MicroRNAs: Biogenesis and Molecular Functions

Xuhang Liu*; Kristine Fortin*; Zissimos Mourelatos

Department of Pathology and Laboratory Medicine, Division of Neuropathology, University of Pennsylvania School of Medicine, Philadelphia, Pa.

Abstract

Small regulatory RNAs are essential and ubiquitous riboregulators that are the key mediators of RNA interference (RNAi). They include microRNAs (miRNAs) and short-interfering RNAs (siRNAs), classes of ~22 nucleotide RNAs. miRNAs and siRNAs bind to Argonaute proteins and form effector complexes that regulate gene expression; in animals, this regulation occurs primarily at the post-transcriptional level. In this review, we will discuss our current understanding of how miRNA and siRNA are generated and how they function to silence gene expression, focusing on animal and, in particular, mammalian miRNAs.

Keywords

microRNA; miRNA; Argonaute; Ago; Dicer; RNA interference; RNAi; miRNP; RISC; siRNA; mRNA transcription.

Corresponding author:
Zissimos Mourelatos M.D., Department of Pathology and Laboratory Medicine, Division of Neuropathology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104
(E-mail: mourelaz@uphs.upenn.edu)

Received: 14 November 2007; accepted 16 November 2007.

* Equal contribution.

doi:10.1111/j.1750-3639.2007.00121.x

An overview of how animal microRNAs (miRNAs) and short-interfering RNAs (siRNAs) are generated and function that concisely captures the essence of the miRNA and RNAi field (24) is shown in Figure 1. Details are discussed in the remainder of this review.

miRNA GENES: ORGANIZATION, TRANSCRIPTION AND NUCLEAR PROCESSING

The official miRNA database (miRBase) currently lists ~540 human miRNAs (31). Most of the miRNA genes are transcribed by RNA Polymerase II (13, 49), although a cluster of human miRNAs interspersed within Alu elements in Chromosome 19 have recently been shown to utilize RNA Polymerase III for their transcription (13). The primary miRNA transcripts are termed pri-miRNAs (47). Many miRNA genes tend to be in close proximity to other miRNAs, and are transcribed as polycistronic transcripts (47). According to the genomic region that a miRNA resides in, miRNAs can be grouped into several categories: intronic miRNAs in protein-coding genes; exonic miRNAs in non-coding genes; intronic miRNAs in non-coding genes (84). Many mammalian miRNAs lie within introns of protein-coding genes and have the same transcription pattern as that of the protein-coding genes in which they reside (84). When located in imprinted regions, some miRNAs (eg, mir-127 and mir-136) have also been found to exhibit parental specific expression pattern as that of imprinted genes. Very interestingly, the prototype of imprinted genes—H19 non-coding RNA—has recently been shown to give rise to miR-675 in both humans and mice (12).

A typical monocistronic pri-miRNA is composed of a double-stranded stem of ~33 base pairs, a terminal loop and two flanking unstructured single-stranded segments (Figure 2). After transcription, a pri-miRNA is first cropped by the microprocessor into a ~70 nucleotide (nt) hairpin-like precursor miRNA (pre-miRNA; Figure 2; (16, 29, 35, 45, 48). The core components of the microprocessor are the RNase III enzyme Drosha (48) and a double-stranded (ds) RNA binding protein termed DGCR8/Pasha (16, 29, 35). Drosha liberates the double-stranded stem from the remainder of the pri-miRNA by cleaving proximal and distal of the stem, and thus generates a pre-miRNA that has a 5′ monophosphate and a 3′2-nt hydroxyl overhang (Figure 2) (48). Very interestingly, DGCR8 has recently been shown to be a heme-binding protein (22). Binding to heme dimerizes DGCR8, substantially enhances the affinity of DGCR8 to pri-miRNA and the cleavage of pri-miRNA by microprocessor (22). Cleavage of a pri-miRNA by microprocessor begins with DGCR8 recognizing the ssRNA-dsRNA junction typical of a pri-miRNA (36). Then, Drosha is brought close to its substrate through interaction with DGCR8 and cleaves the stem of a pri-miRNA ~11 nt away from the two single-stranded segments (36). Although microprocessor is already sufficient for conversion of a pri-miRNA into a pre-miRNAs in vitro, cleavage of pri-miRNA in vivo does not depend on Drosha and DGCR8 only, but also on other accessory proteins, such as the
RNA binding protein hnRNP A1 (32) and the p68 and p72 RNA helicases (26). Gene-targeting experiments have demonstrated the importance of p68 and p72 for the biogenesis of a subset of mouse miRNAs (26). In p68 or p72 knock-out embryos (knock-out of p68 in mice causes embryonic lethality while knock-out of p72 causes neonatal death), the expression of a subset of miRNAs is severely compromised (26). In another recent study, hnRNP A1 has been shown to be specifically required for the processing of pri-miR-18a in a context-dependent manner (32).

