Sorting of Drosophila small silencing RNAs partitions microRNA* strands into the RNA interference pathway

MEGHA GHILDYAL,1,4 JIA XU,2,4 HERVÉ SEITZ,1,5,6 ZHIPING WENG,3 and PHILLIP D. ZAMORE1

1Howard Hughes Medical Institute, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA
2Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215, USA
3Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA

ABSTRACT

In flies, small silencing RNAs are sorted between Argonaute1 (Ago1), the central protein component of the microRNA (miRNA) pathway, and Argonaute2 (Ago2), which mediates RNA interference. Extensive double-stranded character—as is found in small interfering RNAs (siRNAs)—directs duplexes into Ago2, whereas central mismatches, like those found in miRNA/miRNA* duplexes, direct duplexes into Ago1. Central to this sorting decision is the affinity of the small RNA duplex for the Dcr-2/R2D2 heterodimer, which loads small RNAs into Ago2. Here, we show that while most Drosophila miRNAs are bound to Ago1, miRNA* strands accumulate bound to Ago2. Like siRNA loading, efficient loading of miRNA* strands in Ago2 favors duplexes with a paired central region and requires both Dcr-2 and R2D2. Those miRNA and miRNA* sequences bound to Ago2, like siRNAs diced in vivo from long double-stranded RNA, typically begin with cytidine, whereas Ago1-bound miRNA and miRNA* disproportionately begin with uridine. Consequently, some pre-miRNA generate two or more isoforms from the same side of the stem that differentially partition between Ago1 and Ago2. Our findings provide the first genome-wide test for the idea that Drosophila small RNAs are sorted between Ago1 and Ago2 according to their duplex structure and the identity of their first nucleotide.

Keywords: microRNA; miRNA; siRNA; Argonaute; RNA interference; RNAi; small RNA sorting

INTRODUCTION

In animals, microRNAs (miRNAs) regulate the stability and rate of translation of mRNAs, whereas small interfering RNAs (siRNAs) silence transposons, defend against viral pathogens, and regulate mRNA expression (Ghildiyal and Zamore 2009). Both miRNAs and siRNAs derive from longer double-stranded RNA (dsRNA) precursors, which are cleaved by RNase III dsRNA-specific endonucleases. miRNA production begins in the nucleus, where long primary miRNAs, transcribed by RNA polymerase II, are converted into ~65 nucleotides (nt) pre-miRNA hairpins by the RNase III ribonuclease, Drosha, aided by a double-stranded RNA-binding domain (dsRBD) partner protein (Lee et al. 2003; Denli et al. 2004; Gregory et al. 2004; Han et al. 2004; Zeng et al. 2005; Han et al. 2006). A minority of pre-miRNAs—mirtrons—correspond to entire introns and are excised from their primary transcripts by the pre-mRNA splicing pathway (Okamura et al. 2007; Ruby et al. 2007a). Pre-miRNAs are then exported to the cytoplasm (Yi et al. 2003; Bohnsack et al. 2004; Lund et al. 2004), where they are processed by a second RNase III enzyme, Dicer, together with its dsRBD partner protein, into ~22-nt-long miRNA/miRNA* duplexes (Grishok et al. 2001; Hutvágner et al. 2001; Ketting et al. 2001; Chendrimada et al. 2005; Fo¨rstemann et al. 2005; Jiang et al. 2005; Maniataki and Mourelatos 2005; Saito et al. 2005; Lee et al. 2006; Melo et al. 2009).

siRNA production also requires Dicer, which excises 21-nt siRNA duplexes, comprising a guide and passenger strand, from long dsRNA formed by the base pairing of complementary sense and antisense transcripts, convergently transcribed miRNAs, or by the intramolecular base pairing of long, self-complementary RNAs. Such endogenous dsRNAs...
yield endo-siRNAs. Similarly, exogenous dsRNA, introduced experimentally or by viral infection, are converted by Dicer to exo-siRNAs.

In *Drosophila melanogaster*, Dicer-1 (Dcr-1) converts pre-miRNAs into miRNA/miRNA* duplexes; Dicer-2 (Dcr-2) converts long dsRNA into 21-nt siRNA duplexes (Zamore et al. 2000; Bernstein et al. 2001; Lee et al. 2004). The use of different Dicer proteins to generate miRNAs and siRNAs may minimize competition between the two pathways, so that an RNAi defense to viral infection does not perturb miRNA production.

All small silencing RNAs function bound to Argonaute proteins. Argonaute proteins display nucleotides 2–8 of the small RNA guide in a prehelical geometry that confers on this region special importance in target recognition: the majority of the binding energy for target binding is contributed by this “seed” sequence (Jackson et al. 2003; Lewis et al. 2003; Haley and Zamore 2004; Rajewsky and Socci 2004; Grun et al. 2005; Krek et al. 2005; Lewis et al. 2005; Lim et al. 2005; Jackson et al. 2006; Ameres et al. 2007; Grimson et al. 2007; Wang et al. 2008; Parker et al. 2009). In flies, miRNAs are loaded from miRNA/miRNA* duplexes into Argonaute1 (Ago1), whereas siRNAs are loaded from guide/passenger duplexes into Argonaute2 (Ago2) (Hammond et al. 2000; Okamura et al. 2004; Rand et al. 2004; Matranga et al. 2003; Miyoshi et al. 2005; Kim et al. 2005; Leuschner et al. 2006). Two binary choices accompany loading of small RNAs into Argonautes proteins in *Drosophila*: the choice of Ago1 versus Ago2 and the selection of one of the two strands of the duplex as a miRNA or guide strand (Khvorova et al. 2003; Schwarz et al. 2003).

