

Published in final edited form as:

Mol Cell. 2008 August 8; 31(3): 309–312. doi:10.1016/j.molcel.2008.07.008.

An Inside Job for siRNAs

Daniel E. Golden¹, Vincent R. Gerbasi¹, and Erik J. Sontheimer^{1,*}

¹Department of Biochemistry, Molecular Biology and Cell Biology, Northwestern University, 2205 Tech Drive, Evanston, IL 60208, USA

Abstract

Among the three main categories of small silencing RNAs in insects and mammals—siRNAs, miRNAs, and piRNAs—siRNAs were thought to arise primarily from exogenous sources, whereas miRNAs and piRNAs arise from endogenous loci. Recent work in flies and mice reveals several classes of endogenous siRNAs (endo-siRNAs) that contribute to functions previously reserved for miRNAs and piRNAs, including gene regulation and transposon suppression.

All cells face a wide range of threats and regulatory demands, some from within and some from without. Among the many tools available to meet these challenges is a collection of pathways that use small (~20–30 nucleotide) RNAs to recognize target nucleic acids and present them to specific effector complexes that generally inhibit gene expression (Farazi et al., 2008). A long and ever-expanding roster of such RNAs has emerged during the last 15 years, but in animals, most can be subsumed under the three main headings of microRNAs (miRNAs), Piwi-interacting RNAs (piRNAs), and short interfering RNAs (siRNAs). Although they share some common features, each RNA category can differ from the others in various ways including length, precursor structure, cofactor requirement, modification state, sequence bias, and regulatory function, and the differences can themselves vary between species.

Another crucial distinction is molecular origin. From the time that miRNAs and piRNAs were first defined (in 1993 and 2006, respectively), it has been clear that they are encoded within cellular genomes and produced endogenously (Figure 1). The threats and regulatory demands that they are called upon to meet are, for the most part, internal: miRNAs regulate the expression of large numbers of endogenous genes (though viruses can also get in on the action) (Kloosterman and Plasterk, 2006), and piRNAs suppress the potentially hazardous mobility of transposons in the germline (Aravin et al., 2007). Accordingly, mutants that lack protein components of the miRNA and piRNA pathways exhibit severe developmental defects and sterility, respectively. In contrast, siRNAs in animals have mostly been considered extragenomic in origin. Some animals, such as *C. elegans*, can employ RNA-dependent RNA polymerase (RdRP) enzymes to generate siRNAs and their precursors, and in many cases, the RdRP-dependent siRNAs correspond to endogenous loci (Pak and Fire, 2007; Sijen et al., 2007; and references therein). However, the genomes of many other animals, such as insects and mammals, apparently lack RdRP coding potential, consistent with a lack of “transitive RNAi” in these organisms. Although there have been some hints of genome-derived siRNAs in insects and mammals, most characterized siRNAs were either virally encoded or experimentally induced. Thus, in cells from organisms such as *Drosophila*, mice, and humans, the classic RNA interference (RNAi) pathway that proceeds through siRNAs was thought to be dedicated primarily to protection against external threats. Consistent with this,

Drosophila mutants that lack key protein components of the siRNA-based silencing pathway are viable, fertile, and highly susceptible to virus infection (Marques and Carthew, 2007).

Discovery of a Broad Population of Endogenous siRNAs

Despite these considerations, the presumption against the possibility of abundant, endogenous siRNAs (endo-siRNAs) in animals such as flies and mice was largely based on negative evidence. In a flurry of eight papers, several groups have now identified endo-siRNA populations in these species (Chung et al., 2008; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a; Okamura et al., 2008b; Tam et al., 2008; Watanabe et al., 2008). These results point to previously unknown roles for endo-siRNAs in gene regulation and transposon taming in both the soma and the germline.

