microPrimer: the biogenesis and function of microRNA

Tingting Du and Phillip D. Zamore*

Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA 01605, USA

*Author for correspondence (e-mail: phillip.zamore@umassmed.edu)

Development 132, 4645-4652
Published by The Company of Biologists 2005
doi:10.1242/dev.02070

Summary

Discovered in nematodes in 1993, microRNAs (miRNAs) are non-coding RNAs that are related to small interfering RNAs (siRNAs), the small RNAs that guide RNA interference (RNAi). miRNAs sculpt gene expression profiles during plant and animal development. In fact, miRNAs may regulate as many as one-third of human genes. miRNAs are found only in plants and animals, and in the viruses that infect them. miRNAs function very much like siRNAs, but these two types of small RNAs can be distinguished by their distinct pathways for maturation and by the logic by which they regulate gene expression.

microHistory

The first miRNA, lin-4, was identified in 1993 in a genetic screen for mutants that disrupt the timing of post-embryonic development in Caenorhabditis elegans (Lee et al., 1993). Cloning of the locus revealed that lin-4 produces a 22-nucleotide non-coding RNA, rather than a protein-coding mRNA (Lee et al., 1993). lin-4 represses the expression of lin-14, which encodes a nuclear protein (Lee et al., 1993; Wightman et al., 1993) whose concentration must be reduced for worms to progress from their first larval stage to the second (Rougvie, 2005). The negative regulation of lin-14 by lin-4 requires partial complementarity between lin-4 and sites in the 3’-untranslated region (UTR) of lin-14 mRNA (Ha et al., 1996; Olsen and Ambros, 1999). It was not until 2000 that a second miRNA, let-7, was discovered, again in worms (Reinhart et al., 2000). let-7 functions in a manner similar to lin-4, repressing the expression of the lin-4 I and hbl-1 miRNAs by binding to their 3’ UTRs (Reinhart et al., 2000; Slack et al., 2000; Lin et al., 2003; Vella et al., 2004). let-7 is conserved throughout metazoans (Pasquinelli et al., 2000), and the discovery of let-7 (Reinhart et al., 2000), together with the subsequent large-scale searches for additional miRNAs, established miRNAs as a new and large class of ribo-regulators (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001), and fueled speculation that tiny RNAs are a major feature of the gene regulatory networks of animals. Now more than 1600 miRNAs have been identified in plants, animals and viruses (Lai et al., 2003; Lim et al., 2003a; Lim et al., 2003b). The human genome alone may contain 800-1000 miRNAs, a large portion of which may be specific to primates (Bentwich et al., 2005; Berezikov et al., 2005; Xie et al., 2005).

microMaturation

miRNAs are transcribed by RNA polymerase II as primary miRNAs (pri-miRNAs), which range from hundreds to thousands of nucleotides in length (Cai et al., 2004; Lee et al., 2004; Parizotto et al., 2004). Most miRNAs are transcribed from regions of the genome that are distinct from previously annotated protein-coding sequences (Fig. 1). Some miRNA-
ar are characteristics of RNase III cleavage of dsRNA. Thus, Drosha cleavage defines either the 5′ or the 3′ end of the mature miRNA. (The mature miRNA resides in the 5′ arm of some pre-miRNA and in the 3′ arm in others.) The pre-miRNA is then exported from nucleus to cytoplasm by Exportin 5/RanGTP, which specifically recognizes the characteristic end structure of pre-miRNAs (Yi et al., 2003; Bohnsack et al., 2004; Lund et al., 2004; Zeng and Cullen, 2004).

