accumulation rate after Dox induction (Figure 6B and 6E).

Since miRNA transcription is driven by the potent VP16 transcriptional activator (rtTA) in this inducible system, we expected that maximal transcription rates of miRNA expression would be approached quickly and robustly, as it appears so for inducing luciferase expression [46]. However, examination of the induction of the precursor revealed a complication in these experiments. Pre-miRNA signals were generally more difficult to quantitate over background, but in most cases it appeared that pre-miRNA levels had not reached steady-state expression even at the 96 hour time point (Figure 6C and 6F). This observation seemingly appears inconsistent with the observed decay or processing rate of precursors (Figure 4), however others utilizing the Tet-On system have observed that inducible expression of GFP does not reach steady state after 96 hours [51], and the transcription rate of other mRNAs under inducible control appear to fluctuate and increase during the 24 hour period after Dox addition [52]. One

proposed explanation is that the rtTA is unstable without Dox, thus some time is needed to accumulate enough rtTA in order to then stimulate expression from the stable expression cassette which is likely a multi-copy, illegitimate integrant [53]. Fluctuating amounts of rtTA, apparently independent of Dox concentration, could likely cause fluctuating induction rates, which would explain the difficulty in reaching steady state as well as explain the initial lag in accumulation of miRNA. Future time course trials that extend time points while maintaining subconfluent growth will be needed to reveal the induction kinetics in a clearer fashion.

DISCUSSION

We demonstrate inducible expression of miRNAs in mammalian cells, and we illustrate the utility of this system for investigating questions of miRNA biology. The level of miRNA expression is robust, and this may facilitate our determination of miRNA half-life and accumulation rate. The results presented here have implications for understanding miRNA metabolism, and may additionally relate to the pharmacokinetics of siRNAs used in cell culture.

Characterization of Cells as Tools for Studying miRNA Biology

Misexpression of some miRNAs in plants and animals have resulted in gross phenotypic changes [54-58]. Our inducible miRNA cell lines constitute a misexpression of two miRNAs not usually found in HeLa cells, but visual inspection of cells in the induced and uninduced states after several weeks did not indicate any overt differences (data not shown). We attempted to confirm the functionality of the expressed miRNAs by testing luciferase reporter genes with a 3' UTR that contain a perfectly complementary site for *mir-1* or for *mir-124* [59]. Despite repeated attempts, no significant repression in luciferase activity could be differentiated between induced

or uninduced states. Cells were cultured for at least a week either with or without Dox to insure maximal expression or suppression of miRNAs.

An explanation for the failure of luciferase reporter gene silencing in these cell lines could be that they are not pure clonal populations, either because ring cloning did not effectively select single clones or miRNA-expression is being lost in a subpopulation of cells due to the multi-component nature of the Tet Inducible System. To test if a subclone of cells could be culled for homogenous expression of the miRNA, limiting dilution experiments were conducted, and subclones were screened for repression of the luciferase reporter. Despite repeated attempts, we have not yet successfully refined these cell lines into homogenous populations functional for reporter gene expression (data not shown). Epigenetic instability has been proposed as a mechanism for other occurrences where Tet Inducible lines become heterogeneous [60], but the construction of the miRNA-expressing cell lines followed the suggestions on avoiding this technical difficulty. A possible remedy towards selecting homogenous miRNA-expressing cell lines would be to insert into the population an engineered negative selection marker, like thymidine kinase (TK), which would contain the miRNA binding site. Upon addition of ganciclovir, cells that fail to express the miRNA and allow for TK expression will die from toxicity of TK reacting with ganciclovir, while miRNA-expressing cells will be positively selected.

Despite the heterogeneity of these cell lines, the fact that miRNAs accumulate with proper heterogeneity patterns lends support to the functionality of these miRNAs expressed in the cells. We can also assume that the miRNA-expressing population in these cell lines must not be less than 10% of the entire population based on results from measuring number of miRNA-expressing subclones from limiting dilution experiments (data not shown). Furthermore, we

believe the miRNA expression levels are not saturating in these cells, because we measured some of the miRNA levels against molecular standards, and we calculate an average maximal abundance of 4,200 miRNA molecules per cell (data not shown). Even if actual numbers of miRNA molecules per cell were higher (an order of magnitude), they would still be within the range of other miRNAs expressed in animals [12]. Thus, we assume that the half lives we observe for these two miRNAs are likely to be physiologically relevant.

These inducible cell lines might offer an opportunity to examine other aspects of miRNA biology. For example, the kinetic framework of miRNA biogenesis *in vivo* could be further elucidated by determining the ratio of pri-miRNA to pre-miRNA during a time course of Dox induction to reveal the processing rates of Drosha. The expression of miR-1 and miR-124 in these cells presents an additional system for validating potential predicted mRNA targets for gene downregulation, as has been demonstrated previously [59]. The simple mode of miRNA inducibility coupled with the biochemical tractability of these cell lines should also provide a platform for mechanistic studies in the translational repression of target mRNAs.

Implications of a Long miRNA Half-life

The stability of miRNAs would explain the high steady-state levels in cells. The corollary of this conclusion is that the transcription rate of miRNAs is not so radically higher than of typical mRNAs, which is consistent with hypotheses that miRNA transcription is driven by RNA Polymerase II, although the possibility of RNA Polymerase III transcription remains open [2-5]. The half life of >24 hours seen for miRNAs is rather long when compared to the average 8-hour half lives of most mammalian mRNAs [50]. However, other stable RNAs have also been observed, such as rRNA and the globin mRNA in erythrocytes, which have half lives

greater than 60 hours [61]. The encasement of miRNAs in proteins of the miRNP likely protects

the single-stranded miRNA from cytoplasmic nucleases. Accordingly, the miRNP, and not

necessarily the miRNA itself, appears extremely stable because observations in *Drosophila*

lysate of siRNAs in the RISC indicate near irreversible association between siRNA and RISC,

while exogenous single-stranded siRNAs are rapidly degraded in both *Drosophila* and HeLa cell

lysates within 2 minutes and 20 minutes, respectively [34, 62, 63].

There is much promise in utilizing siRNAs as therapeutic agents to knock down genes

implicated in a disease pathway or as a prophylactic against virus infection [64]. The half lives

of siRNAs in cells could yield insight into their pharmacokinetics. One kinetic analysis of

reporter gene silencing by transfected siRNAs reported a functional stability curve for

transfected siRNAs in the rapidly dividing HeLa cell, with downregulation of reporter gene

levels persisting to 6 days after siRNA transfection [63].   A second kinetic analysis on two

different siRNAs transfected into quiescent cells suggests the detection and activity of some

siRNAs can persist over 20 days [65]. Both studies suggest long lasting effects of RNAi in

transfection experiments, however, the question of *in vivo* siRNA stability still remains open

because nucleic acid metabolism in lipid-based transfection experiments are poorly

characterized, and amounts of siRNAs submitted to each cell may begin in such excess that

physiological siRNA stabilities might be distorted. Our data indicating half lives for miRNAs

exceeding 24 hours suggest that siRNAs loaded within the RISC also possess such long half

lives. Because our data indicate the long physiological stability of miRNAs, we would predict

that the predominant means of miRNA and siRNA clearance is by cell division. In the kinetic

study by Song and colleagues, the siRNA directed at an endogenous mRNA persisted, while the

second siRNA targeting a viral gene not usually expressed in mammalian cells disappeared more

rapidly, implying that presence of a target mRNA stabilized the first siRNA [65]. If this hypothesis holds true, then perhaps in HeLa cells there may be "inadvertent" target mRNAs that might account for stabilizing miR-1 and miR-124 regardless of the fact that these miRNAs are not normally expressed in these cells.

A recent genetic screen in *C.elegans* uncovered the evolutionarily conserved RNase, *eri-1*, which degrades siRNAs and dampens the sensitivity of certain cells to RNAi [66]. Mutants lacking the RNase show enhanced sensitivity to RNAi of neuronal genes and an enduring accumulation of injected siRNAs, while *in-vitro* translated worm and human *eri-1* exhibit *in vitro* siRNase activity [66]. The capacity for *eri-1* mutants to preserve siRNA accumulation intriguingly suggests that ERI-1 may be actively stimulating the turnover of siRNAs and perhaps miRNAs. However, *in vitro* experiments showing no RNase activity of ERI-1 on single-stranded siRNAs and our data indicating a long half life for miRNAs suggests that ERI-1 does not actively turn over siRNAs or miRNAs already incorporated into RISC/miRNP, but ERI-1 might degrade siRNA duplexes and miRNA/miRNA* duplexes that remain unincorporated into silencing complexes (Figure 7). Determining the presence and scope of activity of endogenous HeLa *eri-1* might lend further support to our proposed model.

Why should miRNAs and the miRNP be so stable in nature? Stability facilitates the miRNAs' molecular abundance, and this abundance may be necessary because the predominant mechanism of gene silencing for miRNAs in animals is to inhibit productive translation. While mRNA cleavage by miRNAs is catalytic, translational repression probably involves stoichiometric binding to the mRNA [29, 30]. Conservation of miRNA stability might be questioned in invertebrates, where animal development is considerably faster, and where temporally regulated miRNAs do show complete turnover within 24 hours in the developing

Drosophila embryo [67]. However, the different model systems have informed each other quite well with regards to miRNAs, so such pronounced reductions seen in the *Drosophila* embryo might reflect quick dilution of miRNAs during rapid cell division or cell death during early animal development.

MATERIALS AND METHODS

Construction of Expression Plasmids

Genomic fragments of human mir-1d and mir-124a-1 were amplified by PCR with the following primers: mir-1d primers: 31.24 (sequence: AAAAGGATCCGAGAGATGGATTCAGGGATGG), 30.29 (sequence: CCCCATCGATTGTCTGGTGAGCACTTCCAC); mir-124a-1 primers: 30.52 (sequence: AAAAGGATCCCTCATTGTCTGTGTGATTGG), 30.49 (sequence: CCCCATCGATTCAGCTTCTG-TTTCTCTCCC). Genomic fragments were cloned into the inducible plasmid pTRE-TIGHT (Clontech) to form the plasmids pTRE-m1d and pTRE-m124a.

Construction of Inducible Cell Lines

HeLa cells that stably expressed either the chimeric VP16-Tet repressor (tTA) or the chimeric VP16-reverse Tet repressor (rtTA) were purchased from Clontech, and are referred in the text as Tet-Off cells (expressing tTA) and Tet-On cells (expressing rtTA). Tet-Off and Tet-On cells at 80% confluency in a 10-cm plate were transfected with 2 μg of either pTRE-m1d or pTRE-m124a and 0.2 μg or 0.1μg of a linearized DNA fragment containing the hygromycin resistance marker (Clontech). Two days after transfection, cells were split to multiple 6-well plates and selected for ~3-weeks in media containing 300 μg/ml hygromycin (Tet-Off cells were cultured with the addition of 1μg/ml Dox). Between 24-48 clones per cell line and construct were isolated by ring cloning, and were cultured in the absence or presence of Dox (1μg/ml) for 2-3 days before assaying for miRNA expression.

Clones were screened for miRNA expression by a modified RNA Invader® Assay (Third Wave Technologies) [68]. Cells were directly lysed in culture plates, and lysates were added to Invader® Assay

reactions specific for detecting miR-1d and miR-124a. Fluorescent signals were detected on a Cytofluor

fluorimeter (PerSeptive Biosystems), and clones exhibiting greater than 4 fold induction or repression

were culled and propagated in 100 μg/ml G418 and 100 μg/ml hygromycin.

Time Course Studies

All cell cultures were grown in DMEM containing L-glutamine and glucose (4.5g/L)

(MediaTech) and supplemented with 10% Tet-Approved FBS, Penicillin (10 IU/ ml), and Streptomycin

(10 μg/ml). All time courses were performed in 10-cm plates and initiated by adding Dox to the media at

1μg/ ml. The initial 8-day time courses were conducted by first plating out the night before Dox addition

about $2x10^6$ cells for 12 hour time points, $1x10^6$ cells for 24 hour time points, and $5x10^5$ cells for a third

set of plates that is eventually split 1:4 every two days for the 48 hour, 96 hour, and 192 hour time points.

Cells were cultured between 70 to >95% confluence. In the 8-day long time course, Dox-treated and

mock treated plates reaching time points were subjected to direct cell lysis with Tri-Reagent (Sigma) and

total RNA extraction.

For the 12-hour and 150-minute time courses, $1x10^6$ cells in 6-well dishes were plated the night

before, and Dox was added to the media at successive time points before all samples were lysed

simultaneously with Tri-Reagent. In the 4-day time courses, $1x10^6$ cells were plated per time point the

night before Dox addition. Cells were split 1:2 every day, with all splits kept for experiments with the

Tet-Off cells, while only 1 of the 2 splits were kept for experiments with the Tet-On cells. Cells were

cultured between 30-70% confluent, and were never allowed to exceed 80% confluence. The harvest

procedure for the 4-day long time course consisted of trypsinizing cells and collecting in Dox-containing

media, incubating on ice while cell number was determined, and finally centrifuging and freezing the cell

pellets. Total RNA was extracted later by Tri-Reagent lysis.

Northern Analysis

Northern blots were created and probed essentially as described in [41], where 30 μg of total

RNA was loaded per lane. The exceptions to the standard protocol are noted for the Northern blots in

Figure 4B and Figure 5. A riboprobe was used in probing the blot in Figure 4B in an attempt to improve the sensitivity for miRNA precursor. In Figure 5, a 3-mm thick, 15% denaturing polyacrylamide gel that included a 5% stacking gel layer was used to accommodate all the total RNA from cells at the later time points of the 4-day long time course. Pilot experiments with radiolabeled small RNA markers doped with carrier HeLa RNA indicated that up to 1 mg of total HeLa RNA could be loaded and resolved adequately on this type of gel. A 1-hour transfer on a Semi-Dry apparatus also indicated uniform and near-complete transfer of small RNAs from this modified gel to the membrane, while RNAs as large as or larger than tRNA showed minor variations in transfer completeness.

The total cellular RNA yields from the 4-day time courses ranged from an average of ~28 μg at t=0 hour to ~280 μg at t=96 hour for experiments with the Tet-Off cells (total accumulation of dividing cells), while the average total RNA yield per cell across all time points was 20 pg per cell, varying no more than 2 fold from time point to time point. The inconsistency of RNA accumulation with cell doubling rates may be due to lower extraction yields with increased cell matter that was processed in constant amounts of extraction reagent. To ensure equal RNA loading across the lanes on the modified denaturing gel, total cellular RNA from HeLa S3 cells was added as carrier to early time point samples to equalize RNA loading amounts. RNA integrity and resolution were qualitatively assessed by ethidium staining.

Luciferase Reporter Assays

Reporter assays were performed as described previously [59].

ACKNOWLEDGEMENTS

REFERENCES

1. Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

2. Lee, Y., Jeon, K., Lee, J.T., Kim, S., and Kim, V.N. (2002). MicroRNA maturation: stepwise processing and subcellular localization. Embo J *21*, 4663-4670.

3. Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. (2003). The nuclear RNase III Drosha initiates microRNA processing. Nature *425*, 415-419.

4. Chen, C.Z., Li, L., Lodish, H.F., and Bartel, D.P. (2004). MicroRNAs modulate hematopoietic lineage differentiation. Science *303*, 83-86 Published online December 84, 2003 2010.1126/science.1091903.

5. Zeng, Y., and Cullen, B.R. (2003). Sequence requirements for micro RNA processing and function in human cells. RNA *9*, 112-123.

6. Bohnsack, M.T., Czaplinski, K., and Gorlich, D. (2004). Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. RNA *10*, 185-191.

7. Lund, E., Güttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. (2004). Nuclear export of microRNA precursors. Science *303*, 95-98 Published online on November 20, 2003,2010.1126/science.1090599.

8. Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev. *17*, 3011-3016.

9. Ketting, R.F., Fischer, S.E.J., Bernstein, E., Sijen, T., Hannon, G.J., and Plasterk, R.H.A. (2001). Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in C. elegans. Genes Dev. *15*, 2654-2659.

10. Hutvágner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. (2001). A cellular function for the RNA-interference enzyme Dicer in the maturation of the *let-7* small temporal RNA. Science *293*, 834-838.

11. Grishok, A., Pasquinelli, A.E., Conte, D., Li, N., Parrish, S., Ha, I., Baillie, D.L., Fire, A., Ruvkun, G., and Mello, C.C. (2001). Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control C. elegans developmental timing. Cell *106*, 23-34.

12. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003a). The microRNAs of *Caenorhabditis elegans*. Genes Dev *17*, 991-1008.

13. Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell *115*, 209-216.

14. Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell *115*, 199-208.

15. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843-854.

16. 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*, 855-862.

17. Reinhart, B.J., Slack, F.J., Basson, M., Bettinger, J.C., Pasquinelli, A.E., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature *403*, 901-906.

18. Slack, F.J., Basson, M., Liu, Z., Ambros, V., Horvitz, H.R., and Ruvkun, G. (2000). The *lin-41* RBCC gene acts in the *C. elegans* heterochronic pathway between the *let-7* regulatory RNA and the LIN-29 transcription factor. Molecular Cell *5*, 659-669.

19. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K., and Slack, F.J. (2004). The C. elegans microRNA let-7 binds to imperfect let-7 complementary sites from the lin-41 3'UTR. Genes Dev *18*, 132-137.

20. Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. (2004). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci U S A *101*, 360-365.

21. Olsen, P.H., and Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. Developmental Biology *216*, 671-680.

22. Nelson, P.T., Hatzigeorgiou, A.G., and Mourelatos, Z. (2004). miRNP:mRNA association in polyribosomes in a human neuronal cell line. RNA *10*, 387-394.

23. Hutvágner, G., and Zamore, P.D. (2002). A microRNA in a multiple-turnover RNAi enzyme complex. Science *297*, 2056-2060.

24. Llave, C., Xie, Z., Kasschau, K.D., and Carrington, J.C. (2002). Cleavage of Scarecrow-like mRNA targets directed by a class of Arabidopsis miRNA. Science *297*, 2053-2056.

25. Tang, G., Reinhart, B.J., Bartel, D.P., and Zamore, P.D. (2003). A biochemical framework for RNA silencing in plants. Genes Dev *17*, 49-63.

26. Xie, Z., Kasschau, K.D., and Carrington, J.C. (2003). Negative feedback regulation of Dicer-like1 in Arabidopsis by microRNA-guided mRNA degradation. Curr Biol *13*, 784-789.

27. Pfeffer, S., Zavolan, M., Grasser, F.A., Chien, M., Russo, J.J., Ju, J., John, B., Enright, A.J., Marks, D., Sander, C., and Tuschl, T. (2004). Identification of Virus-Encoded MicroRNAs. Science *304*, 734-736.

28. Yekta, S., Shih, I.H., and Bartel, D.P. (2004). MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

29. Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev *18*, 504-511.

30. Doench, J.G., Peterson, C.P., and Sharp, P.A. (2003). siRNAs can function as miRNAs. Genes Dev. *17*, 438-442.

31. Zeng, Y., Yi, R., and Cullen, B.R. (2003). MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. Proc Natl Acad Sci U S A *100*, 9779-9784.

32. Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

33. Hammond, S.C., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates posttranscriptional gene silencing in Drosophila cells. Nature *404*, 293-296.

34. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

35. Nykänen, A., Haley, B., and Zamore, P.D. (2001). ATP requirements and small interfering RNA structure in the RNA interference pathway. Cell *107*, 309-321.

36. Ishizuka, A., Siomi, M.C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev *16*, 2497-2508.

37. Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev *16*, 2491-2496.

38. Pham, J.W., Pellino, J.L., Lee, Y.S., Carthew, R.W., and Sontheimer, E.J. (2004). A Dicer-2-Dependent 80S Complex Cleaves Targeted mRNAs during RNAi in Drosophila. Cell *117*, 83-94.

39. Lee, Y.S., Nakahara, K., Pham, J.W., Kim, K., He, Z., Sontheimer, E.J., and Carthew, R.W. (2004). Distinct Roles for Drosophila Dicer-1 and Dicer-2 in the siRNA/miRNA Silencing Pathways. Cell *117*, 69-81.

40. Xie, Z., Johansen, L.K., Gustafson, A.M., Kasschau, K.D., Lellis, A.D., Zilberman, D., Jacobsen, S.E., and Carrington, J.C. (2004). Genetic and Functional Diversification of Small RNA Pathways in Plants. PLoS Biol *2*, E104.

41. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science *294*, 858-862.

42. Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862-864.

43. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

44. Mourelatos, Z., Dostie, J., Paushkin, S., Sharma, A., Charroux, B., Abel, L., Rappsilber, J., Mann, M., and Dreyfuss, G. (2002). miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. Genes Dev *16*, 720-728.

45. Loflin, P.T., Chen, C.Y.A., Xu, N.H., and Shyu, A.B. (1999). Transcriptional pulsing approaches for analysis of mRNA turnover in mammalian cells. Methods-a Companion to Methods in Enzymology *17*, 11-20.

46. Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766-1769.

47. Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Current Biology *12*, 735-739.

48. Clement, J.Q., Qian, L., Kaplinsky, N., and Wilkinson, M.F. (1999). The stability and fate of a spliced intron from vertebrate cells. RNA *5*, 206-220.

49.   Hargrove, J.L., Hulsey, M.G., and Beale, E.G. (1991). The kinetics of mammalian gene expression. Bioessays *13*, 667-674.

50.   Hargrove, J.L., and Schmidt, F.H. (1989). The role of mRNA and protein stability in gene expression. Faseb J *3*, 2360-2370.

51.   Qu, Z., Thottassery, J.V., Van Ginkel, S., Manuvakhova, M., Westbrook, L., Roland-Lazenby, C., Hays, S., and Kern, F.G. (2004). Homogeneity and long-term stability of tetracycline-regulated gene expression with low basal activity by using the rtTA2S-M2 transactivator and insulator-flanked reporter vectors. Gene *327*, 61-73.