In *Drosophila*, several features that distinguish siRNAs from their miRNA and piRNA relatives (see Farazi et al. [2008] for review) were exploited to search for endo-siRNAs. First, siRNAs are typically ~21 nt long, slightly shorter than most miRNAs (~22 nt) and piRNAs (24–26 nt); thus, front-end size fractionation or back-end computational filtering can be used to enrich for siRNA sequences. Second, the different categories of RNAs associate with distinct effector proteins. All of the ~20–30 nt regulatory RNAs exert their functions in association with proteins of the Argonaute superfamily (Hutvagner and Simard, 2008), and in most animals, two subfamilies (Ago and Piwi) exist that bind siRNAs/miRNAs and piRNAs, respectively. Of the five Argonaute proteins in *Drosophila*, Ago2 is primarily devoted to siRNA function, and Ago2 association therefore can be considered a diagnostic feature of siRNAs. Third, miRNAs arise from intramolecular stem-loop structures of ~60–70 nt that lack perfect Watson-Crick complementarity, whereas siRNAs can be processed from a broader range of duplex structures that are perfectly base paired, or nearly so. The structures of piRNA precursors are not as well defined but are apparently single-stranded. Hence, once a small RNA has been mapped to the genome, the sequence context of the locus can be used to categorize the small RNA. Finally, piRNAs and siRNAs carry 2'-O-methyl modifications at their 3' termini, whereas miRNAs do not, rendering the latter susceptible to periodate oxidation and β -elimination.

To enrich for endo-siRNAs, Kawamura et al. (2008) and Czech et al. (2008) immunoprecipitated Ago2 from *Drosophila* cells and tissues and analyzed the coprecipitating RNAs, most of which had the expected length of 21 nt. The two data sets were in excellent agreement. Most of the detected RNAs corresponded to transposon-derived sequences, but both groups also observed numerous endo-siRNAs that were not related to mobile genetic elements (see below). In three separate reports, Lai and colleagues used bioinformatic tools, small RNA database searches, and some deep sequencing of their own to draw similar conclusions. Okamura et al. (2008a, 2008b) homed in on regions of the *Drosophila* genome predicted to generate duplex structures of various kinds (both intra- and intermolecular) and then examined small RNA sequence databases to identify ~21 nt RNAs from these sites. Similarly, Chung et al. (2008) mined the deep sequencing datasets for ~21 nt sequences corresponding to transposon fragments, and uncovered many examples. In all three of these reports, the small RNAs were periodate-resistant and dependent upon Ago2 for their accumulation, consistent with their designation as endo-siRNAs. Finally, Ghildiyal et al. (2008) adopted yet another approach: they depleted miRNAs from somatically derived small RNA samples by periodate oxidation and β -elimination and then deeply sequenced the resulting siRNA-enriched libraries. Collectively, these six papers document numerous ~21 nt, terminally modified, Ago2-associated or -dependent endo-siRNAs in *Drosophila* cells and tissues.

Genomic Context and Biogenesis of *Drosophila* Endo-siRNAs

Like piRNAs, the endo-siRNA pool segregates into subpopulations that arise from unique or repetitive genomic sequences. The latter comprises a very large subset and mostly corresponds to mobile genetic elements. Although the need for transposon suppression is greatest in the germline, mounting evidence points toward mobility in the soma as well, and the siRNA population reflects this, with transposon transcripts apparently serving both as endo-siRNA source and target. Although all of these studies profiled somatic small RNA populations, Czech et al. (2008) examined germline tissue as well and again observed abundant transposon-derived siRNAs. Thus, it appears that siRNAs, as well as piRNAs, contribute to transposon suppression in the germline. Interestingly, some genomic regions previously characterized as piRNA clusters also give rise to endo-siRNAs in both germline and somatic tissues (Chung et al., 2008; Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008). The features that allow these loci to generate both endo-siRNAs and piRNAs are currently unknown. Unlike piRNAs, the transposon-derived endo-siRNAs exhibit little strand or sequence bias, as expected for RNAs with double-stranded precursors.

As for the nonrepetitive siRNAs in *Drosophila* cells, most appear to be born from two classes of genomic sites: structured or hairpin RNA (hpRNA) loci and convergent transcripts (Czech et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008; Okamura et al., 2008a, 2008b) (Figure 1, red arrows). The latter generally consist of adjacent transcription units of opposite orientation that overlap within their 3' UTRs and have the capacity to anneal and form extensive regions of perfect duplex RNAs. These would serve as processing substrates for Dicer-2, the predominant siRNA-generating enzyme in flies (Filipowicz, 2005). Intriguingly, one participant in such an siRNA-generating arrangement is none other than *ago2*, though whether the corresponding siRNAs induce some form of RNAi autoregulation remains unknown. Two other protein-coding genes—*klarsicht* and *thickveins*—also spawned large numbers of endo-siRNAs, apparently from sense and antisense transcripts that arise in *cis* (Czech et al., 2008; Okamura et al., 2008a). In each of these two cases, however, the siRNA-generating region is limited to a discrete portion of an intron rather than the 3'UTR, indicating that the antisense transcript initiates internally rather than in a flanking gene. The roles of these siRNAs are not yet established, though their apparent tissue specificity (Czech et al., 2008) could reflect regulatory functions.