Although fly miRNAs are overwhelmingly associated with Ago1 and siRNAs with Ago2, small RNA production and Argonaute loading are uncoupled (Fürstemann et al. 2007; Tomari et al. 2007). Instead, miRNA and siRNA duplexes are actively partitioned between Ago1 and Ago2 according to their structure. Extensive double-stranded character directs duplexes, such as siRNAs, into Ago2, which mediates RNAi, whereas bulges and mismatches, like those found in miRNA/miRNA* duplexes, are sorted into Ago1 (Kawamata et al. 2009). Central to this sorting decision is the affinity of the small RNA duplex for the Dcr-2/R2D2 heterodimer, which loads small RNAs into Ago2 (Liu et al. 2003; Tomari et al. 2004a,b, 2007). Central mismatches reduce binding of small RNA duplexes by the Dcr-2/R2D2 heterodimer, antagonizing Ago2 loading and promoting loading into Ago1 (Fürstemann et al. 2007; Tomari et al. 2007; Kawamata et al. 2009). The function of the Dcr-2/R2D2 heterodimer in Ago2 loading is separate and distinct from its role in dicing siRNAs from long dsRNA: Dcr-2 bearing a glycine to arginine substitution (G31R) in its helicase domain cannot dice, but can still load siRNA into Ago2 (Lee et al. 2004).

Increasingly, this simple picture of small RNA strand choice is at odds with the intracellular abundance, processing accuracy, and evolutionary conservation of miRNA* strands. First, some evolutionarily conserved miRNAs are less abundant than their miRNA* strands, which appear to be evolving regulatory functions (Ruby et al. 2007b). Second, miRNA* 5’ ends are far more precisely defined than their 3’ ends, suggesting selective pressure to generate an accurate seed region—implying that they have regulatory targets (Ruby et al. 2007b; Okamura et al. 2008c; Seitz et al. 2008). Third, there is mounting evidence that some miRNA*s may have regulatory potential (Ro et al. 2007; Okamura et al. 2008c; Lin et al. 2009), and fly miRNA* strands are evolutionarily conserved, albeit not to the same extent as miRNAs (Okamura et al. 2008c). Thus, miRNA* strands may regulate gene expression, rather than serve merely as carriers for loading the miRNA strand. Such a mechanism would make small RNA biogenesis more efficient, with each pre-miRNA producing two different regulatory small RNAs. Nonetheless, miRNAs are typically far more abundant than their miRNA* counterparts, and regulation by low abundance Ago1–small RNA complexes has not been reported in flies.

Here, we show that while most *Drosophila* miRNAs are bound to Ago1 in vivo, most miRNA* strands accumulate bound to Ago2. Partitioning of miRNAs into Ago1 and Ago2 provides a wide-scale in vivo test for the previously proposed principles for small RNA sorting in flies: miRNAs and miRNA* strands are sorted between the two Argonaute proteins according to the structure of their small RNA duplex, a process that requires both Dcr-2 and R2D2. Like the exo-siRNAs that direct RNAi, miRNA* strands bound to Ago2 typically begin with cytidine, whereas Ago1-bound miRNAs usually begin with uridine. Thus, the identity of the first nucleotide of a small RNA plays a role in its sorting in flies, as previously reported for plants. Finally, miRNA*’s bound to Ago2 are more abundant than siRNAs that direct RNAi, suggesting that they function to silence target RNAs.

RESULTS

miRNAs and miRNA*s partition differentially between Ago1 and Ago2

We used high throughput sequencing of 18–29-nt RNA from fly heads to determine the small RNA profile and distribution of small RNAs between Ago1 and Ago2 in this complex somatic structure (Supplemental Table 1). Unlike other fly tissues, heads express little, if any, Piwi-interacting RNA, allowing us to focus on small RNAs bound to Ago1 or Ago2 (Ghildiyal et al. 2008). One of the ~1.6 million genome-matching small RNAs sequenced (excluding annotated noncoding RNAs such as 2S ribosomal RNA), 90.2% were derived from pre-miRNAs (Fig. 1A). In parallel, we used an Ago1 monoclonal antibody (Miyoshi et al. 2005) to immunoprecipitate Ago1-associated small RNAs from fly head extracts. Nearly 97% of the >5.03 million
small RNA reads associated with Ago1 were miRNAs; only 2.2% were miRNA* strands (Fig. 1A).

Ago2-loaded guide strands acquire a 3’ terminal 2’-O-methyl modification after their corresponding passenger strand is discarded (Horwich et al. 2007; Saito et al. 2007). To enrich for Ago2-loaded small RNAs, we oxidized the 18–29 nt RNAs prior to library preparation, a treatment that excludes from the library most Ago1-loaded small RNAs, which bear 2’3’ hydroxyl termini, but allows sequencing of Ago2-loaded small RNAs, because their 2’-O-methyl modification protects them from reaction with NaIO₄ (Ghildiyal et al. 2008; Seitz et al. 2008). In general, the pre-miRNA-derived small RNAs associated with Ago1 correlated well with the total small RNA profile \( (r = 0.91 \text{ for miRNAs}; r = 0.70 \text{ for miRNA* strands}) \), supporting the view that the majority of small RNAs in fly heads accumulate because they are bound to Ago1. However, a global fit of the sum of the miRNA and miRNA* species detected in the Ago1 immunoprecipitation and the miRNA and miRNA* species detected in the library prepared from oxidized RNA more closely recapitulated the total small RNA profile \( (r = 0.91 \text{ for miRNAs}; r = 0.85 \text{ for miRNA* strands}) \), suggesting that Ago2-bound miRNA and/or miRNA* species are a significant component of the total pre-miRNA-derived small RNA population.

siRNAs were previously identified as the major class of Ago2-associated endogenous small RNAs in flies (Chung et al. 2008; Czech et al. 2008; Ghildiyal et al. 2008; Kawamura et al. 2008; Okamura et al. 2008a,b). Yet, the population of Ago2-associated small RNAs contained more miRNA plus miRNA* combined (53.2%) than endo-siRNAs (33.2%) (Fig. 1A). Thus, the identity of the Dicer paralog that generates a small RNA does not determine the Argonaute protein into which it is loaded. Compared to the total small RNA population—where miRNAs represented \(^{-87.5}\%\) of all small RNAs, but miRNA* reads were just 2.6%—miRNAs were underrepresented (39.4%) and miRNA*s were overrepresented among the Ago2-associated small RNA sequences. The abundance of pre-miRNA-derived small RNAs associated with Ago2 calls into question the prevailing view that Ago2 is restricted to the RNAi pathway.

