Published in final edited form as:
Mol Cell. 2013 September 12; 51(5): . doi:10.1016/j.molcel.2013.08.014.

Bacterial Argonaute samples the transcriptome to identify foreign DNA

Ivan Olovnikov[1,2], **Ken Chan**[1], **Ravi Sachidanandam**[3], **Dianne K. Newman**[1,4], and **Alexei A. Aravin**[1,&]

[1]California Institute of Technology, Division of Biology, 147-75, 1200E California Blvd. Pasadena, CA 91125, USA

[2]Institute of Molecular Genetics, Russian Academy of Sciences, Kurchatov sq. 2, Moscow, 123182, Russia

[3]Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029, USA

[4]Howard Hughes Medical Institute, Division of Biology, 147-75 1200E California Blvd. Pasadena, CA 91125, USA

summary

Eukaryotic Argonautes bind small RNAs and use them as guides to find complementary RNA targets and induce gene silencing. Though homologs of eukaryotic Argonautes are present in many bacteria and archaea their small RNA partners and functions are unknown. We found that the Argonaute of *Rhodobacter sphaeroides* (RsAgo) associates with 15-19 nt RNAs that correspond to the majority of transcripts. RsAgo also binds single-stranded 22-24 nt DNA molecules that are complementary to the small RNAs and enriched in sequences derived from exogenous plasmids as well as genome-encoded foreign nucleic acids such as transposons and phage genes. Expression of RsAgo in the heterologous *E. coli* system leads to formation of plasmid– derived small RNA and DNA and plasmid degradation. In a *R. sphaeroides* mutant lacking RsAgo, expression of plasmid-encoded genes is elevated. Our results indicate that RNAi-related processes found in eukaryotes are also conserved in bacteria and target foreign nucleic acids.

Introduction

Argonaute proteins are key players in RNA interference (RNAi) and related gene silencing phenomena in diverse eukaryotic species. Argonautes form tight complexes with small (19 to 31 nt) RNA partners and use them as guides to recognize target RNA molecules by Watson-Crick base pairing between the small and target RNAs (Meister, 2013). Argonaute proteins harbor an RNaseH-like Piwi domain that is capable of endonucleolytic cleavage of target RNA leading to its subsequent degradation and post-transcriptional silencing. Endonucleolytic activity of Argonautes requires the DEDH tetrad that forms the catalytic center of the Piwi domain (Liu et al., 2004; Nakanishi et al., 2012). In some Argonautes these residues are mutated which correlates with their lack of endonuclease activity.

[&]To whom correspondence should be addressed (aaa@caltech.edu).

Argonautes that are deficient in endonuclease activity can nevertheless induce silencing of complementary RNA targets through recruitment of additional proteins. Particularly, Argonaute cleavage is not required for silencing of the majority of miRNA targets in mammals and Drosophila (Cheloufi et al., 2010; Liu et al., 2004).

Beyond post-transcriptional silencing, Argonaute proteins associated with small RNA partners are able to induce transcriptional gene repression through establishing repressive chromatin structure on cognate genomic loci (Castel and Martienssen, 2013; Olovnikov et al., 2012). In *S. pombe* Argonaute/small RNA complexes directly interact with nascent RNA transcripts followed by recruitment of silencing complexes to chromatin (Verdel et al., 2004). Similarly, the requirement for target transcription was shown or postulated in other cases of small RNA-induced transcriptional repression in plants (Herr et al., 2005) and Metazoa (Le Thomas et al., 2013; Shpiz et al., 2011; Sienski et al., 2012) suggesting that even when Argonaute induce transcriptional repression it binds target RNA and not DNA molecules.

In eukaryotes, several classes of Argonaute-associated small RNA have been described that differ in both their mechanism of biogenesis and cellular functions. Two prominent classes, siRNA and miRNA, are processed by RNaseIII-type enzymes from double-stranded RNA or single-stranded molecules with hairpin structures, respectively. Other types of small RNA, particularly piRNA in Metazoa, are processed from single-stranded precursors that lack recognizable secondary structure without the involvement of RNaseIII-type enzymes (Vagin et al., 2006). Overall, eukaryotes demonstrate a great diversity in small RNA biogenesis mechanisms with some pathways requiring distinct secondary structures of the precursor RNA molecules while others generate small RNA by sampling specific portions of or even the entire transcriptome.

Studies of diverse small RNA pathways in eukaryotic species showed that they play two major roles in cells: regulation of expression of host genes and protection against foreign genetic material. In Metazoa and plants, miRNA and some siRNA target messages of multiple genes to regulate their expression and fine-tune output of gene networks (Baek et al., 2008; Tam et al., 2008). In many species siRNAs repress expression of double-stranded RNA viruses (Li et al., 2002; Lu et al., 2005). siRNAs in plants and piRNAs in Metazoa are involved in repressing endogenous transposable elements that, though they are not infectious, are considered to be ‘selfish’ genomic elements capable of both multiplication and damaging the host genome (Haag and Pikaard, 2011; Herr et al., 2005; Luteijn and Ketting, 2013; Vagin et al., 2006).

Although studies of natural small RNA pathways have been restricted to eukaryotes, Argonaute proteins are also present in many bacterial and archaeal species (Makarova et al., 2009). Furthermore, the structure of several bacterial and archaeal Argonautes was solved and used for modeling the structure of their eukaryotic orthologs (Ma et al., 2005; Parker et al., 2005; Song et al., 2004; Wang et al., 2008b; Yuan et al., 2005). *In vitro* the Argonaute protein of the eubacterium *Aquifex aeolicus* has the highest affinity to single-stranded DNA oligonucleotides while binding affinities to single- and double-stranded RNA are significantly lower (Ma et al., 2005). Similarly, Argonaute from the archaeon *Archaeoglobus fulgidus* binds single- and double-stranded short DNA and DNA-RNA hybrids more efficiently than single- and double-stranded RNA (Yuan et al., 2005). Furthermore, *in vitro*, bacterial members of the Argonaute family that have slicer activity are able to use nucleic acid guides (preferentially DNA) for endonucleolytic cleavage of complementary RNA targets (Wang et al., 2008a; Wang et al., 2008b; Yuan et al., 2005). The bioinformatic analysis of Ago operons and their relationship with other genome defense systems suggested that bacterial and archaeal Argonautes might play a role in protection of

the host genome (Makarova et al., 2009), however, experimental analysis of Argonaute functions and their native small RNA partners in either bacteria or archaea was lacking.

Here we identified the natural nucleic acid partners of Argonaute protein (RsAgo) from the alphaproteobacterium *Rhodobacter sphaeroides*. Surprisingly, we found that in native cells RsAgo associates with a complex population of both small RNA and small DNA species. Analyses of nucleic acid partners of RsAgo suggest that associated small RNAs are generated by broad sampling of the bacterial transcriptome, and formation of small DNA is dependent on small RNAs. RsAgo-associated small DNAs are enriched in sequences derived from foreign nucleic acids and RsAgo deficiency leads to increased expression of plasmid-encoded genes implicating a role for bacterial Argonaute in the pathway that targets foreign nucleic acids.

Results

Identification of nucleic acid partners of *Rhodobacter sphaeroides* Argonaute

To identify the natural nucleic acid partners of bacterial Argonautes *in vivo* we expressed and purified 6xHis-tagged Argonaute protein from host strain *R. sphaeroides* ATCC17025. RsAgo is a 777 amino-acid protein and has the same domain structure as eukaryotic Argonautes, namely PAZ, Mid and Piwi domains responsible for binding small nucleic acid guides and endonucleolytic cleavage (slicing) of the target (Fig. 1A). However, like the majority of prokaryotic Argonautes, RsAgo is likely not active as a slicer nuclease due to a substitution of the critical DEDH residues responsible for the nuclease activity to GGHE residues (Makarova et al., 2009). Similar to many other bacterial Argonautes, RsAgo resides within one operon with a predicted DNA nuclease (Fig. 1A) (Makarova et al., 2009).