In the cytoplasm, a second RNase III, Dicer, together with its dsRBDD protein partner, Loquacious (Loqs) in Drosophila or the trans-activator RNA (tar)-binding protein (TRBP) in humans, makes a pair of cuts that defines the other end of the mature miRNA, liberating an ~21-nucleotide RNA duplex (Bernstein et al., 2001; Grishok et al., 2001; Hutvagner et al., 2001; Ketting et al., 2001; Chendrimada et al., 2005; Forstemann et al., 2005; Jiang et al., 2005; Saito et al., 2005). This RNA duplex has essentially the same structure as a double-stranded siRNA, except that the mature miRNA is only partially paired to the miRNA* – the small RNA that resides on the side of the pre-miRNA stem opposite the miRNA – because the stems of pre-miRNAs are imperfectly double stranded. From the miRNA/miRNA* duplex, one strand, the miRNA, preferentially enters the protein complex that represses target gene expression, the RNA-induced silencing complex (RISC), whereas the other strand, the miRNA* strand, is degraded. The choice of strand relies on the local thermodynamic stability of the miRNA/miRNA* duplex – the strand whose 5′ end is less stably paired is loaded into the RISC (Khvorova et al., 2003; Schwarz et al., 2003). This thermodynamic difference arises, in part, because miRNAs tend to begin with uracil, and, in part, because miRNA/miRNA* duplexes contain mismatches and bulges that favor the miRNA strand being loaded into the RISC.

**Plant microMaturation**

miRNA maturation in plants differs from the pathway in animals because plants lack a Drosha homolog (Fig. 2B). Instead, the RNase III enzyme DICER-LIKE 1 (DCL1), which is homologous to animal Dicer, is required for miRNA maturation (Park et al., 2002; Reinhart et al., 2002; Papp et al., 2003; Xie et al., 2004). In plants, DCL1 is localized in the nucleus and can make both the first pair of cuts made by Drosha and the second pair of cuts made by animal Dicer. As for animal Dicer, a dsRNA-binding domain protein partner, HYL1, has been implicated in DCL1 function in plant miRNA maturation (Papp et al., 2003; Vazquez et al., 2004a). The resulting miRNA/miRNA* duplex is exported from the nucleus by HASTY (HST), the plant ortholog of Exportin 5, and completes its assembly into the RISC in the cytoplasm (Peragine et al., 2004; Park et al., 2005). Unlike animal miRNAs, which end with free 2′, 3′ hydroxyl groups, plant miRNAs have a methyl group on the ribose of the last nucleotide. The terminal methyl group is added by the S-adenosyl methionine (SAM)-dependent methyltransferase HEN1, and the modification of the miRNA by HEN1 either protects the miRNA from further modification or degradation, or may facilitate its assembly into the RISC (Boutet et al., 2003; Yu et al., 2005). In plants, RNA-dependent RNA polymerases may use small RNAs as primers to synthesize double-stranded RNA from aberrant single-stranded transcripts, raising the possibility that the terminal methoxy modification on miRNA serves to prevent miRNA from acting as primers.

**The RISC directs gene silencing**

The RISC carries out small RNA-directed gene silencing in both the miRNA and the RNAi pathways in plants and animals (Hammond et al., 2000; Hammond et al., 2001; Hutvagner and Zamore, 2002; Zeng et al., 2002; Doench et al., 2003). When the small RNA guide in the RISC pairs extensively to a target mRNA, the RISC functions as an endonuclease, cleaving the mRNA between the target nucleotides paired to bases 10 and 11 of the miRNA or siRNA. The core component of every RISC is a member of the Argonaute (Ago) protein family, whose members all contain a central PÁZ domain (named after the family member proteins Piwi, Argonaute and Zwille) and a carboxy terminal PIWI domain. Structural studies show that the PIWI domain binds to small RNAs at their 5′ end, whereas the PÁZ domain binds to the 3′ end of single-stranded RNAs (Song et al., 2003; Yan et al., 2003; Lingel et al., 2004; Parker et al., 2004; Ma et al., 2005; Parker et al., 2005).
Moreover, the three-dimensional structure of the PIWI domain closely resembles that of RNase H, the enzyme that cleaves the RNA strand of an RNA-DNA hybrid (Song et al., 2004; Nowotny et al., 2005), and both structural and biochemical studies have confirmed that Argonaute is the target-cleaving endonuclease of the RISC (Liu et al., 2004; Rand et al., 2004; Song et al., 2004; Bamberger and Baulcombe, 2005; Qi et al., 2005; Rivas et al., 2005). Notably, a subpopulation of Argonaute proteins do not retain all the amino acids that are crucial for RISC catalytic activity, and thus cannot cleave a target RNA even when the small RNA guide is sufficiently complimentary to the target (Liu et al., 2004; Meister et al., 2004b; Rivas et al., 2005). Notably, a subpopulation of Argonaute proteins do not retain all the amino acids that are crucial for RISC catalytic activity, and thus cannot cleave a target RNA even when the small RNA guide is sufficiently complimentary to the target (Liu et al., 2004; Meister et al., 2004b; Rivas et al., 2005). The RISC-associated proteins include the putative RNA-binding protein VIG (Vasa intronic gene), the Fragile-X related protein in Drosophila, the exonuclease Tudor-SN, and several putative helicases (Caudy et al., 2002; Hutvagner and Zamore, 2002; Ishizuka et al., 2002; Mourelatos et al., 2002; Caudy et al., 2003). The molecular function of these proteins in RNA silencing is not known.