52.   Audibert, A., Weil, D., and Dautry, F. (2002). In vivo kinetics of mRNA splicing and transport in mammalian cells. Molecular and Cellular Biology *22*, 6706-6718.

53.   Farmer, A (CLONTECH). (personal communication).

54.   Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25-36.

55.   Aukerman, M.J., and Sakai, H. (2003). Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. Plant Cell *10*, 10.

56.   Chen, X. (2003). A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. Science, Published online September 11 2003; 2010.1126/science.1088060.

57.   Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

58.   Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature *425*, 257-263.

59.   Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

60.   Izumi, M., and Gilbert, D.M. (1999). Homogeneous tetracycline-regulatable gene expression in mammalian fibroblasts. J Cell Biochem *76*, 280-289.

61.   Volloch, V., and Housman, D. (1981). Stability of globin mRNA in terminally differentiating murine erythroleukemia cells. Cell *23*, 509-514.

62.   Schwarz, D.S., Hutvágner, G., Haley, B., and Zamore, P.D. (2002). Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. Mol. Cell *10*, 537-548.

63.   Chiu, Y.L., and Rana, T.M. (2003). siRNA function in RNAi: a chemical modification analysis. RNA *9*, 1034-1048.

64.   Dykxhoorn, D.M., Novina, C.D., and Sharp, P.A. (2003). Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol *4*, 457-467.

65.   Song, E., Lee, S.K., Dykxhoorn, D.M., Novina, C., Zhang, D., Crawford, K., Cerny, J., Sharp, P.A., Lieberman, J., Manjunath, N., and Shankar, P. (2003). Sustained small interfering RNA-mediated human immunodeficiency virus type 1 inhibition in primary macrophages. J Virol *77*, 7174-7181.

66.   Kennedy, S., Wang, D., and Ruvkun, G. (2004). A conserved siRNA-degrading RNase negatively regulates RNA interference in C. elegans. Nature *427*, 645-649.

67.   Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

68.   Allawi, H.T., Dahlberg, J.E., Olson, S., Lund, E., Olson, M., Ma, W.P., Takova, T., Neri, B.P., and Lyamichev, V.I. (2004). Quantitation of microRNAs using a modified Invader® assay. RNA *in press*.

Table 1. Calculation of miRNA half-lives. Assuming first-order decay kinetics, linear regression exponential fits were applied to the data in Figure 5B and 5C. Only fits displaying a negative rate constant, k, are tabulated. Half lives were calculated by the equation $t^{1/2}=-0.69/k$.

mir-1d^{OFF} line A1				mir-124a^{OFF} line A1			
trial	$t^{1/2}$ (hours)	k	R^2	trial	$t^{1/2}$ (hours)	k	R^2
TC-IV	38	-0.0179	0.61	TC-II	70	-0.0099	0.92
TC-V	313	-0.0075	0.20	TC-III	51	-0.0136	0.97
				TC-IV	29	-0.024	0.89
				TC-V	144	-0.0048	0.46

FIGURE LEGENDS

Figure 1. An Inducible System for miRNA Expression. A Tet-regulable expression cassette is stably transfected into Tet On or Tet Off cells, providing opposite forms of transcriptional regulation upon Dox addition.

Figure 2. Inducible Expression of human miR-1d. Northern blot analysis of total RNA from 8-day time courses are on the left, while quantitation of the signals are displayed on the right. Closed diamonds are time points where Dox is added to media, while open squares are mock-treated time points. Normalized counts represent miRNA signal divided by U6 snRNA signal. (A) Examination miR-1d^{ON} (line C2). (B) Examination of miR-1d^{OFF} (line A1).

Figure 3. Inducible Expression of human miR-124a. Annotations are essentially as described for Figure 2. (A) Examination miR-124a^{ON} (line A2). (B) Examination of miR-124a^{OFF} (line A1).

Figure 4. Timescale of pre-miRNA turnover. Northern analysis of time courses extending to 12 (A) and 150 minutes (B). Because of the high background from the riboprobe used to probe the blot in (B), the precursor signals were not quantitated.

Figure 5. Measuring miRNA decay by assaying all accumulated RNA. (A) A representative Northern blot where all total RNA from harvested Tet-Off cells is loaded in each lane, supplemented with carrier HeLa S3 total RNA so that each lane contains ~300 μg total RNA.

This blot examined time courses of mir-124a^{OFF} (line A1). No signals were detected for miR-1d nor miR-124a in lanes with HeLa S3 total RNA alone (data not shown). (B) Single exponential fits to measurements of miR-124a. Time course trials are labeled as TC-#. (C) Single exponential fits to measurements of miR-1d from Northern analyses of time courses of miR-1d^{OFF} (line A1).

Figure 6. Measuring miRNA induction. (A) Northern blot of miR-1d from time courses of induction with Dox of mir-1d^{ON} (line C2). (B) Quanatitation of mature miR-1d signal. (C) Quantitation of pre-miR-1d signal. (D) Northern blot of miR-124a from time courses of induction with Dox of mir-124a^{ON} (line A2). (E) Quanatitation of mature miR-124a signal. (F) Quantitation of pre-miR-124a signal. Time points are hours after Dox addition. All quantitation values were normalized against respective U6 snRNA signals. Time course trials are labeled as TC-#.

Figure 7. A model of miRNA biogenesis and metabolism (adapted from [1]).

Figure 1.

Tet-On cells

Chimeric VP16
transcription activator –
reverse Tet Repressor
bound to doxycycline

miRNA genomic
segment (~500 bp)

Tetracycline -
inducible promtoter

Expression
of pri-miRNA

Tet-Off cells

Chimeric VP16
transcription activator –
Tet Repressor bound
to doxycycline

Transcription
of pri-miRNA
is halted.

Figure 2.

A.

miR-1d On (line C2)

B.

miR-1d Off (line A1)

160

Figure 3.

A.

miR-124a On (line A2)

B.

miR-124a Off (line A1)

Figure 4.

A.

miR-1d Off

Hours after Dox addition

M   0   3   6   9   12

60—

precursor

24—
21—

mature

miR-124a Off

Hours after Dox addition

M   0   3   6   9   12

60—

precursor

24—
21—
18—

mature

B.

miR-124a Off
Minutes after Dox addition

M   0   10   20   40   60   90   120   150

60—

precursor

24—
21—

mature

Figure 5.

A.

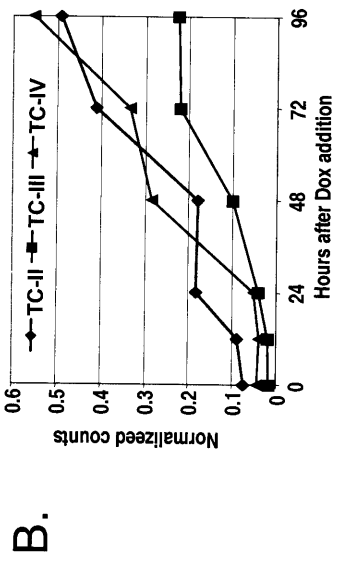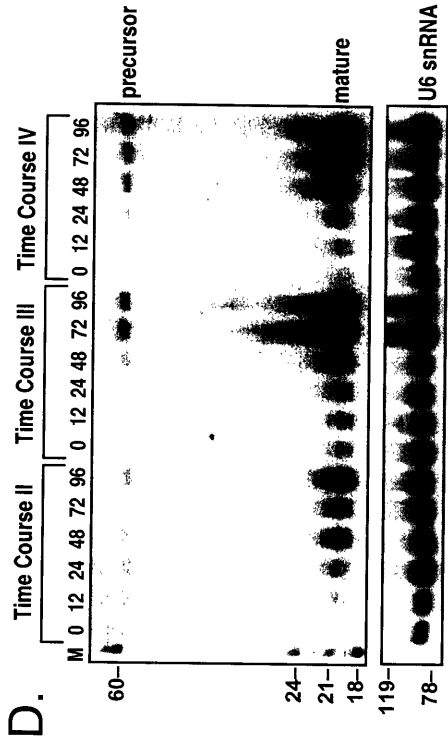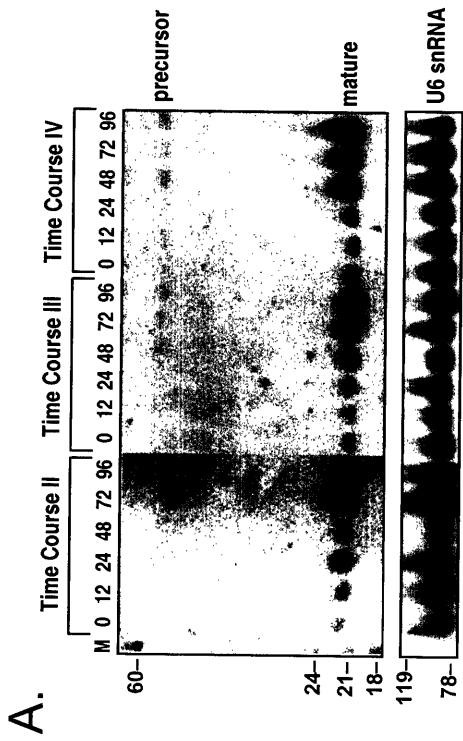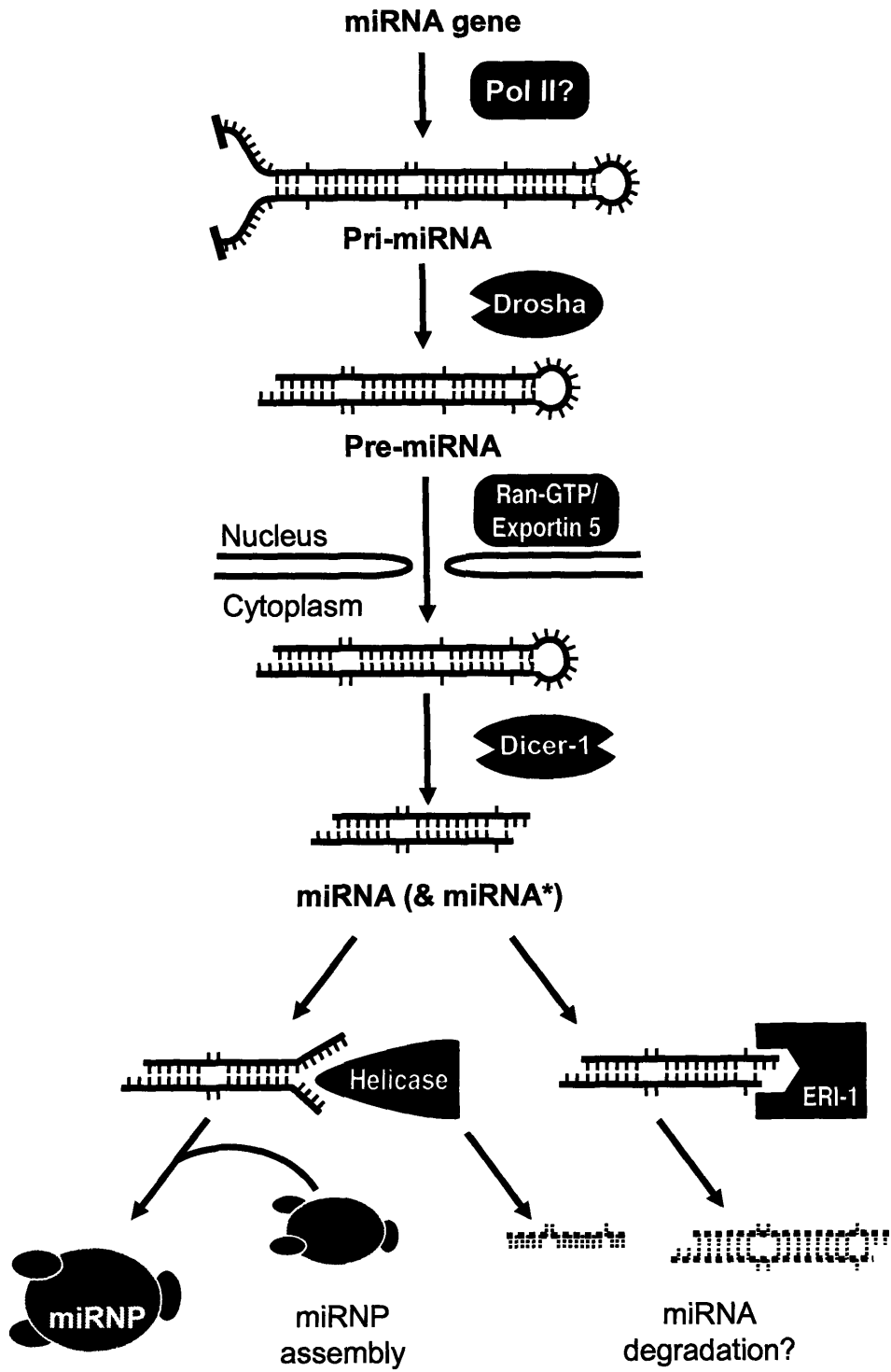| Time Course III	Time Course IV	Time Course V
0 12 24 48 72 96	0 12 24 48 72 96	0 12 24 48 72 96

B.

C.

Figure 6.

164

Figure 7.

Future Directions

Despite significant progress in understanding the biogenesis and functions of microRNAs (miRNAs), there are still many outstanding questions in the field. Topics of continuing investigation range from miRNA gene finding efforts, to identifying more miRNA target genes, to dissecting the biochemical mechanism of inhibiting productive translation, and to determining the phenotypic consequences of perturbing miRNA function. Each of these lines of research benefits from extensive headway made in the last 4 years, with some of the groundwork laid by studies described in this thesis. However, hurdles that will be encountered in future studies will demand creativity distinct from past studies.

The size of the miRNA gene class is quite extensive, but there is a desire to determine the totality of miRNAs within organisms and within eukaryotic lineages. Although estimates for the total number of miRNA genes have been placed for model invertebrates such as *Caenorhabditis elegans* and *Drosophila melanogaster*, the numbers of confirmed miRNAs have not reached these limits (78 confirmed out of 110 estimated *D.melanogaster* miRNAs, 103 confirmed out of 120 estimated *C.elegans* miRNAs) [1-9]. Refinements or composites of computational methods like MirScan and MirSeeker could present additional candidates for validation, and perhaps RNA samples from various organism growth conditions or shorter intervals during development might better present candidate miRNAs for detection. For example, *mir-247* is only detectable on Northern blots in the L3 stage or when worms are starved and induced to develop into dauers [8]. Alternatively, cloning of new small RNAs might be extended by the development of subtractive hybridization methods to remove the most abundant miRNAs. The ten most abundantly cloned miRNAs in *C.elegans* represent 54% of the total 3423 clones sequenced, yet comprise only 10% of the total miRNA genes known to exist in the *C.elegans* genome [8]. Obtaining complete lists of miRNAs in flies and worms should bring to light interesting but rare

miRNAs like *lsy-6* [10]. Lists of vertebrate miRNAs, although extensive, are also quite incomplete, having depended on just human, mouse, and pufferfish genomes [11, 12]. Future analysis will surely benefit from several additional mammalian, avian and tunicate genomes that are in later phases of completion [13-16].

Because the miRNA gene family is so large and seemingly pervasive in metazoans and plants, the evolution of this gene class is a compelling problem to investigate. The conservation of *let-7* across bilaterian animals is astounding, but also curious is lack of *let-7* detected in other non-bilateral metazoans like cniderian and porfiran species [17]. Perhaps *let-7* has diverged in these species similarly as how *lin-4* may have diverged from *mir-125b* in vertebrates [18]. A few miRNAs like *mir-1* and *mir-124* are perfectly conserved in primary sequence like *let-7* across vertebrates and invertebrates, while most other miRNAs show overall sequence similarity but significant divergence in the 3' end of the miRNA [1, 8]. The insight that the RNA-induced silencing complex (RISC) appears more dependent on target recognition with the "seed" evokes the question as to what combination of targets common to vertebrates and invertebrates could be enforcing the complete primary sequence conservation of a select few miRNAs [19-21]. In land plants, the conservation of miRNAs like *MIR166* appears to run deep into different lineages; this conservation can be explained by the vital function of downregulating HD-Zip genes through miR166 base-pairing extensively to and cleaving the target site [22-26]. Because so few animal target sites approach the degree of complementarity of plant targets to miRNAs, the puzzle of such well-conserved miRNAs in animals is particularly confounding. Although plant and animal miRNA sequences are poorly conserved, it has been suggested that miRNAs might have arisen during the advent of multicellularity in organisms, since single-celled organisms like *Schizosaccaromyces pombe* and *Trypansome brucei* appear to lack miRNAs but possess siRNAs

[27-30]. It is debatable if an organism with miRNAs that is an ancestor to plants and animals can be defined, but if no miRNAs can be identified in other fungal species that have pseudo-multi-cellular form, this could further argue that miRNAs convergently evolved in plants and animals. Convergent evolution might also reconcile why the preponderance of transcription factors in predicted plant targets is not seen in predicted animal targets [20, 25, 31-34]. Since plants critically depend on miRNAs to downregulate specific transcription factor messages, the convergent evolution hypothesis might assume animal cells evolved miRNAs separately to "tune" a broader range of targets; and animals may have depended much less on diminishing transcription factors early in the evolution of RNA silencing pathways [35]. Alternatively, our abilities to predict and then validate true animal miRNA targets may not be sophisticated enough to pinpoint the actual scope of targets, or perhaps the range of plant targets could potentially be as wide as animal targets [33].

Addressing miRNA conservation and other functional questions heavily depends on refining algorithms for predicting miRNA targets. Currently the most sensitive algorithms for predicting animal targets simply search for multiple base-pairing elements that are conserved in the 3' untranslated regions (UTRs) of homologous genes from multiple genomes [20, 31, 32]. Refinement might be accomplished by testing numerous candidate UTRs in reporter gene assays in cell culture, and then reexamine the miRNA-responsive UTR-reporter constructs for motifs or structural elements that might enhance the specificity of prediction algorithms. Structural elements and motifs within 5' and 3' UTRs are well known to exert translational repression effects, like the iron response elements in the 5' UTR of Ferritin mRNA, or cytoplasmic poly-adenylation elements in the 3' UTR of *cyclinB* mRNA [36, 37]. The scenario of RISC recognizing motifs in addition to seed interactions is quite possible because the Fragile-X

protein, FMRP, is associated with RISC and miRNAs, and is known to bind RNAs with G-quartet structures [38-40]. In addition to motif refinements, prediction algorithms could be enhanced to find targets regulated by combinations of miRNAs. This approach is definitely biologically relevant but potentially computationally and experimentally intensive, so the simplest and most practical increment would be to examine combinations of two miRNAs. A presumption by current analyses of target predictions is that regulation amongst members of a highly-related miRNA family might be hard to distinguish since the seed pairing is the same [20]. However, sizeable families of miRNAs are conserved in worms, flies and vertebrates, suggesting that an exquisite specificity not well recognized by scientists is well utilized by organisms. Perhaps current predictions originally tested against a single miRNA might be retested against combinations of miRNAs that are members within a gene family to see if there is some synergism imparted by the combination of the two miRNAs.

A better understanding of miRNA target recognition might provide insight into the mechanism of miRNA inhibition of productive translation, or vice versa. Despite what is known about RISC-directed mRNA cleavage, there is no model that best explains how the inhibition of productive translation occurs. One has to take into account that mRNA stability is not affected, and the expression block occurs after polysomes form on the mRNA [41-43]. Perhaps RISC affixed to a site it cannot cleave serves as a block that might retard protein elongation rates, or prolonged RISC interaction selectively localizes polysomes away from translation co-factors. Perhaps translation and ribosome movement is not altered at all, and instead protein half-life is altered, either by recruitment of a peptidase, or a translation cofactor that stimulates active degradation of the translated protein. Addressing these speculations requires an experimental system that is both competent in protein translation and amenable to modulation of miRNA

levels. *Drosophila* embryo and Hela cell *in vitro* translation lysates are possible systems for

investigation [44, 45], since exogenous siRNAs can program these lysates at least for RNAi [46-

48]. The Hela cell lines with inducible miRNA expression described in this thesis, however,

could also be a potentially valuable system for investigating the mechanism of translation

repression. Biochemical purifications of a reporter mRNA specifically regulated by a *mir-1* or

mir-124 site might be compared between the induced or uninduced states of miRNA expression.

This experiment could examine modifications to the reporter mRNA or perhaps lead to the

identification of factors that localize to the reporter mRNA after miRNA induction. The reporter

peptide might also be compared between miRNA expressing and non-expressing cells to see if

co-translational or post-translational modifications might be destabilizing the protein.

Substantiated by the overt defects in *lin-4*, *let-7*, *bantam*, *jawD*, and *MIR172* mutants [49-

54], many labs are attempting to perturb miRNA expression or regulation in order to dissect

miRNA function regardless of whether mechanism or target genes are known. Since the

majority of miRNA genes in model organisms have been cloned or mapped, reverse genetics is a

feasible approach towards ascertaining the phenotype of miRNA loss-of-function. One such

endeavor to knock out miRNA genes in *C.elegans* is described in the appendix of this thesis,

where libraries of mutagenized worms are screened in high-throughput fashion to identify

deletion mutants. Analogous reverse genetics approaches are likely being applied to other

animals and plants, and hopefully phenotypes will present themselves in a fashion similar to the

obvious defects of *lin-4* and *let-7*. However, it is possible that many phenotypes might be as

subtle as *lsy-6*, which superficially grows and develops normally under lab conditions, but

misregulation of its target gene, *cog-1*, only causes a chemotaxis defect [10, 55]. The

complicated chemotaxis assays or specific GFP-tagged transgenes used to characterize *lsy-6* are

methods that are not commonly included in a simple survey of a deletion mutant. An additional complication that genetic approaches must consider is redundancy of other miRNAs, especially amongst family members with very similar seed sequences. Double, triple or multiple knock out of genes may be needed to elicit a phenotype.

