Ancestral Roles of Small RNAs: An Ago-Centric Perspective

Leemor Joshua-Tor and Gregory J. Hannon

Watson School of Biological Sciences, Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724

Correspondence: hannon@cshl.edu and leemor@cshl.edu

SUMMARY

RNAi has existed at least since the divergence of prokaryotes and eukaryotes. This collection of pathways responds to a diversity of “abberant” RNAs and generally silences or eliminates genes sharing sequence content with the silencing trigger. In the canonical pathway, double-stranded RNAs are processed into small RNAs, which guide effector complexes to their targets by complementary base pairing. Many alternative routes from silencing trigger to small RNA are continuously being uncovered. Though the triggers of the pathway and the mechanisms of small RNA production are many, all RNAi-related mechanisms share Argonaute proteins as the heart of their effector complexes. These can act as self-contained silencing machines, binding directly to small RNAs, carrying out homology-based target recognition, and in some cases cleaving targets using an endogenous nuclease domain. Here, we discuss the diversity of Argonaute proteins from a structural and functional perspective.

Outline

1 Introduction
2 Argonautes in three clades
3 Ago structure reveals function
4 Multiple modes of RNA binding
5 Small RNA guided cleavage
6 Comparative studies of argonaute biology hint at ancestral function

References
1 INTRODUCTION

The evolution of genomes capable of sustaining their own replication was almost certainly followed by the rapid emergence of parasites capable of exploiting this innovation. Hosts able to control or combat such parasites would have a strong selective advantage. Thus, the deeply rooted conflict between hosts and those seeking to exploit them has driven the development of a panoply of immune strategies that permit discrimination of self from nonself at the genomic level and provide strategies for selectively negating the propagation of pathogens at the expense of their host.

Small regulatory RNAs are pervasive throughout eukaryotes and components of small RNA pathways are also found in select archaea and eubacteria (Hock and Meister 2008; Chen 2009; Karginov and Hannon 2010). In multiple kingdoms of life, small RNAs play key roles in responses to both endogenous nucleic acid pathogens, such as viruses, and to resident genomic parasites, namely transposons (Aravin et al. 2007). Though the precise architecture of these small RNA-based immune systems varies, their sharing of core components and of general regulatory strategies strongly suggests genome defense as an ancestral function of small RNAs. Numerous observations suggest later diversification of the biological roles of small RNAs, extending these to gene regulation at nearly every conceivable level and even to control of genomic architecture and content.

The diversity of small RNA families seems to be growing at an alarming rate, catalyzed in large part by the application of increasingly deep next-generation sequencing. At times, it seems as if we will find that nearly every position in every eukaryotic genome will be soon be reflected within at least one small RNA species. However, here we will confine our discussion to the classes of small RNAs that are united under the broad umbrella of RNA interference pathways. Though small RNAs may enter these pathways through a surprising number of different avenues (Kim et al. 2009), all pathways falling under the RNAi rubric share one common and immutable feature. They exert their functions through an effector complex, called RISC or the RNA-induced silencing complex, that has an Argonaute-family protein at its core.

Argonaute proteins were first linked to RNAi through genetic studies in Caenorhabditis elegans. A screen for RNAi-resistant mutants uncovered RDE-1 (RNAi-defective-1) as being essential for gene silencing in response to exogenously introduced double-stranded RNA (Tabara et al. 1999). This was a representative of a little-studied family of proteins, which had previously only been noted because of developmental phenotypes that arose upon their inactivation. In fact, the Argonaute designation came from the appearance of an Arabidopsis mutant whose morphology was reminiscent of a squid (Bohmert et al. 1998). However, even from the beginning, it was clear that Argonaute proteins were members of a highly conserved and quite large family of proteins, with multiple members in nearly every eukaryotic genome (Cerutti et al. 2000). Later studies placed Ago proteins within the emerging biochemical framework of RNA silencing mechanisms, showing that Argonaute was the component of RISC, which directly interacted with the small RNA species that contribute target specificity to the complex (Hammond et al. 2001).