In addition to the classical biogenesis pathway that depends on microprocessor, a subclass of pri-miRNAs, pre-miRNA/intertrons (mirtrons), have recently been shown to depend on the RNA splicing machinery for their biogenesis in Drosophila, Caenorhabditis elegans and mammals (6, 68, 85). Mirtrons are derived from certain debranched introns that fold into hairpin structures with 5′ monophosphates and 3′ 2-nucleotide, 3′ overhang. In the case of mirtrons, the immediate precursor RNA is termed pre-miRNA, adopts a hairpin structure and has a 5′-phosphate and a 2-nucleotide, 3′ overhang. In the case of siRNAs, the precursor is long dsRNA. Both are processed by the Dicer nuclease into duplexes termed miRNA or siRNA duplexes. The mature miRNA or siRNA is then released from the miRNA or siRNA duplex and binds to an Ago protein, to form a core effector complex that is commonly known as miRNP or RISC. miRNAs or siRNAs then act as specificity determinants to deposit the Ago proteins that they are bound to (RISCs, RNA-induced silencing complex; miRNPs) onto their RNA targets, which are typically mRNAs. The deposited miRNP/RISC silences gene expression by suppressing the translation of the targeted mRNA and/or by destabilizing the targeted mRNA. Ago, Proteins that bind directly to miRNAs and siRNAs and are the core protein components of miRNPs/RISCs; Dicer, dsRNA endonuclease; miRNP, RiboNucleoProtein containing miRNA and an Ago protein.

After a pre-miRNA is released from a pri-miRNA in the nucleus, it is exported by RanGTP and Exportin-5 (Exp-5) to the cytoplasm (Figure 2) (9, 57, 100). After a pre-miRNA is exported to the cytoplasm, RanGTP is hydrolyzed by RanGAP to RanGDP, and the pre-miRNA is released from Exp-5. Exp-5 is also important for stabilizing pre-miRNAs in the nucleus. When Exp-5 is knocked down by siRNAs, the levels of pre-miRNAs are reduced not only in the cytoplasm, but also in the nucleus, suggesting that binding of pre-miRNAs to Exp-5 protects them from degradation (100).

**CYTOPLASMIC PROCESSING OF PRE-miRNAs AND ASSEMBLY OF miRNAs INTO EFFECTOR COMPLEXES**

In the cytoplasm, another RNase III endonuclease termed Dicer is responsible for dicing pre-miRNAs into short RNA duplexes termed miRNA duplexes (Figures 2 and 3) (7). The RNA strand of the miRNA duplex that is complementary to the mature miRNA is depicted with a star symbol (miRNA*; Figure 3). Dicer can also process long dsRNA into small RNA duplexes (33), and these small RNAs (33) are termed siRNA duplexes (19). In addition to two RNase III signature domains, mammalian Dicer has a N-terminal ATPase/helicase domain, a DUF 283 domain, a PAZ domain and a C-terminal dsRNA binding domain (dsRBD) (76, 102, 103). Biochemical experiments have revealed that both PAZ domain and dsRBD are essential for the interaction of Dicer with pre-miRNAs and long dsRNAs (103). PAZ domain functions in recognizing the 2-nucleotide 3′ overhang signature generated by Drosha, while dsRBD is critical for binding long dsRNAs (103). After capturing pre-miRNAs or long dsRNAs with its PAZ domain and dsRBD, Dicer
dimerizes its two RNase III domains intra-molecularly to form a single processing center, and cuts the stem of pre-miRNAs or long dsRNAs ~22 nt away from their termini at positions separated by 2 nts, which generates 3’ 2-nt termini (103) (Figures 2 and 3). In the case of long dsRNA, subsequent to the first cleavage, Dicer can use its PAZ domain to place it again at the termini of the dsRNA and dice processively from the termini (76, 102). Although the functions of the C-terminal domains of Dicer are clear now, the roles of the N-terminal ATPase/helicase domain and DUF 283 domain of Dicer still remain elusive.