If phenotypes are too subtle for a general survey to reveal, expression data on miRNA localization might focus the analysis to particular tissues or cell types. Such tissue-specific expression for miRNAs is available in mice [18], but for other model organisms that are smaller and more difficult to dissect, examining tissue-specific expression is not as technically feasible. More sensitive and high-throughput methods for detecting miRNAs from small samples might surmount some of the technical limitations, such as signal amplification from a sensitive fluorescence based assay [56], or from a prototype microarray for miRNAs [57]. Alternatively, reporter transgenes like the *bantam* sensor might potentially reveal miRNA localization in a dynamic fashion during live animal development [31, 52], but a serious technical consideration is determining the proper promoter that drives ubiquitous but not saturating expression of the sensor gene. Finally, *in situ* hybridization techniques may become more widespread for detecting miRNA localization, as this method shows promise in plant sections [26, 58, 59], but the actual specificity of probes and conditions for in situ hybridization of miRNAs await rigorous demonstration with null mutants and comparison with sensor experiments.

Two indirect but potentially illuminating approaches that could guide functional studies would be to misexpress miRNAs or to interfere with miRNA function with antisense 2-O-methyl RNA (2OMe) oligos. The mutants, *jawD*, *MIR172, bantam*, and *mir-14*, exhibit defects from miRNA misexpression [52-54, 60], and a microrarray approach successfully identified target genes for *jawD* [53]. While microarray experiments have proven well in determining plant

miRNA targets, their utility for finding animal targets may be questionable if the dominant

mechanism for animal miRNA function does not affect mRNA levels. The ability to interfere

with RISC function through 2OMe oligos has been demonstrated recently *in vitro* as well as *in*

vivo [61, 62]. Phenotypes generated by this antisense interference method must be interpreted

with caution since miRNA genetics is still in its infancy and the toxicity or side effects of 2OMe

oligos *in vivo* have not been clearly determined. However, synthetic defects could be revealing

and might be more attainable with this technique because multiple 2OMe oligos directed against

a whole miRNA gene family might overcome functional redundancy.

The strong interest in miRNA function is rooted not only in the novelty of miRNA

research, but also in the potential ramifications for global gene network regulation in multi-

cellular organisms by small RNA pathways. The excitement in miRNA research is currently

strongest in academic circles, but will soon permeate into agricultural and clinical investigations

as well. Proper growth of several crop plants clearly depends on proper miRNA function [29],

and plant viruses can exert their pathogenicity through interference of miRNA pathways [63-66],

so the impact of miRNAs on agriculture is apparent. No specific human disease has yet been

definitively linked to defects in miRNA function, but hints of miRNAs tying into clinical cases

are beginning to emerge. A few studies suggest miRNA levels might be perturbed in cancer

samples [67, 68]; miRNA mutants in flies antagonize apoptosis and are hypothesized to be

oncogenes [52, 60]; the protein implicated in Fragile X mental retardation is linked to RISC

function [40, 69, 70]; and PPD proteins are linked to tumorigenesis and stem cell maintenance

[71]. Regardless of the agricultural or clinical implications of miRNAs, basic research on

miRNA will continue to reveal surprises and remind us of the macroscopic influence that

microRNAs have on eukaryotic development.

REFERENCES

1. Ambros, V., Lee, R.C., Lavanway, A., Williams, P.T., and Jewell, D. (2003). MicroRNAs and Other Tiny Endogenous RNAs in C. elegans. Curr Biol *13*, 807-818.

2. Aravin, A.A., Lagos-Quintana, M., Yalcin, A., Zavolan, M., Marks, D., Snyder, B., Gaasterland, T., Meyer, J., and Tuschl, T. (2003). The small RNA profile during Drosophila melanogaster development. Dev Cell *5*, 337-350.

3. Grad, Y., Aach, J., Hayes, G.D., Reinhart, B.J., Church, G.M., Ruvkun, G., and Kim, J. (2003). Computational and experimental identification of C. elegans microRNAs. Mol Cell *11*, 1253-1263.

4. Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. Science *294*, 853-858.

5. Lai, E.C., Tomancak, P., Williams, R.W., and Rubin, G.M. (2003). Computational identification of Drosophila microRNA genes. Genome Biol *4:R42*, 1-20.

6. Lau, N.C., Lim, L.P., Weinstein, E.G., and Bartel, D.P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. Science *294*, 858-862.

7. Lee, R.C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis elegans*. Science *294*, 862-864.

8. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003a). The microRNAs of *Caenorhabditis elegans*. Genes Dev *17*, 991-1008.

9. Ohler, U., Yekta, S., Lim, L.P., Bartel, D.P., and Burge, C.B. (2004). Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. RNA, in press.

10. Johnston, R.J., and Hobert, O. (2003). A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans. Nature *426*, 845-849.

11. Bartel, D.P. (2004). MicroRNAs: Genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

12. Lim, L.P., Glasner, M.E., Yekta, S., Burge, C.B., and Bartel, D.P. (2003b). Vertebrate microRNA genes. Science *299*, 1540.

13.	Kirkness, E.F., Bafna, V., Halpern, A.L., Levy, S., Remington, K., Rusch, D.B., Delcher, A.L., Pop, M., Wang, W., Fraser, C.M., and Venter, J.C. (2003). The dog genome: survey sequencing and comparative analysis. Science *301*, 1898-1903.

14.	Holland, L.Z., and Gibson-Brown, J.J. (2003). The Ciona intestinalis genome: when the constraints are off. Bioessays *25*, 529-532.

15.	Andersson, L., and Georges, M. (2004). Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet *5*, 202-212.

16.	Gibbs, R.A., Weinstock, G.M., Metzker, M.L., Muzny, D.M., Sodergren, E.J., Scherer, S., Scott, G., Steffen, D., Worley, K.C., Burch, P.E., Okwuonu, G., Hines, S., Lewis, L., DeRamo, C., Delgado, O., Dugan-Rocha, S., Miner, G., Morgan, M., Hawes, A., Gill, R., et al. (2004). Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature *428*, 493-521.

17.	Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M., Maller, B., Srinivasan, A., Fishman, M., Hayward, D., Ball, E., Degnan, B., Muller, P., Spring, J., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., and Ruvkun, G. (2000). Conservation across animal phylogeny of the sequence and temporal regulation of the 21 nucleotide *let-7* heterochronic regulatory RNA. Nature *408*, 86-89.

18.	Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. (2002). Identification of tissue-specific microRNAs from mouse. Current Biology *12*, 735-739.

19.	Doench, J.G., and Sharp, P.A. (2004). Specificity of microRNA target selection in translational repression. Genes Dev *18*, 504-511.

20.	Lewis, B.P., Shih, I., Jones-Rhoades, M.W., Bartel, D.P., and Burge, C.B. (2003). Prediction of mammalian microRNA targets. Cell *115*, 787-798.

21.	Lai, E.C. (2002). MicroRNAs are complementary to 3'UTR motifs that mediate negative post-transcriptional regulation. Nature Genetics *30*, 363-364.

22.	Floyd, S.K., and Bowman, J.L. (2004). Gene regulation: ancient microRNA target sequences in plants. Nature *428*, 485-486.

23.	Emery, J.F., Floyd, S.K., Alvarez, J., Eshed, Y., Hawker, N.P., Izhaki, A., Baum, S.F., and Bowman, J.L. (2003). Radial Patterning of Arabidopsis Shoots by Class III HD-ZIP and KANADI Genes. Curr Biol *13*, 1768-1774.

24.   McConnell, J.R., Emery, J., Eshed, Y., Bao, N., Bowman, J., and Barton, M.K. (2001). Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. Nature *411*, 709-713.

25.   Rhoades, M.W., Reinhart, B.J., Lim, L.P., Burge, C.B., Bartel, B., and Bartel, D.P. (2002). Prediction of plant microRNA targets. Cell *110*, 513-520.

26.   Juarez, M.T., Kui, J.S., Thomas, J., Heller, B.A., and Timmermans, M.C. (2004). microRNA-mediated repression of rolled leaf1 specifies maize leaf polarity. Nature *428*, 84-88.

27.   Reinhart, B.J., and Bartel, D.P. (2002). Small RNAs correspond to centromere heterochromatic repeats. Science *297*, 1831.

28.   Reinhart, B.J., Weinstein, E.G., Rhoades, M.W., Bartel, B., and Bartel, D.P. (2002). MicroRNAs in plants. Genes Dev *16*, 1616-1626.

29.   Bartel, B., and Bartel, D.P. (2003). MicroRNAs: At the Root of Plant Development? Plant Physiol *132*, 709-717.

30.   Djikeng, A., Shi, H., Tschudi, C., and Ullu, E. (2001). RNA interference in Trypanosoma brucei: cloning of small interfering RNAs provides evidence for retroposon-derived 24-26-nucleotide RNAs. RNA *7*, 1522-1530.

31.   Stark, A., Brennecke, J., Russell, R.B., and Cohen, S.M. (2003). Identification of Drosophila microRNA targets. PLOS Biol. *1*, E60.

32.   Rajewsky, N., and Socci, N.D. (2004). Computational identification of microRNA targets. Dev Biol *267*, 529-535.

33.   Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets. Mol Cell *In Press*.

34.   Enright, A.J., John, B., Gaul, U., Tuschl, T., Sander, C., and Marks, D.S. (2003). MicroRNA targets in Drosophila. Genome Biology *5*, R1.

35.   Bartel, D.P., and Chen, C.-Z. (2004). Micromanagers of Gene Expression: The Potentially Widespread Influence of Metazoan MicroRNAs. Nat Rev Genet *5*, 396-400.

36.   Kuersten, S., and Goodwin, E.B. (2003). The power of the 3' UTR: translational control and development. Nat Rev Genet *4*, 626-637.

37.   Wilkie, G.S., Dickson, K.S., and Gray, N.K. (2003). Regulation of mRNA translation by 5'- and 3'-UTR-binding factors. Trends Biochem Sci *28*, 182-188.

38. Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T., and Darnell, R.B. (2001). Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. Cell *107*, 489-499.

39. Brown, V., Jin, P., Ceman, S., Darnell, J.C., O'Donnell, W.T., Tenenbaum, S.A., Jin, X., Feng, Y., Wilkinson, K.D., Keene, J.D., Darnell, R.B., and Warren, S.T. (2001). Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. Cell *107*, 477-487.

40. Jin, P., Zarnescu, D.C., Ceman, S., Nakamoto, M., Mowrey, J., Jongens, T.A., Nelson, D.L., Moses, K., and Warren, S.T. (2004). Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. Nat Neurosci *7*, 113-117.

41. Nelson, P.T., Hatzigeorgiou, A.G., and Mourelatos, Z. (2004). miRNP:mRNA association in polyribosomes in a human neuronal cell line. RNA *10*, 387-394.

42. Olsen, P.H., and Ambros, V. (1999). The *lin-4* regulatory RNA controls developmental timing in *Caenorhabditis elegans* by blocking LIN-14 protein synthesis after the initiation of translation. Developmental Biology *216*, 671-680.

43. Kim, J., Krichevsky, A., Grad, Y., Hayes, G.D., Kosik, K.S., Church, G.M., and Ruvkun, G. (2004). Identification of many microRNAs that copurify with polyribosomes in mammalian neurons. Proc Natl Acad Sci U S A *101*, 360-365.

44. Bergamini, G., Preiss, T., and Hentze, M.W. (2000). Picornavirus IRESes and the poly(A) tail jointly promote cap-independent translation in a mammalian cell-free system. RNA *6*, 1781-1790.

45. Castagnetti, S., Hentze, M.W., Ephrussi, A., and Gebauer, F. (2000). Control of oskar mRNA translation by Bruno in a novel cell-free system from Drosophila ovaries. Development *127*, 1063-1068.

46. Tuschl, T., Zamore, P.D., Lehmann, R., Bartel, D.P., and Sharp, P.A. (1999). Targeted mRNA degradation by double-stranded RNA *in vitro*. Genes Dev. *13*, 3191-3197.

47. Zamore, P.D., Tuschl, T., Sharp, P.A., and Bartel, D.P. (2000). RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals. Cell *101*, 25-33.

48. Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R., and Tuschl, T. (2002). Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. Cell *110*, 563-574.

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

50. Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. Cell *75*, 843-854.

51. Reinhart, B.J., Slack, F.J., Basson, M., Bettinger, J.C., Pasquinelli, A.E., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. (2000). The 21 nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. Nature *403*, 901-906.

52. Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. (2003). bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. Cell *113*, 25-36.

53. Palatnik, J.F., Allen, E., Wu, X., Schommer, C., Schwab, R., Carrington, J.C., and Weigel, D. (2003). Control of leaf morphogenesis by microRNAs. Nature *425*, 257-263.

54. Aukerman, M.J., and Sakai, H. (2003). Regulation of Flowering Time and Floral Organ Identity by a MicroRNA and Its APETALA2-Like Target Genes. Plant Cell *10*, 10.

55. Hobert, O., Johnston, R.J., Jr., and Chang, S. (2002). Left-right asymmetry in the nervous system: the Caenorhabditis elegans model. Nat Rev Neurosci *3*, 629-640.

56. Allawi, H.T., Dahlberg, J.E., Olson, S., Lund, E., Olson, M., Ma, W.P., Takova, T., Neri, B.P., and Lyamichev, V.I. (2004). Quantitation of microRNAs using a modified Invader® assay. RNA *in press*.

57. Krichevsky, A.M., King, K.S., Donahue, C.P., Khrapko, K., and Kosik, K.S. (2003). A microRNA array reveals extensive regulation of microRNAs during brain development. RNA *9*, 1274-1281.

58. Kidner, C.A., and Martienssen, R.A. (2004). Spatially restricted microRNA directs leaf polarity through ARGONAUTE1. Nature *428*, 81-84.

59. Chen, X. (2003). A MicroRNA as a Translational Repressor of APETALA2 in Arabidopsis Flower Development. Science, Published online September 11 2003; 2010.1126/science.1088060.

60. Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A. (2003). The Drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism. Curr Biol *13*, 790-795.

61. Hutvagner, G., Simard, M.J., Mello, C.C., and Zamore, P.D. (2004). Sequence-specific inhibition of small RNA function. PLoS Biol *2*, E98.

62. Meister, G., Landthaler, M., Dorsett, Y., and Tuschl, T. (2004). Sequence-specific inhibition of microRNA- and siRNA-induced RNA silencing. RNA *10*, 544-550.

63. Chen, J., Li, W.X., Xie, D., Peng, J.R., and Ding, S.W. (2004). Viral Virulence Protein Suppresses RNA Silencing-Mediated Defense but Upregulates the Role of MicroRNA in Host Gene Expression. Plant Cell.

64. Kasschau, K.D., Xie, Z., Allen, E., Llave, C., Chapman, E.J., Krizan, K.A., and Carrington, J.C. (2003). P1/HC-Pro, a viral suppressor of RNA silencing, interferes with *Arabidopsis* development and miRNA function. Dev Cell *4*, 205-217.

65. Mallory, A.C., Reinhart, B.J., Bartel, D.P., Vance, V.B., and Bowman, L.H. (2002). A viral suppressor of RNA silencing differentially regulates the accumulation of short interfering RNAs and microRNAs in tobacco. PNAS.

66. Papp, I., Mette, M.F., Aufsatz, W., Daxinger, L., Schauer, S.E., Ray, A., van der Winden, J., Matzke, M., and Matzke, A.J. (2003). Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. Plant Physiol *132*, 1382-1390.

67. Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., and Croce, C.M. (2002). Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A *99*, 15524-15529.

68. Michael, M.Z., SM, O.C., van Holst Pellekaan, N.G., Young, G.P., and James, R.J. (2003). Reduced accumulation of specific microRNAs in colorectal neoplasia. Mol Cancer Res *1*, 882-891.

69. Caudy, A.A., Myers, M., Hannon, G.J., and Hammond, S.M. (2002). Fragile X-related protein and VIG associate with the RNA interference machinery. Genes Dev *16*, 2491-2496.

70. Ishizuka, A., Siomi, M.C., and Siomi, H. (2002). A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. Genes Dev *16*, 2497-2508.

71. Carmell, M.A., Xuan, Z., Zhang, M.Q., and Hannon, G.J. (2002). The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. Genes Dev *16*, 2733-2742.

The three-dimensional architecture of the class I ligase ribozyme

NICHOLAS H. BERGMAN,[1,3] NELSON C. LAU,[1] VALERIE LEHNERT,[2,4] ERIC WESTHOF,[2] and DAVID P. BARTEL[1]

[1]Whitehead Institute for Biomedical Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

[2]Institut de Biologie Moléculaire et Cellulaire, Unité Propre de Recherche 9002 du Centre National de la Recherche Scientifique, Université Louis Pasteur, 67084 Strasbourg-Cedex, France

ABSTRACT

The class I ligase ribozyme catalyzes a Mg^{++}-dependent RNA-ligation reaction that is chemically analogous to a single step of RNA polymerization. Indeed, this ribozyme constitutes the catalytic domain of an accurate and general RNA polymerase ribozyme. The ligation reaction is also very rapid in both single- and multiple-turnover contexts and thus is informative for the study of RNA catalysis as well as RNA self-replication. Here we report the initial characterization of the three-dimensional architecture of the ligase. When the ligase folds, several segments become protected from hydroxyl-radical cleavage, indicating that the RNA adopts a compact tertiary structure. Ribozyme folding was largely, though not completely, Mg^{++} dependent, with a $K_{1/2[Mg]} < 1$ mM, and was observed over a broad temperature range (20°C –50°C). The hydroxyl-radical mapping, together with comparative sequence analyses and analogy to a region within 23S ribosomal RNA, were used to generate a three-dimensional model of the ribozyme. The predictive value of the model was tested and supported by a photo-cross-linking experiment.

Keywords: RNA structure; RNA catalysis; molecular modeling

INTRODUCTION

The class I ligase (Fig. 1A) catalyzes a reaction similar to that of biological RNA polymerases: attack by the 3'-OH of a small substrate RNA on a 5'-triphosphate, forming a new 3'–5' linkage with concomitant release of pyrophosphate (Ekland et al. 1995). Comparisons between this reaction and RNA polymerization were extended by experiments showing that engineered derivatives of the ligase are able to perform short primer-extension reactions (Ekland and Bartel 1996). More recently, variants of the ribozyme have been shown to catalyze template-directed polymerization of up

to 14 nt (Johnston et al. 2001), supporting the idea that early in the origin of life, RNA might have catalyzed its own replication (Bartel 1999; Joyce and Orgel 1999).

Recent studies have begun to define the reaction kinetics of the ligase, using ribozymes in both multiple- and single-turnover formats. In a multiple-turnover format, the ribozyme catalyzes ligation with a k_{cat} greater than 2 sec^{-1} at pH 8.0, a rate exceeding those of other multiple-turnover ribozyme-catalyzed reactions (Bergman et al. 2000). This speed is due in large part to a very fast chemical step (k_c). Studies examining k_c in a single-turnover format have shown that this step can reach rates exceeding 10 sec^{-1} at pH 9.0 (N.H. Bergman, C.C. Yen, and D.P. Bartel, in prep.). The studies of self-ligation also showed that the majority of ligase molecules fold accurately and quickly, with a folding rate of about 1.0 sec^{-1}, as measured by attainment of an active structure (Glasner et al. 2002). This suggests that the alternative folding pathways (and accompanying misfolding) seen in some other ribozymes (Pan and Sosnick 1997; Russell and Herschlag 1999) are less prevalent in the case of the ligase.