Tagged RsAgo complex purified from *R. sphaeroides* cells harvested at stationary phase contains small nucleic acid species with major sizes around 15-24 nt and ~45 nt (Fig 1B, S1). The majority of the RsAgo-bound nucleic acids are eliminated by RNase A treatment indicating that these are small RNAs. A minor fraction of 20-24 nt species was resistant to RNase A treatment but susceptible to DNase I, indicating that RsAgo associates with short DNA molecules. To gain further insight into the nature of Ago-associated nucleic acids, we cloned and sequenced two size fractions of RsAgo-bound RNAs (15-24 nt and ~ 45 nt) using an established method that relies on the presence of a 5′ phosphate and 3′ hydroxyl in the cloned molecules (Aravin et al., 2007; Lau et al., 2001). Sequence analysis of 45 nt long RsAgo-associated RNA showed that this fraction consists almost exclusively of the 3′ halves of two types of $tRNA^{met}$ cut 1nt upstream of the anticodon (data not shown). It is difficult to determine the biological significance of co-purification of RsAgo with tRNA halves; this association could be caused by contamination by abundant RNA species, though it should be noted that mammalian Argonaute was also found in complex with tRNA fragments (Burroughs et al., 2011; Maute et al., 2013). Sequence analysis of the shorter (15-24 nt) RNA fraction revealed RNA population between 15 and 19 nt in length (Fig. 2A). These small RNAs show strong enrichment for uridine residue at the first position (Fig. 2B), a distinctive feature of several classes of eukaryotic Argonaute-associated small RNAs (Brennecke et al., 2007; Lau et al., 2001). A strong enrichment of a pyrimidine (U or C) residue at position 2 was also apparent. The RsAgo-bound smRNA population was moderately complex: the roughly 35 million sequenced reads correspond to about 1.7 million unique sequences (Table S1). The diversity of the 15-19 nt small RNA library is corroborated by the absence of nucleotide biases at any position except of the first and second nucleotides.

Presence of small RNA population in wild-type cells depends on RsAgo expression

Purification of Ago-RNA complexes from bacterial cell lysate takes several hours during which cellular RNA is exposed to nucleases that might generate small RNA fragments. These could possibly be loaded into RsAgo (Riley et al., 2012). To verify that RsAgo-bound small RNA species are present in native cells we purified total RNA from two *R. sphaeroides* strains: ATCC17025 (referenced as 'strain 25' later in the text) that expresses Argonaute endogenously and ATCC17029 (referenced as 'strain 29') that lacks the gene encoding Ago. We also transformed both strains with a plasmid expressing RsAgo or an empty control plasmid lacking Ago. Deep sequencing of the 13-30 nt range of small RNA from non-transformed strain 25 revealed a prominent peak at 15-19 nt corresponding to the size of RsAgo-associated small RNAs (Fig. 2C). Importantly, this peak was absent in strain 29 that lacks endogenous Argonaute protein. Analysis of sequence composition showed that while small RNAs cloned from strain 29 do not show any bias for uridine at the first base position, small RNAs from strain 25 are strongly enriched in 5′U (Fig. 2D). Expression of RsAgo from the plasmid further increased the bias for 5′U in 15-19 nt RNA in strain 25. These results indicate that an endogenous small RNA population with features identical to purified RsAgo-bound small RNA is present in the *R. sphaeroides* strain that expresses non-tagged RsAgo on a physiological level. Interestingly, expression of RsAgo from the plasmid in strain 29 lacking the whole Ago operon leads to the appearance of 15-19 nt 5′U-rich RNA in the total RNA population (Fig. 2C,D). Furthermore, expression of RsAgo significantly changes the small RNA profile in strain 29: without RsAgo expression, 50% of small RNA in the size range of 13 to 30 nt are mapped to rRNA genes and likely represent degradation products of abundant rRNA species, compared to only 15% after RsAgo expression. This result suggests that expression of RsAgo changes the small RNA profile by binding to a select pool of small RNAs and that the putative nuclease that is encoded by the second gene in the operon (Rsph17025_3695) is not required for generation of RsAgo-bound small RNA.

RsAgo-associated small RNAs match the majority of cellular transcripts

To further analyze RsAgo-associated small RNAs, we mapped them to the *R. sphaeroides* genome (strain 25) and the pSRKKm-RsAgo plasmid used for protein expression. Small RNAs mapped to all six *R. sphaeroides* circular chromosomes as well as the expression plasmid with 54.3% of genomic and 98.7% of plasmid nucleotides covered by at least one read. 60.3% of the small RNA reads map to the largest 3.2 Mb chromosome, while 24.4% of RNA reads map to the 8 kb expression plasmid (Fig. 3A). Normalized to chromosome length RsAgo-bound small RNAs show a strong bias for the expression plasmid. This enrichment of plasmid-derived sequences was also clear in strain 25 transformed with the 'empty' pSRKKm plasmid lacking the RsAgo gene, demonstrating that it is not caused by the propensity of Argonaute to generate small RNAs in cis (Fig. S2A,B).

Annotation of small RNAs shows that the majority map to protein-coding and non-coding genes in sense orientation (Fig. 3B). The distribution of small RNAs over genes is not uniform: the top 10 genes, including three located on the expression plasmid (RsAgo itself, Kan and lacI), contribute 37% of all RNA reads. Among the top genes located on the host chromosomes were rRNAs and the IS4 family of DNA transposons that has more than 20 nearly identical copies scattered over the six chromosomes (Fig. 3C). To understand the relationship between the RsAgo-bound small RNA population and normal cellular transcripts, we analyzed the transcriptome of strain 25 using RNA-Seq of total and rRNA-depleted RNA samples. Comparison of long RNA transcripts and short RsAgo-bound RNAs shows a significant but relatively weak correlation between gene expression and the amount of small RNAs mapping to each gene (Pearson correlation coefficient $r \sim 0.35$, $p < 2.2\times10^{-16}$, Fig. 3D). Many genes are greatly depleted or enriched in the small RNA population relative to the amount of their longtranscripts (Fig. 3E). Interestingly, the RsAgo

gene expressed from the plasmid was among the top 30 genes with the ratio of small to long RNA 30-fold higher than median. In contrast, non-coding RNAs, such as RNaseP and SRP were among the 10 most depleted genes with the ratio of small to long RNA being 246- and 15,500-fold lower than median (Fig. 3E, Fig.S2C).

The most parsimonious interpretation of the sense bias of small RNAs and the general correlation between their amount and that of long genic transcripts is that small RNAs are processed from full-length mRNAs or their degradation products by nuclease activity. However, the significant variation in the ratio of small to long RNA indicates that the efficiency of such processing varies for different transcripts. Except for the depletion of non-coding RNA genes and enrichment for plasmid-encoded genes (Fig. 3E), we were unable to identify distinct gene features that could explain their enrichment or depletion in the small RNA population. RsAgo-associated small RNAs are derived from many different transcripts and many different positions inside each transcript. We did not detect any strong sequence or structure motif upstream or downstream of mature small RNAs in potential precursor transcripts, implying that there is no requirement for distinct primary or secondary structure for processing of small RNA from long precursors (data not shown). Overall, our results show that the population of RsAgo-bound small RNAs represents a sampling of the whole bacterial transcriptome with significant enrichment for individual transcripts and selection against others such as non-coding RNAs.

RsAgo-associated small DNAs are complementary to small RNAs and gene transcripts

To analyze the small DNA species associated with RsAgo we developed a method for directional cloning of short single-stranded DNA molecules using bridged ligation that is dependent on the presence of a 5′ phosphate and 3′ hydroxyl groups in the cloned molecule (Fig. S3). Analysis of small DNA deep sequencing data revealed that they are 22-24 nt long (Fig. 2A). The complexity of the small DNA population is comparable to that of small RNAs: ~16 million sequenced reads corresponds to ~1.1 million unique sequences. Similar to small RNA, RsAgo-bound small DNA map to all *R. sphaeroides* circular chromosomes and the expression plasmid, but the bias for the expression plasmid is even more pronounced: 87.6% of DNA reads map to the 8 kb expression plasmid, while 9.1% map to the largest 3.2 Mb chromosome (Fig. 4A).

Annotation of small DNAs shows that, similar to small RNAs, the majority map to protein-coding genes (Fig. 4B). A significant portion of small DNAs mapped to non-annotated loci; the majority of these sequences match non-coding regions of the expression plasmid. The orientation of small DNAs is opposite to that of small RNA species: for plasmid-mapped sequences 95% of small DNAs are in antisense orientation to genes, while 97% of small RNAs are in the sense orientation (Fig. 4C, D). A similar bias for antisense orientation is apparent in small DNAs mapped to host chromosomes (Fig. S4B). The strong orientation bias of small DNAs that map to the genomic strand corresponding to the antisense strand of genes suggests that they are generated in an RNA- or, more generally, transcription-dependent fashion.

