



The microRNA world: small is mighty

Peter Nelson¹, Marianthi Kiriakidou^{1,2}, Anup Sharma¹, Elsa Maniatakis¹ and Zissimos Mourelatos¹

¹Department of Pathology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6100, USA

²Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6100, USA

A new paradigm of RNA-directed gene expression regulation has emerged recently, profound in scope but arresting in the apparent simplicity of its core mechanism. Cells express numerous small (~22 nucleotide) RNAs that act as specificity determinants to direct destruction or translational repression of their mRNA targets. These small RNAs arise from processing of double-stranded RNA by the Dicer nuclease and incorporate with proteins that belong to the Argonaute family. Small RNAs might also target and silence homologous DNA sequences. The immense potential of small RNAs as controllers of gene networks is just beginning to unfold.

A mighty flame could begin as a faraway spark. Such is the case with MICRORNAs (miRNAs; see Glossary), an evolutionary conserved class of small (~22 nucleotide; nt) noncoding RNAs, whose discovery has recently created much heat and light. The preceding spark, however, occurred ten years ago with the discovery by the Ambros laboratory that *lin-4*, a gene required for proper development in *Caenorhabditis elegans*, encodes not a protein but a 22nt noncoding RNA: the founding member of miRNAs [1]. At the same time, experiments in fungi [2] and an even earlier fortuitous discovery in plants [3,4] provided further glimpses into the fascinating phenomenon of RNA-mediated gene silencing. More recently, there has been remarkable progress and the emergence of a unified mechanistic view of gene expression regulation directed by small RNAs. Historical accounts detailing these important early discoveries have been presented [5]. In this review, we will attempt to present a working model of how miRNAs control gene expression based on recent developments, and discuss what, in our view, constitute the most pressing, unsolved issues.

Discovery of miRNAs and connections with RNA-directed gene silencing mechanisms

Genetic screens performed by the Ambros laboratory to characterize the heterochronic gene pathway (i.e. the temporal progression of developmental events in *C. elegans*) uncovered a ~22nt noncoding RNA as the product of the *lin-4* gene [1]. *lin-4* RNA repressed the protein levels of *lin-14*, a gene that functions in the same developmental pathway. The *lin-4* RNA had the potential to bind, with partial antisense complementarity, to sequences found in the 3'-untranslated region (3'-UTR)

of *lin-14* mRNA and repress its translation [1]. For some time, *lin-4* was the only example of an endogenous ~22nt regulatory RNA. The situation changed dramatically with the discovery by the Ruvkun laboratory of *let-7*, a second ~22nt RNA that also functioned in the heterochronic gene pathway [6]. Similarly to *lin-4*, *let-7* RNA recognized sequences present in the 3'-UTR of its *lin-41* mRNA target and repressed LIN-41 protein levels; *lin-4* and *let-7* RNAs were named small temporal (st) RNAs [6]. A key observation was that *let-7* was evolutionary conserved and expressed in all bilaterians tested [7]. This finding suggested that other ~22nt regulatory RNAs might exist in diverse organisms and could be derived, like stRNAs, from double-stranded RNA (dsRNA)-hairpin precursors. Meanwhile, studies in plants, fungi and worms were uncovering a central role for RNA in mediating silencing of genes homologous to the RNA trigger, a phenomenon generally known as RNA INTERFERENCE (RNAi). The Fire and Mello laboratories demonstrated that dsRNA was the critical trigger for target gene silencing [8], and the Baulcombe laboratory discovered that a species of small

Glossary

Argonaute : family of proteins that contain two conserved domains termed PAZ and PIWI. Multiple paralogs are typically present in each organism. Argonaute proteins are crucial for the maturation and function of microRNAs (miRNAs) and are essential for RNA interference. eIF2C2 is a human Argonaute protein.

Dicer : RNase III-type nuclease that also contains RNA helicase, PAZ and double-stranded RNA (dsRNA)-binding domains. Dicer processes linear, dsRNA into small interfering RNA (siRNA) duplexes and also excises mature miRNAs from pre-miRNAs.

Micro-ribonucleoproteins (miRNPs) : ribonucleoprotein complex containing miRNAs, an Argonaute protein (eIF2C2) and the proteins Gemin3 (an RNA helicase) and Gemin4.

MicroRNAs : ~22nt noncoding RNAs derived from endogenous genes. Processed from one of the strands of longer (~75nt) hairpin-like precursors termed pre-miRNAs. miRNAs assemble in complexes termed miRNPs and recognize their mRNA targets by antisense complementarity. If the complementarity is extensive, the target mRNA is cleaved and the miRNA acts as an siRNA; if the complementarity is partial, the translation of the target mRNA is repressed.

miRgonaute/siRgonaute : argonaute protein bound to miRNAs or siRNAs. RNAi-induced silencing complexes (**RISC**) : Multisubunit nuclease that directs target RNA destruction in **RNA interference (RNAi)**. The core components of RISCs are siRNAs and Argonaute proteins.

RNA interference : initially defined as a technique in which experimental introduction in *C. elegans* of dsRNA homologous to a target mRNA led to degradation of the targeted mRNA. More broadly defined as degradation of target mRNAs by homologous siRNAs.

Small interfering RNAs : ~22 to 25nt RNAs derived from processing of linear dsRNA. siRNAs assemble in complexes termed RISCs and target homologous RNA sequences for endonucleolytic cleavage. Synthetic siRNAs also incorporate in RISCs and cleave homologous RNA sequences.