Like proteinaceous RNA polymerases, the ligase has a near absolute requirement for Mg^{++} ions (Glasner et al. 2002). Furthermore, these protein enzymes and the ligase
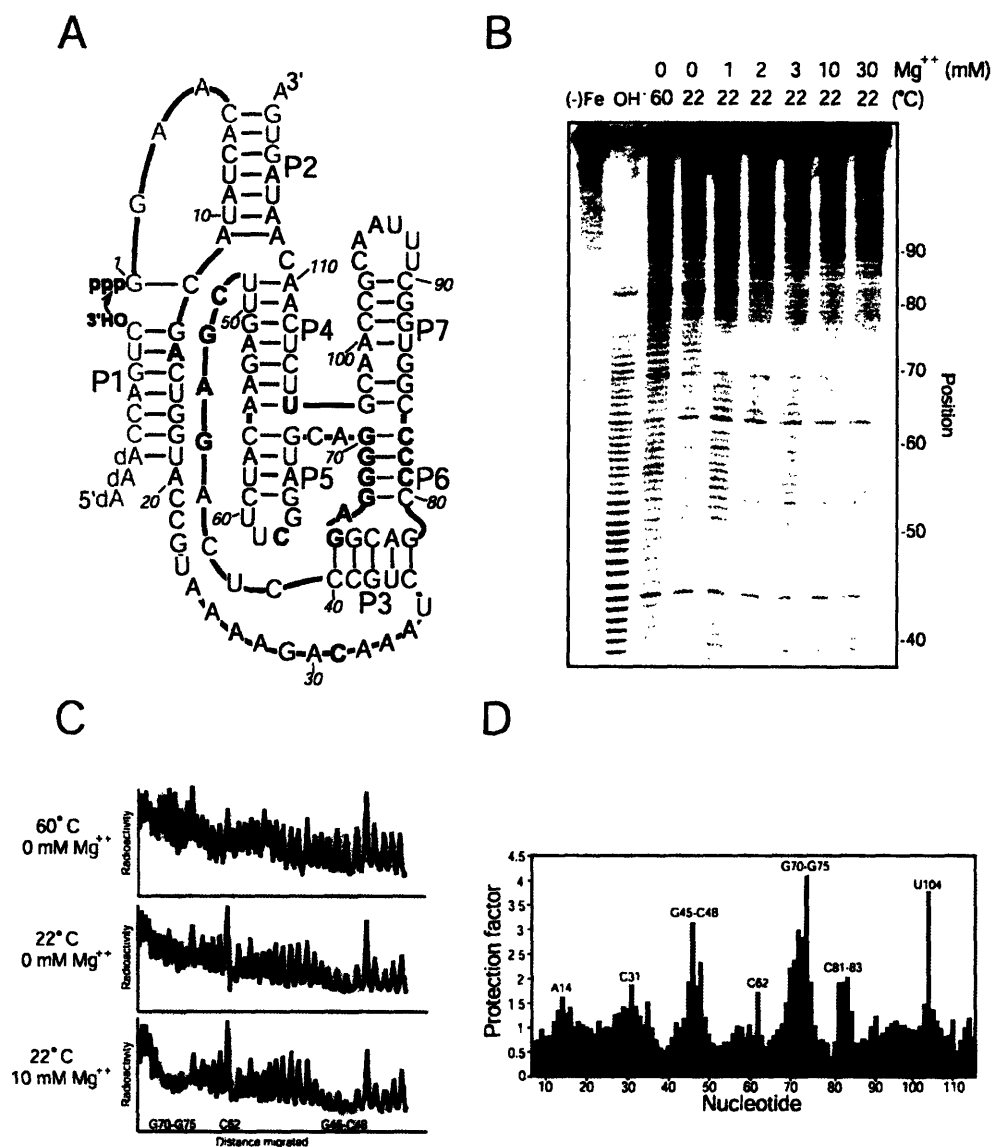
FIGURE 1. Hydroxyl-radical cleavage of the Class I ligase. (*A*) The class I ligase. The ligase promotes the attack of the 3'-OH of the substrate on its own 5'-α-phosphate, forming a new 3'-5' linkage with release of pyrophosphate. Paired regions are designated P1–P7. Residues protected from hydroxyl-radical cleavage are colored red and reflect those highlighted in panel *C*. Residues colored blue are those for which solvent accessibility was not measured because they were too near to the ends of the RNA. (*B*) Gel showing hydroxyl-radical cleavage in the presence of different Mg^{++} concentrations. Background cleavage in the absence of iron is shown [(−)Fe], along with a lane showing cleavage under denaturing conditions (60°C, 0 mM Mg^{++}). The gel was run for 90 min and used to collect data at positions 40–80. Nucleotides were identified by comparison to a ladder generated by partial digest of radiolabeled product using RNase T1 (not shown) and alkaline hydrolysis (OH^{-}). (*C*) Radioactivity profiles of lanes from the gel in panel *B*. Regions of substantial protection are shaded. (*D*) Protection factors for nt 7–115 in the presence of 10 mM Mg^{++} at 22°C. Protection factors are defined as the ratio of cleavage under denaturing conditions (0 mM Mg^{++}, 60°C) to cleavage under experimental conditions. Protection factors exceeding 1.5 are labeled and colored red.

show a similar stereospecific response to sulfur substitution at the reactive phosphate, which is consistent with the idea that one of these essential metal ions may be bound in the same position relative to substrate in both catalysts (Eckstein 1985; Glasner et al. 2000). This finding leaves open the possibility that the ligase uses the same mechanism as that proposed for general protein-catalyzed phosphoryl transfer,

and prompts questions of whether the ribozyme might also share structural features with analogous protein enzymes.

To better understand the relationship between the structure and function of the ligase, we have begun to characterize its tertiary structure. The solvent accessibility at each position along the sugar–phosphate backbone of the ribozyme was measured using hydroxyl radical probing (Ce-

lander and Cech 1990, 1991). These measurements identified portions of the RNA that are protected from cleavage when the ligase is folded. Together with comparative sequence analysis, these data were used herein to model the ribozyme in three dimensions. A photo-cross-linking experiment showed that the model successfully predicted three-dimensional proximity between ribozyme segments that were far apart in the secondary structure. The model provides a framework for future structural studies and suggests strategies for crystallization of the ligase.

RESULTS AND DISCUSSION

Hydroxyl-radical probing

Hydroxyl radicals, generated by either chemical methods (Tullius et al. 1987; King et al. 1993) or synchrotron radiation (Sclavi et al. 1998), have been used successfully to examine the structure of catalytic RNAs, ribosomal RNAs, and protein–nucleic acid complexes (Tullius and Dombroski 1986; Latham and Cech 1989; Celander and Cech 1991; Joseph and Noller 2000). The radicals, when produced in solution with RNA, attack the ribose moieties in the nucleic acid backbone, causing strand cleavage (Wu et al. 1983; Hertzberg and Dervan 1984). This cleavage is independent of sequence and secondary structure, and is instead dependent on the solvent accessibility of each position in the RNA backbone (Celander and Cech 1990). In beginning investigations of the tertiary structure of the ligase, hydroxyl-radical probing was used to define portions of the ribozyme that are internalized by tertiary structure and thus protected from cleavage.

At 22°C and 10 mM Mg^{++}, a substantial number of nucleotides (17 out of 109 tested) became protected from hydroxyl-radical cleavage (protection factor >1.5; Fig. 1), indicating that the ligase assumes a compact structure in the presence of Mg^{++}. The most striking protections were seen for nucleotides G45–C48, which make up the most conserved part of a joining region connecting helix P3 with helix P4, and for nucleotides G70–G75, which comprise the 5' arm of helix P6 (Fig. 1). Interestingly, the opposite arm of this helix (C81–C83) also had high protection factors, implying that this short helix is nestled within the ribozyme core. In addition to these segments, several isolated nucleotides were protected. Nt A14 and C31 were consistently protected, though at a much weaker level than segments G45–C48 or G70–G75. Nt U104 was strongly protected, perhaps because of the structure inherent in the junction of helices P4,

P5, P6, and P7. Finally, nt C62 was strongly protected even in the absence of Mg^{++}, perhaps because of local structure in the UUCG tetraloop (Cheong et al. 1990).

Temperature and Mg^{++} dependence of ribozyme tertiary structure

The fold of the ribozyme was assayed with respect to Mg^{++} and temperature, using the average protection factors from segments G45–C48 and G70–G75 as a measure of the degree to which molecules assumed the native tertiary structure (Fig. 2). In the presence of 10 mM Mg^{++}, the ligase showed some native structure at 10°C, and was well structured in the range 20°C–40°C. The group I intron shows significant misfolding at lower temperatures (Russell and Herschlag 1999), and the relatively lower protection seen at 10°C might reflect a similar phenomenon. However, the pattern of protection did not change at low temperature, so if a misfolded form is more populated, it is not sufficiently compact to detectably alter the protection pattern. At 50°C, protection factors were again slightly lower, and at 60°C most of the protection from hydroxyl-radical cleavage had disappeared (Fig. 2A). At 60°C and in the absence of Mg^{++}, the ribozyme had no detectable tertiary structure.

When assayed at 20°C, the ligase was essentially completely folded at Mg^{++} concentrations as low as 1 mM (Fig. 2B), indicating that the [Mg^{++}]$_{1/2(folding)}$ is below 1 mM. (Assaying for folding with less than 1 mM Mg^{++} could not be performed with this procedure because of trace amounts of free EDTA in probing reactions.) In comparison, the [Mg^{++}]$_{1/2(catalysis)}$ is much higher (40–50 mM), suggesting that a native structure can be achieved without Mg^{++} ions bound in every catalytically useful binding site (Glasner et al. 2002).

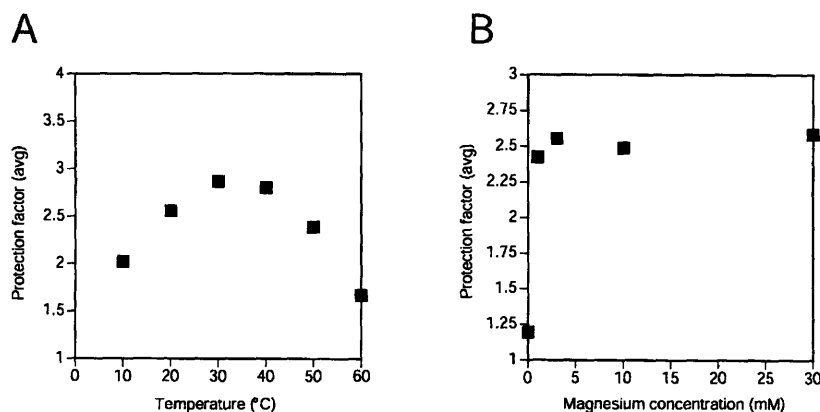Interestingly, removing Mg^{++} ions from the ribozyme

FIGURE 2. Temperature and Mg^{++} dependence of solvent accessibility in the class I ligase. (A) Protection factors were averaged for G45–C48 and G70–G75, and the average is plotted for experiments in which hydroxyl-radical cleavage was performed at varying temperatures in the presence of 10 mM Mg^{++}. (B) Averaged protection factors are shown for the positions noted in panel A in experiments probing the ligase at different Mg^{++} concentrations (at 22°C).

solution by adding EDTA did not completely denature the ribozyme's tertiary fold. Rather, the areas protected from cleavage when the ribozyme was folded in the presence of Mg^{++} were also partially protected by simply dropping the temperature from 60°C to 22°C in the absence of Mg^{++} (Fig. 1C). The presence of Mg^{++} changed the magnitude of each protection, but not the overall pattern of protection, suggesting that similar structure was present under both conditions. These data suggest that upon removal of Mg^{++}, the ribozyme stays folded into near native conformations.

Three-dimensional model

It had been shown that the hairpin loops capping helices P5 and P7 were not important for catalysis and that 5 nt could

be removed from the J1/3 junction without substantially impairing catalysis (Ekland and Bartel 1995). Thus, we initiated the modeling with a minimal ligase of 112 nt. Later, the whole ligase was assembled with residues 20–24 added in helical continuity with helix P1. Several models were built that accommodated the stereochemical constraints imposed by the two pseudoknots (helices P2 and P3) present in the secondary structure. Two of these models were consistent with the information contained in the Fe-EDTA protection data. However, one was favored because residues known to be crucial for activity were positioned near the catalytic site and because its topology was more conducive for folding (Fig. 3). In the alternative model, folding would have been problematic because the path of J3/4 was such that a knot would have been created with the formation of P2.
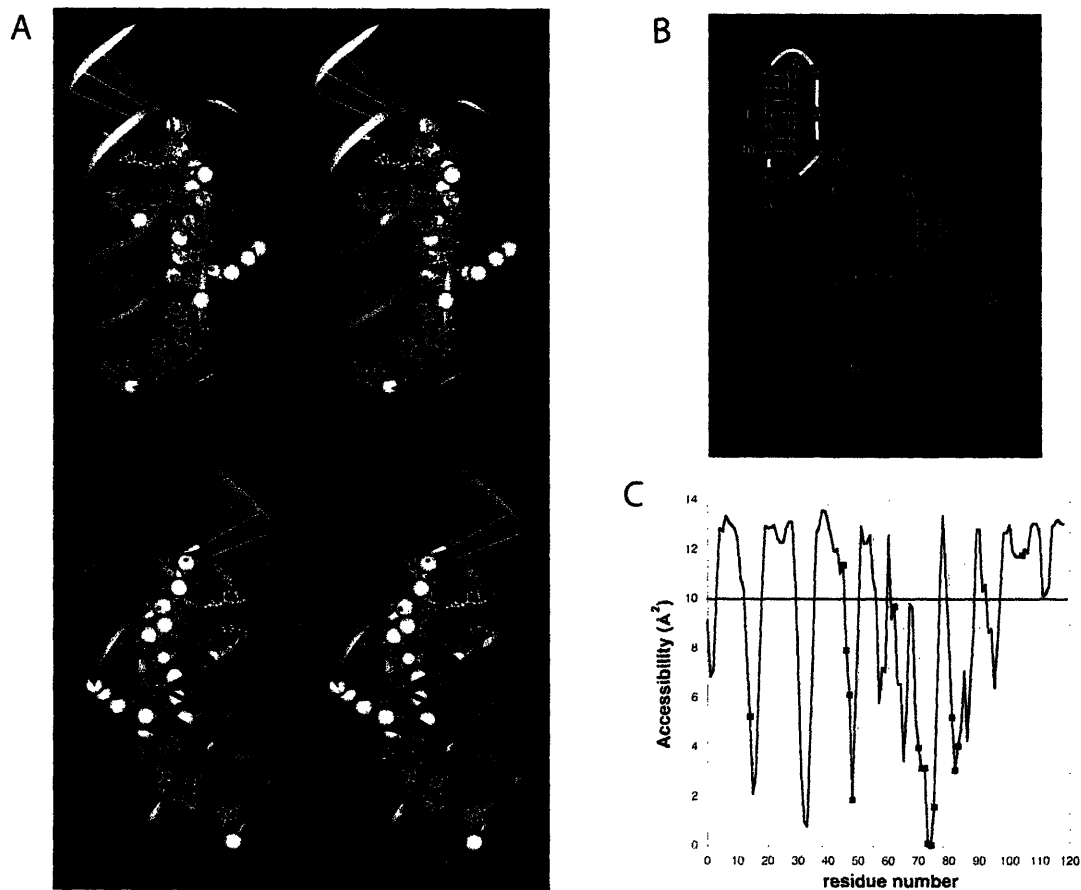

FIGURE 3. A three-dimensional model for the class I ligase. (A) Two stereo views of the model. White spheres indicate those residues protected from hydroxyl-radical cleavage. Residues of particular interest are highlighted with all-atom representation. These include the residues of the proposed tandom G·A pairs and residues that both were unpaired in previous representations of the ligase secondary structure (Fig. 1A) and were invariant among 25 active variants of the ligase (Ekland and Bartel 1995), all colored the same as the proximal paired regions. Also shown in all-atom representation are the four residues comprising the 2 bp flanking the ligation junction, colored according to the identity of the atoms. (B) A revised secondary structure diagram of the ligase that better reflects the arrangement of helices proposed by the tertiary model. Watson–Crick pairing is the same as in the original secondary structure (Fig. 1A), except for one additional base pair, G47:C111, near the catalytic site. The color scheme reflects that of the ribbon representations in panel A. (C) Solvent accessibilities of the C4' atoms in the modeled structure, as calculated using NACCES and a 1.4 Å sphere radius with an averaging over a window of three residues. A cut-off of 10 Å² between accessibility and nonaccessibility is indicated by a horizontal line.

The key element in the modeling of the ligase is the four-way junction constituted by helices P4, P5, P6, and P7. Numerous possibilities for the arrangement of these helices were explored. A search of the crystal structure of the 50S ribosomal subunit revealed a four-way junction in domain VI of the 23S rRNA of *Haloarcula marismortui* with similarities to the ligase four-way junction (Ban et al. 2000). The ligase junction is characterized by the presence of 2 nt between helices P5 and P6, the first one being variable and the second one invariably an A in the 25 isolates selected from a diverse pool of class I ligase variants (Ekland and Bartel 1995). The four-way junction in the 23S rRNA also contains single-stranded residues between the two stacks of helices: on one side two As and on the other side a single A (next to a non-Watson–Crick pair). In our favored model of the ligase (Fig. 3), the four-way junction was based on the framework of the 23S rRNA four-way junction, in that the helices interrupted by the two As correspond to P5 and P6 in the ligase four-way junction. That choice led to P4/P5 and P6/P7 stacks, the four-way junction being crossed so that loops L5 and L7 are brought in the same region of space.

The favored model is consistent with the Fe-EDTA mapping; these protections fit well with a calculated solvent accessibility profile based on the model (Fig. 3C). The model can also rationalize much of the conservation in sequence and pairing potential previously observed among the 25 isolates previously selected from the pool of ligase variants (Ekland and Bartel 1995). For example, the importance of pseudoknot P3 and junction J3/4 in the formation of the catalytic site is apparent.

It was rather natural to form an additional Watson–Crick pair between nt G47 and C111, both invariant among the 25 isolates. To test this predicted pairing possibility, mutants were constructed that disrupted and then restored the potential for Watson–Crick pairing. The ligation rates of the G47U mutant was 140-fold slower than the parent, and the rate of the C111A mutant was 41-fold slower, but the rate of the double mutant (G47U, C111A) was only 17-fold slower, supporting the idea that these residues are paired. Extending P4 with the G47:C111 pairing creates a bulged nucleotide, C48, which was invariant among the 25 isolates. In the model, this residue points towards the reacting G1.

The relative positioning of J3/4 with respect to the 5′ end of the molecule, which carries the reactive 5′-pppG1, suggested the presence of tandem sheared G·A base pairs, G2:A46 and A3:G45, which would constitute the third pseudoknot in the ligase ribozyme. Residues G45 and A46 in J3/4 are conserved among the 25 isolates previously selected from the pool of variants (Ekland and Bartel 1995). Evidence for the importance of G2 and A3 comes from studies of an engineered variant of the ribozyme able to perform short primer-extension reactions (Glasner et al. 2000). For example, this ribozyme utilizes the pppGGA trinucleotide substrate 1300 times more efficiently than the pppG single-

nucleotide substrate. G2 and A3 had been removed in the design of this ribozyme and thus the trinucleotide substrate restored in *trans* analogs of these 2 nt. The proposed G·A tandem explains the increased activity with the trinucleotide substrate, suggesting how G2 and G3 could position and anchor the reactive pppG.

In J1/3, several adenine residues are invariant among the 25 previously selected isolates (Ekland and Bartel 1995). These invariant As were modeled to make contacts in the shallow groove with P1 (A-minor type) and cradle between helices P1 and P3 after touching the 5′ end of helix P5, with residue C31 being the turning residue. The chosen path for J1/3 was such that the invariant adenine A34 faces the reactive pppG from the other side of C48.

Photo-cross-linking

The model predicts that stems P5 and P7 project outward from the core of the ribozyme in a nearly parallel fashion, positioning L5 next to L7. To test this prediction, we investigated whether a single 4-thiouridine (4SU) placed at the position of one loop could produce a cross-link to the adjacent loop. The most straightforward way to incorporate 4SU involves breaking the ribozyme into two strands of RNA at either of the loops. Cross-linking experiments were conducted using a ribozyme containing a break in L5 because a break in this loop leads to only a threefold drop in activity, whereas a break in L7 leads to a sixfold drop. 4SU was incorporated at the 3′ terminus of the upstream fragment (Fig. 4A), then the two-piece ribozyme was reconstituted, folded, and irradiated. Several catalytically active cross-linked molecules were identified based on their ability to ligate a radiolabeled substrate to themselves following irradiation (Fig. 4B). One of these cross-links was to the other strand of the two-piece ribozyme (Fig. 4C). This cross-link mapped to a short segment of L7 (Fig. 4D). Thus, molecules constrained such that the end of P5 is near to L7 retain ligation activity, as predicted by the model.

Conclusion

With the data presented here, we have been able to gain the first view of the tertiary structure of the class I ligase. As modeled, the structure of the ligase is quite compact, with extensive coaxial stacking and the P1 substrate helix lying on the four-way junction made of helices P4, P5, P6, and P7. In a general sense, the model is similar to the X-ray structures of several recently characterized RNAs, in that all are based on parallel arrangements of coaxially stacked helices (Cate et al. 1996; Ferre-D'Amare et al. 1998). In the case of the ligase, these stacks seem to be held close by the pseudoknotted secondary structure and conserved joining regions that wrap around much of the ribozyme.
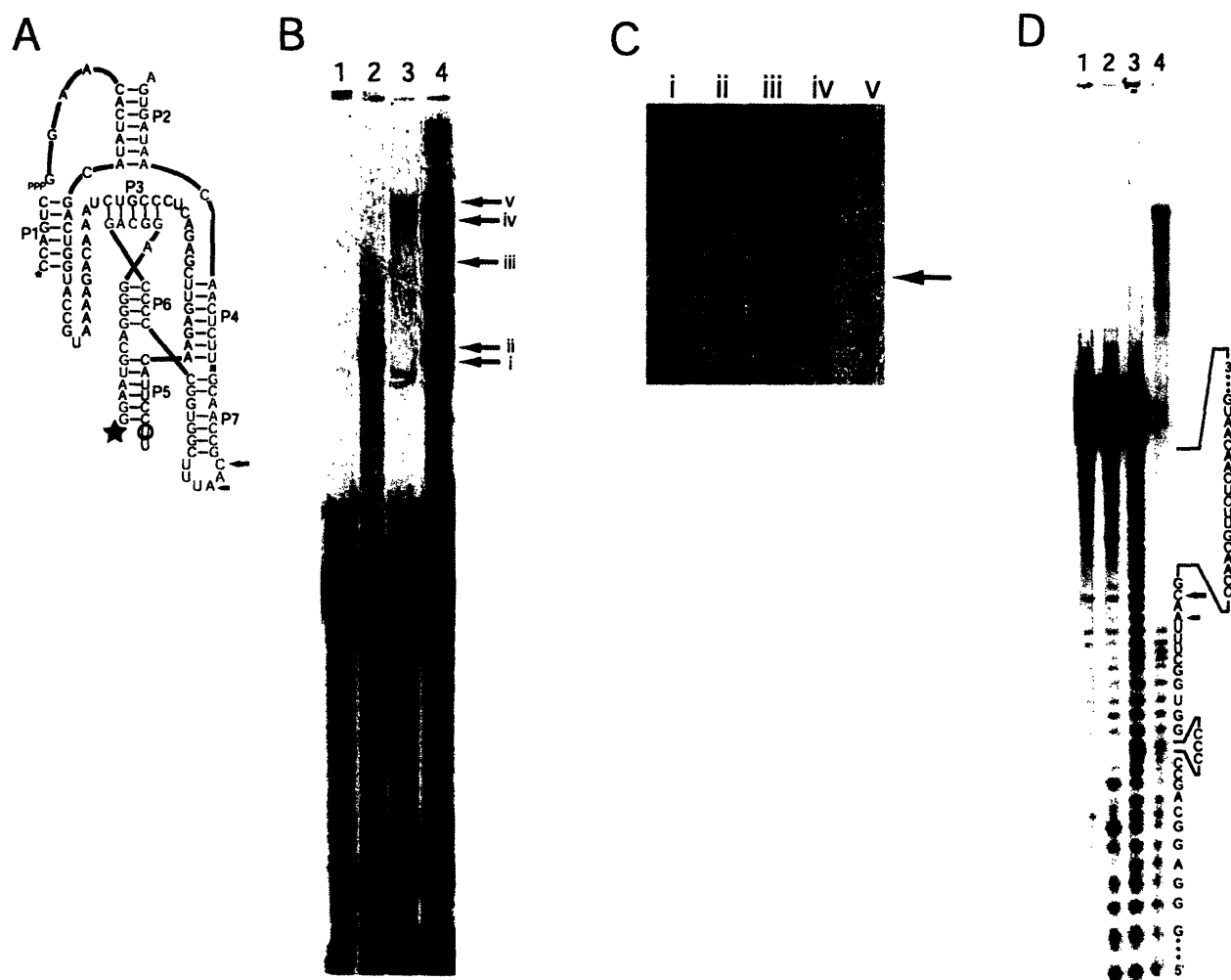
A   B   C   D

FIGURE 4. Crosslinking analyses of the class I ligase. (*A*) The secondary structure of the two-piece ribozyme used for the cross-linking study. The break in L5 facilitated insertion of a single 4SU residue (circled). The black sequence is fragment A, and the red sequence is fragment B. Stars indicate the sites of radiolabeling on the two different strands comprising the ribozyme (black ^{33}P, red ^{32}P). Arrows mark cross-linking sites, as mapped in panel *D*. (*B*) Separation of cross-linked ribozymes. (Lanes *1,3*) The two-piece ribozyme lacks 4SU. (Lanes *2,4*) The two-piece ribozyme contains 4SU. (Lanes *1,2*) The RNA was not irradiated. (Lanes *3,4*) Ribozymes were irradiated for 1 h before radiolabeled substrate was added. Lanes *3* and *4* contain about twice as much sample as lanes *1* and *2*. The arrows with Roman numerals point to products that were excised and eluted from gel slices for subsequent relabeling. (*C*) Relabeling of cross-linked RNAs. The lanes marked with Roman numerals correspond to the bands in panel *A*. Only the RNA in lane *iii* was appreciably relabeled. (*D*) Mapping of the cross-links within an active ribozyme. (Lane *1*) Unmodified, 5'-labeled fragment B RNA (red strand in *A*). (Lanes *2,3*) Digests of labeled fragment B RNA by RNase T1, and partial alkaline hydrolysis, respectively. (Lane *4*) The partial alkaline hydrolysis of the relabeled crosslinked RNA shown in panels *B* and *C*. Arrows mark the residues with the most pronounced transitions in ladder intensity, indicating the sites of the cross-links within the sequence of fragment B.