To gain further insight into the relative positions of small RNA and DNA molecules we performed genome-wide analysis of the distances between the ends of small RNAs and small DNAs that mapped within 30 nt of each other. This analysis revealed a striking enrichment for complementary RNA/DNA pairs with both ends of DNA protruding 3 nt over small RNA sequence (Fig. 4E). Interestingly, while the 3′ DNA overhang is almost invariably 3 nt, the length of the 5′ DNA overhang is less precise (Fig. 4E), suggesting that the mechanisms of 5′ and 3′ end formation are different and specific. Small DNA species did not show significant nucleotide biases in the first and second residues where small RNAs had strong biases. However, when the 23-24 nt small DNA sequences were aligned

by their 3′ end, a strong enrichment for adenine and purine residues was seen 4 and 5 nt away from the 3′ end, respectively (Fig. 4F, Fig. S4C). This result supports the analysis of relative positions of small RNA and DNA molecules that indicates that the majority of small RNAs and small DNAs are complementary to each other with the small DNA having a 3 nt overhang on each end (Fig. 4G). Indeed, mapping has shown that 79.9% of small RNA reads are fully complementary to at least one sequence in small DNA library. To emphasize this relationship, we propose to refer to single-stranded small RNA and DNA such as found in the RsAgo complex as DNA-interacting (di)RNA and RNA-interacting (ri)DNA, respectively. Considering that diRNA are ~2-fold more abundant in purified RsAgo complexes than riDNA (Fig. 1B) it is clear that some RsAgo complexes are loaded with small RNA alone. Other complexes might be loaded with riDNA or a diRNA-riDNA duplex.

RsAgo-associated small DNAs are enriched in foreign sequences

The amount of riDNA mapped to each gene correlates with the amount of diRNA (Pearson correlation coefficient $r \sim 0.73$, $p < 2.2\times10^{-16}$, Fig. S5A,B). However there are significant variations in the ratio of riDNA to diRNA for individual genes: for example the five genes encoded on the expression plasmid have on average 15-fold higher ratio compared to median. Analysis of genes with high ratio of riDNA/diRNA shows that this set is enriched in sequences that are annotated as transposons and phage genes (Fig. 5A). Many, but not all of these genes are present in multiple copies in the *R. sphaeroides* genome. Similarly, genes located on the expression plasmid that have 30 to 60 copies per cell and genes of unknown origin that are present as multi-copy sequences in the genome are enriched in riDNA (Fig. 5A). To test if the disproportionally large amount of riDNA derived from some genes is caused by greater availability of substrate DNA due to its high copy-number, we normalized the amount of small DNA to the number of positions to which each sequence can be mapped in the genome. Even after this normalization, genes with a high ratio of riDNA to diRNA are enriched in phage genes and multi-copy genes of unknown origin (Fig. S5C). Importantly, even single-copy phage genes have high ratio of riDNA to diRNA, while multi-copy host genes do not have this bias. Average phage genes have a ratio of riDNA to diRNA that is 6.6- or 5-fold higher than that of an average single-copy host gene, before and after normalization to copy number, respectively (Fig. 5B, Fig.S5D). Together these results suggest that sequences that can be classified as ‘foreign DNA’ and not simply multi-copy sequences generate disproportionally more riDNA. Considering that ~90% of the riDNA map to the expression plasmid, it is plausible that discrimination of foreign DNA against host genes is achieved by recognition of extra-chromosomal DNA molecules as a substrate for small DNA processing. In agreement with this hypothesis, transposons and DNA phages that show enrichment in riDNA have extra-chromosomal DNA stages in their life cycles.

Expression of RsAgo in *E. coli* causes degradation of plasmid DNA

In order to further understand the ability of RsAgo to recognize foreign nucleic acids we expressed it in the heterologous *E. coli* system. As a control we also expressed the point mutant, denoted as RsAgo-YK, with two amino acid substitutions in the 5′ end binding pocket that should render it unable to bind 5′ phosphorylated nucleic acids (Ma et al., 2005; Parker et al., 2005). Purification of wild-type RsAgo from *E. coli* cells showed that it binds both small RNA and DNA species of the same sizes as in *R. sphaeroides* (Fig. 6A) suggesting that either RsAgo is autonomous in producing and loading small nucleic acids or it requires factors that are conserved between the two species. As expected, the RsAgo-YK mutant was devoid of both small RNA and DNA species (data not shown). Cloning of small RNA and DNA isolated from RsAgo expressed in *E. coli* showed that they are almost exclusively derived from the expression plasmid (90.2% and 99.2% of small RNA and DNA reads, respectively) (Fig. 6B). Induction of expression of the wild-type protein, but not the

RsAgo-YK mutant, leads to significant decrease in plasmid yield (Fig. 6C, S6). Furthermore, plasmid DNA isolated from cells expressing wild-type RsAgo was strongly degraded (Fig. 6D) suggesting that the riDNAs found in the RsAgo complex are generated by excision from plasmid DNA.

RsAgo represses plasmid-encoded genes in *R. sphaeroides*

To gain further insight into the function of *R. sphaeroides* Argonaute, we used directional insertion of the omega transcription termination cassette upstream of the Piwi domain of RsAgo to generate a mutant of strain 25 deficient in RsAgo expression (Fig. S11). RT-PCR of the RsAgo transcript confirmed the lack of RsAgo expression in mutant cells (Fig. S7). Furthermore, cloning of small RNAs from RsAgo-deficient cells showed the absence of the 15-19 nt peak, which is present in wild-type 25 cells indicating that RsAgo protein is indeed functionally impaired (Fig. 7A). Measuring growth of the wild type and RsAgo-deficient cells under standard conditions did not reveal significant differences between the two strains indicating that RsAgo is not required for general cellular fitness (Fig. S7D). To study the effect of RsAgo deficiency on gene expression, we used RNA-Seq to profile the transcriptome of wild-type and mutant cells grown under identical conditions. Duplicate experiments analyzed by the DESeq (Anders and Huber, 2010) showed no significant reproducible difference in gene expression between the two strains (Fig. S7E). Taken together, these data indicate that despite being loaded with small RNAs derived from a variety of chromosomal transcripts, RsAgo does not strongly affect gene expression. RsAgo-associated diRNA and riDNA are particularly enriched in sequences derived from the exogenously introduced plasmid (Fig. 3A, 4A) suggesting that RsAgo might specifically affect expression of genes encoded on foreign extrachromosomal DNA. To test this hypothesis we introduced the pSRKTc-Fluc plasmid encoding the firefly luciferase and lacI genes into wild-type strain 25 and RsAgo mutant cells. Measurement of luciferase activity in three independent experiments showed that its expression was ~ 2-fold higher in the RsAgo mutant cells (Fig. 7B). Furthermore, abundance of both luciferase and lacI plasmid-derived transcripts were also ~2-fold higher in mutant cells as measured by RT-qPCR, while the amount of plasmid DNA measured by PCR was identical (Fig. 7B). These data indicate that RsAgo is able to repress gene expression from an exogenous plasmid.