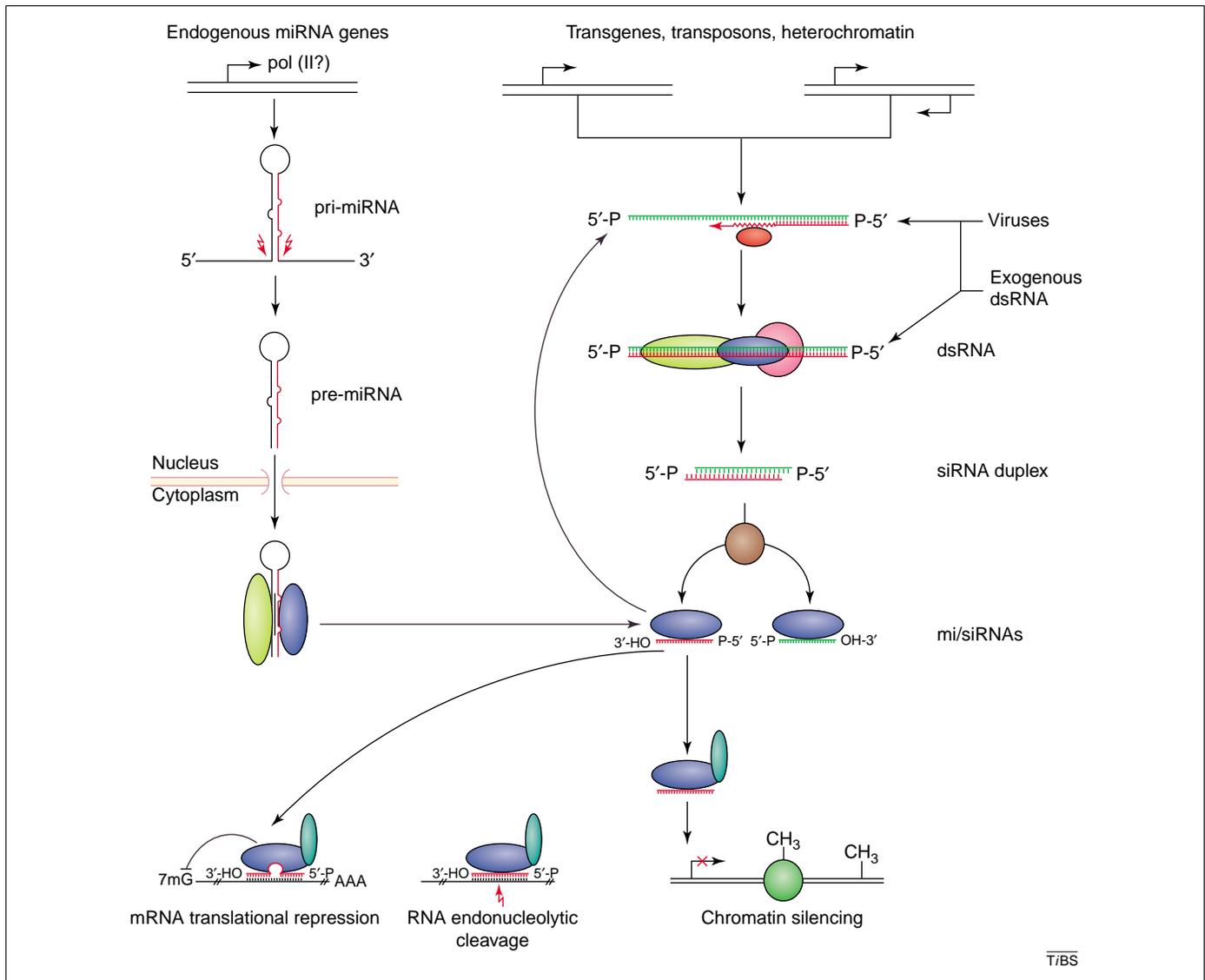


Figure 1. Proposed unified model of small interfering RNAs/microRNAs (mi/siRNA)-mediated gene expression regulation. miRNA biogenesis (left): transcription of endogenous miRNA genes (most likely by RNA polymerase II) generates pri-miRNAs that are processed in the nucleus into pre-miRNAs. pre-miRNAs are exported to the cytoplasm and are processed by Dicer (light green), possibly in conjunction with Argonaute proteins (dark blue), into mature miRNAs, which bind to Argonaute proteins to form miRgonaute ribonucleoproteins and also associate with additional proteins to form miRNPs. In plants, miRNAs are processed in the nucleus. siRNA biogenesis (right): RNAs derived from transgenes, transposons and heterochromatic repeats are substrates for RNA-dependent RNA polymerase (RdRP; red) and are converted to double-stranded RNA (dsRNA), which is processed by Dicer into siRNA duplexes. Other Dicer substrates might include exogenous dsRNA (i.e. experimentally introduced) or foreign nucleic acids (such as viral RNAs). Dicer might cooperate with Argonaute proteins (such as RDE-1) and dsRNA-binding proteins (such as RDE-4; pink) to generate siRNAs. siRNA duplexes are unwound and single-stranded siRNAs bind to Argonaute proteins to form siRgonaute ribonucleoproteins and also associate with other proteins to form RISCs. mi/siRNA function (bottom): we hypothesize that mi/siRNAs recognize their cognate RNA targets in the form of mi/siRgonaute ribonucleoproteins. If the complementarity between a mi/siRNA is partial, the translation of the target mRNA is repressed (left), whereas if the complementarity is extensive, the target RNA is destabilized by endonucleolytic cleavage (*center*). siRgonautes might also recognize homologous DNA and silence chromatin by histone (green) and DNA methylation (right). In addition to Argonaute proteins, mi/siRNAs might bring with them, or recruit once bound to their targets, other, as yet unidentified, factors (depicted as 'effector specific factors'; turquoise). These hypothetical factors might play crucial roles in mi/siRNA function. In some organisms (such as worms, plants or fission yeast) siRNAs might bind to their homologous RNAs and act as primers for RdRP to generate additional dsRNA.