2 ARGONAUTES IN THREE CLADES

A phylogenic analysis of Argonautes from Homo sapiens (Hs), Arabidopsis thaliana (At), C. elegans (Ce), C. briggsae (Cb), Drosophila melanogaster (Dm), and Schizosaccharomyces pombe (Sp), representatives of the plant, animal, and fungal kingdoms, suggested that these proteins can be divided into three clades (Fig. 1) (Tolia and Joshua-Tor 2007). In many ways, this diversity mirrors the breadth of biological processes, which Argonaute proteins can impact and mechanisms through which they can act. The Ago-like subfamily proteins, defined based on their similarity to AtAgo1, are present in animals, plants, and fission yeast and are involved in transcriptional and posttranscriptional silencing and in genome organization (Joshua-Tor 2006). They show broad expression patterns being present in nearly every tissue and cell type examined in these representative species. They most often interact with what have become known as canonical small RNAs, which based upon their mechanisms of biogenesis average around 22 nucleotides in length and have 5′ phosphate and 3′ hydroxyl groups (Zamore et al. 2000; Bernstein et al. 2001). These include the well-known microRNAs and small interfering RNAs, both derived from double-stranded RNA precursors. The Piwi subfamily is an animal-specific clade that is expressed almost exclusively in gonadal tissues, often only in germline cells, where they form the core of animal transposon silencing pathways (Aravin et al. 2007). Piwi proteins interact with piRNAs (Piwi-interacting) RNAs, which are generated via a pathway distinct from canonical small RNAs. They have 5′ P and 3′ OH ends, suggesting nucleolytic processing from longer precursors, but are often larger than RNAs of the canonical classes (23–29 nt). The Wago clade is constituted entirely of worm Argonautes. It is unusually diverse, contributing 18 of the 27 family members in the C. elegans genome (Tolia and Joshua-Tor 2007). Many that have been characterized so
far bind small RNAs generated by direct synthesis from RNA templates (Yigit et al. 2006). Thus, Wago proteins are generally occupied by “secondary siRNAs” which bear 5′ triphosphate and 3′ hydroxyl termini. The biological roles of Wago proteins are still emerging; however, it is clear that they are important for silencing responses to exogenously delivered dsRNAs, for small RNA-based regulation of endogenous genes, and chromosome structure and segregation (Tabara et al. 1999; Yigit et al. 2006; Claycomb et al. 2009; Conine et al. 2010; Gu et al. 2009).

3 AGO STRUCTURE REVEALS FUNCTION

A continuous accumulation of genetic and biochemical evidence pointed to central roles for Ago proteins in virtually all RNAi-related processes. Characterization of the RISC, particularly from Drosophila and mammals demonstrated that it acted as a multiple turnover enzyme that could cleave RNA substrates in a manner determined by the sequence of its bound RNA, an activity termed “Slicer” (Martinez and Tuschl 2004; Schwarz et al. 2004). How
Argonaute proteins contributed to this activity, however, remained a mystery.

Sequence conservation of Ago proteins pointed to the presence of two discrete domains, which were termed PAZ and PIWI (Cerutti et al. 2000). Unfortunately, sequence alone provided no insight into the functions of these regions nor could their roles be inferred by comparisons to other better-characterized proteins. A true understanding of their function and consequently the functions of the Ago proteins themselves came only after structures of PAZ domains and then full-length Argonautes emerged.

The first PAZ domain structures were accompanied by biochemical studies demonstrating that it formed an RNA binding module, which specifically interacts with the 3′ end of the small RNA that guides Ago target specificity (Lingel et al. 2003; Song et al. 2003; Yan et al. 2003). Most of the contacts to the RNA are made by highly conserved aromatic residues—two tyrosines and a histidine—that contact the phosphate that bridges the two terminal bases of the small RNA (Lingel et al. 2004; Ma et al. 2004).