The structured or hairpin RNA (hpRNA) loci, in contrast, are generally derived from noncoding RNAs that are initially single-stranded but that fold back on themselves to form intramolecular duplexes that, despite some deviations from perfect Watson-Crick complementarity, can be processed by Dicer-2. (These structures are distinct from miRNA precursor hairpins that are processed by Drosha and Dicer-1, but not Dicer-2.) Many such hpRNA loci appear to exist, though all four groups focused on two apparent noncoding RNA genes (dubbed *esi-1* and *esi-2* by Czech et al. [2008]) that yield a particularly large number of endo-siRNAs. The presumed *esi-1* and *esi-2* precursor transcripts give rise to a nonrandom population of endo-siRNAs—specific 21 nt sequences from within these RNAs appear frequently, and most others not at all. A few of these endo-siRNAs are highly complementary to particular protein-coding genes in the *Drosophila* genome, suggesting possible functions (see below).

In addition to the endo-siRNAs that precisely match the *Drosophila* genome, two other populations were also noted. Curiously, Kawamura et al. (2008) found that nearly 20% of their Ago2-associated endo-siRNAs from S2 cells match the *Drosophila* genome at all nucleotide positions save for a single A-to-G transition. This prompted the authors to speculate that the endo-siRNAs might be substrates for the adenosine deaminase RNA-editing enzyme ADAR, though the functional significance of the proposed editing is not known. Separately, when analyzing Ago2-associated RNAs from S2 cells, Czech et al. (2008) observed large numbers

of siRNAs from the flock house virus genome, consistent with the reported role of RNAi in battling this virus that is known to infect many S2 cultures (Marques and Carthew, 2007).

What precise pathways give rise to the pool of endo-siRNAs? Given that siRNAs from exogenous sources are known to function through the effector protein Ago2, all four groups tested whether endo-siRNA accumulation and function also depend on Ago2, and all found that they do. Similarly, all four groups reported that the role of Dicer-2 in the biogenesis of exogenous siRNAs is recapitulated with endo-siRNAs. Beyond Ago2 and Dicer-2, however, the scenario takes an unanticipated turn. Dicer-2 heterodimerizes with the double-stranded RNA binding domain (dsRBD) protein R2D2, and exogenous siRNAs generally depend upon R2D2 for loading into Ago2 (Filipowicz, 2005) (Figure 1). In contrast, Dicer-1 heterodimerizes with a different dsRBD protein, Loquacious (Loqs), which is thought to function primarily in the miRNA pathway. Surprisingly, new results reveal that Loquacious, not R2D2, is broadly important for the accumulation of mature endo-siRNAs (Chung et al., 2008; Czech et al., 2008; Okamura et al., 2008a, 2008b), including those derived from the presumably perfectly base-paired, convergent transcripts. Czech et al. (2008) also demonstrate that Dicer-2 is present in anti-Loqs immunoprecipitates. There is little to tell us what dictates Dicer-2's apparent partnership with Loqs when faced with endo-siRNA precursors and R2D2 in the case of exogenous double-stranded RNAs. Nonetheless, these and other recent results (Kalidas et al., 2008) indicate that the relationships between *Drosophila* Dicers and their dsRBD partner proteins are more malleable than previously appreciated, and disentangling these newly recognized complexities is now an important goal.

The Functional Meaning of *Drosophila* Endo-siRNAs

With the existence of a large population of *Drosophila* endo-siRNAs firmly established, each lab turned to the obvious question: what are they there for? The answer is most definitive in the case of the transposon-derived endo-siRNAs. Given that these small RNAs arise from a very large number of genomic sites, it is not possible to mutate them directly, but mutations or knockdowns involving their protein cofactors, such as Dicer-2 and Ago2, can still provide inroads into functional significance. Reassuringly, the reports broadly agree that transposons are derepressed (at least at the mRNA level) when the siRNA pathway is compromised. Transcripts from a subset of mobile elements increase 2- to 10-fold in *ago2* mutant heads (Chung et al., 2008) and ovaries (Czech et al., 2008), *dicer-2* mutant heads (Chung et al., 2008; Ghildiyal et al., 2008; Kawamura et al., 2008), and *dicer-2* and *ago2* knockdown S2 cells (Chung et al., 2008; Ghildiyal et al., 2008), as foreshadowed by earlier microarray analyses (Rehwinkel et al., 2006). No significant increase in transposon transcripts was apparent after *dicer-1* knockdown (Ghildiyal et al., 2008). These results correlate well with the reductions in endo-siRNAs observed in similar samples. Thus, the endo-siRNA pathway apparently contributes to transposon repression, either on its own in somatic tissues or in collaboration with the piRNA pathway in the germline. The comparatively severe effects of piRNA pathway mutations on fertility suggest a dominant role in transposon taming, especially in the male germline, and this may be due in part to the piRNA-specific feed-forward amplification loop that facilitates an adaptive response to transposon mobilization (Aravin et al., 2007).