~22–25nt RNAs, later named SMALL INTERFERING RNAs (siRNAs) by the Tuschl laboratory [9], were found in plants undergoing RNA-mediated gene silencing [10]. These observations suggested that the numerous manifestations of RNA-mediated gene silencing observed in diverse organisms might have a common mechanism: dsRNA (endogenous or experimentally introduced) might give rise to ~22nt RNAs that target and silence homologous mRNA sequences. The Tuschl, Bartel and Ambros laboratories, and soon thereafter the Dreyfuss laboratory, reported the existence in different organisms of >100 endogenous ~22nt RNAs. These RNAs are derived from longer

hairpin-like precursors and were named microRNAs [11–14]. stRNAs are thus the two founding members of this large class of small regulatory RNAs.

Small is mighty: function of mi/siRNAs

There has been tremendous progress since these initial reports on miRNAs were published, and rather than exhaustively report on all available data on miRNAs, we will attempt to provide a synthesis and a working model of their biogenesis, function and intersection with RNA-mediated gene silencing pathways (Figure 1).

miRNA genes

A total of ~300 miRNAs have now been identified from nematodes, flies, plants, fish, mice and humans [11–27]. A miRNA registry with a searchable database has been created [25]. Each miRNA is derived from the stem of a hairpin-like ~75nt precursor (pre-miRNA), which is usually present as a single copy in the genome. Some plant miRNAs are derived from longer hairpin precursors [26,27]. The precursors of some miRNAs are present in two or more copies in the same or in different chromosomes. A few miRNAs can be derived from two or more slightly different precursors. The miRNA genes are typically found in intergenic areas but can also be found in sense or antisense orientation within introns of known genes. Some miRNAs are clustered in the genome and are most likely to be transcribed together as long polycistronic RNAs. The total number of miRNAs in each organism is unknown but is estimated to represent ~1% of all genes [19–21]. This figure is based on extensive cloning efforts and on predictions made by genome-wide searches with computational algorithms designed to detect hairpin-forming, putative pre-miRNAs. Almost all human miRNAs are (~100%) conserved in the mouse [17] and one-third of *C. elegans* miRNAs have vertebrate homologs [20–22]. Some miRNAs are expressed in specific tissues and the expression of some miRNAs is temporally regulated. Quantitative analysis of miRNA levels in *C. elegans* and HeLa cells shows that, overall, miRNAs are fairly abundant, ranging from 1000 to 60 000 molecules per cell [20].

miRNA biogenesis

Pre-miRNA transcripts are thought to be processed from longer transcripts, termed pri-miRNAs. This finding is based on several observations. The most direct evidence is the demonstration that the precursor of miR-30 is part of a ~600nt transcript and that the precursors of the miR-23/27/24–2 and miR-17/18/19a/20/19b-1 are polycistronically transcribed as longer RNA transcripts [28]. Some miRNAs are found within longer expressed sequence tag sequences (ESTs), some of which are polyadenylated. The RNA polymerase(s) or regulatory elements that control miRNA transcription are unknown. However, the genes of at least some miRNAs might be transcribed by RNA polymerase II. An interesting example is miR-155; this is derived from a ~1.4 kb noncoding RNA termed BIC [16], which is spliced and polyadenylated, indicating that it is an RNA polymerase II transcript. Pri-miRNAs are processed in the nucleus, and mature pre-miRNAs are exported to the cytoplasm for further processing [28]. The nuclease(s) that processes pri-miRNAs and the factors that export them to the cytoplasm are unknown. Exportin-t, a tRNA transport factor, has been proposed as a likely pre-miRNA export factor [29]. Another candidate transport factor might be exportin-5, which has recently been shown to export VA1 RNA, a ~160nt adenovirus RNA that is produced in the nucleus and accumulates in the cytoplasm of infected cells [30]. Exportin-5 binds to a hairpin-like RNA motif in VA1 RNA that has a striking resemblance to pre-miRNAs. Such motifs are found in many viral and

cellular RNAs, raising the possibility that pre-miRNAs might be among the exportin-5 cargo [30].