Discussion

The finding of nucleic acid partners of bacterial Ago protein sheds new light on the evolution and function of RNAi pathways. RsAgo-bound small RNAs (diRNAs) are shorter (15-19 nt) compared to all known classes of Argonaute-associated RNA in eukaryotes (19-31 nt). However, similar to many eukaryotic small RNAs, RsAgo-bound RNAs have a strong bias for uridine at the first position, a feature that is likely determined by the structure of the 5′ binding pocket of the Argonaute (Frank et al., 2012). Unlike miRNAs and siRNAs in eukaryotes, RsAgo-associated diRNAs are derived from RNA precursors that lack distinct secondary structure. This fact, together with a remarkable sequence diversity of RsAgo-bound RNAs, make them similar to another class of eukaryotic RNA, piRNA that is generated from single-stranded non-structured RNAs of any sequence (Brennecke et al., 2007; Muerdter et al., 2012). Indeed, RsAgo-associated diRNAs can be mapped to the majority of cellular transcripts in sense orientation with a bias against structural non-coding RNAs. Therefore diRNAs are either directly processed from mRNAs or generated from the products of their degradation (Fig. 7C). The discrimination against structural RNA favors the latter explanation, sampling of degradation products (so called degradome), as structural RNAs are significantly more stable compared to mRNAs. The strong bias for uridine at the first position might be introduced by the processing machinery, or, alternatively, the result of RsAgo selecting 5′U-RNAs from a pool of available sequences. diRNAs with correct size

and 5′U-bias are formed upon RsAgo expression in the heterologous *E. coli* system arguing against involvement of species-specific proteins in small RNA processing. Overall, generation of diRNAs can be explained without postulating the existence of a complex processing machinery. In the simplest scenario, RsAgo would bind products of mRNA degradation that possess 5′ phosphate. Subsequently, a nuclease would trim the bound RNA from the 3′ end until it reaches the region protected by the footprint of the RsAgo protein, the mechanism proposed for formation of the 3′ end of Metazoan piRNAs (Kawaoka et al., 2011).

The finding of small single-stranded DNA (riDNA) species associated with RsAgo is very intriguing as small DNAs have not been found in association with Argonaute proteins in eukaryotes. Natural short single-stranded DNAs, to our knowledge, have not been found in any cellular pathway, though short DNA sequences, Okazaki fragments, paired with long DNA are synthesized during replication of the lagging DNA strand. riDNAs found in association with RsAgo complex have two distinct and surprising features. First, they are largely complementary to diRNAs and gene transcripts. Second, the riDNAs have a peculiar arrangement with the complementary diRNAs: they form duplexes that have 3 nt DNA overhang on each side of the duplex (Fig. 4G). Both features strongly argue that riDNAs are generated in a RNA-dependent fashion as only transcribed RNA can provide such strand information. It is unlikely that riDNAs are generated by direct reverse transcription of riRNAs as DNAs are longer and harbor nucleotide residues that match genomic sequences but are lacking in complementary RNA molecules (Fig. 4G).

The observed arrangement of diRNAs and riDNAs resembles the ping-pong cycle in the piRNA pathway where formation of secondary piRNAs is guided by primary piRNA (Brennecke et al., 2007). However, the ping-pong cycle uses the endonucleolytic (slicer) activity of Argonaute to generate secondary piRNA. As a result of slicer cleavage, the 5′ end of the secondary piRNA is shifted by 10 nt compared to the 5′ end of the primary piRNA. The residues necessary for slicer activity are mutated in RsAgo and we did not detect cleavage activity in an *in vitro* assay (data not shown). Equally important, observed arrangement of diRNA/riDNA pairs is not compatible with Argonaute-dependent cleavage. Instead we propose that recognition of DNA by small RNA-loaded RsAgo leads to DNA cleavage by an unknown nuclease(s) on both sides of the targeted region leading to formation of riDNA with 3 nt overhands (Fig. 7C). Indeed, we observe strong degradation of the plasmid upon expression of RsAgo in *E. coli* cells that correlates with production of small DNAs mapping almost exclusively to the plasmid. Argonautes and associated small RNA can cause transcriptional silencing and even DNA elimination of cognate sequences in eukaryotes (Kataoka and Mochizuki, 2011), however, it is believed that Argonaute complexes target nascent transcripts followed by recruitment of chromatin-modifying machinery to the locus. In contrast, our model proposes direct targeting of DNA by RsAgo/small RNA complexes.

Generation of riDNAs by cutting them out of double-stranded DNA will lead to damage of genomic DNA unless it is tightly controlled and restricted to specific sequences. Indeed, we observed that riDNA are highly enriched in sequences that can be characterized as foreign to the cell: artificially introduced plasmid as well as phages and transposons. Therefore, we postulate the existence of an additional mechanism that allows small RNA-loaded RsAgo complexes to discriminate and target foreign DNA while avoiding normal genomic substrates. The specific molecular determinant that allows such discrimination should be a subject for future studies. The extra-chromosomal nature of DNA can be used to discriminate proper targets: indeed plasmids are extrachromosomal by definition and both phages and transposons go through an extrachromosomal step during their life-cycle. However, the extra-chromosomal nature by itself is not sufficient to explain the observed

specificity as endogenous plasmids similar in size to the artificially introduced plasmid do not serve as preferred substrate for riDNA generation (Fig. 4A). High rate of DNA replication might be a feature for recognition of proper targets as replicating DNA provides readily available single-stranded DNA substrate for recognition by small RNA-loaded RsAgo. According to this model, effective RsAgo targeting requires both a high level of gene expression (to generate small RNA) and a high level of DNA replication (to provide single-stranded DNA target). Such a combination is not typical for host genes and would allow RsAgo to discriminate against invasive DNA.

The enrichment of riDNAs in foreign sequences implicates RsAgo in repression of foreign genomic elements. Indeed, expression of genes encoded on a foreign (but not endogenous) plasmid is suppressed by RsAgo (Fig. 7B). What is the molecular mechanism of gene silencing by RsAgo? Though we did not observe strong degradation of plasmid DNA at the physiological expression level of RsAgo in *R. sphaeroides* cells (data not shown), it is possible that a low level of plasmid DNA damage associated with excision of riDNAs is sufficient to cause moderate repression (Fig. 7C). Alternatively, the repression might be achieved by transcriptional interference as elongating RNA polymerase collides with RsAgo/diRNA complexes that are bound to the template strand of the plasmid DNA. According to both models, that are not mutually exclusive, riDNAs are by-products of a rather rare event of DNA degradation and do not play an independent role in the silencing process. Alternatively, after their generation antisense riDNA might be loaded into RsAgo and repress complementary plasmid transcripts post-transcriptionally (Fig. 7C). Unlike Argonautes of *Aquifex aeolicus* and *Thermus thermophilus* that are able to use DNA guides to cleave RNA in vitro (Wang et al., 2008a; Yuan et al., 2005), RsAgo lacks slicer activity and would need to employ different mechanism. Independently of the specific molecular mechanism, which future studies will address, our findings suggest that bacterial Argonautes and their nucleic acid partners play a function similar to that of CRISPR, metazoan piRNA and plant siRNA, namely, protection of the genome against invading genetic material.

Experimental Procedures

Bacterial strains

Rhodobacter sphaeroides strains ATCC17025 and ATCC17029 were kindly provided by Timothy Donohue (University of Wisconsin–Madison). Cells were grown on Sistrom's minimal medium A at 30C under aerobic conditions. Kanamycin was used at concentration 25 μg/ml for *R. sphaeroides* and 25 μg/ml for E. coli BL21(DE3), Tetracyclin at 1 μg/ml for *R. sphaeroides*.

RsAgo expression and purification in R. sphaeroides

N-6xHis-tagged ORF of RsAgo (Rsph17025_3694) was amplified from genomic DNA of strain ATCC17025 and cloned into broad-host-range expression vector pSRKKm (Khan et al., 2008). Plasmid was mobilized into *R. sphaeroides* by biparental mating with *E. coli* BW29427. Protein synthesis was induced by 1mM IPTG for 5-10 hours. RsAgo was isolated using Talon beads (Clontech). A detailed protocol is provided in the Supplemental Experimental Procedures.

RsAgo expression in E. coli

For expression of RsAgo in E. coli BL21(DE3) vector pSRKKm-RsAgo was used. For experiments shown in Fig. 6 C, D and Fig. S10, 6xHis-Flag-tagged RsAgo was cloned into vector pET30a(+), which contains T7 RNA polymerase-driven promoter. RsAgo-YK isoform impaired in small RNA binding contained mutations Y463G and K467G in the

RsAgo MID domain. As a control Flag-tagged GFP was cloned into pET30a(+). Protein synthesis was induced with 1mM IPTG for 5 hours at 37C.