After Dicer processing, the miRNA duplex is unwound and the mature miRNA binds to an Argonaute (Ago) protein in a process that is referred to as miRNA loading or assembly, while the miRNA* is degraded (Figures 2 and 3).

The Argonaute family is a diverse family of proteins, each containing characteristic domains termed PAZ and PIWI. The Argonaute family can be phylogenetically divided into the Ago and Piwi protein families based on similarities to Arabidopsis AGO1 and Drosophila Piwi proteins, respectively (14). miRNAs bind Ago proteins whereas Piwi proteins bind a newly discovered class of small RNAs known as piwi-interacting RNAs (piRNAs), which are almost exclusively expressed in the germline [reviewed in (3, 40)].

The miRNA/Ago ribonucleoprotein that is formed represents the core component of the effector complexes that mediate miRNA function and is known as miRNP (65). siRNA duplex unwinding also leads to loading of one strand to an Ago protein, and the resultant complex is referred to as RNA-induced silencing complex (RISC) (34). The siRNA strand that is bound to Ago
proteins is termed the guide strand, whereas the opposite strand is referred to as the passenger strand (Figure 4) [reviewed in Tomari and Zamore (91)]. miRNPs and RISCs are functionally equivalent.

A primary determinant of which the two strands of an miRNA duplex or an siRNA duplex will be loaded on Ago proteins is the inherent thermodynamic asymmetry of the miRNA or siRNA duplex. The RNA strand whose 5′ end is less stably bound to the opposite strand will be loaded to Ago proteins and forms the mature miRNA or siRNA (Figure 4) (91, 92).

The factors and mechanisms that sense the thermodynamic asymmetry of siRNA duplexes have been elucidated in Drosophila melanogaster. Flies contain two Dicer proteins (Dcr-1 and Dcr-2) with distinct functions. Dcr-1 is responsible for miRNA production and is an essential protein that is required for fly development (50), whereas Dcr-2 is required for siRNA production from long dsRNA, but flies lacking Dcr-2 develop normally (50). After siRNA duplexes are generated by Dcr-2, they interact with the RISC loading complex (RLC) that is comprised of Dcr-2 and its dsRNA binding partner R2D2 (Figure 4) (54, 91, 92). The Dcr-2/R2D2 heterodimer examines the base-pairing extent of siRNA duplexes and of miRNA/miRNA* duplexes (the latter generated by Dcr-1), and sorts them into different D. melanogaster (dm) Ago proteins.

miRNA/miRNA* duplexes that typically have mismatches and thus have limited double-stranded areas, have low affinity to R2D2, and are preferentially loaded into dmAgo1-RISC (25, 93). In contrast, with extensive complementarity between its two strands, siRNAs have high affinity to R2D2 and are loaded into dmAgo2 (Figure 4) (25, 93). Once a siRNA is bound by the Dcr-2/R2D2 heterodimer, R2D2 senses the thermodynamic differences in the base-pairing stabilities of the 5′ ends of siRNA duplexes, and binds to the end with the greatest double-stranded character (Figure 4) (92). During loading of miRNA/miRNA* or siRNA duplexes into Ago proteins, the miRNA or the guide strand of a siRNA is dissociated from the miRNA* or the passenger strand differentially. The passenger strand of a siRNA is cleaved by dmAgo2, and is released from the guide strand (63, 78). On the contrary, many miRNA/miRNA* duplexes have mismatches in the central region (positions 9–11) of the duplexes, and thus dmAgo2 cannot release the miRNA* strand by cleavage. Instead, many miRNA/miRNA* duplexes have mismatches in the seed region (positions 2–8 of the mature miRNA), and the binding affinity in the seed region of a miRNA/miRNA* duplex is much weaker than that of an siRNA duplex. Accordingly, a miRNA* strand is separated from a miRNA strand by a bypass mechanism, although the detail of the bypass mechanism is still unclear at this point.