In the cytoplasm of animal cells, DICER, an RNA III-like enzyme that processes dsRNA into siRNAs, also excises mature miRNAs from pre-miRNAs [28,31,32]. In RNAi, Dicer cleavage of long, linear dsRNAs generates 21 to 23nt siRNA duplexes with 5' phosphates and 2nt 3'-terminal overhangs [9,33]. RDE-4, a dsRNA binding protein, is also required for dsRNA processing during RNAi in *C. elegans* but does not participate in pre-miRNA processing [34]. Typically, both strands of the siRNA duplex are stable [10] and incorporate in RNAI-INDUCED SILENCING COMPLEXES (RISC) [9]. In *C. elegans*, there is evidence for preferential accumulation of siRNAs antisense to the target RNA, implying that the target RNA might stabilize siRNAs [35]. Only one strand of the two strands of a pre-miRNA accumulates (typically) in cells as mature miRNA [11–14,31]. How this is accomplished is one of the key questions in the field. It is possible that the pre-miRNA contains all of the information necessary to direct both proper processing by Dicer and incorporation of the miRNA into MICRO-RIBONUCLEOPROTEINS (miRNPs). Dicer might cleave both strands of the precursor but only one (the miRNA) accumulates, while the other strand is degraded. Alternatively, Dicer might liberate the miRNA by cutting only one strand of the pre-miRNA. The former possibility is favored by the observation that ~22nt RNAs derived from the strands opposite of the true miRNAs (termed miRNAs*) [12], are occasionally cloned from preparations of small RNAs from total RNA [20]. If Dicer cleaves both strands of the pre-miRNA but only one accumulates as miRNA, how is it selected? One possibility is that the RNA that incorporates in the downstream effector complex (miRNP) accumulates as miRNA while the RNA from the opposite strand is degraded. Another possibility is that both strands incorporate in miRNPs but that interactions of the miRNA with its RNA target(s) is required for miRNA accumulation; in this case, the RNA target determines which is the authentic miRNA. This latter possibility, however, is less likely because *in vitro* processing of pre-*let-7* in *Drosophila* embryo extracts produces *let-7* only [31]. It was also reported recently that recombinant human Dicer protein prepared from baculovirus-transduced insect cells is able to cleave synthetic pre-*let-7* *in vitro* [36]. In that study, purified human Dicer was incubated with *in vitro* transcribed pre-*let-7* RNA, and the products of the reaction were analyzed by northern blotting with probes recognizing *let-7* and its complement, corresponding to the opposite strand of the pre-*let-7* stem. Only a single ~22nt RNA was detected corresponding to *let-7*, implying that Dicer recognizes the authentic miRNA and cleaves the precursor only in one strand [36]. These results, however, should be interpreted with caution because recombinant Dicer was not sufficiently pure, representing only 25% of the total protein [36]. It is thus possible that other factors required for proper pre-miRNA processing, and normally present in insect cells which contain the miRNA/RNAi machinery, copurified with recombinant human Dicer. A physical association between Dicer and ARGONAUTE proteins has been demonstrated [37] and it is thought that dsRNA or pre-miRNA

processing by Dicer is coupled with RISC and miRNP formation, respectively. Overall, these findings indicate that target RNA might not be required for miRNA selection but do not exclude a role for target RNA in stabilization of miRNAs. How miRNAs are selected and stabilized remains unknown. Interestingly, in plants, miRNAs are processed in the nucleus [38].

miRNA function and intersection with RNA-mediated gene silencing pathways

Following Dicer processing, mature human miRNAs associate with a ~15S miRNP complex, and in particular are bound to eIF2C2, a member of the Argonaute family of proteins ([14] and, A. Sharma *et al.*, unpublished). The Argonaute proteins are ~100 kDa proteins that contain two conserved domains of unknown biochemical function, known as the PAZ and PIWI domains (reviewed in [39]). Argonaute proteins are evolutionary conserved, and multiple paralogs exist in many organisms, with the exception of the fission yeast *Schizosaccharomyces pombe*, which has a single member. Humans have at least seven Argonaute proteins, whereas *C. elegans* has 24 [39]. Genetic screens in various organisms have implicated Argonaute proteins in RNAi, in the maturation and function of tRNAs and in development (reviewed in [39]). Biochemical studies using *Drosophila* and human cells have shown that Argonaute proteins are also core components of RISC, which contain the endonuclease that cleaves the target RNA in RNAi. In particular, *Drosophila* RISCs contain AGO2 [37], whereas human RISCs contain the two Argonaute proteins eIF2C1 and eIF2C2 [40]. Interestingly, siRNAs are present as single-stranded forms in RISCs [40] and this is similar to the presence of single-stranded miRNAs in miRNPs [14]. We would like to propose the general terms 'MIRGONAUTE' and 'SIRGONAUTE' to designate a ribonucleoprotein composed of an Argonaute protein complexed with mature miRNAs and siRNAs, respectively. Because Dicer processing of dsRNA produces siRNA duplexes, an (as yet unknown) helicase might unwind the two strands before, or concurrent to, their incorporation in RISCs. Many putative helicases have been implicated in RNAi and, although their functions remain unknown, it is reasonable to expect that they mediate RNA unwinding or RNP restructuring of si/miRNA-containing complexes (reviewed in [41]).

It was originally thought that siRNAs and miRNAs reside in different complexes that have distinct and nonoverlapping physiological roles: siRNAs are derived from dsRNA that are in turn derived from RNAs perceived by cells as harmful (such as viruses, transposons and overexpressed transgenes). siRNAs base-pair with these RNAs and cleave them. By contrast, miRNAs are derived from endogenous genes and regulate the expression of endogenous mRNAs not by cleaving them but by repressing their translation. One of the biggest recent breakthroughs is the demonstration that, in reality, siRNAs and miRNAs are functionally equivalent and that a miRNA can act as a siRNA. Immunopurified human miRNPs containing *let-7*, an endogenous miRNA, act as enzymes and cleave *in vitro* a target RNA that contains a fully complementary site to *let-7* [42]. Furthermore, many plant

miRNAs have extensive complementarity to their target mRNAs and act as siRNAs by cleaving their mRNA targets [43–45]. These findings indicate that a miRNA-directed endonuclease, similar to the siRNA-directed RISC endonuclease, is present in miRNPs and that the degree of complementarity between the miRNA and its target determines its function: if the complementarity is partial, the translation of the target mRNA is repressed, whereas if it is extensive, the target mRNA is cleaved and the miRNA acts as a siRNA. The converse is also true; a siRNA might repress the translation of its target mRNA and act as a miRNA if the complementarity with its target is made partial [46].