Intriguingly, Ghildiyal et al. (2008) close with a guarded but potentially tantalizing description of “piRNA-like” RNAs from *ago2* null mutant heads. Whether these 24–27 nt species represent true piRNAs is not yet known, but if so, this observation could imply that endo-siRNAs somehow participate in limiting piRNA function outside of the germline. The interplay between endo-siRNAs and piRNAs in transposon control and, in particular, the involvement of clusters that give rise to both types of silencing RNAs are now ripe for detailed analysis.

The unique endo-siRNAs that are processed from structured loci and convergent transcripts raise the possibility of a pervasive role in host gene regulation, but so far the evidence for this is less compelling. The accumulation of specific hpRNA-derived endo-siRNAs naturally prompted a complementarity-based search for potential targets, and in the case of one *esi-2* endo-siRNA, a clear candidate emerged: the *mus308* gene that has been implicated in DNA damage responses (Czech et al., 2008; Okamura et al., 2008b). Native endo-siRNAs can direct Ago2-catalyzed slicer activity in vitro (Kawamura et al., 2008; Okamura et al., 2008b). Moreover, sensor assays, 5'-RACE detection of apparent natural *mus308* cleavage products, and mild *mus308* derepression in *dicer-2* and *ago2* mutant tissue all suggest that *esi-2*-directed *mus308* silencing may be meaningful in vivo (Czech et al., 2008; Okamura et al., 2008b). Nonetheless, the potential functions of most other hpRNA-derived endo-siRNAs remain to be defined. The situation is even murkier for the siRNAs derived from convergent transcripts: these endo-siRNAs are of relatively low abundance, and little or no upregulation of convergently transcribed loci is observed in endo-siRNA-defective mutants (Czech et al., 2008; Okamura et al., 2008a). Czech et al. (2008) even consider the possibility that these endo-siRNAs represent little more than the “noise” that could be inherent to *Drosophila* silencing in vivo. Nonetheless, the possibility that the convergent-transcript-derived siRNAs could autoregulate their own source messages under some circumstances still has considerable appeal, especially in light of a recent report that 3'-UTR length can be modulated in a manner that affects responsiveness to RNA silencing pathways (Sandberg et al., 2008).

A Broad Class of Endo-siRNAs in Mouse Oocytes

Two additional reports demonstrate that *Drosophila* is not unique among apparently RdRP-negative organisms in the existence of endo-siRNA pathways. Prompted in part by the puzzling male-specific sterility of mutations in Piwi-class proteins, despite the need to limit transposition in both the male and female germlines, Tam et al. (2008) and Watanabe et al. (2008) profiled the total small RNA population in mouse oocytes and uncovered a much broader class of endo-siRNAs than previously recognized (Watanabe et al., 2006). Many endo-siRNAs match transposon sequences, indicating that piRNAs and endo-siRNAs may collaborate in germline transposon suppression in mice as it does in flies; why endo-siRNAs would suffice in the female, but not the male, germline, as indicated by the Piwi family mutant phenotypes, remains unknown.

Another subclass of mouse oocyte endo-siRNAs points toward an unexpected role of pseudogenes in regulating their intact, functional counterparts (Tam et al., 2008; Watanabe et al., 2008). Certain clusters of antisense and sense endo-siRNAs (including some spanning exon-exon junctions) correspond to pseudogenes and their cognate protein-coding genes, respectively, and in some cases the protein-coding gene is upregulated in *Dicer* null oocytes. Thus, pseudogenes may be more useful than previously appreciated, thanks to their potential to step into regulatory roles via RNA silencing pathways.