In general, Ago2 was significantly depleted of miRNAs and enriched for miRNA* sequences \( (P \leq 2.2 \times 10^{-16}) \). Conversely, Ago1 was significantly depleted of miRNA* sequences and enriched for miRNAs \( (P \leq 2.2 \times 10^{-16}) \). For some of these—especially miRNAs—more of a particular small RNA was present in Ago1 than in Ago2, but more of that small RNA was associated with Ago2 than would be expected by chance. In all, 26 miRNAs and 49 miRNA* strands were significantly \( (P \leq 0.01) \) enriched in Ago2, whereas 71 miRNAs and 9 miRNA*s were significantly \( (P \leq 0.01) \) enriched in Ago1 (Fig. 1B). Of the 49 miRNA* strands enriched in Ago2, 32 had their corresponding miRNA enriched in Ago1, while 15 had their miRNA enriched in Ago2. Among the examples illustrated in Figure 2, the miRNAs bantam and mir-308 were enriched in Ago1, whereas bantam* and mir-308* were enriched in Ago2. Supplemental Table 2 reports the enrichment or depletion of individual miRNAs and miRNA* species between the two Drosophila Argonaute proteins.

Although generally less abundant than miRNAs bound to Ago1, miRNA* isoforms (i.e., all of the species derived from the same side of the stem of a single pre-miRNA and sharing a common seed) bound to Ago2 were equally or more abundant than other small RNAs that exert their regulatory functions through Ago2, including the well-studied...
exo-siRNAs derived from an inverted repeat transgene that fully silences the white gene via the RNAi pathway (Lee and Carthew 2003). The median abundance for miRNA* isoforms enriched in Ago2 was more than twice that of the median abundance for white exo-siRNAs bound to Ago2, and 18 miRNA* were more abundant than the single most abundant white exo-siRNA detected in the same fly heads. (These 18 miRNA*s are outliers whose abundance was too large to display on the box plot in Fig. 1C.) In fact, the abundance of a single miR-8* isoform alone (2748 parts per million [ppm]), was nearly two-thirds of the aggregate abundance of all antisense white exo-siRNAs (4273 ppm), whose concentration in heads is sufficient to phenocopy a strong loss-of-function white mutation. Summing the isoforms of each miRNA*, 25 miRNA*s were more abundant than all antisense white exo-siRNAs combined.

The siRNA-loading machinery sorts miRNA* strands into Ago2

Apart from its function in producing siRNAs, Dcr-2 acts with its double-stranded RNA-binding domain protein partner, R2D2, to both load small RNA duplexes into Ago2 and determine the identity of guide and passenger strands. Thus, both Dcr-2 and R2D2 are required to load Ago2 with siRNAs derived from exogenous dsRNA (exo-siRNAs), such as those derived from a long inverted repeat transcript designed to silence white mRNA expression (Lee et al. 2004; Förstemann et al. 2005). At least one Drosophila miRNA, miR-277, which associates equally with Ago1 and Ago2 in cultured S2 cells, requires Dcr-2 and R2D2 to load it into Ago2, even though miR-277 requires Dcr-1 to liberate it from pre-miR-277 (Förstemann et al. 2007). Likewise, those miRNA and miRNA* sequences that were enriched in Ago2 required Dcr-2 and R2D2 for their loading (Fig. 3A). The median extent of Ago2 loading of these miRNAs declined 2.7-fold in dcr-2L811fsX and 3.3-fold in r2d2 mutant heads, compared to wild-type; loading of miRNA* into Ago2 declined 2.1-fold in dcr-2L811fsX and 3.1-fold in r2d2L811fsX. In contrast, the overall abundance of the miRNA sequences that were enriched in Ago1 was unaltered in dcr-2 or r2d2 mutant heads. R2D2 is stabilized by its association with Dcr-2 (Liu et al. 2003; Förstemann et al. 2007). Consequently, dcr-2L811fsX flies are also deficient in R2D2. For miRNA and miRNA* that were preferentially loaded into Ago2, the effect of the absence of Dcr-2 and R2D2 on Ago2 loading were well correlated (r = 0.828) (Fig. 3B,C). As expected, the abundance of miRNA and miRNA* that were preferentially loaded into Ago1 were largely unchanged in these two mutants.