Almost all of the nucleotides that were protected from hydroxyl-radical cleavage are internalized in the modeled tertiary structure (Fig. 3). In addition, cross-linking data places L5 and L7 close together, supporting the model's predicted arrangement for the four-helix junction formed by helices P4-P5-P6-P7. The model suggested the presence of an additional Watson–Crick pair, G47:C111, which was subsequently supported experimentally. The model incorporates also a tandem of sheared G·A pairs that can rationalize previous data on the polymerase ribozyme showing the dramatic effect of GA nucleotides in the extended sub-

strate pppGGA compared to the single substrate pppG (Glasner et al. 2000).

Taken together, these results suggest that our view of the architecture of the ligase ribozyme is accurate and should be informative for future studies of the ligase. Indeed, the model was considered when designing the experiments that successfully generated a polymerase ribozyme with the ligase core as its catalytic domain (Johnston et al. 2001). For example, the model predicted that very long primer–template duplexes (analogous to long extensions of P1) would not clash with the ribozyme core. Indeed, the polymerase

ribozyme that emerged from these experiments was able to accommodate very long primer–template duplexes, making it a promising starting point for the eventual demonstration of RNA-catalyzed RNA replication.

MATERIALS AND METHODS

Ribozyme and substrate RNAs

The class I ligase ribozyme (GenBank no. U26413) was transcribed in vitro from a plasmid template linearized with EarI as described previously (Bergman et al. 2000). Transcripts were purified on 6% polyacrylamide/8 M urea gels and stored in water at −20°C. The RNA components of ligases containing breaks in either L5 or L7 were transcribed from synthetic DNA templates and purified in 10% polyacrylamide/8 M urea gels.

The ribozyme fragment with a single 4-thiouridine residue (fragment A) was prepared as follows: a synthetic DNA template for transcribing the 5′-terminal 52 nt (fragment A1) of fragment A was synthesized by standard phosphoramidite chemistry, except that 2-O-methyl phosphoramidites were used for the last two positions of the template to reduce the heterogeneity at the 3′-end of the RNA typical of RNA transcribed in vitro (Kao et al. 1999). Radiolabeled fragment A1 RNA was prepared by adding [^{33}P]α-UTP to the transcription reaction. The 3′-terminal 11 nt of fragment A (fragment A2) was purchased as an RNA oligonucleotide (5′-GAACAUUCC-[4SU]-U, Dharmacon Research). It included a terminal U nucleotide because 4SU-Controlled Pore Glass support beads were not available. Fragments A1 and A2 were ligated to form fragment A as follows: fragment A1 (12 μM), fragment A2 (32 μM), and a 20-nt-long DNA oligonucleotide that spans the junction of fragments A1 and A2 (24 μM) were heated to 80°C in 1 mM EDTA and 10 mM Tris-HCl (pH 7.5) for 2 min and allowed to cool slowly to room temperature (Moore and Sharp 1992). T4 DNA ligase buffer, 60 μM ATP, and 3.7 units/μL of T4 DNA ligase (USB) were added, and the ligation reaction was incubated overnight in the dark at room temperature. Fully ligated products were separated in 10% polyacrylamide/8 M urea gels. Ligation efficiency was ~60%, about twice that seen when fragment A was transcribed from a DNA template that lacked the 2′-methoxy substitutions.

The substrate for the ribozyme reaction was a synthetic RNA–DNA hybrid (5′-aaaCCAGUC, DNA bases lowercase; Bergman et al. 2000). It was radiolabeled using T4 polynucleotide kinase and either [^{32}P]γ-ATP or [^{33}P]γ-ATP (NEN). When substrate was used for cross-linking experiments, an additional 15-min "chase" phosphorylation reaction containing excess unlabeled ATP was performed to ensure that nearly all substrate molecules had a 5′-phosphate. RNA concentrations were measured spectrophotometrically at 260 nm, assuming an extinction coefficient that was the sum of those for the individual nucleotides.

Ribozyme assays

Ribozyme reactions in which the parent ligase and derivatives were compared were performed in 50 mM MES (pH 6.0), 60 mM MgCl$_2$, 200 mM KCl, and 600 μM EDTA at 22°C. In all cases, the ribozyme was heated (80°C, 2 min, in H$_2$O) and then cooled (22°C, 2 min) just prior to initiation of the reaction. Ligation

reactions were initiated by addition of buffer, salts, and trace ^{32}P-labeled substrate. Aliquots were taken at appropriate time points and added to 2 volumes stop solution containing 120 mM EDTA and 8 M urea. Product and substrate were separated using 20% polyacrylamide gels and quantified by phosphorimaging. Ligation rates were calculated as described previously (Bergman et al. 2000).

Hydroxyl-radical probing

About 30% of the ligase molecules were not active upon initial folding (Schmitt and Lehman 1999; Bergman et al. 2000). To avoid probing the fold of misfolded ribozymes, ligase molecules were incubated with radiolabeled substrate, so that those that had assumed the active fold acquired the radiolabel and became visible in our analysis. Ribozyme (1 μM final concentration) was incubated with trace radiolabeled substrate (<100 nM final concentration) in buffer containing 50 mM BES (pH 7.0), 10 mM MgCl$_2$, 200 mM KCl, and 0.1 mM EDTA. The ribozyme was heated (80°C, 2 min, in H$_2$O) and then cooled (22°C, 2 min) just prior to initiation of the reaction. Ligation reactions were incubated for 10 min at 22°C, at which point the reaction was diluted 10-fold into 50 mM BES (pH 7.0) buffer with MgCl$_2$ sufficient to bring the concentration of Mg^{++} to that indicated. When a final concentration of 0 mM Mg^{++} was desired, the RNA was diluted 10-fold into a solution containing 50 mM BES (pH 7.0) and 5 mM EDTA.

Labeled ligation product was subjected to hydroxyl-radical cleavage by adding to the RNA solution 0.1 volumes of a solution containing 20 mM (NH$_4$)$_2$Fe(II)(SO$_4$)$_2$, 20 mM ascorbic acid, and 22 mM EDTA. Solutions of (NH$_4$)$_2$Fe(II)(SO$_4$)$_2$ and ascorbic acid were prepared fresh before each experiment. Cleavage reactions were typically for 15 min at 22°C. For probing at other temperatures, the RNA was also allowed to equilibrate for 15 min at the desired temperature prior to addition of the Fe/Ascorbate/EDTA solution. In the range of 10°C–60°C, changing temperature did not noticeably affect the overall extent of cleavage. After addition of 2 volumes of a solution containing 8 M urea and 25 mM EDTA, cleaved RNAs were separated in 10% polyacrylamide gels and quantified by phosphorimaging (BAS2000, Fuji). Because a single gel provided accurate data on only ~50 nt, reactions were typically loaded several times and electrophoresed for times ranging from 45 min to 4 h to access most portions of the ribozyme. Individual positions were identified by comparison to partial alkaline hydrolysis and partial RNase T1 digestion ladders. Cleavage at each position was normalized to allow for differences in gel loading and cleavage efficiency (although cleavage was always done at levels in which a very small percentage of the ligase molecules were cleaved), and protection factors were calculated for nt 7–115. A protection factor was defined as the amount of cleavage at position N under denaturing conditions (60°C, 0 mM Mg^{++}) divided by the amount of cleavage at position N under folded or experimental conditions (Pan 1995). The level of protection usually varied less than 30% from day to day, and the same positions were protected in each experiment.

Modeling of the class I ligase

Molecular modeling was performed as described previously (Westhof 1993; Massire and Westhof 1998), using the programs

MANIP with FRAGMENT. The model was refined with the restrained least-squares program NUCLIN-NUCLSQ. Figures were produced using the program DRAWNA (Massire et al. 1994). The coordinates of the modeled structure of the ribozyme have been deposited in the Protein Data Bank and are available under accession number 1QXI (RCSB020188).

Isolation and mapping of active, 4SU–cross-linked ribozymes

The 4SU cross-linking agent was chosen because it has been used successfully in a variety of settings (Sontheimer and Steitz 1993; Dontsova et al. 1994; Christian and Harris 1999) and because it generates photo-dependent cross-links without a long linker arm (for review, see Favre et al. 1998). A 4SU was appended to the 3' end of the upstream half of the ribozyme (fragment A), and the phosphates were removed from the downstream half of the ribozyme (fragment B; Fig. 4A). Full ribozymes were assembled by mixing equimolar fragment A and fragment B RNAs, then heating to 80°C in water and cooling to 22°C. Reaction buffer (10 mM $MgCl_2$, 200 mM KCl, 600 µM EDTA, and 50 mM MES at pH 7.0) was added (bringing the ribozyme concentration to 1.2 µM), and solutions containing split ribozymes were then placed in a microtiter plate that was cooled to 4°C and irradiated with a UV transilluminator (UVP) set at 302 nM. A polystyrene petri dish was used to filter out wavelengths lower than 300 nM. After irradiation for 1 h, ^{33}P-labeled substrate was added (200 nM final concentration) and ribozymes were allowed to react for 1 h in the dark at room temperature. Note that this substrate was prepared with a final phosphorylation step using an excess of unlabeled ATP to block the free 5'-OH of the substrate. After allowing the ligation reaction to proceed for 1 h at 22°C, 2 volumes of a stop solution containing 8 M urea and 25 mM EDTA were added, and the RNAs were resolved in a 10% or 12% polyacrylamide/8 M urea gel. Control experiments were performed in parallel omitting irradiation and/or the 4SU substitution.

When the 4SU-containing ribozyme mixture was analyzed following irradiation, five bands were detected above the major band corresponding to the uncross-linked, reacted ribozyme (Fig. 4B, lane 4). Quantitation of the five bands indicated that each band contained approximately the same amount of radioactivity, and each was less than 1% of counts corresponding to the uncross-linked, ligated product. Two of the five bands were observed in the control reaction in which UV irradiation was omitted (Fig. 4B, lane 2), and the counts from these bands did not increase with irradiation. No bands were seen in control lanes showing ribozymes constructed without a 4SU residue, so it appeared that the three slowest migrating bands contained UV- and 4SU-dependent cross-links. All five cross-linked RNAs were excised from the gel, eluted, and precipitated in ethanol and then used as substrates in a kinase reaction, this time using [^{32}P]-γATP instead of [^{33}P]-γATP. This second radiolabeling identified products that cross-linked to the downstream RNA strand (fragment B), because only this type of cross-link would contain a free 5'-hydroxyl for labeling.

Only one of the five isolated RNAs was efficiently radiolabeled (Fig. 4C, lane iii), and this product was purified from a 6% polyacrylamide/8 M urea gel. Inactive, cross-linked molecules should migrate differently in gels of different acrylamide percentages, so

changing the gel percentage between the first and second purifications allowed the active cross-linked ribozymes to be better separated from inactive molecules. The relabeled product was then subjected to partial alkaline hydrolysis, and this reaction was run in a 10% polyacrylamide/8 M urea gel (Fig. 4D). The point at which cross-linking occurred was mapped by comparison to partial alkaline hydrolysis and partial RNase T1 ladders generated from uncross-linked fragment B labeled with [^{32}P]γ-ATP (Fig. 4D).

REFERENCES

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **289:** 905–920.

Bartel, D.P. 1999. Re-creating an RNA replicase. In *The RNA world,* 2nd ed. (eds. R.F. Gesteland et al.), pp. 143–162. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bergman, N.H., Johnston, W.K., and Bartel, D.P. 2000. Kinetic framework for ligation by an efficient RNA ligase ribozyme. *Biochemistry* **39:** 3115–3123.

Cate, J.H., Gooding, A.R., Podell, E., Zhou, K., Golden, B.L., Kundrot, C.E., Cech, T.R., and Doudna, J.A. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* **273:** 1678–1685.

Celander, D.W. and Cech, T.R. 1990. Iron(II)-ethylenediaminetetraacetic acid catalyzed cleavage of RNA and DNA oligonucleotides: Similar reactivity toward single- and double-stranded forms. *Biochemistry* **29:** 1355–1361.

———. 1991. Visualizing the higher order folding of a catalytic RNA molecule. *Science* **251:** 401–407.

Cheong, C., Varani, G., and Tinoco Jr., I. 1990. Solution structure of an unusually stable RNA hairpin, 5'GGAC(UUCG)GUCC. *Nature* **346:** 680–682.

Christian, E.L. and Harris, M.E. 1999. The track of the pre-tRNA 5' leader in the ribonuclease P ribozyme–substrate complex. *Biochemistry* **38:** 12629–12638.

Dontsova, O., Tishkov, V., Dokudovskaya, S., Bogdanov, A., Doring, T., Rinke-Appel, J., Thamm, S., Greuer, B., and Brimacombe, R. 1994. Stem-loop IV of 5S rRNA lies close to the peptidyltransferase center. *Proc. Natl. Acad. Sci.* **91:** 4125–4129.

Eckstein, F. 1985. Nucleoside phosphorothioates. *Annu. Rev. Biochem.* **54:** 367–402.

Ekland, E.H. and Bartel, D.P. 1995. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23:** 3231–3238.

———. 1996. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382:** 373–376.

Ekland, E.H., Szostak, J.W., and Bartel, D.P. 1995. Structurally complex and highly active RNA ligases derived from random RNA

sequences. *Science* **269**: 364–370.

Favre, A., Saintome, C., Fourrey, J.L., Clivio, P., and Laugaa, P. 1998. Thionucleobases as intrinsic photoaffinity probes of nucleic acid structure and nucleic acid–protein interactions. *J. Photochem. Photobiol. B* **42**: 109–124.

Ferre-D'Amare, A.R., Zhou, K., and Doudna, J.A. 1998. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**: 567–574.

Glasner, M.E., Yen, C.C., Ekland, E.H., and Bartel, D.P. 2000. Recognition of nucleoside triphosphates during RNA-catalyzed primer extension. *Biochemistry* **39**: 15556–15562.

Glasner, M.E., Bergman, N.H., and Bartel, D.P. 2002. Metal ion requirements for structure and catalysis of an RNA ligase ribozyme. *Biochemistry* **41**: 8103–8112.

Hertzberg, R.P. and Dervan, P.B. 1984. Cleavage of DNA with methidiumpropyl-EDTA-iron(II): Reaction conditions and product analyses. *Biochemistry* **23**: 3934–3945.

Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E., and Bartel, D.P. 2001. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* **292**: 1319–1325.

Joseph, S. and Noller, H.F. 2000. Directed hydroxyl radical probing using iron(II) tethered to RNA. *Methods Enzymol.* **318**: 175–190.

Joyce, G.F. and Orgel, L.E. 1999. Prospects for understanding the origin of the RNA world. In *The RNA world*, 2nd ed. (eds. R.F. Gesteland et al.), pp. 49–77. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Kao, C., Zheng, M., and Rudisser, S. 1999. A simple and efficient method to reduce nontemplated nucleotide addition at the 3 terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**: 1268–1272.

King, P.A., Jamison, E., Strahs, D., Anderson, V.E., and Brenowitz, M. 1993. 'Footprinting' proteins on DNA with peroxonitrous acid. *Nucleic Acids Res.* **21**: 2473–2478.

Latham, J.A. and Cech, T.R. 1989. Defining the inside and outside of a catalytic RNA molecule. *Science* **245**: 276–282.

Massire, C. and Westhof, E. 1998. MANIP: An interactive tool for modelling RNA. *J. Mol. Graph. Model* **16**: 197–205, 255–257.

Massire, C., Gaspin, C., and Westhof, E. 1994. DRAWNA: A program for drawing schematic views of nucleic acids. *J. Mol. Graph* **12**: 201–206.

Moore, M.J. and Sharp, P.A. 1992. Site-specific modification of pre-mRNA: The 2'-hydroxyl groups at the splice sites. *Science* **256**: 992–997.

Pan, T. 1995. Higher Order folding and domain analysis of the ribozyme from *Bacillus subtilis* ribonuclease P. *Biochemistry* **34**: 902–909.

Pan, T. and Sosnick, T.R. 1997. Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat. Struct. Biol.* **4**: 931–938.

Russell, R. and Herschlag, D. 1999. New pathways in folding of the *Tetrahymena* group I RNA enzyme. *J. Mol. Biol.* **291**: 1155–1167.

Schmitt, T. and Lehman, N. 1999. Non-unity molecular heritability demonstrated by continuous evolution *in vitro*. *Chemis. Biol.* **6**: 857–869.

Sclavi, B., Woodson, S., Sullivan, M., Chance, M., and Brenowitz, M. 1998. Following the folding of RNA with time-resolved synchrotron X-ray footprinting. *Methods Enzymol.* **295**: 379–402.

Sontheimer, E.J. and Steitz, J.A. 1993. The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science* **262**: 1989–1996.

Tullius, T.D. and Dombroski, B.A. 1986. Hydroxyl radical "footprinting": High-resolution information about DNA-protein contacts and application to lambda repressor and Cro protein. *Proc. Natl. Acad. Sci.* **83**: 5469–5473.

Tullius, T.D., Dombroski, B.A., Churchill, M.E., and Kam, L. 1987. Hydroxyl radical footprinting: A high-resolution method for mapping protein-DNA contacts. *Methods Enzymol.* **155**: 537–558.

Westhof, E. 1993. Modeling the three-dimensional structure of ribonucleic acids. *J. Mol. Struct.* **286**: 203–210.

Wu, J.C., Kozarich, J.W., and Stubbe, J. 1983. The mechanism of free base formation from DNA by bleomycin. A proposal based on site specific tritium release from Poly(dA.dU). *J. Biol. Chem.* **258**: 4694–4697.
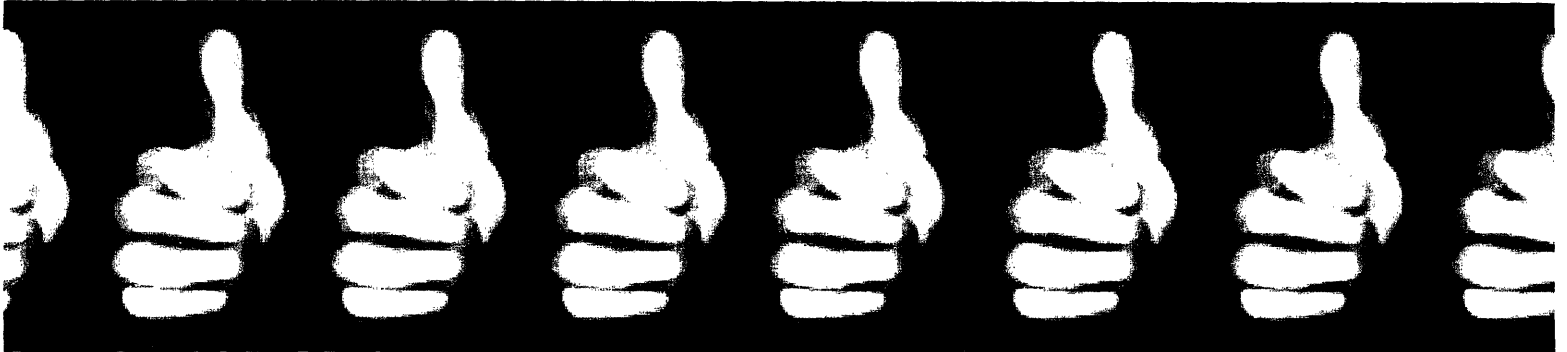
CENSORS of the
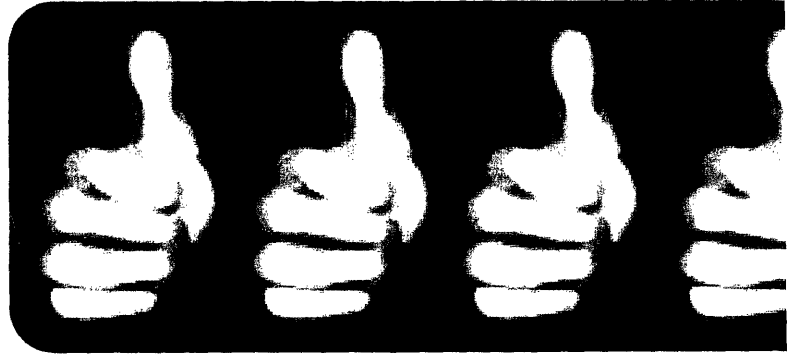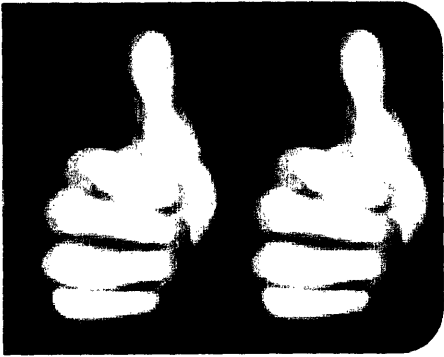
Biologists have
been surprised
to discover
that most animal
and plant cells contain
a built-in system
to silence
individual genes
by shredding the
RNA they produce.
Biotech companies
are already working
to exploit it

Genome

By Nelson C. Lau and David P. Bartel

O bserved on a microscope slide, a living cell appears serene. But underneath its tranquil facade, it buzzes with biochemical chatter. The DNA genome inside every cell of a plant or animal contains many thousands of genes. Left to its own devices,

the transcription machinery of the cell would express every gene in the genome at once: unwinding the DNA double helix, transcribing each gene into single-stranded messenger RNA and, finally, translating the RNA messages into their protein forms.