Humans and other mammals contain a single Dicer gene and miRNP, and RISC assembly has many similarities but also important differences to RISC assembly in flies. In humans, miRNP assembly is accomplished by a protein complex termed the miRNA RISC Loading Complex (miRLC) (Figure 5) (59). The miRLC is a multiprotein complex whose core components are Ago and Dicer proteins (30, 59). The miRLC is devoid of miRNAs and processes miRNAs from pre-miRNAs, and loads mature miRNAs to Ago proteins (59). The miRLC is then disassembled and the core miRNP (miRNA-Ago ribonucleoprotein) is generated (Figure 5) (59). However, the details of miRNP assembly in humans are unknown.
AGO PROTEINS, THE MAJOR MEDIATORS OF miRNA AND siRNA FUNCTION

The major mediators of miRNA and siRNA function are Ago proteins, which are ~95 kDa proteins that contain N-terminal and Mid-domains in addition to the PAZ and PIWI domains. Ago proteins are highly conserved from metazoans to fission yeast. Most organisms possess multiple paralogs, the exception being the fission yeast, Schizosaccharomyces pombe, which has a single representative (14). Humans and mice contain four Ago proteins (Ago1–4).

Structural and biochemical analyses have shown that the ~130-amino-acid PAZ domain contains an oligonucleotide-binding fold that allows the protein to bind the single-stranded 2-nt 3′ terminal overhangs characteristic of small RNAs processed by Dicer (52). Furthermore, crystallographic studies of archaeal Ago family members have demonstrated that the PIWI domain of these proteins adopts a three-dimensional fold similar to that of the RNase H family of endonucleases (70, 89, 101). RNase H proteins are characterized by a structural motif consisting of three β sheets surrounded by α helices (98), and they cleave RNA in RNA–DNA duplexes. Interestingly, only certain Ago proteins, including the Ago2 protein in humans, possess the triad of residues required for catalytic activity. The archaeal crystal structures, together with mutational analysis, have revealed that two aspartate residues (~18 kDa) and one histidine residue (H807) are required for the catalytic function of the human Ago2 protein (53, 83). Interestingly, only mammalian Ago2 is catalytically active, whereas the other Ago proteins are not, even though some of them contain the catalytic triad.

Co-crystals of archaeal Ago proteins complexed with siRNAs have demonstrated that PIWI domains contain a conserved structural element that serves to bind and anchor the 5′ phosphate of small RNAs (58, 71). Additionally, recent analyses of the Ago Mid-domain have revealed the presence of sequence motifs similar to the cap-binding domain of the eukaryotic initiation factor eIF4E (42); the significance of this domain is discussed below.

FUNCTIONS OF miRNPs AND RISCs

miRNAs base-pair with miRNA recognition elements (MREs) found in their mRNA targets (typically in the 3′ untranslated region—3′ UTR) and deposit their bound Ago proteins onto mRNA targets. The result is translational repression of the targeted mRNA, often followed by miRNA destabilization or endonucleolytic cleavage of the targeted mRNA. The exact molecular function is dependent upon how extensive the complementarity of the miRNA or siRNA is with its mRNA target and which Ago protein is deposited on the mRNA target. If an miRNA or siRNA bound to Ago2 pairs with extensive complementarity with a cognate mRNA target, then the mRNA is cleaved at a position across from nucleotides 10 and 11 of the miRNA (or siRNA), while the miRNA remains intact (Figure 6) (19, 38, 53). This cleavage event produces 5′-phosphate and 3′-hydroxyl terminal products, characteristic of other RNase H-like enzymes (61, 87). The target mRNA is subsequently degraded via routine cellular pathways (Figure 6). Target mRNA cleavage by miRNAs is the major mechanism of regulation by plant miRNAs (18, 55). In animals, however, there are very few examples of miRNAs that regulate their mRNA targets by cleavage (99); rather, the predominant silencing mode of animal miRNAs is to repress the translation of their mRNA targets and/or to destabilize them without endonucleolytic cleavage (1, 23).