si/miRNA-directed target RNA destruction

The RISC/miRNP endonuclease cleaves a single phosphodiester bond on the target RNA, located across from the middle of the guide siRNA [47]. These properties make RISC/miRNP endonuclease a unique and extremely versatile enzyme, and the identification and characterization of its catalytic subunit(s) represents a most pressing issue. It is reasonable to propose that some si/miRgonaute ribonucleoproteins might actually be the elusive RISC/miRNP endonuclease. The only common protein between miRNPs and human RISCs is the Argonaute protein eIF2C2 [14,40], and *Drosophila* RISCs contain the Argonaute protein AGO2 [37]. Moreover, the molecular mass of purified RISCs is ~90–160 kDa and is composed of single-stranded siRNAs and the Argonaute proteins eIF2C2 and eIF2C1 (each with a molecular mass of ~100 kDa) [40]. These findings suggest that some si/miRgonautes might in fact represent the RISC/miRNP endonucleases. Obviously, further work is required to evaluate this hypothesis, including experiments to address the mi/siRNA binding properties of Argonaute proteins and the stoichiometry of mi/siRNAs in miRNPs/RISCs. Ultimate proof requires the reconstitution of RISC activity by recombinant proteins and synthetic siRNAs.

Although the central role of RISC/miRNP endonuclease as the executor of target RNA destruction in *Drosophila* and human RNAi is firmly established, a different mechanism might be operating to destroy targeted RNAs during RNAi in other organisms such as *C. elegans*. This mechanism, termed transitive RNAi, postulates that siRNAs generated after Dicer processing of dsRNA bind to the homologous target RNA and act not as guides to direct cleavage of the target RNA but rather as primers that direct a RNA-dependent RNA polymerase (RdRP) to generate more dsRNA, which is processed again by Dicer [48]. In this model, Dicer is both the initiator and executor molecule of RNAi. Interestingly, there are no known homologs of RdRP in the genome of humans or flies and no RdRP activity has been detected in biochemical experiments designed to detect such activity in the context of RNAi, suggesting that RdRP does not have a role in human or fly RNAi [40,49]. However, the central role of RdRP in RNA silencing pathways has been repeatedly demonstrated in other organisms, including worms, plants and fungi [5]. A central role for RdRP has been postulated for the generation of dsRNA from 'aberrant' single-stranded RNAs (ssRNAs; i.e. RNAs generated by transgenes etc.).

How cells sense these 'aberrant' ssRNAs is unknown and many putative mechanisms can be envisioned. One possibility is that they are sensed by exclusion: if these ssRNAs are not channeled in any of the pathways that process physiological RNAs, they become substrates for RdRP and are converted to dsRNAs, which are then processed by Dicer to generate siRNAs. RdRP might not require siRNA primers in the initial conversion of ssRNAs to dsRNAs (or even at later stages [50]) but might instead rely on other properties of 'aberrant' ssRNAs, such as their abundance. Intriguingly, in plants, two classes of siRNAs have been described [51]. The first class has a size of ~22nt (same size with mature plant miRNAs) and mediates target RNA degradation. The siRNAs of the second class are slightly longer, averaging at ~25nt, and are dispensable for target RNA degradation but required for homologous DNA silencing; siRNAs derived from retrotransposons belong to the long siRNA class [51]. Presumably, different Dicer proteins (*Arabidopsis* contains four Dicer paralogs) generate the two classes of siRNAs [50,51]. This is in contrast to the situation in humans, where there is a single Dicer gene.

siRNA-directed DNA silencing

RNA-mediated gene silencing might also operate at the DNA level, and among the most remarkable recent developments was the demonstration that the same factors that participate in target RNA destruction also silence chromatin. In fission yeast, silencing of centromeric heterochromatic repeats is accomplished by methylation of the ninth lysine of histone H3 (H3K9) and subsequent binding of the Swi6 (heterochromatin protein 1), resulting in transcriptional repression [52]. This silencing requires Dicer as well as the single fission yeast Argonaute protein ago1 and RdRP [52]. siRNAs corresponding to heterochromatic repeats (but not miRNAs) have been cloned from fission yeast, implying that chromatin silencing rather than post-transcriptional control of mRNAs is the main, if not the sole, purpose of RNA-mediated silencing in the fission yeast [53]. In *Arabidopsis*, the Argonaute protein AGO4, along with long siRNAs, are required for retro-element silencing through DNA and histone (H3K9) methylation [54]. In *Tetrahymena*, the Argonaute protein TIWI is required for DNA elimination, which, in turn, requires histone methylation and ~28nt RNAs [55]. Collectively, these studies assign a central role for siRNAs and associated factors in DNA silencing in certain organisms. It will be of great interest to investigate whether RNAi-mediated chromatin silencing operates in mammals.

miRNA-directed translational repression

Animal miRNAs are thought to bind with partial antisense complementarity sequences in the 3'-UTR of their target mRNAs. What is the fate of the mRNA if the complementarity between a miRNA and its mRNA target is partial? By contrast to RNAi, the stability of the target mRNA is not affected but rather its expression is blocked after translation initiation. This novel regulatory mechanism was first demonstrated for the *C. elegans lin-14* mRNA, which is regulated by the *lin-4* miRNA [56]. The stability

or polyadenylation status of *lin-14* is not affected, nor is its translation blocked at the level of initiation, which constitutes the most common mechanism of translational control. *lin-14* and *lin-4* are found to be associated with polyribosomes, suggesting that expression of LIN-14 protein is controlled at a level after translation initiation [56]. *lin-28* mRNA is also regulated by *lin-4* miRNA at a level after translation initiation, indicating that this type of regulation is common among miRNAs [57]. The mechanism of miRNA-mediated translational repression is unknown, but it might involve inhibition of translational elongation or termination, or destruction of the nascent polypeptide as it emerges from the ribosome. A simple translational elongation arrest is rather unlikely because the polyribosome profiles of unrepressed or miRNA-repressed *lin-14* and *lin-28* mRNAs are identical [56,57]. It is also not known how miRNAs repress the translation of their target mRNAs. miRNAs might transiently associate and modify their mRNA targets, or, alternatively, miRNAs might bind stably to their targets and either bring with them, or recruit once bound, factors that will mediate translational repression. We favor a stoichiometric binding of miRNAs in the form of miRNPs onto repressed mRNA targets ([56] and P. Nelson and Z. Mourelatos, unpublished). Although there are no examples of miRNAs that control processes other than mRNA stability and translation, it is possible that miRNAs have other regulatory roles that remain to be described.