No cell could function amid the resulting cacophony. So cells muzzle most genes, allowing an appropriate subset to be heard. In most cases, a gene's DNA code is transcribed into messenger RNA only if a particular protein assemblage has docked onto a special regulatory region in the gene.

Some genes, however, are so subversive that they should never be given freedom of expression. If the genes from mobile genetic elements were to successfully broadcast their RNA messages, they could jump from spot to spot on the DNA, causing cancer or other diseases. Similarly, viruses, if allowed to express their messages unchecked, will hijack the cell's protein production facilities to crank out viral proteins.

Cells have ways of fighting back. For example, biologists long ago identified a system, the interferon response, that human cells deploy when viral genes enter a cell. This response can shut off almost all gene expression, analogous to stopping the presses. And just

within the past several years, scientists have discovered a more precise and—for the purposes of research and medicine—more powerful security apparatus built into nearly all plant and animal cells. Called RNA interference, or RNAi, this system acts like a censor. When a threatening gene is expressed, the RNAi machinery silences it by intercepting and destroying only the offender's messenger RNA, without disturbing the messages of other genes.

As biologists probe the modus operandi of this cellular censor and the stimuli that spur it into action, their fascination and excitement are growing. In principle, scientists might be able to invent ways to direct RNA interference to stifle genes involved in cancer, viral infection or other diseases. If so, the technology could form the basis for a new class of medicines.

Meanwhile researchers working with plants, worms, flies and other experimental organisms have already learned how to co-opt RNAi to suppress nearly any gene they want to study, allowing them to begin to deduce the gene's purpose. As a research tool, RNAi has been an immediate success, allowing hundreds of laboratories to tackle questions that were far beyond their reach just a few years ago.

JANA BRENNING
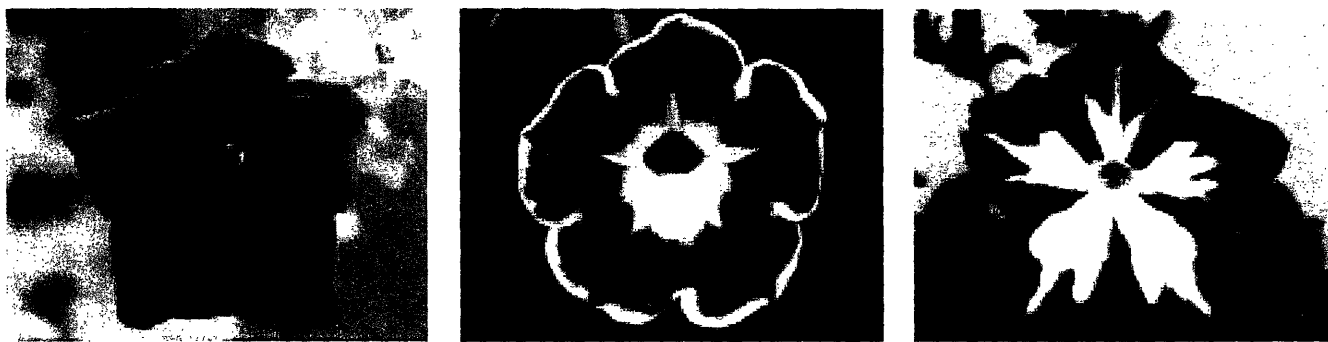
PURPLE PETUNIAS offered the first clues to the existence of gene censors in plants. When extra pigment genes were inserted into normal plants (*left*), the flowers that emerged ended up with areas that strangely lacked color (*center* and *right*).

Whereas most research groups are using RNA interference as a means to an end, some are investigating exactly how the phenomenon works. Other labs (including our own) are uncovering roles for the RNAi machinery in the normal growth and development of plants, fungi and animals—humans among them.

A Strange Silence

THE FIRST HINTS of the RNAi phenomenon surfaced 13 years ago. Richard A. Jorgensen, now at the University of Arizona, and, independently, Joseph Mol of the Free University of Amsterdam inserted into purple-flowered petunias additional copies of their native pigment gene. They were expecting the engineered plants to grow flowers that were even more vibrantly violet. But instead they obtained blooms having patches of white.

Jorgensen and Mol concluded that the extra copies were somehow triggering censorship of the purple pigment genes—including those natural to the petunias—resulting in variegated or even albino-like flowers. This dual censorship of an inserted gene and its native counterpart, called co-suppression, was later seen in fungi, fruit flies and other organisms.

Clues to the mystery of how genes were being silenced came a few years later from William G. Dougherty's lab at Oregon State University. Dougherty and his colleagues started with tobacco plants that had been engineered to contain within their

Overview/ *RNA Interference*

- Scientists have long had the ability to introduce altered genes into experimental organisms. But only within the past few years have they discovered a convenient and effective way to turn off a specific gene inside a cell.
- It turns out that nearly all plant and animal cells have internal machinery that uses unusual forms of RNA, the genetic messenger molecule, to naturally silence particular genes.
- This machinery has evolved both to protect cells from hostile genes and to regulate the activity of normal genes during growth and development. Medicines might also be developed to exploit the RNA interference machinery to prevent or treat diseases.

DNA several copies of the CP (coat protein) gene from tobacco etch virus. When these plants were exposed to the virus, some of the plants proved immune to infection. Dougherty proposed that this immunity arose through co-suppression. The plants apparently reacted to the initial expression of their foreign CP genes by shutting down this expression and subsequently also blocking expression of the CP gene of the invading virus (which needed the coat protein to produce an infection). Dougherty's lab went on to show that the immunity did not require synthesis of the coat protein by the plants; something about the RNA transcribed from the CP gene accounted for the plants' resistance to infection.

The group also showed that not only could plants shut off specific genes in viruses, viruses could trigger the silencing of selected genes. Some of Dougherty's plants did not suppress their CP genes on their own and became infected by the virus, which replicated happily in the plant cells. When the researchers later measured the RNA being produced from the CP genes of the affected plants, they saw that these messages had nearly vanished—infection had led to the CP genes' inactivation.

Meanwhile biologists experimenting with the nematode *Caenorhabditis elegans,* a tiny, transparent worm, were puzzling over their attempts to use "antisense" RNA to inactivate the genes they were studying. Antisense RNA is designed to pair up with a particular messenger RNA sequence in the same way that two complementary strands of DNA mesh to form a double helix. Each strand in DNA or RNA is a chain of nucleotides, genetic building blocks represented by the letters A, C, G and either U (in RNA) or T (in DNA). C nucleotides link up with Gs, and As pair with Us or Ts. A strand of antisense RNA binds to a complementary messenger RNA strand to form a double-stranded structure that cannot be translated into a useful protein.

Over the years, antisense experiments in various organisms have had only spotty success. In worms, injecting antisense RNAs seemed to work. To everyone's bewilderment, however, "sense" RNA also blocked gene expression. Sense RNA has the same sequence as the target messenger RNA and is therefore unable to lock up the messenger RNA within a double helix.

The stage was now set for the eureka experiment, performed five years ago in the labs of Andrew Z. Fire of the Carnegie Institution of Washington and Craig C. Mello of the University of Massachusetts Medical School. Fire and Mello guessed that the previous preparations of antisense and sense RNAs that were being injected into worms were not totally pure. Both mixtures probably contained trace amounts of double-stranded RNA. They

suspected that the double-stranded RNA was alerting the censors.

To test their idea, Fire, Mello and their colleagues inoculated nematodes with either single- or double-stranded RNAs that corresponded to the gene *unc-22*, which is important for muscle function. Relatively large amounts of single-stranded *unc-22* RNA, whether sense or antisense, had little effect on the nematodes. But surprisingly few molecules of double-stranded *unc-22* RNA caused the worms—and even the worms' offspring—to twitch uncontrollably, an unmistakable sign that something had started interfering with *unc-22* gene expression. Fire and Mello observed the same amazingly potent silencing effect on nearly every gene they targeted, from muscle genes to fertility and viability genes. They dubbed the phenomenon "RNA interference" to convey the key role of double-stranded RNA in initiating censorship of the corresponding gene.

Investigators studying plants and fungi were also closing in on double-stranded RNA as the trigger for silencing. They showed that RNA strands that could fold back on themselves to form long stretches of double-stranded RNA were potent inducers of silencing. And other analyses revealed that a gene that enables cells to convert single-stranded RNA into double-stranded RNA was needed for co-suppression. These findings suggested that Jorgensen and Mol's petunias recognized the extra pigment genes as unusual (through a mechanism that is still

GLOWING NEMATODES proved that RNA interference operates in animals as well as plants. When worms whose cells express a gene for a fluorescent protein (*left*) were treated with double-stranded RNA corresponding to the gene, the glow was extinguished (*right*).

normal and viral—and the enzyme RNAse L indiscriminately destroys the messenger RNAs. These responses to double-stranded RNA are considered components of the so-called interferon response because they are triggered more readily after the cells have been exposed to interferons, molecules that infected cells secrete to signal danger to neighboring cells.

Unfortunately, when researchers put artificial double-stranded RNAs (like those used to induce RNA interference in worms and flies) into the cells of mature mammals, the interferon response indiscriminately shuts down every gene in the cell. A deeper understanding of how RNA interference works was needed before it could be used routinely without setting off the interferon alarms. In addition to the pioneering researchers already mentioned, Thomas Tuschl of the Rockefeller University,
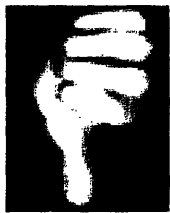
FEW MOLECULES

TWITCH UNCONTROLLABLY

mysterious) and converted their messenger RNAs into double-stranded RNA, which then triggered the silencing of both the extra and native genes. The concept of a double-stranded RNA trigger also explains why viral infection muzzled the CP genes in Dougherty's plants. The tobacco etch virus had created double-stranded RNA of its entire viral genome as it reproduced, as happens with many viruses. The plant cells responded by cutting off the RNA messages of all genes associated with the virus, including the CP genes incorporated into the plant DNA.

Biologists were stunned that such a powerful and ubiquitous system for regulating gene expression had escaped their notice for so long. Now that the shroud had been lifted on the phenomenon, scientists were anxious to analyze its mechanism of action and put it to gainful employment.

Slicing and Dicing Genetic Messages

RNA INTERFERENCE was soon observed in algae, flatworms and fruit flies—diverse branches of the evolutionary tree. Demonstrating RNAi within typical cells of humans and other mammals was considerably trickier, however.

When a human cell is infected by viruses that make long double-stranded RNAs, it can slam into lockdown mode: an enzyme known as PKR blocks translation of all messenger RNAs—both

Phillip D. Zamore of the University of Massachusetts Medical School, Gregory Hannon of Cold Spring Harbor Laboratory in New York State and many others have added to our current understanding of the RNA interference mechanism.

RNAi appears to work like this: Inside a cell, double-stranded RNA encounters an enzyme dubbed Dicer. Using the chemical process of hydrolysis, Dicer cleaves the long RNA into pieces, known as short (or small) interfering RNAs, or siRNAs. Each siRNA is about 22 nucleotides long.

Dicer cuts through both strands of the long double-stranded RNA at slightly staggered positions so that each resulting siRNA has two overhanging nucleotides on one strand at either

NELSON C. LAU and DAVID P. BARTEL have been studying microRNAs and other small RNAs that regulate the expression of genes. Lau is completing a doctoral degree at the Whitehead Institute and the Massachusetts Institute of Technology. Bartel started his research group at the Whitehead Institute in 1994, after earning a Ph.D. at Harvard University. Bartel is also an associate professor at M.I.T. and a co-founder of Alnylam Pharmaceuticals, which is developing RNAi-based therapeutics. Lau and Bartel are among the recipients of the 2002 AAAS Newcomb Cleveland Prize.

NORMAL GENE EXPRESSION

CELL NUCLEUS

DNA

DNA template strand

Complementary base pair

mRNA

Protein

Ribosome
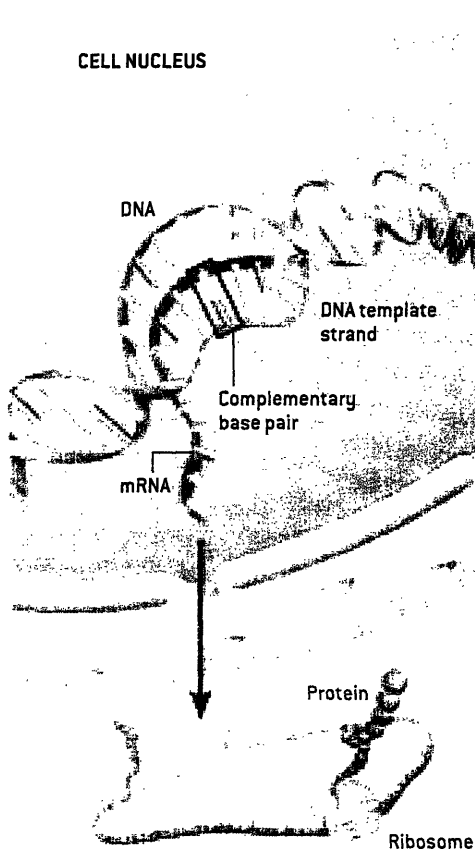
GLOWING CELLS reveal successful translation of a gene (encoding the lamin protein) into protein form.

TRIGGERS OF RNA SILENCING

a Double-stranded RNA from mobile genetic elements or abnormal genes

Dicer

Dicer cleaves RNA

c MicroRNA precursor

Dicer

siRNA or microRNA

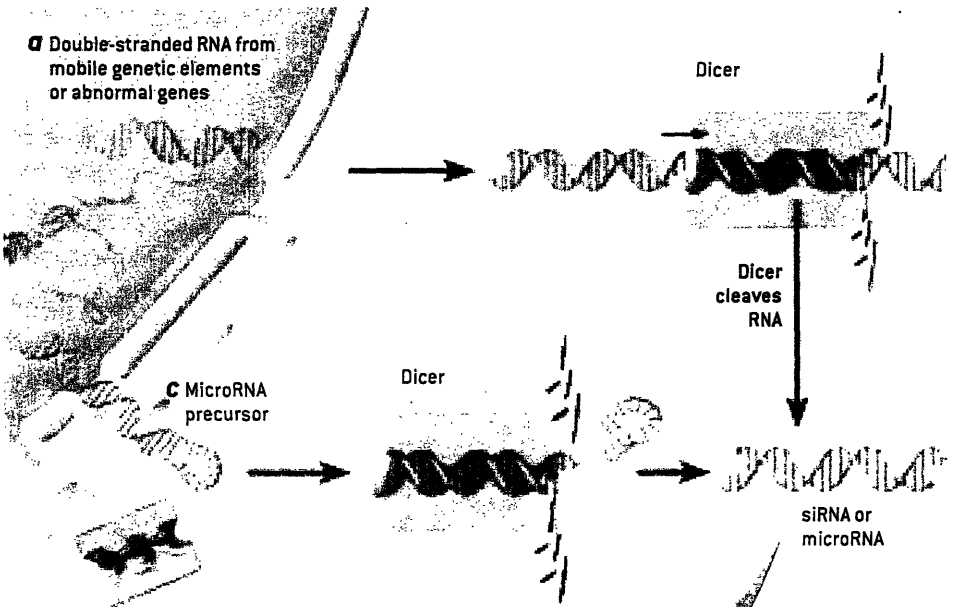## HOW RNAi SUPPRESSES GENE EXPRESSION

Single strand of siRNA or microRNA

Cell unwinds RNA strands

mRNA

---

end [*see box above*]. The siRNA duplex is then unwound, and one strand of the duplex is loaded into an assembly of proteins to form the RNA-induced silencing complex (RISC).

Within the silencing complex, the siRNA molecule is positioned so that messenger RNAs can bump into it. The RISC will encounter thousands of different messenger RNAs that are in a typical cell at any given moment. But the siRNA of the RISC will adhere well only to a messenger RNA that closely complements its own nucleotide sequence. So, unlike the interferon response, the silencing complex is highly selective in choosing its target messenger RNAs.

When a matched messenger RNA finally docks onto the siRNA, an enzyme known as Slicer cuts the captured messenger RNA strand in two. The RISC then releases the two messenger RNA pieces (now rendered incapable of directing protein synthesis) and moves on. The RISC itself stays intact, free to find and cleave another messenger RNA. In this way, the RNAi censor uses bits of the double-stranded RNA as a black-

list to identify and mute corresponding messenger RNAs.

David C. Baulcombe and his co-workers at the Sainsbury Laboratory in Norwich, England, were the first to spot siRNAs, in plants. Tuschl's group later isolated them from fruit fly embryos and demonstrated their role in gene silencing by synthesizing artificial siRNAs and using them to direct the destruction of messenger RNA targets. When that succeeded, Tuschl wondered whether these short snippets of RNA might slip under the radar of mammalian cells without setting off the interferon response, which generally ignores double-stranded RNAs that are shorter than 30 nucleotide pairs. He and his co-workers put synthetic siRNAs into cultured mammalian cells, and the experiment went just as they expected. The target genes were silenced; the interferon response never occurred.
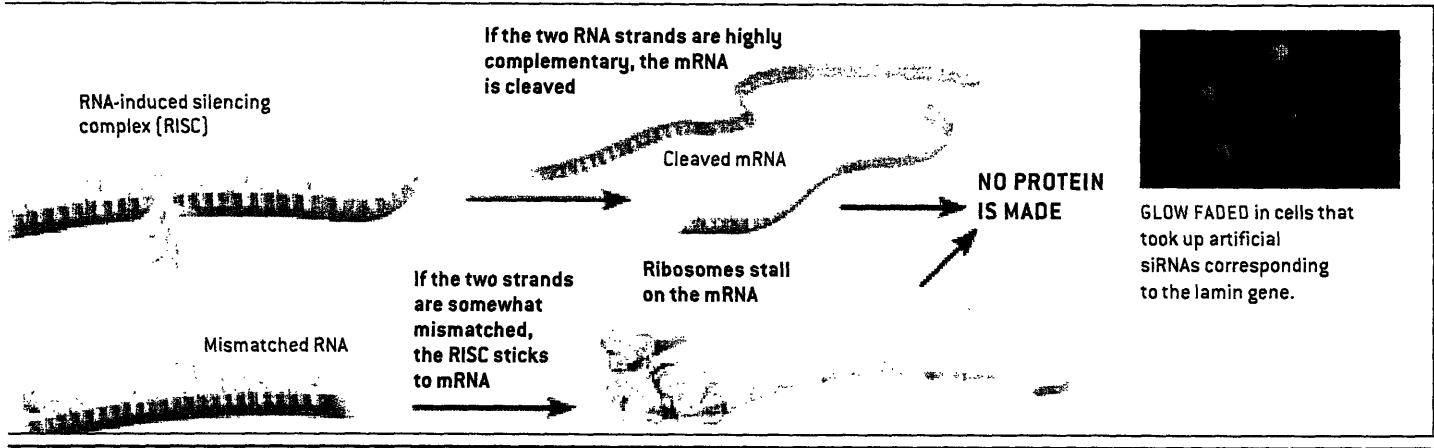
Tuschl's findings rocked the biomedical community. Geneticists had long been able to introduce a new gene into mammalian cells by, for example, using viruses to ferry the gene into cells. But it would take labs months of labor to knock out a gene

b Double-stranded RNA from a replicating virus

d Artificial siRNAs delivered by lipid molecules
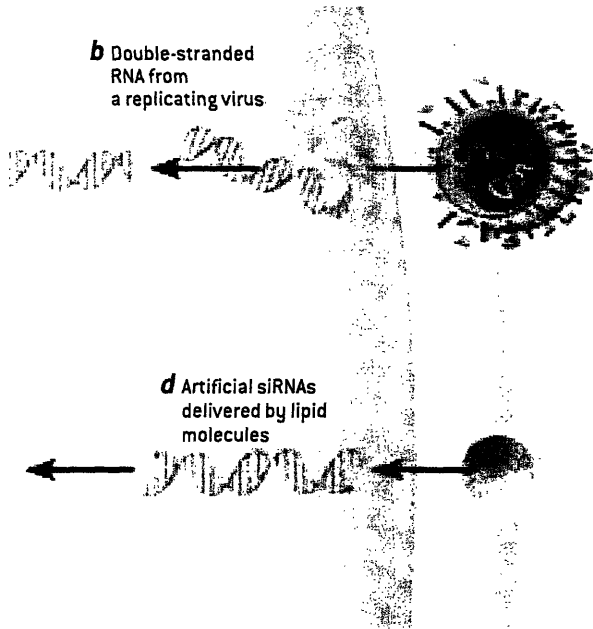
A CELL CAN CENSOR the expression of an individual gene inside it by interfering with the messenger RNA (mRNA) transcribed from the offending gene, thus preventing the RNA from being decoded by ribosomes into active protein, as normally happens (*left panel*). The censorship machinery is triggered by small, double-stranded RNA molecules with ragged ends. An enzyme called Dicer chemically snips such short interfering RNAs (siRNAs) from longer double-stranded RNAs produced by self-copying genetic sequences (*a*) or viruses (*b*). Regulatory RNA sequences known as microRNA precursors (*c*) are also cleaved by Dicer into this short form. And scientists can use lipid molecules to insert artificial siRNAs into cells (*d*).