The median abundance of Ago2-enriched miRNA* sequences in the total RNA library declined ∼2.1-fold in...
FIGURE 3. Association of miRNA* with Ago2 relies on the Ago2-loading machinery. (A) Efficient loading into Ago2 of miRNA and miRNA* strands—measured by their abundance in an oxidized small RNA library—was diminished in heads from dcr-2L811fsX and r2d21 mutants for miRNA and miRNA* normally enriched in Ago2, but the abundance of Ago1-enriched miRNAs was unaltered, as measured in the total small RNA library. Box plots illustrate the fold-change between mutant and wild-type. (B,C) The requirement for Dcr-2 and R2D2 for Ago2 loading was well correlated for miRNA and miRNA* strands preferentially loaded into Ago2. (D) The overall abundance of Ago2-enriched miRNA and miRNA*—measured in the total small RNA library—decline in ago2 mutant heads. Box plots illustrate the fold-change between mutant and wild-type in total small RNA libraries.
the absence of Ago2 (Fig. 3D). In contrast, the median abundance of miRNA enriched in Ago1 was unaltered in ago2 mutants heads, compared to wild-type (median fold change = 1.0), a significant difference from the Ago2-enriched miRNA* (P ≤ 3.1 × 10^-8). These data suggest that in the ago2 mutant, those miRNA* species that normally are loaded into Ago2 become less stable when that Argonaute protein is not available. We envision that these miRNA*/miRNA duplexes, while good substrates for the Ago2-loading machinery, are poor loading substrates for the Ago1-loading machinery. In the absence of Ago2, miRNA*/miRNA duplexes from which the Ago2-enriched miRNA* are normally loaded into Ago2 can no longer be used for this purpose. Instead, they are now used as miRNA/miRNA* duplexes—whose structure typically favors Ago1 loading—to load their miRNA strand into Ago1. The observation that abundance of Ago2-enriched miRNA* sequences declines in ago2 heads supports the earlier proposal that the duplex features that promote Ago2-loading are anti-determinants for Ago1 loading (Tomari et al. 2007; Kawamata et al. 2009).

**miRNA/miRNA* duplex structure determines Argonaute loading**

The Dcr-2/R2D2 heterodimer interprets the structure of a small RNA duplex, sorting centrally paired duplexes into Ago2 and leaving duplexes with an unpaired region centered on guide nucleotide 9 to enter the Ago1 loading pathway (Fürstemann et al. 2007; Tomari et al. 2007). Each small RNA duplex presents two distinct duplexes to the fly sorting machinery. For example, bantam/bantam* displays mismatches at guide positions 9 and 10 when viewed from the 5’ end of the miRNA, but these positions are paired when viewed from the 5’ end of the miRNA* strand (Fig. 2). That is, the bantam/bantam* and bantam*/bantam duplexes are not equivalent.

To evaluate if miRNA/miRNA* duplexes and miRNA/miRNA duplexes generally present distinct structures to the Drosophila Argonaute loading machineries, we calculated the pairing probability for each nucleotide in each miRNA/miRNA* duplex that loads an Ago1- or Ago2-enriched miRNA or miRNA*/miRNA duplex that loads an Ago1- or Ago2-enriched miRNA* (Fig. 4A). Viewed in this way, two significant (p < 0.01) structural differences emerge that distinguish duplexes that load Ago1 from those that load Ago2 (Fig. 4B,C): from the perspective of the loaded strand, Ago1-loading duplexes are more likely to have an unpaired 5’ end and a central unpaired region that spans nucleotide positions 8–11. Conversely, Ago2-loading duplexes more likely have a paired 5’ end and a central region with greater double-stranded character. Ago2-loading duplexes are also more likely to have an unpaired guide 3’ end (Fig. 4D). Remarkably, these differences reflect the “rules” for sorting small RNA duplexes between Ago1 and Ago2 that were inferred previously from biochemical studies (Tomari et al. 2007; Kawamata et al. 2009). Thus, they provide in vivo validation of the hypothesis that Drosophila small RNA duplex structure determines its partitioning between Ago1 and Ago2.

**The 5’ terminal nucleotide of a small RNA reflects its partitioning between Ago1 and Ago2**

*Arabidopsis thaliana* produces ten distinct AGO proteins, and small RNAs are sorted among them according to their first nucleotide. Of the 187 annotated miRNAs in *Arabidopsis*, ~76% begin with uridine, consistent with the idea that a 5’ U steers a small RNA into plant Ago1 (Mi et al. 2008; Montgomery et al. 2008). *Arabidopsis* Ago2 and Ago4 preferentially load small RNAs that begin with an adenosine, whereas Ago5 favors small RNAs that begin with cytidine (Mi et al. 2008; Montgomery et al. 2008). Small RNAs in flies partition between Ago1 and Ago2 according to the structure of the duplex from which they are loaded, yet, as in plants, *Drosophila* miRNAs overwhelmingly begin with U, whereas U is not overrepresented as the first nucleotide of siRNAs (Ghildiyal et al. 2008).

We analyzed the sequence composition of Ago1- and Ago2-loaded miRNA and miRNA* strands present in our small RNA libraries from fly heads. To prevent differential rates of transcription or miRNA precursor processing from skewing our analysis, for each set of small RNAs derived from a common precursor, we weighted the sequence bias of each miRNA or miRNA* isoform by its relative abundance, then averaged the sequence bias among all miRNAs or miRNA* strands, weighting each locus equally (Fig. 5).

Our analysis suggests that the first nucleotide of a fly small RNA reflects its sorting between Ago1 and Ago2. miRNAs expressed in fly heads generally began with U (72%) rather than A (15.2%), C (7.6%), or G (5.2%); for miRNAs bound to Ago1, as judged by their copurification with immunoprecipitated Ago1, 73.5% began with U, whereas 7.1% began with C. Among the miRNA and miRNA* species that were significantly (p ≤ 0.01) enriched in Ago1 relative to the total small RNA pool of fly heads, 83.9% began with U; just 3.4% began with C. In contrast, 49% of miRNAs that were enriched in Ago2 began with U; 21.6% began with C and 21.8% began with A, indicating a selection against a 5’ U.

miRNA* strands showed a distinctly different 5’ sequence bias. The miRNA* detected in fly heads typically began with A (28.2%), C (32.1%), or G (22.1%), rather than U (17.6%). In contrast to this overall 5’ sequence bias, those miRNA* that were significantly enriched in Ago1 began either with A (56.3%) or U (29.2%); the population of miRNA* loaded into Ago1 was depleted of miRNA* isoforms that begin with C.