Small RNA and DNA isolation and sequencing

For total small RNA cloning (13-30 nt range) RNA was isolated using the Amresco Phenol-Free Total RNA Purification Kit after fixation of the cell culture with Ambion RNAlater reagent. Small RNA and DNA species were extracted from purified RsAgo complex using proteinase K treatment followed by neutral phenol:chloroform extraction. Small RNAs were cloned according to published protocol (Brennecke et al., 2007; Lau et al., 2001) that requires the presence of 5′ phosphate and 3′ hydroxyl termini in RNA using linkers and primers from the Illumina TrueSeq Small RNA Sample Prep kit. Half of the sample of RsAgo-bound small RNA was radioactively labeled using sequential treatment with CIP and PNK prior to ligation of the adapters allowing cloning of RNA molecules with other termini. No such treatment was used during cloning of small RNAs from total RNA samples.

To clone small DNA we used bridged (splinted) ligation approach (Fig. S4). A simultaneous 5′ and 3′ linker ligation reaction was performed in 15 μl volume and contained varying amounts of small DNA (1:3-1:5 mixture of 5`-$^{32}$P labeled: non-labeled DNA), 100 pmoles of 5` and 3` linker and 5` and 3` bridge oligonucleotides, 5% PEG8000, 1x T4 DNA ligase buffer (50 mM Tris-HCl, 10 mM MgCl2,1 mM ATP, 10 mM DTT, pH 7.5) and 1 μl of T4 DNA ligase (NEB, 400,000 units/ml). The reaction was incubated at room temperature for 1 to 10 hours, followed by electrophoresis in 12% urea-PAGE. Standard Illumina primers were used to create indexed libraries.

Details on cloning procedures are given in the Supplemental Experimental Procedures.

Generation of RsAgo mutant strain ATCC17025

To create the mutagenesis vector a 1 kb homology arm corresponding to 5` half of RsAgo CDS was joined with an omega transcription termination cassette (Prentki and Krisch, 1984) and cloned into the suicide vector pK18*mobsacB*, containing kanamycin resistance gene (Schafer et al., 1994) (Fig. S11). After conjugation cells were selected on kanamycin and correct insertion was verified with PCR.

Long RNA sequencing and analysis

For regular RNA sequencing samples were processed according to the Illumina TrueSeq RNA prep kit. rRNA depletion was performed using the RiboZERO gram-negative bacterial rRNA depletion kit (EpiBio). To profile the transcriptomes of wild-type and RsAgo mutant strain 25, duplicate RNA-Seq libraries were prepared from rRNA-depleted RNA isolated from two independent experiments.

Sequence analysis

For sequence analysis of small RNA and DNA we used the small RNA dashboard server (Olson et al., 2008) and Galaxy tools (Blankenberg et al., 2010).

Accession numbers

RNA-seq, small RNA and small DNA data were deposited in the Gene Expression Omnibus database under accession number ###.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Katalin Fejes Tóth and members of the Aravin lab for helpful discussion and comments on the manuscript. We thank Chia-Hung Wu, Shannon Park and Gargi Kulkarni from Dianne Newman's lab for guidance with bacterial experiments and Konstantin Piatkov for helpful suggestions. We thank Timothy Donohue (University of Wisconsin–Madison), Shulin Chen (WSU), Rebecca Parales (UC, Davis) and Carlos Ríos-Velázquez (UPR-Mayagüez) for provided reagents and bacterial strains. We are grateful to Todd Lowe (UCSC) for providing access to the genome browser. We thank Sailakshmi Subramanian (Mount Sinai), Georgi Marinov and Sergei Manakov (Caltech) for help with bioinformatic and statistical analysis and Igor Antoshechkin (Caltech) for help with RNA sequencing. We thank Gregory Donaldson and Vishnu Manoranjan for assistance with experiments. IO is a CEMI (Center for Environmental Microbiology Interactions) fellow at Caltech. This work was supported by grants from the National Institutes of Health (R01 GM097363, R00 HD057233 and DP2 OD007371A) and the Searle Scholar Award to AAA. DKN is an Investigator of the Howard Hughes Medical Institute.

References

Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106. [PubMed: 20979621]

Aravin AA, Sachidanandam R, Girard A, Fejes-Toth K, Hannon GJ. Developmentally regulated piRNA clusters implicate MILI in transposon control. Science. 2007; 316:744–747. [PubMed: 17446352]

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. The impact of microRNAs on protein output. Nature. 2008; 455:64–71. [PubMed: 18668037]

Blankenberg D, Von Kuster G, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. Galaxy: a web-based genome analysis tool for experimentalists. Curr Protoc Mol Biol. 2010; Chapter 19:11–21. Unit 19 10.

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. Cell. 2007; 128:1089–1103. [PubMed: 17346786]

Burroughs AM, Ando Y, de Hoon MJ, Tomaru Y, Suzuki H, Hayashizaki Y, Daub CO. Deep-sequencing of human Argonaute-associated small RNAs provides insight into miRNA sorting and reveals Argonaute association with RNA fragments of diverse origin. RNA biology. 2011; 8:158–177. [PubMed: 21282978]

Castel SE, Martienssen RA. RNA interference in the nucleus: roles for small RNAs in transcription, epigenetics and beyond. Nature reviews Genetics. 2013; 14:100–112.

Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. Nature. 2010; 465:584–589. [PubMed: 20424607]

Frank F, Hauver J, Sonenberg N, Nagar B. Arabidopsis Argonaute MID domains use their nucleotide specificity loop to sort small RNAs. The EMBO journal. 2012; 31:3588–3595. [PubMed: 22850669]

Haag JR, Pikaard CS. Multisubunit RNA polymerases IV and V: purveyors of non-coding RNA for plant gene silencing. Nature reviews Molecular cell biology. 2011; 12:483–492.

Herr AJ, Jensen MB, Dalmay T, Baulcombe DC. RNA polymerase IV directs silencing of endogenous DNA. Science. 2005; 308:118–120. [PubMed: 15692015]

Kataoka K, Mochizuki K. Programmed DNA elimination in Tetrahymena: a small RNA-mediated genome surveillance mechanism. Advances in experimental medicine and biology. 2011; 722:156–173. [PubMed: 21915788]

Kawaoka S, Izumi N, Katsuma S, Tomari Y. 3′ end formation of PIWI-interacting RNAs in vitro. Molecular cell. 2011; 43:1015–1022. [PubMed: 21925389]

Khan SR, Gaines J, Roop RM 2nd, Farrand SK. Broad-host-range expression vectors with tightly regulated promoters and their use to examine the influence of TraR and TraM expression on Ti plasmid quorum sensing. Appl Environ Microbiol. 2008; 74:5053–5062. [PubMed: 18606801]

Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25. [PubMed: 19261174]

Lau NC, Lim LP, Weinstein EG, Bartel DP. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. Science. 2001; 294:858–862. [PubMed: 11679671]

Le Thomas A, Rogers AK, Webster A, Marinov GK, Liao SE, Perkins EM, Hur JK, Aravin AA, Toth KF. Piwi induces piRNA-guided transcriptional silencing and establishment of a repressive chromatin state. Genes & development. 2013; 27:390–399. [PubMed: 23392610]

Li H, Li WX, Ding SW. Induction and suppression of RNA silencing by an animal virus. Science. 2002; 296:1319–1321. [PubMed: 12016316]

Liu J, Carmell MA, Rivas FV, Marsden CG, Thomson JM, Song JJ, Hammond SM, Joshua-Tor L, Hannon GJ. Argonaute2 is the catalytic engine of mammalian RNAi. Science. 2004; 305:1437–1441. [PubMed: 15284456]

Lu R, Maduro M, Li F, Li HW, Broitman-Maduro G, Li WX, Ding SW. Animal virus replication and RNAi-mediated antiviral silencing in Caenorhabditis elegans. Nature. 2005; 436:1040–1043. [PubMed: 16107851]

Luteijn MJ, Ketting RF. PIWI-interacting RNAs: from generation to transgenerational epigenetics. Nature reviews Genetics. 2013; 14:523–534.

Ma JB, Yuan YR, Meister G, Pei Y, Tuschl T, Patel DJ. Structural basis for 5′-end-specific recognition of guide RNA by the A. fulgidus Piwi protein. Nature. 2005; 434:666–670. [PubMed: 15800629]

Makarova KS, Wolf YI, van der Oost J, Koonin EV. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. Biology direct. 2009; 4:29. [PubMed: 19706170]

Maute RL, Schneider C, Sumazin P, Holmes A, Califano A, Basso K, Dalla-Favera R. tRNA-derived microRNA modulates proliferation and the DNA damage response and is down-regulated in B cell lymphoma. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:1404–1409. [PubMed: 23297232]

Meister G. Argonaute proteins: functional insights and emerging roles. Nature reviews Genetics. 2013; 14:447–459.