**microMechanism**

miRNAs regulate their target genes via two main mechanisms: target mRNA cleavage and ‘translational repression’.

In plants, most miRNAs have perfect or near perfect complementarity to their mRNA targets (Rhoades et al., 2002). Upon binding to their mRNA targets, the miRNA-containing RISCs function as endonucleases, cleaving the mRNA (Llave et al., 2002; Tang et al., 2003). Single miRNA-binding motifs are found both in the coding regions, such as the miR-166-targeting site in the PHABULOSA mRNA, and in the untranslated regions of miRNA-regulated plant mRNAs, such as the miR-156-targeting site in the SPL4 mRNA, albeit mainly in coding sequences, perhaps because cleavage here most strongly inactivates translation of the mRNA into functional protein (Rhoades et al., 2002). At least eight animal miRNA, miR-127, miR-136, miR-196, miR-431, miR-433-3p, miR-433-5p, and two viral miRNAs, miR-BART2 and SVmiRNA, also act to cleave their targets (Mansfield et al., 2004; Pfeffer et al., 2004; Yekta et al., 2004; Davis et al., 2005; Sullivan et al., 2005).

In contrast to plant miRNAs, the complementarity between animal miRNAs and their targets is usually restricted to the 5′ region (nucleotides 2-8 or 2-7) of the miRNA, i.e. to the 3′ region of the target site (Lewis et al., 2003; Lai, 2004; Brennecke et al., 2005; Lewis et al., 2005; Xie et al., 2005). This 5′ miRNA region has been called the ‘seed region’ to describe its disproportionate contribution to target-RNA...
binding. Because the seed region of a miRNA is so short, miRNA are predicted to regulate surprisingly large numbers of genes; the complete complement of human miRNAs may regulate as many as one-third of the human protein-coding genes (Lewis et al., 2005; Xie et al., 2005)! In the absence of extensive complementarity between the miRNA and the target, binding of the RISC blocks translation of the target miRNA into protein, rather than catalyzing its cleavage into two pieces (Olsen and Ambros, 1999). Recent results suggest that regulation by miRNAs can directly target mRNA degradation through a pathway that is distinct from small RNA-directed endonucleolytic cleavage (Bagga et al., 2005; Lim et al., 2005).

In human cells, the core component of the RISCs, Argonaute proteins, together with miRNAs that are targeted for silencing by miRNAs are concentrated in cytoplasmic foci called Processing bodies (P-bodies) (Liu et al., 2005; Pillai et al., 2005; Sen and Blau, 2005). (P-bodies are also known as cytoplasmic bodies or GW-bodies.) miRNAs may initially block translational initiation, causing the miRNA-programmed RISC and the target mRNA to be re-localized to the P-body (Pillai et al., 2005). In C. elegans, a P-body protein, AIN-1, interacts with the miRNA-programmed RISC component ALG-1, an Argonaute protein, and is sufficient to localize ALG-1 to the P-body (Ding et al., 2005). Thus, P-body-mediated miRNA-directed regulation may be a general mechanism among animals.