Ago2-load miRNA* strands showed the opposite bias: they typically began with C. Nearly 58% of miRNA* strands
enriched in Ago2 and detected in the oxidized library began with C, 15.2% began with A, and just 7.7% began with U, a sequence bias significantly different from the composition of nucleotides 2–18 of the same small RNAs ($P = 6.7 \times 10^{-10}$, Fisher’s exact test) and from the first nucleotide bias of miRNA* overall ($P = 6.6 \times 10^{-7}$) and of those miRNA*s

FIGURE 4. Pairing profiles of Ago1- and Ago2-loaded small RNA guides. (A) Box plots illustrate the predicted double-stranded character of each nucleotide position, 1–19, for all Ago1- or Ago2-enriched miRNA or miRNA* strands. (B) The Wilcoxon test $P$-value for each comparison was used to identify nucleotide positions that were significantly different between Ago1-enriched miRNA plus miRNA* compared with Ago2-enriched miRNA plus miRNA*. The red line indicates $P = 0.01$. Gray circles, nonsignificant; black circles, significant. (C) Box plots illustrate the differences in double-stranded character for each position that was significantly different in double-stranded character between Ago1- and Ago2-loaded miRNA plus miRNA* in B. (D) The data in A–C suggest that miRNA duplexes with less stable 5′ ends and central mismatches act as guides for Ago1 and miRNA duplexes with less stable 3′ ends act as guides for Ago2.
miRNAs and miRNA* s show an Argonaute-specific first nucleotide bias. miRNAs and miRNA* s associated with Ago1 or Ago2 differ in the bias of their first nucleotide. miRNAs generally begin with uridine; this bias increased for the subset of miRNA that were Ago1-bound (measured in the Ago1 immunoprecipitate library), and increased further for the subset of Ago1-enriched miRNAs (measured in the total small RNA library). In contrast, Ago2-enriched miRNAs were depleted of 5’ uridine in the oxidized small RNA library, miRNA* strands generally began with adenosine or cytidine. All miRNA* strands detected in the oxidized library (i.e., loaded in Ago2) or those enriched in Ago2, were significantly more likely to begin with cytidine, whereas those miRNA* s enriched in Ago1 were depleted of a 5’ cytidine. A 5’ cytidine bias was also observed for white exo-siRNAs and was diminished in r2d2’ mutant defective in Ago2 loading.

loaded into Ago1 (P ≤ 0.017). Overall, 40% of the Ago1-enriched miRNA or miRNA* species began with U, whereas 23% of the Ago2-enriched miRNA or miRNA* species began with C.

Essentially identical sequence biases for both miRNA and miRNA* were present in independent small RNA libraries from male and female heads, in libraries prepared from three distinct genetic backgrounds (Oregon R, dcr-2Δ161fsX/Cyo, or r2d2/Cyo), in libraries of Ago2-associated small RNAs that were prepared using either oxidation or oxidation followed by β-elimination, and in libraries processed and sequenced using two different high throughput technologies: pyrosequencing (“454”) or sequencing-by-synthesis (Illumina Genome Analyzer). Together, these data suggest that, in flies, a 5’ terminal U promotes Ago1 loading, but discourages association with Ago2, whereas a 5’ terminal C directs a small RNA away from Ago1 and toward Ago2.

To further test this hypothesis, we analyzed the 5′ nucleotide composition of exo-siRNAs derived from a P-element transgene expressing a long inverted repeat corresponding to exon 3 of the white miRNA. We compared the overall population of white exo-siRNAs with those white exo-siRNAs bound to Ago2, as inferred from their presence in an oxidized small RNA library. Because the white exo-siRNA species are transcribed and diced from a common transcript, differences in their steady-state abundance likely reflect, at least in part, their different propensities to load into an Argonaute protein. Supporting this view, white exo-siRNAs levels decline >10-fold in vivo in a dcr-2Δ161fsX, r2d2’, or ago2414 mutant (T Du and PD Zamore, unpubl.). We therefore weighted each first nucleotide according to the abundance of the corresponding white exo-siRNA species.

Like Ago2-enriched miRNA*, exo-siRNAs isolated from fly heads typically began with C (39.8%), rather than A (17.3%), G (20.5%), or U (22.4%), a sequence bias significantly different from that of the corresponding strand of the dsRNA from which they are derived (P ≤ 1.8 × 10^-9). Among the white exo-siRNAs in the library prepared from oxidized small RNA—i.e., small RNAs bound to Ago2—47% began with C. Supporting the view that the strong C-bias of exo-siRNAs reflects their association with Ago2, the 5′ C bias was not observed among the 17-fold lower amount of exo-siRNAs that remained in an r2d2’ mutant. r2d2’ mutant flies are defective in loading exo-siRNAs into Ago2 and do not silence white expression (Fürstemann et al. 2005).

To further test the idea that the first nucleotide of a small RNA duplex influences its sorting between Ago1 and Ago2 in flies, we examined the loading of small RNA duplexes in vitro, using a previously described UV cross-linking assay (Tomari et al. 2007). We synthesized two miRNA duplexes, one corresponding to the authentic Drosophila let-7 miRNA/miRNA* duplex, which begins with a 5′ U, and a second in which the initial U of let-7 was changed to a 5′ C (Fig. 6A). In parallel, we also synthesized two siRNA duplexes in which the guide strand was either authentic let-7 (paired to its reverse complement) or let-7 bearing a 5′ C instead of a U (Fig. 6A). Each miRNA or guide strand was 5′ 32P-radiolabeled, so that cross-linking identified the proteins, including Ago1 and Ago2, to which it bound when incubated in Drosophila embryo lysate.