Conclusions and Perspectives

Clearly, the scope of RNA silencing in biology is still expanding, and the list of new questions grows in parallel. The capacity of endo-siRNAs to regulate their targets posttranscriptionally seems clear, but could they operate at the level of chromatin as well as suggested by Kawamura et al. (2008)? How independent are the operations of the endo-siRNA and piRNA pathways in the germline? What is the significance of the apparent *lack* of endo-siRNAs in mouse embryonic stem cells (Calabrese and Sharp, 2006)? Given the limited reproductive phenotypes of siRNA-defective *Drosophila* mutants, what is the true importance of endo-siRNA-based transposon control, and under what conditions (if any) might the endo-siRNA pathway become essential?

The functional interplay between endogenous and exogenous siRNAs is also obscure. Does one class ever come close to saturating the RNAi machinery (for instance, exogenous siRNAs during an acute virus infection), thereby limiting the capacity of the other class to exert its effects? The answer to this question could have practical ramifications: if circumstances exist in which endo-siRNAs limit the availability of the silencing machinery for exogenous siRNAs, then finding ways to circumvent these limitations could be useful in enhancing the effects of siRNA-based therapeutics.

In insects and mammals, endo-siRNAs are apparently authorized to keep tabs on domestic miscreants such as transposons, whereas exogenous siRNAs are called upon to thwart foreign opponents such as viruses. The degree to which the endogenous and exogenous siRNAs impinge upon each other's turf remains to be fully explored, but it seems certain that they will have found ways to cooperate to the benefit of the entire organism.

REFERENCES

- Aravin AA, Hannon GJ, Brennecke J. *Science* 2007;318:761–764. [PubMed: 17975059]
- Calabrese JM, Sharp PA. *RNA* 2006;12:2092–2102. [PubMed: 17135486]
- Chung WJ, Okamura K, Martin R, Lai EC. *Curr. Biol* 2008;18:795–802. [PubMed: 18501606]
- Czech B, Malone CD, Zhou R, Stark A, Schlingeheyde C, Dus M, Perrimon N, Kellis M, Wohlschlegel JA, Sachidanandam R, et al. *Nature* 2008;453:798–802. [PubMed: 18463631]
- Farazi TA, Juranek SA, Tuschl T. *Development* 2008;135:1201–1214. [PubMed: 18287206]
- Filipowicz W. *Cell* 2005;122:17–20. [PubMed: 16009129]
- Ghildiyal M, Seitz H, Horwich MD, Li C, Du T, Lee S, Xu J, Kittler EL, Zapp ML, Weng Z, Zamore PD. *Science* 2008;320:1077–1081. [PubMed: 18403677]
- Hutvagner G, Simard MJ. *Nat. Rev. Mol. Cell Biol* 2008;9:22–32. [PubMed: 18073770]
- Kalidas S, Sanders C, Ye X, Strauss T, Kuhn M, Liu Q, Smith DP. *Mech. Dev* 2008;125:475–485. [PubMed: 18299191]
- Kawamura Y, Saito K, Kin T, Ono Y, Asai K, Sunohara T, Okada TN, Siomi MC, Siomi H. *Nature* 2008;453:793–797. [PubMed: 18463636]
- Kloosterman WP, Plasterk RH. *Dev. Cell* 2006;11:441–450. [PubMed: 17011485]
- Marques JT, Carthew RW. *Trends Genet* 2007;23:359–364. [PubMed: 17467114]
- Okamura K, Balla S, Martin R, Liu N, Lai EC. *Nat. Struct. Mol. Biol* 2008a;15:581–590. [PubMed: 18500351]
- Okamura K, Chung WJ, Ruby JG, Guo H, Bartel DP, Lai EC. *Nature* 2008b;453:803–806. [PubMed: 18463630]
- Pak J, Fire A. *Science* 2007;315:241–244. [PubMed: 17124291]
- Rehwinkel J, Natalin P, Stark A, Brennecke J, Cohen SM, Izaurralde E. *Mol. Cell. Biol* 2006;26:2965–2975. [PubMed: 16581772]
- Sandberg R, Neilson JR, Sarma A, Sharp PA, Burge CB. *Science* 2008;320:1643–1647. [PubMed: 18566288]
- Sijen T, Steiner FA, Thijssen KL, Plasterk RH. *Science* 2007;315:244–247. [PubMed: 17158288]
- Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, Hannon GJ. *Nature* 2008;453:534–538. [PubMed: 18404147]
- Watanabe T, Takeda A, Tsukiyama T, Mise K, Okuno T, Sasaki H, Minami N, Imai H. *Genes Dev* 2006;20:1732–1743. [PubMed: 16766679]
- Watanabe T, Totoki Y, Toyoda A, Kaneda M, Kuramochi-Miyagawa S, Obata Y, Chiba H, Kohara Y, Kono T, Nakano T, et al. *Nature* 2008;453:539–543. [PubMed: 18404146]