**microFunctions**

miRNAs function in a broad range of biological processes in plants and animals (Kidner and Martienssen, 2005; Alvarez-Garcia and Miska, 2005). The first insight into their function came from phenotypic studies of mutations that disrupt core components of the miRNA pathway. dicer mutants show diverse developmental defects, including abnormal embryogenesis in Arabidopsis, delayed germ-line stem-cell (GSC) division in Drosophila, germ-line defects in C. elegans, abnormal embryonic morphogenesis in zebrafish and stem-cell differentiation defects in mice (Knight and Bass, 2001; Park et al., 2002; Bernstein et al., 2003; Wienholds et al., 2003; Giraldez et al., 2005; Hatfield et al., 2005). Similarly, the disruption of Argonaute function causes widespread developmental defects, such as defective stem-cell maintenance and failure to form axillary meristem in an Arabidopsis mutant for PINHEAD/ZWILLE (PNH/ZLL) or ARGONAUTE 1 (AGO1), a stem-cell self-renewal defect in Drosophila pivi mutants, and defective early development in C. elegans alg-1 and alg-2 mutants (Bohmert et al., 1998; Cox et al., 1998; Moussian et al., 1998; Grishok et al., 2001). Arabidopsis plants mutant for ZIPPY (ZIP), an Argonaute gene, and HASTY (HST), which encodes the miRNA export receptor, exhibit a precocious vegetative phenotype and produce abnormal flowers (Peragine et al., 2004). Overall, these phenotypes suggest that at least a subset of miRNAs play important roles in early development.

**Target prediction**

The functional characterization of miRNAs relies largely on the identification of their regulatory targets. In plants, because miRNAs are almost perfectly complementary to their targets, target prediction is straightforward (Rhoades et al., 2002), and automated plant miRNA target prediction can now be performed online (Zhang, 2005). At least half of the predicted plant miRNA targets are transcription factors, although transcription factors represent only 6% of Arabidopsis protein-coding genes (Riechmann et al., 2000; Rhoades et al., 2002; Jones-Rhoades et al., 2004). Typically, many members of a family of related transcription factors are coordinately repressed by a single miRNA. These miRNA-regulated transcription factors control developmental patterning, cell proliferation, and environmental and hormonal responses (Kidner and Martienssen, 2005). DCLI and AGO1 themselves are also miRNA targets, suggesting a negative-feedback mechanism in which miRNAs tune their own expression (Rhoades et al., 2002a; Xie et al., 2003; Vaucheret et al., 2004).

The bioinformatic prediction of animal miRNA targets is more complex because animal miRNAs display only modest complementarity to their targets. Different algorithms have been developed to predict animal miRNA targets, using at least some of the following criteria: (1) perfect or nearly perfect sequence pairing. Because depletion by RNAi of the seed region at the 5’ end of the miRNAs and the 3’ UTR of the target mRNA; (2) putative miRNA-binding site conservation between closely related species; (3) multiple miRNA-binding sites in a single target; and (4) lack of a strong secondary structure at the miRNA-binding site on the target (Enright et al., 2003; Lewis et al., 2003; Stark et al., 2003; Kirikakidou et al., 2004; Rajewsky and Socci, 2004; Brennecke et al., 2005; Krek et al., 2005; Lewis et al., 2005; Zhao et al., 2005). The computational prediction of animal miRNA targets suggests that the logic of miRNA regulation differs between animals and plants (see Box 1). In animals, miRNA has been proposed to fine-tune the expression of hundreds of genes, but to dramatically downregulate the expression of a much smaller number of transcripts (Bartel, 2004); such dramatic downregulation of transcript levels appears to be widespread for plant miRNAs. Moreover, animal miRNAs, but perhaps not plant miRNAs, may act combinatorially, with several miRNAs binding a single transcript. Thus, one miRNA might be expressed early in development, reducing the steady-state level of protein synthesis from a targeted mRNA by just a bit, a tuning function. The subsequent expression of additional miRNAs targeting the same mRNA would lower its expression still further (Bartel, 2004). Cell culture experiments suggest that when multiple miRNAs bind the same target, they act cooperatively, reducing mRNA translation by more than the sum of their individual effects (Doench et al., 2003).