Early studies in the worm C. elegans demonstrated that the small non-coding RNAs lin-4 and let-7 (now known to be miRNAs) represented imperfect, complementary matches to sequences within the 3′ UTRs of their respective target mRNAs (46, 80). These interaction between the miRNA and target mRNA further resulted in decreased target protein levels (96) without affecting the stability of the mRNA (69). This profile of a significant reduction in protein level without a proportionate reduction in target mRNA levels became a hallmark of miRNA function.

We now know that the vast majority of animal miRNAs base-pair with imperfect complementarity with their mRNA targets. Experimental and bioinformatics approaches have shown that the most important determinant of target RNA recognition by a miRNA is perfect or near-perfect complementarity between the proximal (5′) region of the miRNA and the mRNA, also known as the “seed” region or the “nucleus” (Figure 7A) (10, 17, 41, 44, 51, 77). Base-pairing between the 3′ portion of the miRNA and the mRNA target is not always essential for repression, but strong base-pairing within this region can partially compensate for weaker seed matches or enhance repression (11, 41). Additionally, multiple MREs for the same, or different, miRNAs within the same 3′ UTR can function cooperatively to enhance repression (17, 43). Spacing of the seed sites within the 3′ UTR may play a significant role in the cooperative action of miRNAs (86). Finally, sequences adjacent to MREs (94) and the secondary structure of the 3′ UTR of the target mRNA affect binding of miRNAs; miRNPs cannot efficiently “unwind” structured RNA areas, and thus miRNPs cannot bind to sites that are embedded in such structured areas (2, 56). Applying such miRNA binding rules to computational algorithms predicts that a large fraction (~1/3) of the human mRNA transcriptome may be targeted by miRNAs [reviewed in Bartel and Chen (5)]. However, additional binding configurations between miRNAs and their targets may be functional, thus raising the fraction of miRNAs that may potentially be regulated by miRNAs (64, 82).

Regulation of translation efficiency may occur during initiation, elongation or termination [for review see Gebauer and Hentze (27)]. In general, most translation repressors target translation ini-
Recent studies have lent additional support to the model of miRNA-mediated repression occurring at a post-initiation step (60, 67, 73) (Figure 7B). miRNA-mediated translational repression results in decreased levels of the targeted protein. Thus, the model of miRNA function occurring at a post-initiation step would require the release of and subsequent destruction of the resulting polypeptide. To date, no study has identified the accumulation of a nascent polypeptide corresponding to mRNAs repressed by miRNAs (67).

A second model proposes that miRNA-mediated translational repression occurs at the initiation step (Figure 7B). Using human cells and reporter constructs targeted by either endogenous or exogenous miRNAs, two groups initially discovered that the m7G cap of an mRNA is required for efficient translational repression (37, 75). miRNA-repressed messages sedimented in light fractions in polysome analyses, indicating that their translation was inhibited at the level of initiation (8, 75). Ago proteins are necessary for miRNA-mediated translational repression. This was effectively shown in a study that enabled Ago proteins, expressed as fusion proteins to the phage λN-peptide, to be artificially tethered to a reporter construct bearing several binding sites for the λN-peptide in its 3′UTR (74). This system, which is presumably miRNA-independent, results in repression of the reporter construct translation without significantly affecting the reporter mRNA level (74). Interestingly, Ago proteins have recently been shown to bear sequence similarities to the cap-binding domain of eIF4E (42). Specifically, the two conserved tryptophan residues (W56 and W102) within eIF4E that are crucial for binding of the m7G cap are conserved within Ago proteins as phenylalanine residues (F470 and F505 in human Ago2) (42). Furthermore, using the Ago tethering system described above demonstrated that the ability for Ago2 protein to repress translation is dependent upon these residues (42). This study provides a molecular model to explain the cap-dependent inhibition of translation initiation by Ago2 protein by proposing that binding of Ago2 to the m7G cap of an mRNA target precludes binding of eIF4E, thus inhibiting initiation of translation (42). Recent studies using in vitro translation systems to study the function of miRNAs provide strong evidence that miRNAs repress the initiation of translation in an m7G cap-dependent manner (62, 90, 95). In particular, one study shows that, in vitro, translation initiation is the first step that is inhibited by a miRNA, and that there is a competition for access to the m7G cap of mRNAs by miRNPs, which repress initiation, and the eIF4E/G complex that promotes initiation (62). Interestingly, miRNA-dependent mRNA deadenylation is also recapitulated in vitro (95).