miRNA targets

In plants, miRNAs are extensively base-paired with their target mRNAs. This finding has facilitated computational prediction of plant miRNA targets. Sixty-one targets have been identified, including targets for 18 out of the 19 known plant miRNAs (reviewed in [58]). Many of these targets code for transcription factors that control development [43,44]. Interestingly, miR-162 targets the *dcl1* mRNA (which codes for a plant Dicer protein) [45] and miR-168 might target the plant *AGO1* mRNA (which codes for a plant Argonaute protein) [43], suggesting that miRNAs might provide negative feedback regulation to genes that control their processing and function.

In contrast to plants, and despite a large number of known animal miRNAs, very few of their mRNA targets have been discovered. The two *C. elegans* stRNAs, *lin-4* and *let-7*, control the expression of genes whose protein products are required for proper developmental timing. The *C. elegans lin-4* miRNA regulates the expression of *lin-14*, *lin-28* and possibly *hbl-1* (the *C. elegans* ortholog of the *Drosophila hunchback* gene) mRNAs, whereas *let-7* controls *lin-41* and possibly *hbl-1* mRNAs [1,6,59–61]. The *Drosophila bantam* miRNA controls development and cell growth by repressing the expression of the pro-apoptotic gene *hid* [62]. Similarly, the *Drosophila* miR-14 acts as an apoptosis inhibitor, probably by repressing the expression of pro-apoptotic genes [63]. Two mouse miRNAs (miR-127 and miR-136) show perfect antisense complementarity with the coding region of a retrotransposon-like gene (*Rtl1*), suggesting that some animal miRNAs might cleave their targets [64]. Interestingly, these miRNAs and their putative mRNA target are imprinted. *Hes1*, a

transcription factor, has recently been described as a target for miR-23 in a human neuronal cell line [65]. Extrapolating from the few known examples of miRNA targets, it seems likely that each miRNA controls a few mRNAs whose protein products function in the same pathway. Identification of mRNA targets for animal miRNAs represents one of the most pressing issues in the field because knowledge of miRNA targets will uncover the genes regulated by miRNAs and will provide clues about the function of each miRNA. Bioinformatic approaches to identify mRNA targets for animal miRNAs are complicated by the partial complementarity between a miRNA and its target. Characterization of the rule(s) that control miRNA–target mRNA binding might allow the design of computational algorithms that will predict with confidence animal miRNA targets.

miRNAs and human diseases

Several human diseases have already surfaced in which miRNAs or their machinery might be implicated. One of them is spinal muscular atrophy (SMA), a pediatric neurodegenerative disease caused by reduced protein levels or loss-of-function mutations of the survival of motor neurons (SMN) gene [66]. Two proteins (Gemin3 and Gemin4) that are part of the SMN complex are also components of miRNPs [14]. It remains to be seen whether miRNA biogenesis or function is dysregulated in SMA and what effect this has on pathogenesis.

Another neurological disease linked to mi/siRNAs is fragile X mental retardation (FXMR) caused by absence or mutations of the fragile X mental retardation protein (FMRP). dFMR1, the *Drosophila* homolog of FMRP, is a component of *Drosophila* RISCs/miRNPs [67,68]. This finding is particularly intriguing because FMRP is known to repress the translation of specific mRNAs, and the identification of dFMR1 in RISCs might indicate that miRNAs direct dFMR1 to mRNAs whose translation must be controlled. Whether human FMRP is present in miRNPs, and the possible role of miRNAs in FXMR, remain to be determined. There are additional clues that miRNAs might play a role in other neurological diseases. As yet, these data only indicate the importance of future studies. One intriguing finding is that the miR-224 gene locus lies within the minimal candidate region of two different neurological diseases: early-onset parkinsonism and X-linked mental retardation [18].

Links between cancer and miRNAs have also been described. The most frequent single genetic abnormality in chronic lymphocytic leukemia (CLL) is a deletion localized to chromosome 13q14 (~50% of cases). A recent study determined that two different miRNA (miR-15 and miR-16) genes are present within the minimal 30 kb region of loss in CLL [69]. miR-15 and miR-16 genes are both found in the intron of an inactive gene (LEU2), and are highly expressed in B lymphocytes. Most (~two-thirds) of CLL tumor cells (as well as prostate carcinoma) showed significant reductions in the levels of these miRNAs [69]. More circumstantial evidence links members of the Argonaute family of proteins with other cancers. The region of chromosome 1p on which three members of the Argonaute gene family (EIF2C1, hAgo3

and hAgo4) are physically located (1p34–35) is altered in Wilms' tumors, primitive neuroectodermal tumors and many other types of cancer. Another member of the Argonaute family, Hiwi, is localized to 12q24. Its specific locus displays linkage to the development of testicular germ cell tumors (reviewed in [39]). We anticipate that connections between miRNAs and human diseases will only strengthen in parallel with our knowledge of miRNAs and the gene networks that they control. Moreover, our understanding of the regulation of RNA-mediated gene expression is leading to the development of novel therapeutic approaches that will be likely to revolutionize the practice of medicine [70].