The RNA fragments separate into individual strands (*bottom panel*), which combine with proteins to form an RNA-induced silencing complex (RISC). The RISC then captures mRNA that complements the short RNA sequence. If the match is essentially perfect, the captive message is sliced into useless fragments (*top row*); less perfect matches elicit a different response. For instance, they may cause the RISC to block ribosome movements and thus halt translation of the message into protein form (*bottom row*).

RNA-induced silencing complex (RISC)

If the two RNA strands are highly complementary, the mRNA is cleaved

Cleaved mRNA

NO PROTEIN IS MADE

GLOW FADED in cells that took up artificial siRNAs corresponding to the lamin gene.

Mismatched RNA

If the two strands are somewhat mismatched, the RISC sticks to mRNA

Ribosomes stall on the mRNA

of interest to ascertain the gene's function. Now the dream of easily silencing a single, selected gene in mammalian cells was suddenly attainable. With siRNAs, almost any gene of interest can be turned off in mammalian cell cultures—including human cell lines—within a matter of hours. And the effect persists for days, long enough to complete an experiment.

A Dream Tool

AS HELPFUL AS RNA interference has become to mammal biologists, it is even more useful at the moment to those who study lower organisms. A particular bonus for those studying worms and plants is that in these organisms the censorship effect is amplified and spread far from the site where the double-stranded RNA was introduced. This systemic phenomenon has allowed biologists to exploit RNAi in worms simply by feeding them bacteria engineered to make double-stranded RNA corresponding to the gene that should be shut down.

Because RNA interference is so easy to induce and yet so powerful, scientists are thinking big. Now that complete genomes—all the genes in the DNA—have been sequenced for a variety of organisms, scientists can use RNA interference to explore systematically what each gene does by turning it off. Recently four groups did just that in thousands of parallel experiments, each disabling a different gene of *C. elegans*. A similar genome-wide study is under way in plants, and several consortia are planning large RNAi studies of mammalian cells.

RNA interference is being used by pharmaceutical companies as well. Some drug designers are exploiting the effect as a shortcut to screen all genes of a certain kind in search of promising targets for new medicines. For instance, the systematic silencing of genes using RNAi could allow scientists to find a gene that is critical for the growth of certain cancer cells but not so important for the growth of normal cells. They could then develop a drug candidate that interferes with the protein product of this gene and then test the compound against cancer. Biotech firms have also been founded on the bet that gene silencing by
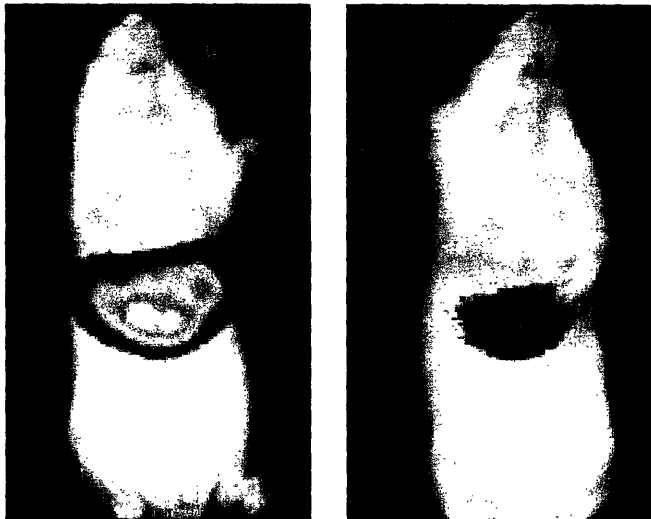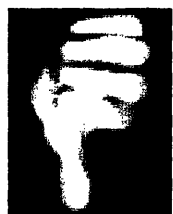
MICE LIGHT UP when injected with DNA containing the luciferase gene (*left*). But scientists took the shine off the mice by also injecting siRNAs that match the gene (*right*), thus demonstrating one way to exploit RNAi in mammals.

therapy. A novel gene that produces a particular siRNA might be loaded into a benign virus that will then bring the gene into the cells it infects. Beverly Davidson's group at the University of Iowa, for example, has used a modified adenovirus to deliver genes that produce siRNAs to the brain and liver of mice. Gene therapy in humans faces technical and regulatory difficulties, however.

Regardless of concerns about delivery, RNAi approaches have generated an excitement not currently seen for antisense and catalytic RNA techniques—other methods that, in principle, could treat disease by impeding harmful messenger RNAs. This excitement stems in part from the realization that RNA interference harnesses natural gene-censoring machinery that evolution has perfected over time.

RNAi could itself become a viable therapy to treat cancer, viral infections, certain dominant genetic disorders and other diseases that could be controlled by preventing selected genes from giving rise to illness-causing proteins.

Numerous reports have hinted at the promise of siRNAs for therapy. At least six labs have temporarily stopped viruses—HIV, polio and hepatitis C among them—from proliferating in human cell cultures. In each case, the scientists exposed the cells to siRNAs that prompted cells to shut down production of proteins crucial to the pathogens' reproduction. More recently, groups led by Judy Lieberman of Harvard Medical School and Mark A. Kay of the Stanford University School of Medicine have reported that siRNAs injected under extremely high pressure into mice slowed hepatitis and rescued many of the animals from liver disease that otherwise would have killed them.

Despite these laboratory successes, it will be years before

Why Cells Have Censors

INDEED, THE GENE-CENSORING mechanism is thought to have emerged about a billion years ago to protect some common ancestor to plants, animals and fungi against viruses and mobile genetic elements. Supporting this idea, the groups of Ronald H. A. Plasterk at the Netherlands Cancer Institute and of Hervé Vaucheret at the French National Institute of Agricultural Research have shown that modern worms rely on RNA interference for protection against mobile genetic elements and that plants need it as a defense against viruses.

Yet RNA interference seems to play other biological roles as well. Mutant worms and weeds having an impaired Dicer enzyme or too little of it suffer from numerous developmental defects and cannot reproduce. Why should a Dicer deficiency cause animals and plants to look misshapen?

One hypothesis is that once nature developed such an effective mechanism for silencing the subversive genes in viruses and mobile DNA sequences, it started borrowing tools from the RNAi tool chest and using them for different purposes. Each cell has the same set of genes—what makes them different from one another is which genes are expressed and which ones are not.

STOPPED VIRUSES IN HUMAN CELLS

RNAi-based therapies can be used in hospitals. The most difficult challenge will probably be delivery. Although the RNAi effect can spread throughout a plant or worm, such spreading does not seem to occur in humans and other mammals. Also, siRNAs are very large compared with typical drugs and cannot be taken as pills, because the digestive tract will destroy them rather then absorb them. Researchers are testing various ways to disseminate siRNAs to many organs and to guide them through cells' outer membranes. But it is not yet clear whether any of the current strategies will work.

Another approach for solving the delivery problem is gene

Most plants and animals start as a single embryonic cell that divides and eventually gives rise to a multitude of cells of various types. For this to occur, many of the genes expressed in the embryonic cells need to be turned off as the organ matures. Other genes that are off need to be turned on. When the RNAi machinery is not defending against attack, it apparently pitches in to help silence normal cellular genes during developmental transitions needed to form disparate cell types, such as neurons and muscle cells, or different organs, such as the brain and heart.

What then motivates the RNAi machinery to hush particular normal genes within the cell? In some cases, a cell may nat-

Efforts to Apply RNA Interference to Medicine

THE MACHINERY for RNA interference was discovered to operate in mammals just two years ago. Yet about 10 companies, including the sampling below, have already begun testing ways to exploit gene censoring to treat or prevent human disease. *The Editors*

COMPANY	PROJECTS	STATUS
Alnylam Pharmaceuticals Cambridge, Mass.	Researching therapeutic applications of RNAi, but specific disease targets not yet announced	Founded in 2002 by Bartel, Tuschl, Sharp and Zamore, the firm has secured initial funding and several patents
Cenix Biosciences Dresden, Germany	Investigating the use of RNAi-based therapies for cancer and viral diseases	With Texas-based Ambion, Cenix is creating a library of siRNAs to cover the entire human genome
Ribopharma Kulmbach, Germany	Attempting to chemically modify siRNAs to make drugs for glioblastoma, pancreatic cancer and hepatitis C	Clinical trials in brain cancer patients are expected to begin this year
Sirna Therapeutics Boulder, Colo.	Testing a catalytic RNA medicine for advanced colon cancer in clinical trials; development of RNAi-based therapeutics is still in early stages	Changed name from Ribozyme Pharmaceuticals in April; recently secured $48 million in venture capital

urally produce long double-stranded RNA specifically for this purpose. But frequently the triggers are "microRNAs"—small RNA fragments that resemble siRNAs but differ in origin. Whereas siRNAs come from the same types of genes or genomic regions that ultimately become silenced, microRNAs come from genes whose sole mission is to produce these tiny regulatory RNAs.

The RNA molecule initially transcribed from a microRNA gene—the microRNA precursor—folds back on itself, forming a structure that resembles an old-fashioned hairpin. With the help of Dicer, the middle section is chopped out of the hairpin, and the resulting piece typically behaves very much like an siRNA—with the important exception that it does not censor a gene with any resemblance to the one that produced it but instead censors some other gene altogether.

As with the RNAi phenomenon in general, it has taken biologists time to appreciate the potential of microRNAs for regulating gene expression. Until recently, scientists knew of only two microRNAs, called *lin-4* RNA and *let-7* RNA, discovered by the groups of Victor Ambros of Dartmouth Medical School and Gary Ruvkun of Harvard Medical School. In the past two years we, Tuschl, Ambros and others have discovered hundreds of additional microRNA genes in worms, flies, plants and humans.

With Christopher Burge at M.I.T., we have estimated that humans have between 200 and 255 microRNA genes—nearly 1 percent of the total number of human genes. The microRNA genes had escaped detection because the computer programs designed to sift through the reams of genomic sequence data had not been trained to find this unusual type of gene, whose final product is an RNA rather than a protein.

Some microRNAs, particularly those in plants, guide the slicing of their mRNA targets, as was shown by James C. Carrington of Oregon State University and Zamore. We and Bonnie Bartel of Rice University have noted that plant microRNAs take aim primarily at genes important for development. By clearing their messages from certain cells during development, RNAi could help the cells mature into the correct type and form the proper structures.

Interestingly, the *lin-4* and *let-7* RNAs, first discovered in worms because of their crucial role in pacing development, can employ a second tactic as well. The messenger RNAs targeted by these microRNAs are only approximately complementary to the microRNAs, and these messages are not cleaved. Some other mechanism blocks translation of the messenger RNAs into productive proteins.

Faced with these different silencing mechanisms, biologists are keeping open minds about the roles of small RNAs and the RNAi machinery. Mounting evidence indicates that siRNAs not only capture messenger RNAs for destruction but can also direct the silencing of DNA—in the most extreme case, by literally editing genes right out of the genome. In most cases, however, the silenced DNA is not destroyed; instead it is more tightly packed so that it cannot be transcribed.

From its humble beginnings in white flowers and deformed worms, our understanding of RNA interference has come a long way. Almost all facets of biology, biomedicine and bioengineering are being touched by RNAi, as the gene-silencing technique spreads to more labs and experimental organisms.

Still, RNAi poses many fascinating questions. What is the span of biological processes that RNA interference, siRNAs and microRNAs influence? How does the RNAi molecular machinery operate at the level of atoms and chemical bonds? Do any diseases result from defects in the RNAi process and in microRNAs? As these questions yield to science, our understanding of the phenomenon will gradually solidify—perhaps into a foundation for an entirely new pillar of genetic medicine. ▨

MORE TO EXPLORE

RNAi: Nature Abhors a Double-Strand. György Hutvágner and Phillip D. Zamore in *Current Opinion in Genetics & Development*, Vol. 12, No. 2, pages 225–232; April 2002.

Gene Silencing in Mammals by Small Interfering RNAs. Michael T. McManus and Phillip A. Sharp in *Nature Reviews Genetics*, Vol. 3, pages 737–747; October 2002.

MicroRNAs: At the Root of Plant Development? Bonnie Bartel and David P. Bartel in *Plant Physiology*, Vol. 132, No. 2; pages 709–717; June 2003. Available at www.plantphysiol.org/cgi/content/full/132/2/709

Isolation and Preliminary Survey

of *C.elegans* miRNA Deletion Mutants

Despite current knowledge on the gene silencing mechanism of microRNAs (miRNAs), the biological roles of many individual miRNAs are largely undetermined. The functions of a few animal miRNAs have been determined through classical genetic studies of mutants where miRNA expression is perturbed, such as *lin-4*, *let-7*, *lsy-6*, *bantam*, and *Drosophila mir-14*. While miRNA target predictions and validation efforts offer a means to ascertain miRNA function, the rigorous standard is to disrupt the function of the miRNA locus and examine the phenotypic consequence. To accelerate the isolation of miRNA mutants, a reverse-genetic approach was applied to *Caenorhabditis elegans*, and over 31 mutants in at least 35 miRNA loci have been isolated. The majority of the mutants are viable and fertile in the homozygous state, and a preliminary phenotype scan has not indicated overt defects. Hypothesizing that functional redundancy may be masking deficiencies; the generation of mutants with multiple disrupted miRNA loci has been initiated. Preliminary analysis of such multiple knock-out mutants suggests defects may be revealed in such animals.

A few genetic screens have successful identified miRNA mutants, but the particular phenotype targeted by each screen has been dissimilar. Screens that examined properly-timed cell divisions identified the heterochronic miRNAs, *lin-4* and *let-7* [1, 2], while a screen examining the loss of asymmetric expression of a GFP-tagged neuronal gene uncovered the *lsy-6* miRNA [3]. A screen for *Drosophila melanogaster* larvae with abnormally large sizes allowed for the isolation of *bantam* [4, 5], while *D.m. mir-14* was found in a modifier screen looking for mutations that enhance the small-eye phenotype in a transgenic fly overexpressing the *Rpr* gene in the eye [6]. Because these miRNA mutants exhibit rather diverse phenotypes (evidenced by different genetic screens that allow for isolation of the mutants), it would be difficult to direct a

particular genetic screen to obtain mutants targeted against the miRNA gene class alone. With the *C.elegans* genome practically sequenced, miRNA loci can be easily mapped, but correlation of genetically-mapped mutants to the physical map of miRNA loci has not particularly revealed many additional miRNA mutants.

A direct method to isolate mutants in miRNA genes is to employ reverse genetics. Although homologous recombination techniques have not been optimized for *C.elegans*, the small size, quick life-span, and simplicity in culturing allows for the generation of a library of randomly mutagenized animals which can then be screened for the deletion of interest. Various methods for constructing libraries of mutagenized worms have been described previously [7-10], but because generating and screening such libraries is such an intensive effort, a *C.elegans* knock-out consortium was established to streamline the process of generating mutant worms for the scientific community. To expedite the hunt for miRNA mutants without the assistance of the *C.elegans* knock-out consortium, it was decided that a library would be generated and screened in-house, with the efforts from collaborations between me from the Bartel lab, Allison Abbot from the Ambros lab, and Ezequiel Alvarez-Saavedra and Eric Miska from the Horvitz lab.

A library of 7×10^6 mutagenized genomes was constructed using a method originally optimized by the Koelle lab [8] and enhanced in the Horvitz lab with the use of robotic liquid handling. A conceptual diagram of the library construction process is illustrated in Figure 1. Worms in the L4 stage are first mutagenized with ethyl methanesulfonate (EMS), which usually causes point mutations but can also induce a varying degree of deletions [11]. The mutagenized parental worms (P_0) self-fertilize to yield progeny (F_1) that are synchronized after hatching by starvation at the L1 arrest. The F_1 progeny are then distributed into several microtiter dishes with food (~20 worms per well), and allowed to self-fertilize again to yield a final distribution of

F_2 progeny. Growth and synchronization of starved F_2 worms are all achieved in the microtiter dishes placed within a humified box. Waiting for the production of F_2 progeny increases the chances that a screened deletion candidate can be represented in a heterozygous form even if the deleted gene is lethal or sterilizes the animal. The library of starved F_2 larvae are then divided in half, where one half is processed with cryoprotectant for live freezing of worms. The other half of the worm library is then lysed, and a half of the lysate is refined into a genomic DNA prep, for which both are archived by freezing. The library generated in this collaboration took almost 6 months to complete with the aid of a customized Tecan Genesis Liquid Handler, and will be referred to as the V3 (version 3) library.

Screening the library for deletions was accomplished by large scale, microtiter-well PCR amplification of samples and then resolving the reactions by agarose electrophoresis. The protocol for PCR screening employs a primary reaction of primers that flank a region of ~1500 bp around the miRNA loci, and includes a "poison primer" that base pairs to the miRNA precursor sequence [12] (Figure 2A). Cycling conditions are designed to favor small amplicons, such as those derived from the desired genomic deletion, or from the extension of the poison primer base-pairing to the wild-type template. Even though favored deletion templates will be far outnumbered in the background of wild-type templates, the secondary PCR using nested primers will only allow the deletion amplicon to be further amplified because most of the wild-type amplicons have been generated from the poison primer, which lack a site for a nested primer and are diluted in the course of setting up the secondary PCR [12]. Each PCR per well samples at least 4000 genomes, and screens are generally conducted on >1500 wells by cycling the reactions at 10 μL volumes in 384-well plates. When secondary PCRs are resolved on agarose gels, a visual scan of the products sometimes indicates a deletion band (Figure 2B).

Candidate wells are retested several times, and amplicons are sequenced to confirm a gene-specific deletion. When a deletion amplicon has been verified, the well of live frozen worms with the corresponding address of the detected amplicon is thawed. Individual worms are isolated and after progeny are produced, the parents are screened by PCR for the individual worm harboring the deleted loci.

To assess how amenable the reverse genetics approach would be in yielding deletions in miRNAs residing in different genomic contexts, a pilot effort was initiated against 5 different miRNA loci (Phase I, Table 1). The successful screening of deletion amplicons for each of the 5 miRNA loci convinced us to enlarge the scope of the deletion hunt to include a second phase of various other miRNA loci targeted for screening of deletions (Phase II, Table 1). Both Phase I and Phase II screenings were actually performed on the V2 library generated earlier in the Horvitz lab in the similar manner that the V3 library was generated. Much of the V2 library was exhausted in the course of screening miRNA-deleted loci of the Phase II effort. When the V3 library was completed, Phase III miRNA loci were included for continued screening and isolation. In a combined effort of about 6 months, our group has isolated at least 31 mutants in over 35 miRNA loci (Table 2), and additional mutants are still coming through the screening pipeline. The size of deletions ranged from 181 bp to 1488 bp, and the average deletion size was ~600 bp. For most mutants targeting a gene cluster, the entire cluster was deleted, except in strain MT13231, where only *mir-38–41* has been deleted but *mir-35–37* are still intact, and in strain MT13016, which only deletes *mir-64* but leaves *mir-65* and *mir-66* intact.

In the course of isolating the various mutants, we were surprised that many deletions could be so readily homozygosed and such worms appeared normal by cursory inspection (Table 2). Only two strains, MT12944 (*mir-50*) and MT13011 (*mir-79*) could not be homozygosed.

However, these strains have not been appreciably backcrossed against wild-type, so a closely linked mutation outside the miRNA loci might be preventing the production of homozygous progeny. The strain MT13443, which deletes the entire *mir-35–41* cluster, had been difficult to self-fertilize and produce progeny because of significant embryonic lethality, however later backcrossing of this strain has alleviated the general embryonic lethality when cultured at 20° C [13]. Overall, the majority of miRNA deletion mutants superficially appeared fertile and quite mobile, a far contrast to the severe morphological defects seen in the heterochronic miRNA mutants, *lin-4* and *let-7*. For example, abnormal movement and the absence of a vulva occur in the *lin-4* mutant [2, 14, 15], late larval lethality and bursting of the vulva is seen in the *let-7* mutant [1], and both *lin-4* and *let-7* are defective in forming a regular adult cuticle[1, 2, 14, 15]. Such defects have not yet been observed for the other miRNA mutants isolated so far.

To characterize the deletion mutants further, we applied a general battery of simple-to-perform assays on mutant strains that had been immediately isolated from thaws and thus have not been backcrossed against wild-type. The battery of assays examined the morphology and development of the mutants, simple behaviors and neurophysiology, and genetic interactions (Table 3). Although only a few hints of phenotypic abnormalities have been garnered from the survey of the mutants, overt defects have generally not been readily detected. Overexpression of miRNAs through transgenes does not actually require the miRNA mutants but have been included in the tasks of phenotypic characterization for the reason that gain of function defects might be observed that would direct the proper focus when characterizing mutants. Such overexpression studies are on going.