Muerdter F, Olovnikov I, Molaro A, Rozhkov NV, Czech B, Gordon A, Hannon GJ, Aravin AA. Production of artificial piRNAs in flies and mice. RNA. 2012; 18:42–52. [PubMed: 22096018]

Nakanishi K, Weinberg DE, Bartel DP, Patel DJ. Structure of yeast Argonaute with guide RNA. Nature. 2012; 486:368–374. [PubMed: 22722195]

Olovnikov I, Aravin AA, Fejes Toth K. Small RNA in the nucleus: the RNA-chromatin ping-pong. Current opinion in genetics & development. 2012; 22:164–171. [PubMed: 22349141]

Olson AJ, Brennecke J, Aravin AA, Hannon GJ, Sachidanandam R. Analysis of large-scale sequencing of small RNAs. Pac Symp Biocomput. 2008:126–136. [PubMed: 18229681]

Parker JS, Roe SM, Barford D. Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. Nature. 2005; 434:663–666. [PubMed: 15800628]

Prentki P, Krisch HM. In vitro insertional mutagenesis with a selectable DNA fragment. Gene. 1984; 29:303–313. [PubMed: 6237955]

Riley KJ, Yario TA, Steitz JA. Association of Argonaute proteins and microRNAs can occur after cell lysis. RNA. 2012; 18:1581–1585. [PubMed: 22836356]

Schafer A, Tauch A, Jager W, Kalinowski J, Thierbach G, Puhler A. Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. Gene. 1994; 145:69–73. [PubMed: 8045426]

Shpiz S, Olovnikov I, Sergeeva A, Lavrov S, Abramov Y, Savitsky M, Kalmykova A. Mechanism of the piRNA-mediated silencing of Drosophila telomeric retrotransposons. Nucleic acids research. 2011; 39:8703–8711. [PubMed: 21764773]

Sienski G, Donertas D, Brennecke J. Transcriptional silencing of transposons by Piwi and maelstrom and its impact on chromatin state and gene expression. Cell. 2012; 151:964–980. [PubMed: 23159368]

Song JJ, Smith SK, Hannon GJ, Joshua-Tor L. Crystal structure of Argonaute and its implications for RISC slicer activity. Science. 2004; 305:1434–1437. [PubMed: 15284453]

Tam OH, Aravin AA, Stein P, Girard A, Murchison EP, Cheloufi S, Hodges E, Anger M, Sachidanandam R, Schultz RM, et al. Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. Nature. 2008; 453:534–538. [PubMed: 18404147]

Vagin VV, Sigova A, Li C, Seitz H, Gvozdev V, Zamore PD. A distinct small RNA pathway silences selfish genetic elements in the germline. Science. 2006; 313:320–324. [PubMed: 16809489]

Verdel A, Jia S, Gerber S, Sugiyama T, Gygi S, Grewal SI, Moazed D. RNAi-mediated targeting of heterochromatin by the RITS complex. Science. 2004; 303:672–676. [PubMed: 14704433]

Wang Y, Juranek S, Li H, Sheng G, Tuschl T, Patel DJ. Structure of an argonaute silencing complex with a seed-containing guide DNA and target RNA duplex. Nature. 2008a; 456:921–926. [PubMed: 19092929]

Wang Y, Sheng G, Juranek S, Tuschl T, Patel DJ. Structure of the guide-strand-containing argonaute silencing complex. Nature. 2008b; 456:209–213. [PubMed: 18754009]

Yuan YR, Pei Y, Ma JB, Kuryavyi V, Zhadina M, Meister G, Chen HY, Dauter Z, Tuschl T, Patel DJ. Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. Molecular cell. 2005; 19:405–419. [PubMed: 16061186]

Highlights

- RsAgo associates with 15-19 nt RNA and 22-24 nt DNA molecules in vivo.
- RsAgo-associated small RNAs correspond to the majority of cellular transcripts.
- Small DNAs are complementary to the small RNAs and enriched in foreign sequences.
- RsAgo degrades plasmid DNA and represses expression of plasmid-encoded genes.

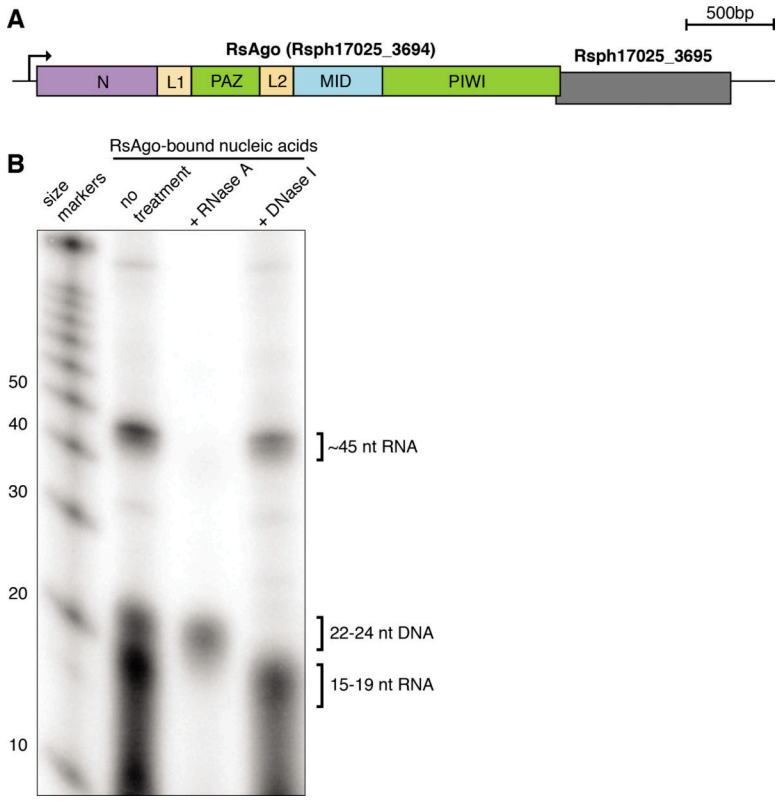

Figure 1. *R. sphaeroides* Argonaute protein associates with small RNA and small DNA molecules

(A) *R. sphaeroides* Argonaute protein and its operon structure. Gene Rsph17025_3695 encodes a predicted nuclease and overlaps with the last 4 nt of the RsAgo coding sequence. PAZ, MID and PIWI are conserved domains shared by all Argonautes, while N (N-terminal), L1 (linker 1) and L2 (linker 2) are not conserved. **(B)** Purified RsAgo is associated with small RNA (15-19 nt) and small DNA (22-24 nt) species. RsAgo-associated nucleic acids were 5′-radioactively labeled, treated with RNase A or DNase I and resolved on a 15% urea PAGE. The ~45 nt band is composed primarily of tRNAmet 3′ fragments. Densitometric measurement of three independent RsAgo samples indicates that the ratio of small RNA to small DNA in RsAgo complexes is 2.12±0.66.

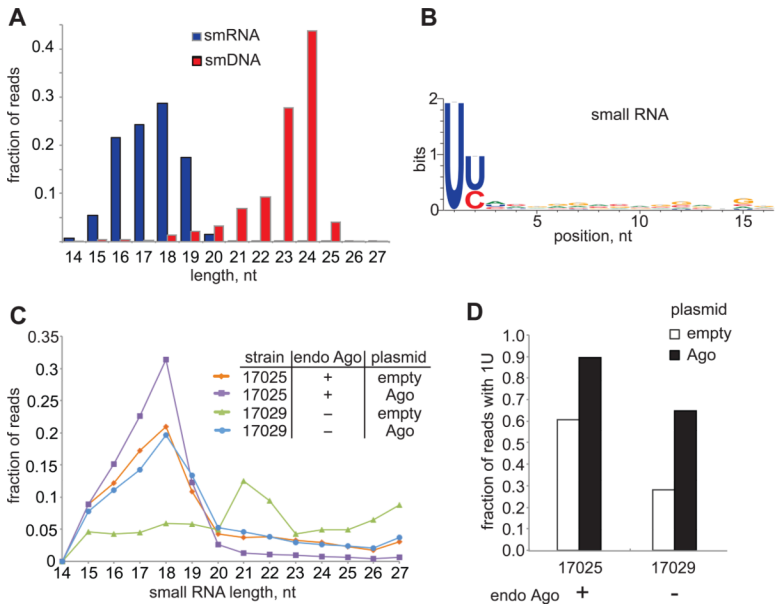

Figure 2. Analysis of RsAgo-associated small RNA

(A) Read length distribution of sequenced small RNA and small DNA associated with RsAgo. Only sequences that could be mapped to the genome or the expression plasmid were considered. **(B)** Nucleotide bias in small RNA extracted from purified RsAgo. 16-19 nt reads were trimmed from the 3′end to 16 nt and analyzed with WebLogo. **(C)** Length profile of total small RNAs extracted from two strains of *R. sphaeroides* - one encoding Ago (17025) and one not (17029) - carrying either an empty plasmid or a plasmid expressing RsAgo. **(D)** Bias for the uridine residue in position 1 in libraries shown in panel C. 61% of reads in total small RNA from wild type strain 17025 lacking expression plasmid had uridine in position 1.