Concluding remarks

The essence, in our view, of gene expression regulation mediated by mi/siRNAs appears surprisingly simple. dsRNA that is recognizable by Dicer (i.e. linear dsRNA longer than ~30 base pairs or the dsRNA stem of pre-miRNAs) is processed to ~22nt mi/siRNAs. Mature mi/siRNAs incorporate with Argonaute proteins to form mi/siRgonaute ribonucleoproteins. In turn, mi/siRNAs base-pair with their RNA targets and bring with them, or recruit once bound to their targets, the factors that will dictate the fate of the targeted RNA: destruction or translational repression. We hypothesize that it is in the form of mi/siRgonautes that mi/siRNAs recognize their RNA targets and that Argonaute proteins are among the key molecules that mediate mi/siRNA function. siRgonautes might also recognize chromatin and guide its silencing by recruiting DNA methylases and chromatin remodeling complexes. This view, of course, represents a working model that will undoubtedly be modified and augmented as the details of the inner workings of the miRNA world unfold. Despite remarkable progress, the mechanisms, targets and extent of miRNA-mediated gene expression regulation remain unknown and their elucidation represents a major challenge for the future.

Acknowledgements

We apologize to those whose work could not be cited because of space limitations. Supported by NIH grants to P.N. (5T32-AG00255), M.K. (5T32-AR07442) and Z.M. (NS02199) and by a University of Pennsylvania Genomics Institute award to Z.M.

References

- 1 Lee, R.C. *et al.* (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854
- 2 Romano, N. and Machino, G. (1992) Quelling: transient inactivation of gene expression in *Neurospora crassa* by transformation with homologous sequences. *Mol. Microbiol.* 6, 3343–3353
- 3 Napoli, C. *et al.* (1990) Introduction of a chimeric chalcone synthase gene into petunia results in reversible co-suppression of homologous genes *in trans*. *Plant Cell* 2, 279–289
- 4 van der Krol, A.R. *et al.* (1990) Flavonoid genes in petunia: addition of a limited number of gene copies may lead to a suppression of gene expression. *Plant Cell* 2, 291–299
- 5 Tijsterman, M. *et al.* (2002) The genetics of RNA silencing. *Annu. Rev. Genet.* 36, 489–519
- 6 Reinhart, B.J. *et al.* (2000) The 21-nucleotide *let-7* RNA regulates developmental timing in *Caenorhabditis elegans*. *Nature* 403, 901–906
- 7 Pasquinelli, A.E. *et al.* (2000) Conservation of the sequence and

- temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408, 86–89
- 8 Fire, A. *et al.* (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811
 - 9 Elbashir, S.M. *et al.* (2001) RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.* 15, 188–200
 - 10 Hamilton, A.J. and Baulcombe, D.C. (1999) A species of small antisense RNA in posttranscriptional gene silencing in plants. *Science* 286, 950–952
 - 11 Lagos-Quintana, M. *et al.* (2001) Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–858
 - 12 Lau, N.C. *et al.* (2001) An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862
 - 13 Lee, R.C. and Ambros, V. (2001) An extensive class of small RNAs in *Caenorhabditis elegans*. *Science* 294, 862–864
 - 14 Mourelatos, Z. *et al.* (2002) miRNPs: a novel class of ribonucleoproteins containing numerous microRNAs. *Genes Dev.* 16, 720–728
 - 15 Park, W. *et al.* (2002) CARPEL FACTORY, a Dicer homolog, and HEN1, a novel protein, act in microRNA metabolism in *Arabidopsis thaliana*. *Curr. Biol.* 12, 1484–1495
 - 16 Lagos-Quintana, M. *et al.* (2002) Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* 12, 735–739
 - 17 Lagos-Quintana, M. *et al.* (2003) New microRNAs from mouse and human. *RNA* 9, 175–179
 - 18 Dostie, J. *et al.* (2003) Numerous microRNPs in neuronal cells containing novel microRNAs. *RNA* 9, 180–186
 - 19 Lim, L.P. *et al.* (2003) Vertebrate microRNA genes. *Science* 299, 1540
 - 20 Lim, L.P. *et al.* (2003) The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* 17, 991–1008
 - 21 Grad, Y.A.J. *et al.* (2003) Computational and experimental identification of *C. elegans* microRNAs. *Mol. Cell* 11, 1253–1263
 - 22 Ambros, V. *et al.* (2003) MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Curr. Biol.* 13, 807–818
 - 23 Aravin, A.A. *et al.* (2003) The small RNA profile during *Drosophila melanogaster* development. *Dev. Cell* 5, 337–350
 - 24 Houbaviy, H.B. *et al.* (2003) Embryonic stem cell-specific microRNAs. *Dev. Cell* 5, 351–358
 - 25 Ambros, V. *et al.* (2003) A uniform system for microRNA annotation. *RNA* 9, 277–279
 - 26 Reinhart, B.J. *et al.* (2002) MicroRNAs in plants. *Genes Dev.* 16, 1616–1626
 - 27 Llave, C. *et al.* (2002) Endogenous and silencing-associated small RNAs in plants. *Plant Cell* 14, 1605–1619
 - 28 Lee, Y. *et al.* (2002) MicroRNA maturation: stepwise processing and subcellular localization. *EMBO J.* 21, 4663–4670
 - 29 Grosshans, H. and Slack, F.J. (2002) Micro-RNAs: small is plentiful. *J. Cell Biol.* 156, 17–21
 - 30 Gwizdek, C. *et al.* (2003) Exportin-5 mediates nuclear export of minihelix-containing RNAs. *J. Biol. Chem.* 278, 5505–5508
 - 31 Hutvagner, G. *et al.* (2001) A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838
 - 32 Ketting, R.F. *et al.* (2001) Dicer functions in RNA interference and in synthesis of small RNA involved in developmental timing in *C. elegans*. *Genes Dev.* 15, 2654–2659
 - 33 Bernstein, E. *et al.* (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* 409, 363–366
 - 34 Tabara, H. *et al.* (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
 - 35 Tijsterman, M. *et al.* (2002) RNA helicase MUT-14-dependent gene silencing triggered in *C. elegans* by short antisense RNAs. *Science* 295, 694–697
 - 36 Provost, P. *et al.* (2002) Ribonuclease activity and RNA binding of recombinant human Dicer. *EMBO J.* 21, 5864–5874
 - 37 Hammond, S.M. *et al.* (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* 293, 1146–1150
 - 38 Papp, I. *et al.* (2003) Evidence for nuclear processing of plant micro RNA and short interfering RNA precursors. *Plant Physiol.* 132, 1382–1390
 - 39 Carmell, M.A. *et al.* (2002) The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* 16, 2733–2742
 - 40 Martinez, J. *et al.* (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* 110, 563–574
 - 41 Hannon, G.J. (2002) RNA interference. *Nature* 418, 244–251
 - 42 Hutvagner, G. and Zamore, P.D. (2002) A microRNA in a multiple-turnover RNAi enzyme complex. *Science* 297, 2056–2060
 - 43 Rhoades, M.W. *et al.* (2002) Prediction of plant microRNA targets. *Cell* 110, 513–520
 - 44 Llave, C. *et al.* (2002) Cleavage of Scarecrow-like mRNA targets directed by a class of *Arabidopsis* miRNA. *Science* 297, 2053–2056
 - 45 Xie, Z. *et al.* (2003) Negative feedback regulation of Dicer-Like1 in *Arabidopsis* by microRNA-guided mRNA degradation. *Curr. Biol.* 13, 784–789
 - 46 Doench, J.G. *et al.* (2003) siRNAs can function as miRNAs. *Genes Dev.* 17, 438–442
 - 47 Elbashir, S.M. *et al.* (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498
 - 48 Sijen, T. *et al.* (2001) On the role of RNA amplification in dsRNA-triggered gene silencing. *Cell* 107, 465–476
 - 49 Schwarz, D.S. *et al.* (2002) Evidence that siRNAs function as guides, not primers, in the Drosophila and human RNAi pathways. *Mol. Cell* 10, 537–548
 - 50 Tang, G. *et al.* (2003) A biochemical framework for RNA silencing in plants. *Genes Dev.* 17, 49–63
 - 51 Hamilton, A. *et al.* (2002) Two classes of short interfering RNA in RNA silencing. *EMBO J.* 21, 4671–4679
 - 52 Volpe, T.A. *et al.* (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297, 1833–1837
 - 53 Reinhart, B.J. and Bartel, D.P. (2002) Small RNAs correspond to centromere heterochromatic repeats. *Science* 297, 1831
 - 54 Zilberman, D. *et al.* (2003) ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* 299, 716–719
 - 55 Mochizuki, K. *et al.* (2002) Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in *tetrahymena*. *Cell* 110, 689–699
 - 56 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. *Dev. Biol.* 216, 671–680
 - 57 Segerson, K. *et al.* (2002) Two genetic circuits repress the *Caenorhabditis elegans* heterochronic gene lin-28 after translation initiation. *Dev. Biol.* 243, 215–225
 - 58 Bartel, B. and Bartel, D.P. (2003) MicroRNAs: at the root of plant development? *Plant Physiol.* 132, 709–717
 - 59 Moss, E.G. *et al.* (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
 - 60 Abrahante, J.E. *et al.* (2003) The *Caenorhabditis elegans* hunchback-like gene lin-57/hbl-1 controls developmental time and is regulated by microRNAs. *Dev. Cell* 4, 625–637
 - 61 Lin, S.Y. *et al.* (2003) The *C. elegans* hunchback Homolog, hbl-1, controls temporal patterning and is a probable microRNA target. *Dev. Cell* 4, 639–650
 - 62 Brennecke, J. *et al.* (2003) Bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* 113, 25–36
 - 63 Xu, P. *et al.* (2003) The Drosophila MicroRNA Mir-14 suppresses cell death and is required for normal fat metabolism. *Curr. Biol.* 13, 790–795
 - 64 Seitz, H. *et al.* (2003) Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat. Genet.* 34, 261–262
 - 65 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
 - 66 Paushkin, S. *et al.* (2002) The SMN complex, an assemblysome of ribonucleoproteins. *Curr. Opin. Cell Biol.* 14, 305–312
 - 67 Caudy, A.A. *et al.* (2002) Fragile X-related protein and VIG associate with the RNA interference machinery. *Genes Dev.* 16, 2491–2496
 - 68 Ishizuka, A. *et al.* (2002) A Drosophila fragile X protein interacts with components of RNAi and ribosomal proteins. *Genes Dev.* 16, 2497–2508
 - 69 Calin, G.A. *et al.* (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
 - 70 McManus, M.T. and Sharp, P.A. (2002) Gene silencing in mammals by small interfering RNAs. *Nat. Rev. Genet.* 3, 737–747