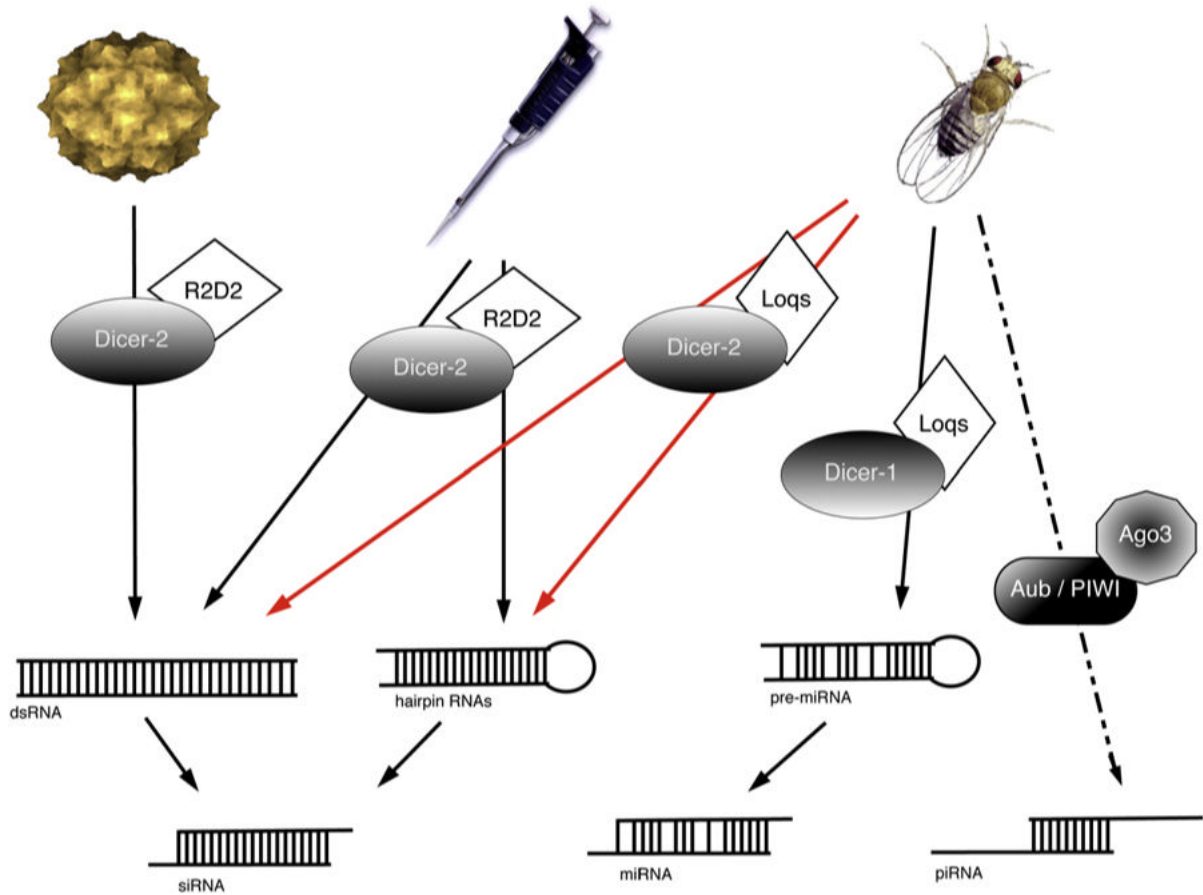


Figure 1. Silencing RNAs Originate from Both Endogenous and Exogenous Sources

SiRNAs from long double-stranded RNAs (dsRNAs) can arise exogenously from viruses such as flock house virus (left) and from experimental introduction (center). In addition, the *Drosophila* genome encodes endogenous miRNAs and piRNAs to regulate gene expression and transposon mobility (right, black arrows). Recent work has revealed numerous endogenous siRNAs processed from structured transcripts, as well as from long dsRNAs derived from convergent transcripts and apparent transposon sense-antisense pairs (right, red arrows). These siRNAs can suppress gene and transposon transcripts. Some of the cofactors required for the biogenesis and function of the respective small RNA classes, including the newly revealed role of Loquacious (Loqs) in the endo-siRNA pathway, are also shown. The biogenesis of piRNAs is poorly understood, as reflected by the dotted line.