The miRNA/miRNA* duplex containing authentic let-7 strand—i.e., let-7 that began with a 5′ U—cross-linked to Ago1 more efficiently than the let-7 variant that began with a 5′ C (Fig. 6A,B); neither duplex detectably loaded its miRNA strand into Ago2. Moreover, when we performed cross-linking in Ago1 immunodepleted lysate, not only was Ago1 cross-linking absent, but no Ago2 cross-link appeared. We conclude that the structure of a miRNA duplex not only favors Ago1 loading, but actively prevents loading of the miRNA into Ago2. Moreover, the interplay between the structure of the miRNA duplex and its 5′ nucleotide determines its distribution between Ago1 and Ago2.

In contrast to the miRNA/miRNA* duplexes, the siRNA duplexes cross-linked mainly to Ago2, although some Ago1 cross-linking was clearly detected. For the siRNA duplexes,
the influence of first nucleotide identity on the efficiency of Ago2 loading was opposite of that observed for the miRNAs: the siRNA duplex whose guide strand began with a 5′ cytidine loaded more efficiently into Ago2 than the siRNA duplex whose guide began with U (Fig. 6A,C). Together, these in vitro data provide strong support for the hypothesis that the enrichment for a 5′ uridine among Ago1-loaded miRNAs and for a 5′ cytidine among Ago2-loaded miRNA*, miRNA*, and siRNAs reflects a direct role for 5′ nucleotide identity in small RNA sorting between Ago1 and Ago2 in *Drosophila*.

**For some miRNA and miRNA*, distinct isoforms load into Ago1 and Ago2**

At least nine *Drosophila* pre-miRNA produce from one side of their stem two small RNAs that partition differentially between Ago1 and Ago2. Such differentially partitioning miRNA or miRNA* isoforms differ at their 5′ ends and therefore present subtly different duplexes to the Argonaute-loading machinery. Moreover, the differentially sorting isoforms have different seed sequences, which would allow them to regulate distinct repertoires of target mRNAs. Figure 7 presents these “seed switching” miRNA and miRNA* isoforms in the context of the duplexes from which they are presumed to be loaded into Ago1 or Ago2. Pre-miR-193 provides a particularly stunning example of such isoform-specific Argonaute loading. This pre-miRNA generates two miR-193 isoforms: one begins with a U and loads into Ago1, whereas a miR-193 isoform that begins at the next nucleotide, an A, loads into Ago2. Pre-miR-193 also generates two miR-193* isoforms. Again, the one that begins with a U loads into Ago1, whereas a less

---

**FIGURE 6.** Ago1 prefers to load miRNAs that begin with a 5′ uridine, while Ago2 prefers siRNAs that begin with a 5′ cytidine. (A) Four small RNA duplexes were incubated with embryo lysate and then cross-linked with shortwave UV to identify small RNA-bound proteins. Representative data are shown. (B) Kinetic analysis of miRNA association with Ago1, monitored by UV cross-linking. (C) Kinetic analysis of siRNA association with Ago2, monitored by UV cross-linking. In B and C, each data point represents the average ± standard deviation for three trials.
abundant isoform that begins at the G that lies immediately 5' to the U loads into Ago2. This small collection of seed switching miRNA and miRNA* gives the impression that the sorting of imperfectly paired small RNA duplexes between Ago1 and Ago2 reflects a complex interplay between structural determinants or anti-determinants and first nucleotide preferences and dislikes.

DISCUSSION

Historically, miRNAs were defined as the more abundant of the small RNAs derived from the two sides of a pre-miRNA stem (Lagos-Quintana et al. 2001; Lau et al. 2001; Lee and Ambros 2001). The miRNA* strand has been proposed to be destroyed during Argonaute loading, explaining its considerably lower abundance (Khvorova et al. 2003; Schwarz et al. 2003). Yet, high depth sequencing has revealed that many miRNA* species are more abundant than some miRNA species, and miRNA/miRNA* ratios may vary dramatically among developmental stages (Ro et al. 2007; Okamura et al. 2008c).

In fly heads and ovaries, several miRNA* strands are more abundant than their annotated miRNA counterparts (Table 1). In our data sets, miR-92a was more abundant than miR-92a* in ovaries (3240 ppm miRNA versus 15 ppm miRNA*), while its miR-92a* was more abundant than miR-92a in heads (24 ppm miRNA versus 106 ppm miRNA*). Likewise, miR-988 (260 ppm miRNA versus 300 ppm miRNA* in heads, but 124 ppm miRNA versus 49 ppm miRNA* in ovaries) and miR-284 (4993 ppm miRNA versus 915 ppm miRNA* in heads; 49 ppm miRNA versus 72 ppm miRNA* in ovaries) showed distinctly different miRNA/miRNA* ratios in ovaries and heads. Such altered ratios may reflect different concentrations of Ago1 and Ago2 or of components of their respective Argonaute-loading machineries in the two organs.