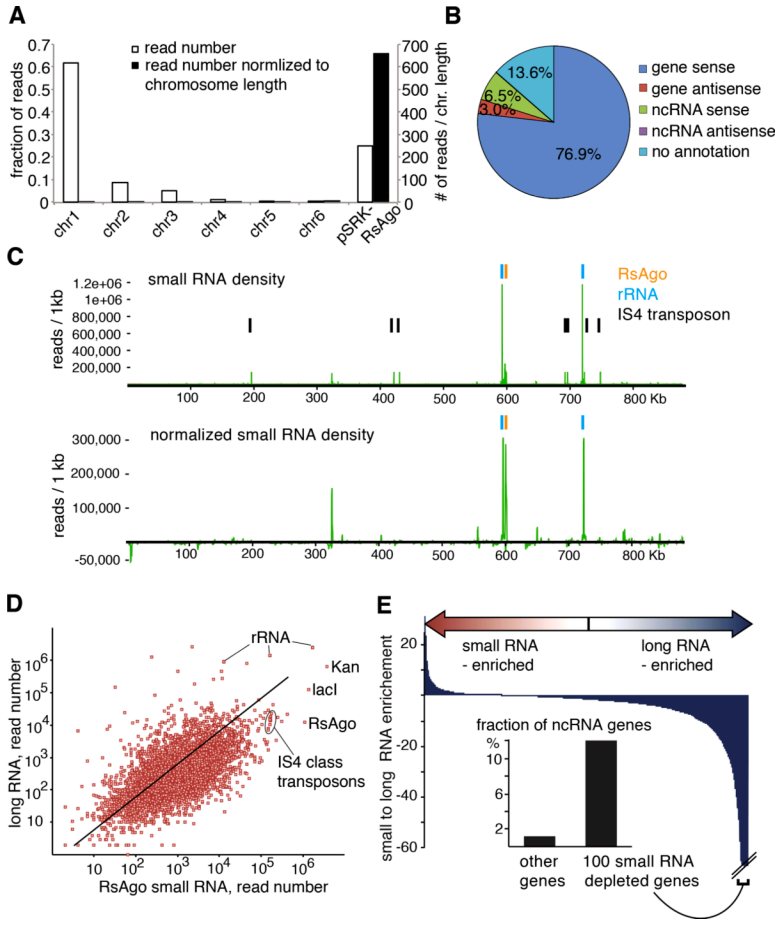

Figure 3. RsAgo-associated small RNAs represent a broad sample of the transcriptome

(A) Distribution of unique small RNA reads from purified RsAgo over the six chromosomes of *R. sphaeroides* strain 25 plotted as a fraction of raw read numbers and read numbers normalized to chromosome length. **(B)** Annotation of RsAgo-associated small RNAs. Genes on the expression plasmid were included in the analysis. **(C)** Density of cloned RsAgo-associated small RNAs along chromosome 2 of *R. sphaeroides* strain 25. The graph includes reads that can be mapped without mismatch to multiple genomic positions. Note that the large number of reads mapping to the RsAgo gene is likely originating from the expression plasmid. In the top graph each genomic position corresponding to a cloned small RNA is given equal weight; on the bottom graph each read is normalized for the number of times it maps to the genome, resulting in proportionally less signal for reads that map many times such small RNA matching IS4 transposon. **(D)** The correlation between long and small RNA abundance for *R. sphaeroides* genes. rRNA-depleted long RNA library was prepared from strain 25 containing the pSRKKm-RsAgo plasmid. **(E)***R. sphaeroides* genes are sorted by their enrichment for small RNAs relative to long transcripts as measured by RNA-Seq. Shown are fold enrichment (positive values) or depletion (negative values) of small to long RNA normalized to the mean small RNA/long RNA ratio for all genes. The 100 genes most strongly depleted in small RNA (right side of the distribution) are enriched in non-coding RNA relative to the rest (12% compared to 1.2%).

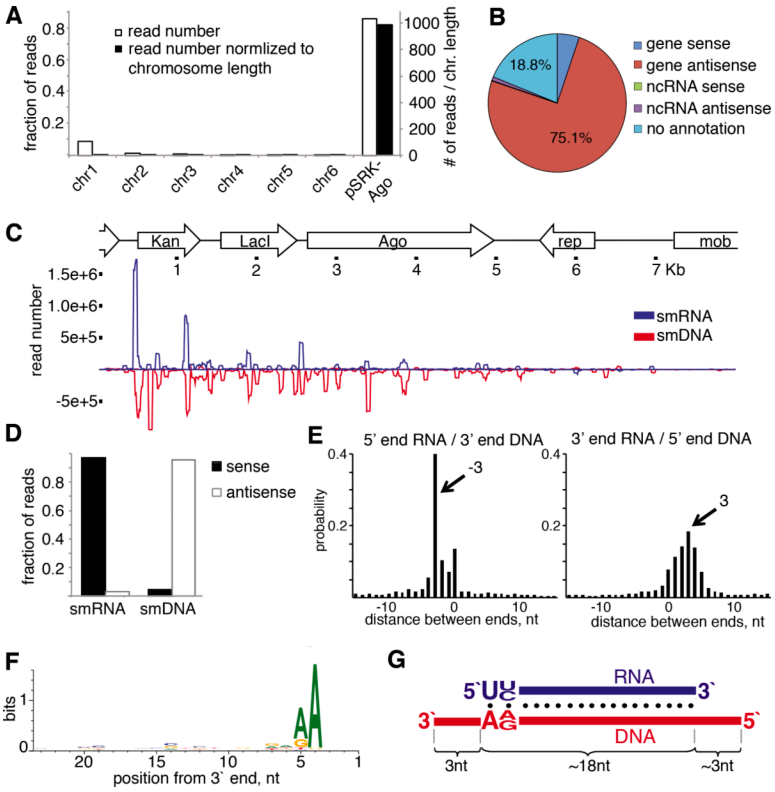

Figure 4. Analysis of RsAgo-associated small DNA

(A) Distribution of unique small DNA reads from purified RsAgo over the six chromosomes of *R. sphaeroides* strain 25 plotted as a fraction of raw read numbers and read numbers normalized to chromosome length. **(B)** Annotation of RsAgo-associated small DNA sequences. **(C)** The profile of RsAgo-associated small RNA and DNA over the expression plasmid pSRKKm-RsAgo. Reads that mapped to the plus or minus strand are shown above and below the axis, respectively. **(D)** Strand bias of RsAgo small RNAs and DNAs mapped to genes encoded on the expression plasmid pSRKKm-RsAgo. **(E)** Relative distance between ends of small RNAs and DNAs mapped to opposite genomic strands of plasmid pSRKKm-RsAgo. Graphs indicate that RNA and DNA have a high tendency to map as pairs with the DNA molecule protruding 3 nt on each side. Note that this distance is more rigid on the left side (5′ end of RNA and 3′ end of DNA) compared to the right side (3′ end of RNA and 5′ end of DNA) **(F)** Nucleotide bias in small DNA aligned by the 3′ end and analyzed with WebLogo. **(G)** Model of small diRNA-riDNA pairs.