But other steps in the mechanism of miRNA-mediated repression may still await discovery. Another recent study focusing on the translational inhibitory protein eIF6 has found, via biochemical
and genetic studies, that this protein might also play an important role in miRNA-mediated translational repression, perhaps at a step downstream of Ago function (15).

The fate of mRNAs targeted by miRNAs is becoming clearer, although no less complex. For example, although early studies in C. elegans proposed that the miRNAs let-7 and lin-4 repressed target mRNAs without affecting their stability, it has since been demonstrated that significant degradation of these mRNAs occurs (4). For many mRNAs, this accelerated degradation appears to be due secondary to miRNA-mediated stimulation of target deadenylation (28, 97) (Figure 7B). However, degradation of targeted messages is not a universal finding; additionally, studies of the endogenous cationic amino acid transporter 1 (CAT-1) mRNA, a target of the liver-specific miR-122, have shown that the repression of some miRNA targets can be reversed under certain conditions (8).

What causes some miRNA targets to be degraded and others preserved is currently unknown, although Ago proteins have been shown to interact with proteins involved in RNA degradation, including the GW182 protein and proteins that are involved in decapping mRNAs (79). Indeed, an emerging topic of study is the relationship between the microRNA pathway and cytoplasmic foci known as Processing bodies (P-bodies) or GW-bodies. P-bodies (for reviews see Eulalio et al, Parker and Sheth (21, 72) are subcellular structures thought to assist in the translational repression and storage or degradation of mRNAs. They are characterized by possessing an abundance of enzymes involved in deadenylation, decapping and degradation of mRNAs. P-bodies are notably preserving an abundance of enzymes involved in deadenylation, decapping and degradation of mRNAs. P-bodies are notably devoid of both ribosomes and translation initiation factors except eIF4E and its transporter, eIF4E-T. P-bodies contain miRNPs and repressed mRNAs, and they likely form as a consequence of mRNA repression, including miRNA-dependent repression (20, 21).

The ultimate fate of a miRNA-targeted mRNA likely depends on many factors in addition to Ago proteins, such as proteins that interact with miRNPs. It also depends on competition between the miRNA-dependent silencing pathways and other cellular pathways. The best example to date comes from studies of the aforementioned miR-122. Under normal conditions, the translation of CAT-1 mRNA is inhibited, and the CAT-1 mRNA is targeted to P-bodies as a result of repression by miR-122 (8). However, various stress conditions have been shown to abrogate this repression, resulting in loss of CAT-1 mRNA from P-bodies and the return of these messages to polysomes (8).

Finally, it should be noted that, to date, virtually nothing is known about the regulation of mammalian miRNAs or Ago proteins. Thus, ultimately, a model where miRNAs function through a variety of mechanisms, including the inhibition of both translation initiation and post-initiation is plausible. An important challenge for the future is to understand how miRNPs are regulated and how they intersect with other cellular pathways that regulate mRNA translation, stability, localization and metabolism.

ACKNOWLEDGMENTS

We apologize to those whose work could not be cited because of space limitations. Supported by NIH Grants to KF (F30NS054396) and ZM (GM0720777, NS053839, P30-HD026979), and by the Philadelphia Foundation to ZM.

REFERENCES