Sorting combines structure and sequence information

In general, miRNAs associate with Ago1 and miRNA* strands associate with Ago2 in Drosophila. It is important to note that our data argues strongly against a model in which miRNA* strands bind Ago2 as a consequence of the corresponding miRNA binding Ago1. First, we can identify

<table>
<thead>
<tr>
<th>TABLE 1. Pre-miRNAs whose miRNA* strands were more abundant than their miRNAs among small RNAs isolated from fly heads and fly ovaries.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-miRNA</strong></td>
</tr>
<tr>
<td>Fly head</td>
</tr>
<tr>
<td>miR-10</td>
</tr>
<tr>
<td>miR-1012</td>
</tr>
<tr>
<td>miR-193</td>
</tr>
<tr>
<td>miR-281-2</td>
</tr>
<tr>
<td>miR-5</td>
</tr>
<tr>
<td>miR-92a</td>
</tr>
<tr>
<td>miR-988</td>
</tr>
<tr>
<td>Fly ovary</td>
</tr>
<tr>
<td>miR-10</td>
</tr>
<tr>
<td>miR-1012</td>
</tr>
<tr>
<td>miR-276b</td>
</tr>
<tr>
<td>miR-281-2</td>
</tr>
<tr>
<td>miR-284</td>
</tr>
</tbody>
</table>
six miRNAs in which both the miRNA and the miRNA* strand are enriched in Ago1 complexes in fly heads. Second, we find 15 miRNAs for which both the miRNA and miRNA* strands are enriched in Ago2 complexes. Our data suggest that each miRNA/miRNA* duplex presents two distinct structures to the sorting machinery: One in which the miRNA is the presumptive guide and one in which the miRNA* assumes that position. Evolution appears to have selected for miRNA/miRNA* duplexes that present sequence and structural features appropriate for loading Ago1, simultaneously favoring Ago2 loading when the same duplex is viewed from the perspective of the miRNA*. Consequently, miRNAs generally load into Ago1, whereas miRNA’s load into Ago2, an Argonaute protein previously thought to act only in the RNAi pathway. miRNA’s are therefore the first class of Drosophila small silencing RNAs produced by Dicer-1, but preferentially loaded into Ago2 (Fig. 8).

miRNA/miRNA* duplexes that preferentially load Ago1 are typically less stably paired at their 5’ ends and contain central mismatches, bulges, or G:U wobble pairs, whereas miRNA*/miRNA duplexes that preferentially load Ago2 possess more stably paired 5’ ends and center, but have less stably paired 3’ ends. In addition to structure, sequence also plays a role in small RNA sorting in flies. Ago1-bound miRNAs begin overwhelmingly with uridine, whereas Ago2-bound miRNA, miRNA*, and siRNA tend to begin with cytidine. Moreover, our in vitro cross-linking experiments show that a 5’ U increased the efficiency of miRNA loading into Ago1, relative to a 5’ C, whereas a 5’ C—in the context of an siRNA duplex—increased the efficiency of Ago2 loading, relative to a 5’ U.

The 5’ terminal nucleotide of a small RNA is anchored in the phosphate-binding pocket of Argonaute proteins and unavailable for base pairing with its RNA target (Ma et al. 2005; Parker et al. 2005). We speculate that the structures of Ago1 and Ago2 discriminate between U and C by making specific hydrogen-bonding contacts with the edges of the first base of a small RNA guide.

The fate of a miRNA/miRNA* duplex, therefore, depends on multiple factors: structure of its duplex, thermodynamic stability of the ends of the duplex and the identity of its 5’ terminal nucleotide. We do not yet know to what extent each factor weighs in the sorting decision.

miRNA loci appear to generate an extraordinary diversity of functional small RNAs. Some miRNA genes are transcribed from both DNA strands, producing two different hairpins from a single genomic locus (Stark et al. 2008; Tyler et al. 2008). A few miRNA have been annotated as producing functional small RNAs—miRNA-5p and miRNA-3p—from a single pre-miRNA, a phenomenon that we suggest may be the rule rather than the exception. Our data argue that the two small RNAs, typically annotated as miRNA and miRNA*, from a single pre-miRNA partition into distinct effector proteins, with the miRNA loading into Ago1 and the miRNA* loading into Ago2. These Ago2-loaded miRNA*s are present at levels comparable to exo-siRNAs. Moreover, Ago2-loaded small RNAs can guide either target cleavage or translational repression (Iwasaki et al. 2009), suggesting that Ago2-loaded miRNA*s function to regulate as yet to be identified target RNAs. Finally, we find that a single arm of a single pre-miRNA hairpin can give rise to several functional RNA isoforms that possess different seed sequence and that associate with different Argonaute proteins that have distinct biological activities. These three layers of functional diversification—multiple small RNAs that partition differently from the two sides of the stem of a single pre-miRNA, different seed isoforms from a single side of a pre-miRNA stem, and distinct partitioning of these RNA seed isoforms—allows a single, compact genomic locus, the miRNA gene, to produce multiple riboregulators, each with a distinct biological activity and target repertoire.

MATERIALS AND METHODS

General methods
Fly strains were wild-type Oregon R, dcr-2<sup>1411X</sup>, r2d2<sup>2</sup>, and ago2<sup>414</sup>. Fly heads were isolated by vigorous shaking of liquid nitrogen-frozen flies in nested, prechilled sieves (U.S. standard sieve, Humboldt MFG), allowing the heads to pass through the top sieve (No. 25), and collecting them on the bottom sieve (No. 40).

Small RNA sequencing
Total RNA was extracted with the mirVana kit (Ambion), then 18–30-nt-long RNA was gel purified. 2S rRNA was depleted as
described (Seitz et al. 2008). A part of the sample was then oxidized using sodium periodate (Horwich et al. 2007) without β-elimination step. Size-selected RNA derived from at least 68 µg total RNA for oxidation; and size-selected RNA derived from at least 7 µg total RNA for untreated. Library preparation was as described previously (Ghildiyal et al. 2008). High-throughput sequencing was by Genome Analyzer II (Illumina). Raw high throughput sequencing data sets have been deposited at the National Center for Biotechnology Information Short Read Archive (www.ncbi.nlm.nih.gov/sites/sra) as GSE18806.