Our current phenotypic survey is broad but cursory, and would have easily missed many subtle defects that might be imparted by loss of miRNA function. For example, the *lsy-6* mutant

who misexpresses the *gcy-5* gustatory receptor in the ASEL neuron instead of the *gcy-7* receptor would otherwise visibly appear normal, except that the worm would be unable to distinguish chloride ions from sodium ions [3, 16]. The importance of *lsy-6* in nematodes is supported by the perfect conservation of the miRNA sequence in *C.briggsae* [3], however even other genetic screens for chemotaxis mutants have yet to also identify *lsy-6*, suggesting that the defect imparted by the *lsy-6* lesion is extremely subtle and its identification might have depended on the sophisticated fluorescent tagging of genes downstream in the regulatory network that establishes neuronal asymmetry. If severe defects seen of *lin-4* and *let-7* are the exception, then determining the phenotypes for most other nematode miRNA mutants will demand other sophisticated measures to inform the investigation. In addition to overexpression studies, localization of miRNA expression in particular cell types will be a key component towards dissecting miRNA function.

An alternate rationalization of the absence of overt phenotypes in individual miRNA mutants would be functional redundancy with other miRNAs, particularly with miRNA family members that share sequence similarity. Fifteen families containing two or more similar *C.elegans* miRNAs exist [17], and deletion mutants have been obtained from at least 12 of these families (Figure 3A). As experimental evidence increases in support of the hypothesis that miRNAs primarily utilize the "seed" interaction for target recognition [18-20], the idea that similar miRNA family members can compensate for deletion of a single miRNA member is particularly appealing. Further support for functional redundancy is indicated in preliminary results with double mutants of *mir-48*, *mir-84*, and *mir-241* deletions which exhibit several strong developmental defects, but are otherwise normal as single deletion mutants [13]. A counter-argument to this hypothesis, however, is that the *let-7* and *lin-4* single mutants exhibit

strong defects despite the presence of clear paralogs (Figure 3A). Additionally, several other

C.elegans miRNAs that lack any other similar miRNAs have been deleted in mutants (Figure

3B), and it is particularly surprising that some of these "orphan" mutants, like *mir-1* and *mir-34*,

are perfectly conserved between invertebrates and vertebrates, like *let-7*, so one might have

expected such conservation would affirm a required dependence of *mir-1* and *mir-34* for normal

worm development. Sorting out the genetic functions of the miRNAs in *C.elegans* will likely

demand long-term analyses, and may require the isolation of additional miRNA mutants and

generation of additional double and triple mutants.

REFERENCES

1. Reinhart, B.J., et al., *The 21 nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans.* Nature, 2000. **403**: p. 901-906.

2. Chalfie, M., H.R. Horvitz, and J.E. Sulston, *Mutations that lead to reiterations in the cell lineages of C. elegans.* Cell, 1981. **24**(1): p. p59-69.

3. Johnston, R.J. and O. Hobert, *A microRNA controlling left/right neuronal asymmetry in Caenorhabditis elegans.* Nature, 2003. **426**(6968): p. 845-9.

4. Hipfner, D.R., K. Weigmann, and S.M. Cohen, *The bantam gene regulates Drosophila growth.* Genetics, 2002. **161**(4): p. p1527-37.

5. Brennecke, J., et al., *bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila.* Cell, 2003. **113**(1): p. 25-36.

6. Xu, P., et al., *The Drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism.* Curr Biol, 2003. **13**(9): p. 790-5.

7. Zwaal, R.R., et al., *Target-selected gene inactivation in Caenorhabditis elegans by using a frozen transposon insertion mutant bank.* Proc Natl Acad Sci U S A, 1993. **90**(16): p. 7431-5.

8. Koelle, M., et al., *C.elegans Gene Knockout Protocols*. 2003.

9. Wei, A., et al., *Efficient isolation of targeted Caenorhabditis elegans deletion strains using highly thermostable restriction endonucleases and PCR*. Nucleic Acids Res, 2002. **30**(20): p. e110.

10. Jansen, G., et al., *Reverse genetics by chemical mutagenesis in Caenorhabditis elegans*. Nat Genet, 1997. **17**(1): p. 119-21.

11. Anderson, P., *Mutagenesis*, in *Caenorhabditis elegans: Modern Biological Analysis of an Organism*, H.F. Epstein and D.C. Shakes, Editors. 1995, Academic Press. p. 31-58.

12. Edgley, M., et al., *Improved detection of small deletions in complex pools of DNA*. Nucleic Acids Res, 2002. **30**(12): p. e52.

13. Abbot, A.L., et al. data not shown.

14. Ambros, V., *A hierarchy of regulatory genes controls a larva-to-adult developmental switch in C. elegans*. Cell, 1989. **57**(1): p. 49-57.

15. Ambros, V. and H.R. Horvitz, *Heterochronic mutants of the nematode Caenorhabditis elegans*. Science, 1984. **226**(4673): p. 409-16.

16. Chang, S., R.J. Johnston, Jr., and O. Hobert, *A transcriptional regulatory cascade that controls left/right asymmetry in chemosensory neurons of C. elegans*. Genes Dev, 2003. **17**(17): p. 2123-37.

17. Lim, L.P., et al., *The microRNAs of Caenorhabditis elegans*. Genes Dev, 2003. **17**: p. 991-1008.

18. Stark, A., et al., *Identification of Drosophila microRNA targets*. PLOS Biol., 2003. **1**(3): p. E60.

19. Lewis, B.P., et al., *Prediction of mammalian microRNA targets*. Cell, 2003. **115**: p. 787-798.

20. Lai, E.C., *MicroRNAs are complementary to 3'UTR motifs that mediate negative post-transcriptional regulation*. Nature Genetics, 2002. **30**(4): p. 363-364.

Table 1. miRNA Loci Selected for Targeted Knock-out Screening

gene	rationale
PHASE I	
mir-35/41	large gene cluster, interesting expression pattern
mir-42/44	gene cluster
mir-50	within an intron
mir-52	far away from ORF, very abundant
mir-84	close to ORF, in *let-7* gene family
PHASE II	
mir-1	conserved, homologs have interesting tissue expression
mir-2	conserved, abundantly expressed
mir-34	conserved, putative interesting localization
mir-45	sequence similarity to *mir-42/44* cluster
mir-48	candidate *lin-58*, *let-7* family member, temporal expression
mir-49	higher expression in embryo
mir-53	sequence similarity to *mir-52*, could also be very abundant
mir-70	temporal expression like *lin-4*
mir-71	early temporal expression
mir-72	conservation beyond *C.briggsae*
mir-79	conservation beyond *C.briggsae*
mir-85	very late larval expression
mir-87	conservation beyond *C.briggsae*
mir-91	conservation beyond *C.briggsae*
mir-238	low expression in embryo, high in larvae and adult
mir-54/56	gene cluster
mir-64-66	gene cluster
mir-73/74	gene cluster
mir-124	conserved, homologs have interesting tissue expression
PHASE III	
mir-237	*lin-4* homolog, temporally regulated expression
mir-248	much higher expression in dauer
mir-244	*mir-9* homolog, Hox cluster location
mir-228	*mir-124* homolog
mir-241 / mir-48	*let-7* family member, temporal expression, only 1.8 kb away from *mir-48*
mir-256	*mir-1* homolog
mir-234	starvation induced expression, homolog of Hs *mir-137*
mir-247	interesting L3 / Dauer specific expression
mir-79	additional mir-79 alleles
mir-50	additional mir-50 alleles
mir-80 / mir-238	in *bantam* gene family, also clustered with *mir-238*
mir-81	in *bantam* gene family
mir-82	in *bantam* gene family
mir-239a,b	gene cluster, same gene family as *mir-238*
mir-61 / mir-250	gene cluster
mir-232	low in embryo, high in larvae and adult
mir-229	longer miRNA, with longer type of hairpin
mir-77	GFP reporter expression in the gut

Table 2. Deletion mutants successfully isolated from mutagenized worm libraries.

gene	allele	strain(s)	deletion size	homozygosed?
mir-1	n4101	MT12954	380	yes, viable
mir-1	n4102	MT12955	825	yes, viable
mir-2	n4108	MT12977	555	yes, viable
mir-34	n4276	MT13406	630	yes, viable
mir-45	n4280	MT13433	1488	yes, viable
mir-48	n4097	MT12927	293	yes, viable
mir-49	n4103	MT12956	571	yes, viable
mir-50	n4099	MT12944 het	1,000	no, possibly lethal
mir-52	n4114	MT12990	150	yes, viable
mir-52	n4100	MT12945	800	yes, viable
mir-53	n4113	MT12989	805	yes, viable
mir-70	n4109	MT12978	738	yes, viable
mir-70	n4110	MT12979	203	yes, viable
mir-71	n4105	MT12968	343	yes, viable
mir-71	n4115	MT12993	181	yes, viable
mir-72	n4130	MT13015	960	yes, viable
mir-79	n4126	MT13011 het	387	no, possibly sterile
mir-84	n4037	MT12494, MT12926 (3x BC)	791	yes, viable
mir-85	n4117	MT12999	538	yes, viable
mir-87	n4124	MT13009 het	615	yes, viable
mir-87	n4123	MT13008 het, MT13221 homo	260	yes, viable
mir-87	n4104	MT12958	514	yes, viable
mir-91	n4107	MT12974	682	yes, viable
mir-238	n4112	MT12983	536	yes, viable
mir-146	n4106	MT12969	530	yes, viable
mir-35/41	nDf48	MT13231	1043	yes, viable
mir-35/41	nDf50	MT13443	1150	yes, mildly embryonic lethal
mir-42/44	nDf49	MT13372	1100	yes, viable
mir-54/56	nDf45	MT12988	150	yes, viable
mir-64/66	n4131	MT13016	652	yes, viable
mir-73/74	nDf47	MT13078	325	yes, viable
mir-124	n4255	MT13292 (2xBC)	212	yes, viable

Table 3. Phenotype survey conducted on miRNA mutants
Morphology and Development
General organ formation
Percent viability
Fertility (brood size)
Dauer formation
Embryonic growth
Male tail structure
Egg laying
Pumping
General movement/speed
Intracellular stains (DAPI, lysotracker, mitotracker)
Cellular stains or immunostains (DiO, DiI, MH27 antibody)
Behavior and Neurophysiology
Nose touch response
Osmotic avoidance
Chemotaxis
Body touch response (mechano-sensation)
Defecation
Aldicarb resistance
Ivermectin resistance
Levamisole resistance
Genetics
Transgene overexpression of miRNAs
Double / triple knock-out crosses
RNAi synthetic knockdowns
Cross into col-19::GFP background (seam cell marker)

Figure Legends

Figure 1. Diagram of the construction of a library of mutagenized worms for deletion screening. See text for details.

Figure 2. PCR screening for miRNA deletions. (A) Schematic of the *mir-84* loci and the design of nest primers used in screening the library of mutagenized worm genomes. Three sets of forward primers (L1, L2, L3) and reverse primers (R1, R2, R3) are selected to flank a region of ~1500 bp, while two poison primers (PR, PF) are selected based on complementarity to the miRNA precursor sequence. The multiple sets of forward and reverse primers are tested in various combinations to optimize on set of outside and nest primers for the actual PCR screen. There is a nearby ORF, B0395.1, colored in green. (B) Typical results of a PCR screen, where a deletion amplicon can be preferentially enriched for detection by agarose electrophoresis. The wild-type amplicon is 1600 nt long, while the deletion amplicon is 800 nt long.

Figure 3. Families of *C.elegans* miRNAs for which deletion mutants have been isolated. (A) Families containing two or more similar miRNA sequences. Orders in the groupings are based on phylogenetic alignments from [17]. Sequences in red are miRNAs that have been successfully deleted in nematode mutants. Brackets denote miRNAs that reside in a cluster, and are often deleted together in the course of isolating the mutant. (B) "Orphan" miRNAs in *C.elegans* where no other miRNA in the genome shares sequence similarity.
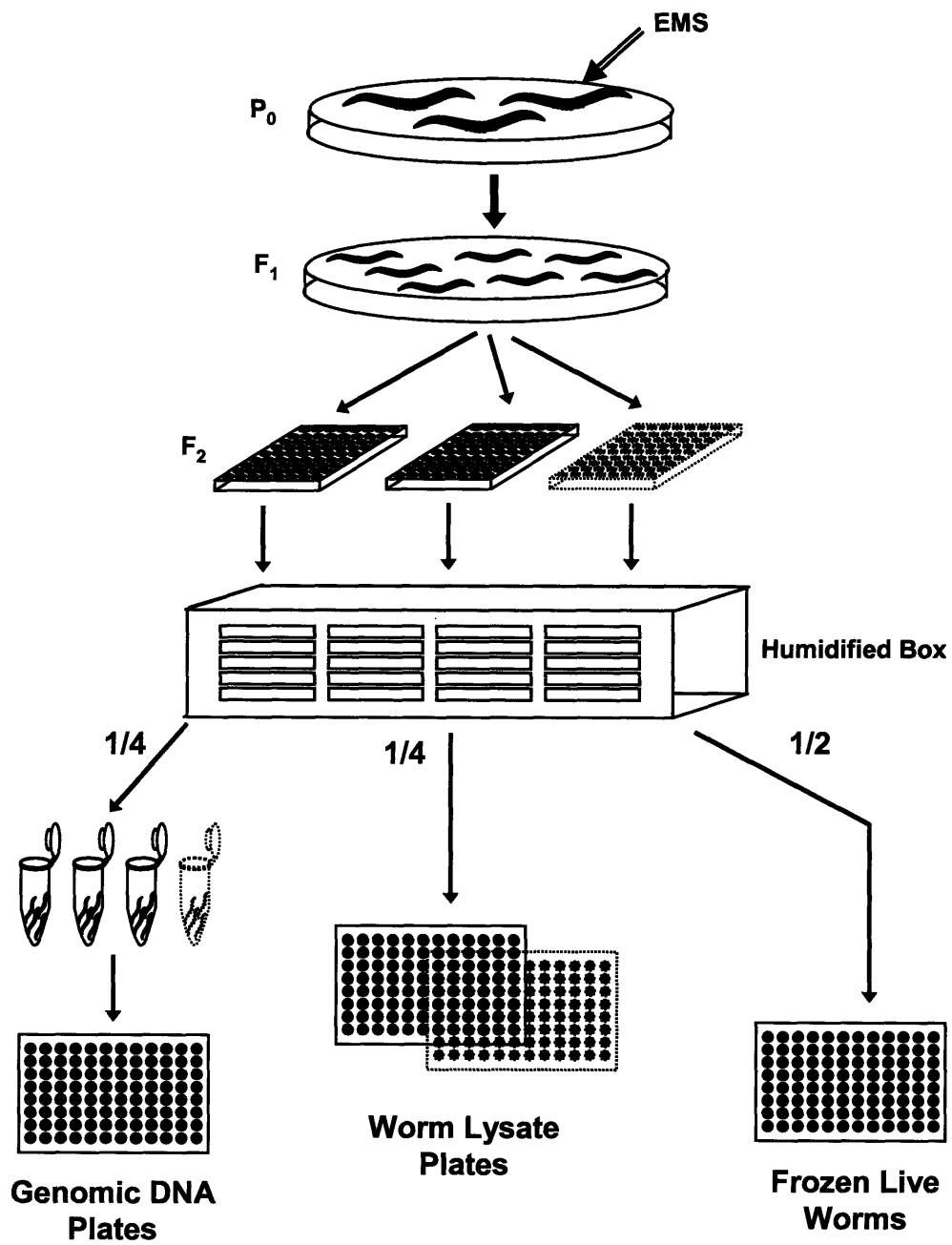
Figure 1.
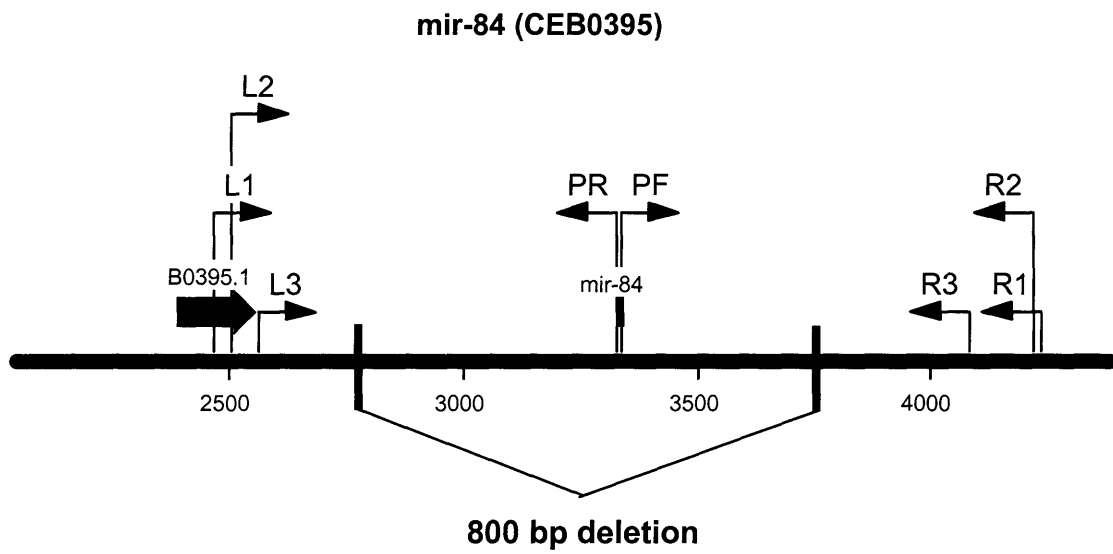

P₀ ... EMS
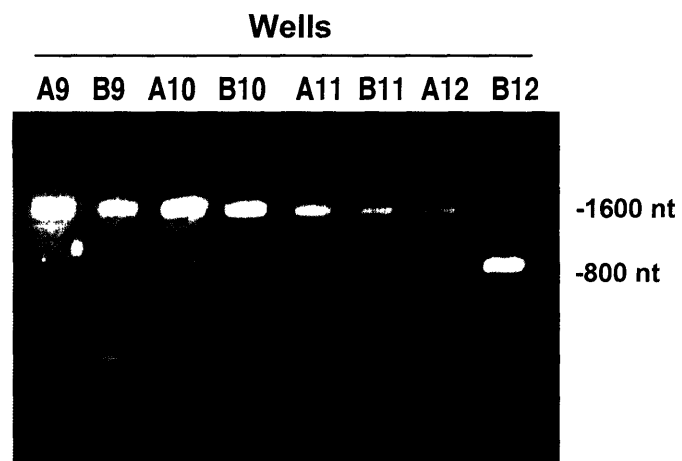
F₁

F₂

Humidified Box

1/4 → Genomic DNA Plates

1/4 → Worm Lysate Plates

1/2 → Frozen Live Worms

Figure 2.

A

mir-84 (CEB0395)

800 bp deletion

B

Wells

A9 B9 A10 B10 A11 B11 A12 B12

-1600 nt

-800 nt

Figure 3.

A

```
lin-4.....TCCCTGAGACCTCAAGTGTGA......        mir-52....CACCCGTACATATGTTTCCGTGCT...
mir-237...TCCCTGAGAATTCTCGAACAGCTT...        mir-53....CACCCGTACATTTGTTTCCGTGCT...
                                             mir-54....TACCCGTAATCTTCATAATCCGAG...─┐
let-7.....TGAGGTAGTAGGTTGTATAGTT.....        mir-55....TACCCGTATAAGTTTCTGCTGAG....─┤
mir-84....TGAGGTAGTATGTAATATTGTA.....        mir-56....TACCCGTAATGTTTCCGCTGAG.....─┘
mir-241...TGAGGTAGGTGCGAGAAATGA......
mir-48....TGAGGTAGGCTCAGTAGATGCGA....

                                             mir-65....TATGACACTGAAGCGTAACCGAA....─┐
mir-2.....TATCACAGCCAGCTTTGATGTGC....        mir-64....TATGACACTGAAGCGTTACCGAA....─┤
mir-43....TATCACAGTTTACTTGCTGTCGC....─┐      mir-63....TATGACACTGAAGCGAGTTGGAAA...─┤
                                    │        mir-66....CATGACACTGATTAGGGATGTGA....─┘
mir-39...TCACCGGGTGTAAATCAGCTTG.....─┐│      mir-229...AATGACACTGGTTATCTTTTCCATCGT
mir-36...TCACCGGGTGAAAATTCGCATG.....─┤│
mir-37...TCACCGGGTGAACACTTGCAGT.....─┤│
mir-35...TCACCGGGTGGAAACTAGCAGT.....─┤│      mir-72....AGGCAAGATGTTGGCATAGC.......
mir-42....CACCGGGTTAACATCTACAG......─┼┤      mir-73....TGGCAAGATGTAGGCAGTTCAGT....─┐
mir-41...TCACCGGGTGAAAAATCACCTA.....─┤│      mir-74....TGGCAAGAAATGGCAGTCTACA.....─┘
mir-40...TCACCGGGTGTACATCAGCTAA.....─┤│
mir-38...TCACCGGGAGAAAAACTGGAGT.....─┘│
                                     │       mir-75....TTAAAGCTACCAACCGGCTTCA.....
mir-45....TGACTAGAGACACATTCAGCT......│       mir-79....ATAAAGCTAGGTTACCAAAGCT.....
mir-44....TGACTAGAGACACATTCAGCT......┘
mir-61....TGACTAGAACCGTTACTCATC......
mir-247...TGACTAGAGCCTATTCTCTTCTT....          mir-124....TAAGGCACGCGGTGAATGCCA.....
                                             mir-228....AATGGCACTGCATGAATTCACGG...

mir-50....TGATATGTCTGGTATTCTTGGGTT...
mir-62....TGATATGTAATCTAGCTTACAG.....         mir-238....TTTGTACTCCGATGCCATTCAGA...
mir-90....TGATATGTTGTTTGAATGCCCC.....         mir-239a...TTTGTACTACACATAGGTACTGG...
```

B

```
mir-1.....TGGAATGTAAAGAAGTATGTA......
mir-34....AGGCAGTGTGGTTAGCTGGTTG.....
mir-49....AAGCACCACGAGAAGCTGCAGA.....
mir-70....TAATACGTCGTTGGTGTTTCCAT....
mir-71....TGAAAGACATGGGTAGTGA........
mir-87....GTGAGCAAAGTTTCAGGTGT.......
```