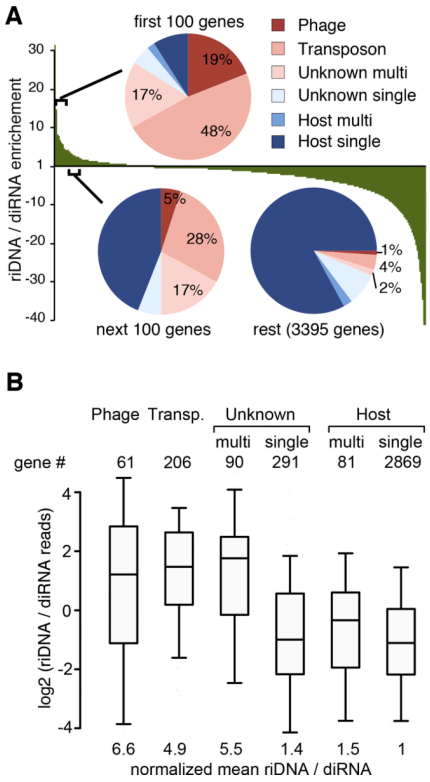

Figure 5. Distribution of diRNA and riDNA over *R. sphaeroides* genes

(A) Fold enrichment (positive values) or depletion (negative values) of riDNA/diRNA ratio normalized to the mean riDNA/diRNA ratio for all genes. Genes are sorted by their enrichment for riDNA relative to diRNA: riDNA-rich genes are on the left, depleted on the right side of the graph. The frequencies of six different gene classes of the *R. sphaeroides* strain 25 genome (single and multi-copy host genes, single- and multi-copy genes of unknown origin and phage- and transposon-related genes) were analyzed for the 100 most DNA-rich genes, the next 100 genes in the distribution and the rest of the distribution (3395 genes). For a similar analysis of small DNA reads normalized to the genomic copy number see Fig. S5C. **(B)** Box plots of riDNA to diRNA ratios for the same gene classes as in panel A. Number of genes in each class is shown above the plot. The box represents the 25th, 50th (the inner line) and the 75th percentiles of the distribution; whiskers are at the 5th and 95th percentile. The mean of the ratio of riDNA to diRNA was calculated for each gene class and normalized to that of host single-copy genes. The same plot with DNA read numbers divided by the number of genome mappings (copy number) in shown on Fig. S5D.

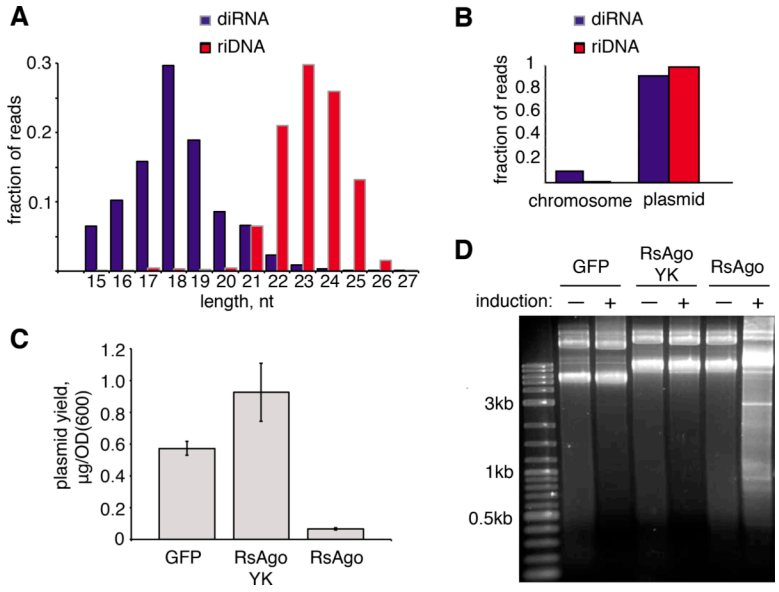

Figure 6. The effects of RsAgo expression in*E. coli*

(A) Read length distribution of sequenced small RNA and small DNA associated with RsAgo expressed in *E. coli* strain BL21(DE3). Only sequences that could be mapped to the genome and the expression plasmid were considered. **(B)** Fractions of small RNA and small DNA reads from RsAgo purified from E. coli BL21(DE3) that map to the host chromosome and expression plasmid **(C)** Plasmid yields from *E. coli* BL21(DE3) cells after induction of expression of GFP, wild-type RsAgo and RsAgo mutant impaired in small RNA binding (RsAgo-YK). Shown is the mean plasmid yield (+/- standard deviation) per cell density unit (OD_{600}). **(D)** RsAgo expression in *E. coli* causes degradation of plasmid DNA. Equal amounts of plasmid DNA from the experiment shown on panel C were resolved on agarose gel and stained with SYBR Gold. The expression of proteins was induced with IPTG and shown on Fig. S6.

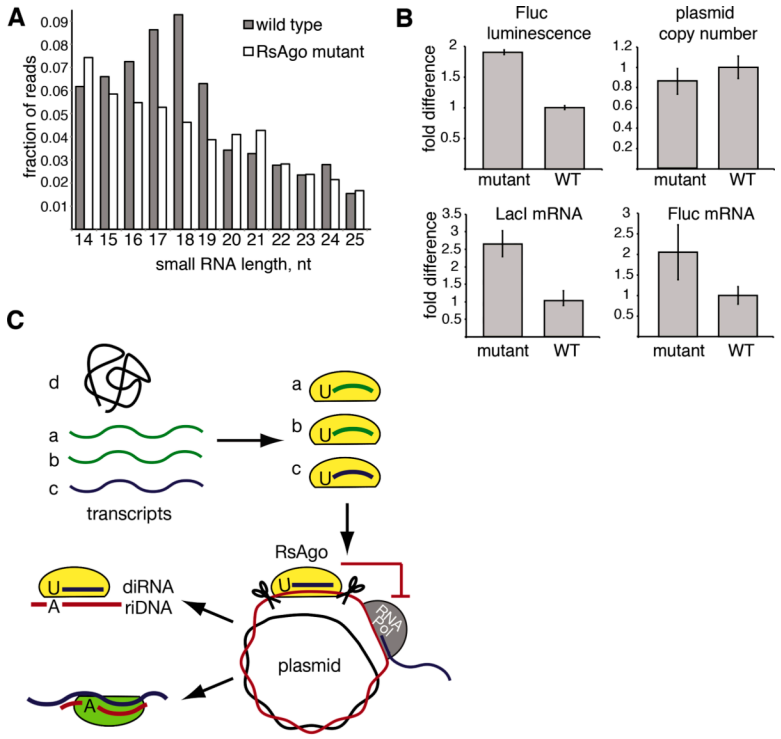

Figure 7. The effects of RsAgo mutation in*R. sphaeroides*

(A) Read length distribution of sequenced total small RNA (13 – 30 nt) from wild-type and RsAgo mutant *R. sphaeroides* strain 25 cells. Only sequences that could be mapped to the genome were considered. **(B)** Comparison of expression of plasmid-encoded genes in wild-type and RsAgo mutant *R. sphaeroides* strain 25. The luciferase (Fluc) and lacI genes are encoded on the plasmid pSRKTc-Fluc. Firefly luciferase activity was quantified by measuring luminescence. Shown is normalized mean ratio of luminescence in induced to non-induced cells +/- SD (n=3). Fluc and lacI mRNA was quantified by RT-qPCR and normalized to the average mRNA level of three host genes (normalized mean +/- SD, n=3). pSRKTc-Fluc plasmid copy number was determined by qPCR. **(C)** The model of RsAgo function and biogenesis of diRNA and riDNA. diRNAs are generated from mRNAs (a-c) or products of their degradation, while structural non-coding RNA (d) is not efficiently processed into small RNA. RsAgo loaded with cognate diRNA (c) recognizes plasmid DNA followed by excision of 22-24 nt single-stranded target DNA fragment leading to damage of DNA template. Binding of plasmid DNA by RsAgo/diRNA complex might also interfere with transcription by RNA polymerase prior to DNA excision. Excised riDNA forms a duplex with guide diRNA in the RsAgo complex and does not play further role in repression. Alternatively, riDNA is loaded into a new RsAgo complex and directs it to the plasmid transcript to induce post-transcriptional repression.