Preparation of fly head extract

Isolated fly heads were transferred to 1.5 mL microcentrifuge tubes, prechilled in liquid nitrogen, and homogenized using a plastic “pellet pestle” (Kontes) in 1 mL ice-cold Lysis buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) containing 5 mM DTT and 1 mg/mL complete “mini” EDTA-free protease inhibitor tablets (Roche Applied Science) per gram of heads. Lysate was clarified by centrifugation at 14,000g for 30 min at 4°C. The supernatant was dispensed into prechilled microcentrifuge tubes, flash frozen in liquid nitrogen, and stored at −80°C. Total protein concentration was determined by Bradford assay.

Immunoprecipitation

For small RNA cloning, immunoprecipitation of Ago1 protein was essentially as described (Miyoshi et al. 2005). Briefly, 40 µL GammaBind beads (GE Healthcare: #17-0885-01) were washed four times with 1 mL of Lysis buffer with DTT and protease inhibitors and containing 0.5% v/v NP-40, then incubated with 40 µL monoclonal anti-Ago1 antibody (Miyoshi et al. 2005) in 1 mL Lysis buffer at 4°C for 3 h. After washing five times with 1 mL of Lysis-IP buffer, the antibody-bound beads were incubated with 910 µL fly head lysate (~4.55 mg total protein) at 4°C for 16 h, and then the supernatant collected and the beads washed five times with 1 mL of RIPA buffer (50 mM Tris [pH 8.0], 1.0% v/v NP-40, 150 mM NaCl, 0.5% v/v DOC, 0.1% v/v SDS, 1X Complete-EDTA-free protease inhibitor cocktail table). Immunoprecipitation efficiency was confirmed by Western blotting.

UV cross-linking

UV cross-linking was performed in embryo lysates prepared as described (Tuschi et al. 1999). Embryo lysates were immunodepleted for Ago1 as described above. UV cross-linking was as previously described (Tomari et al. 2007), except that the samples were ~0.5 cm from the UV lamp.

Computational analyses

For each sequence read, the first occurrence of the hexamer perfectly matching the 5′ end of the 3′ linker was identified. Sequences without a linker match were discarded. The extracted inserts for sequences that contained the 3′ linker were then mapped to the Drosophila melanogaster genome (Release R5.5). Inserts that matched perfectly and completely to the genome were collected using Bowtie (Langmead et al. 2009), and the corresponding genomic coordinates were determined for downstream functional analysis. Sequences corresponding to pre-miRNA hairpins (miRBase, 13.0) or noncoding RNAs (ncRNAs; Supplemental Table S3) were identified using the same suffix tree-based software. Genes were retrieved from FlyBase (R5.5). We manually curated mature miRNA*. Mature miRNA annotations were obtained from miRBase (13.0). We allowed sequence reads to differ in 5′ and 3′ ends from mature miRNA or miRNA* for up to 9 nt. Endogenous siRNA (endo-siRNA) were defined as genome mapping 21-mers detected in the oxidized library and that did not map to ncRNA or miRNA hairpins. Exogenous siRNA (exo-siRNA) were 21-mers detected in the oxidized library and that mapped perfectly to the white inverted repeat. Except for Fisher’s exact test, which requires raw sequence reads, all sequence reads are reported in parts per million reads of sequencing depth, with the sequencing depth defined as total number of linker containing, genome-matching reads excluding ncRNAs.

Fisher’s exact test was applied to each miRNA or miRNA* to identify those that are enriched in Ago1 or Ago2. Take miR-1 as an example, a 2 × 2 contingency table includes the following cells: number of reads of miR-1 detected in the library prepared from the Ago1 immunoprecipitate, number of reads of all other miRNA or miRNA* in this library, number of reads of miR-1 in the library prepared from oxidized small RNA, and number of reads of all other miRNA or miRNA* in the oxidized library. P-values ≤ 0.01 were deemed significant. Furthermore, we required a miRNA or miRNA* enriched in an Argonaute protein to be at least 10 ppm in that Ago. Enrichment score (Fig. 1B) was defined as the number of reads of a particular miRNA or miRNA* in one Argonaute versus the other. A pseudocount (or an uninformied prior in Bayesian statistics) of 10 ppm was used to control noise arising from extremely low abundance. For example, for miR-1 the enrichment score was [(number of miR-1 reads in Ago1 + 10)/(total number of all miRNA reads in Ago1 + 10)]/ [(number of miR-1 in Ago2 + 10)/(total number of all miRNA reads in Ago2 + 10)]. Similarly, the fold change in a mutant compared with the wild type, again using mir-1 as an example, was defined as (number of miR-1 reads in the mutant + 10)/(total number of miR-1 reads in the wild-type + 10), where 10 ppm was the pseudocount.

Pairing probabilities were calculated using RNAcofold (ViennaRNA-1.8.3, http://www.tbi.univie.ac.at/RNA/). For each Argonaute-enriched miRNA or miRNA*, the most abundant isoform for that miRNA or miRNA* was chosen to be the guide strand and the corresponding passenger was taken to be the most abundant isoform of the miRNA* or miRNA from the wild-type untreated experiment (see Supplemental Discussion for empirical support for this approach). Both the guide and passenger were required to pass the aforementioned 10 ppm threshold. The probability per position was the sum of the pairing probabilities for that position. Pairing probability for each position was smoothed by the values of the two neighboring nucleotides. For each position, we tested the significance of the difference between all Ago1-enriched miRNA and miRNA* together and all Ago2-enriched miRNA and miRNA* together using the two-sided Wilcoxon ranked-sum test with 0.01 as the threshold for significance.

To compute first nucleotide bias, we used an egalitarian weighting scheme to account for the difference in transcriptional and processing efficiency for different miRNA and miRNA*. The isoforms for a particular miRNA or miRNA* were weighted by their abundance
Drosophila miRNA* strands loaded in Ago2