miRNA-target relationships can also be identified by beginning with a target miRNA and searching for one or more regulatory miRNAs. In one case, the earlier finding that GY-box, Brd-box and K-box motifs in the 3’ UTR of Notch mRNAs mediate their post-transcriptional repression helped to identify three families of Drosophila miRNAs that are direct regulators of Notch target genes (Lai and Posakony, 1997; Lai et al., 1998; Lai, 2002; Stark et al., 2003; Lai et al., 2005). Another example is the proposal that miR-16 in Drosophila plays a role in AU-rich element (ARE)-mediated mRNA degradation (Jing et al., 2005). Because depletion by RNAi of key RNA silencing proteins – Dcr-1, Ago-1 and Ago-2 – inhibited the rapid mRNA decay normally triggered by AREs, the authors broke with ‘microOrthodoxy’ and proposed that the ARE is a potential miRNA target site, perhaps binding miR-16 in an unconventional mode that does not require seed-sequence pairing.
miRNA profiling has also been used to identify miRNAs with potentially important developmental roles. The rationale is that if a miRNA is highly expressed in a tissue or cell type or at a specific developmental stage, it may reasonably play a regulatory role in specifying tissue or cell identity, or in regulating developmental timing. miRNA expression can be profiled by the cloning and sequencing miRNAs from specific tissues or developmental states (a labor-intensive method that has the benefit of uncovering new miRNAs), or by microarray analysis (a more high-throughput method that can only reveal the expression of known miRNAs). For example, miR-181, which is highly expressed in mouse bone marrow B-lymphoid cells, but not in T-cells, was found to promote hematopoietic differentiation towards the B-cell lineage (Chen et al., 2004). miR-375, an evolutionarily conserved, pancreatic islet-specific miRNA identified by small RNA cloning from glucose-responsive murine pancreatic cell lines, suppresses glucose-induced insulin secretion by repressing myotrophin (Mtpn) expression (Poy et al., 2004). miRNA expression profiling has also identified a zebrafish miRNA that regulates brain morphogenesis, miR-430, whose expression peaks 4 hours after fertilization (when most fish miRNAs are first expressed) and decreases after 24 hours (Chen et al., 2005; Giraldez et al., 2005). miRNA expression profiling has also been used to block miRNA expression in cultured cells, but its broad utility is not yet established (Jing et al., 2005; Lee et al., 2005).

**Box 1. MicroRNAs and siRNAs: the logic of small RNAs**

Small interfering RNAs (siRNAs) are a class of small RNA guides that are distinct from miRNAs (Tomari and Zamore, 2005). These two classes of RNA cannot be distinguished by either their chemical composition or their function. Both are produced by Dicer-mediated cleavage of longer, double-stranded RNA precursors. Consequently, both miRNAs and siRNAs are ~21 nucleotides in length, with 5′-phosphate and 3′-hydroxyl termini. siRNAs and miRNAs are functionally interchangeable: both can direct target mRNA cleavage or translational repression, depending on the degree of complementarity between the small RNA and its target. Nonetheless, the two classes of small RNAs can be distinguished both by their biogenesis pathways and by the logic by which they regulate target genes, especially in animals. miRNAs are processed from small hairpin transcripts, called pre-miRNAs, which are embedded in much longer primary transcripts, the pri-miRNA. Each miRNA corresponds to ~21 nucleotides of one arm of the pre-miRNA stem. The resulting miRNA may regulate tens or hundreds of target RNAs, whose only common features are the short sequences that are complementary to as few as six or seven bases of the miRNA. siRNAs, by contrast, derive from endogenous or exogenous long double-stranded RNAs, typically comprising hundreds or thousands of base pairs. Such precursor double-stranded RNAs typically yield many siRNAs from both strands. The regulatory targets of siRNAs are usually highly homologous to the trigger double-stranded RNA itself, i.e. the trigger and target genes are paralogs. In some cases, such as with heterochromatic siRNAs, which initiate heterochromatin assembly, the trigger and target genes are one-and-the-same. A notable exception is trans-acting siRNAs (tasRNAs) in plants, in which only a few of the many siRNAs generated from the long, double-stranded RNA trigger appear to correspond to regulated target RNAs (Allen et al., 2005; Peragine et al., 2004; Vazquez et al., 2004b).

which may control the balance between differentiation and proliferation during cardiogenesis by regulating the expression of Hand2 mRNA (Wienholds et al., 2005; Zhao et al., 2005).

To date, hundreds of miRNAs have been identified in different organisms, which makes it possible to study their function individually by suppressing their expression in cells. However, genetic depletion, i.e. making miRNA mutants, is labor intensive. 2′-O-methyl antisense oligonucleotides complementary to endogenous miRNAs provide an alternative to genetic mutation. These antisense oligonucleotides transiently block miRNA function (Hutvagner et al., 2004; Meister et al., 2004a). Thus, injecting into worms a 2′-O-methyl oligonucleotide that binds let-7 recapitulates the let-7 mutant phenotype (Hutvagner et al., 2004). In early syncitial Drosophila embryos, where injection of oligonucleotides is straightforward, a panel of 2′-O-methyl oligonucleotides was used to reveal the embryonic loss-of-function phenotypes of 46 miRNAs (Leaman et al., 2005). This study suggests that miRNAs specifically regulate a broad range of developmental events. In another study, Lecellier et al. used antisense locked-nucleic acid (LNA) oligonucleotides, nucleic acid molecules that are modified to dramatically increase their binding affinities, to block miR-32 in cultured human cells, a miRNA proposed to mediate innate anti-viral defense (Lecellier et al., 2005). RNAi itself has also been used to block miRNA expression in cultured cells, but its broad utility is not yet established (Jing et al., 2005; Lee et al., 2005).

Human viruses also express their own miRNAs (Pfeffer et al., 2004; Cai et al., 2005; Pfeffer et al., 2005; Sullivan et al., 2005). Viral miRNAs are proposed to regulate both viral and host gene expression (Pfeffer et al., 2004; Cai et al., 2005), but only viral miRNA targets have been experimentally validated (Pfeffer et al., 2004). Recently, Simian Virus 40-encoded miRNAs have been identified. These viral miRNAs accumulate at late times in infection, and target early viral RNAs for cleavage and reduce viral susceptibility to cytotoxic T cells (Sullivan et al., 2005). Whether viral miRNAs always mediate the cleavage of viral miRNAs or whether they can also act more like animal miRNAs to ‘tune’ gene expression remains unknown.

**microPrognostication**

A dozen years after their discovery, miRNAs represent a large class of regulators of gene expression that control a broad range of physiological and developmental processes in plants and animals. An immediate challenge is to tabulate the functions of each and every miRNA. For this, improved computational and experimental methods for the identification of miRNA targets will be essential. These efforts will no doubt be informed by our expanding knowledge of the mechanism by which miRNAs recognize and regulate their targets. The regulation of miRNA expression and maturation remains largely unknown, but many laboratories have begun to map where miRNAs are expressed and what factors regulate their transcription. Perhaps the most important goal in understanding miRNAs will be to describe how miRNAs function as a network, for it is in studying the coordinate action of multiple miRNAs on a single mRNA target that we are likely to reach a deeper understanding of the logic of these small but powerful ribo-regulators.
References


Primer 4651