The determination of the first full-length Argonaute protein structure from \textit{P. furiosus} (PfAgo) illustrated that Argonautes also contain three additional domains: the amino-terminal, middle (Mid), and PIWI domains (Fig. 2) (Song et al. 2004). The big surprise came upon examination of the PIWI domain located at the carboxyl terminus of the protein. The structure revealed that Argonautes clearly belong to the RNase H family of enzymes. Apart from having the RNase H fold at its core, it also has the two highly conserved aspartates that are invariably present on adjacent \(\beta\)-strands in this class of enzymes. Modeling of an siRNA guide strand and target mRNA further showed this to be consistent with what was known from biochemical experiments regarding the characteristics of the slicing activity of the RNA-Induced Silencing Complex, RISC (Song et al. 2004). These include the dependence on the presence of Mg\(^{++}\), the generation of products similar to those produced by RNase H family enzymes, and the position of the scissile phosphate, or the endonucleolytic cut, opposite the phosphate between the 10th and 11th nucleotide from the 5′ end of the small RNA guide. Thus the identity of the previously unknown Slicer activity was determined to reside in the Argonaute protein itself, defining this protein as a self-contained silencing machine (Liu et al. 2004; Song et al. 2004).

\textbf{Figure 2.} The crystal structure of Argonaute from \textit{P. furiosus} shown as a ribbon diagram. The amino-terminal domain is in blue, the PAZ domain is red, the middle domain is green, and the PIWI domain is in purple. The inter-domain connector is shown in yellow. A close-up of the active site residues coordinating a Mn\(^{++}\) ion (pink ball) is shown in the inset.
The PAZ domain, which resembles an OB (Oligonucleotide Binding) fold domain with a slightly different topology (Lingel et al. 2003; Song et al. 2003; Yan et al. 2003), appears to be rather mobile as evident from the two structures of *Aquifex aeolicus* Argonaute (Yuan et al. 2005; Rashid et al. 2007). In one case the PAZ domain is in a similar orientation to that of the PAZ domain in the *P. furiosus* structure and in the other there is a shift of approximately 24°. Even though crystal contacts could play a role, this disparity does illustrate the conformational space that the PAZ domain may assume. The hypothesis of a flexible PAZ is underscored by molecular dynamics simulations (Rashid et al. 2007) and normal mode analyses of these structures (Ming et al. 2007) showing low frequency motions, especially implicating movement of the PAZ domain in opening the putative binding groove for RNA.

4 MULTIPLE MODES OF RNA BINDING

Within RISC, Ago proteins must exist in a large number of RNA bound configurations. Initially, the small RNA guide must be loaded to prime the complex. In the canonical pathway, it still remains unclear whether all guides are loaded initially as double-stranded RNAs. However, biochemical studies clearly demonstrate that following cleavage into discrete segments, double-stranded guides can enter RISC (Matranga et al. 2005). These can be converted by cleavage or by simple dissociation of the nonguide strand into an active complex, with a single-stranded guide ready to search for targets. In most instances, Ago clade proteins preferentially use a portion of the guide to identify targets. This region, known as the “seed” encompasses nucleotides two to eight of the small RNA (Bartel 2009). Although the remainder of the small RNA and its degree of pairing to its substrate can certainly contribute to both the efficiency of silencing and the mode of RISC action, structural studies have provided clues to the dominant role of the seed in target selection. They have also suggested that cycles of target cleavage and substrate release are accompanied not only by conformational changes in the Ago proteins themselves but also by changes in the way Ago engages its small RNA guide.

Much of this insight has been gained through a series of structures of the *T. thermophilus* Argonaute bound to various combinations of DNA guide strands and RNA and DNA target strands (Wang et al. 2008a; Wang et al. 2008b; Wang et al. 2009). The *thermophilus* protein actually prefers DNA as both target and guide; however the latter fact became apparent over time, causing initial studies to be done with DNA guide/RNA target combinations.

Bound to a 21-mer DNA guide in the absence of target, Argonaute anchors both ends of the small DNA (Wang et al. 2008b). The 3′ end resides within the PAZ domain as predicted based upon prior structural and biochemical studies of other family members (Lingel et al. 2003; Song et al. 2003; Ma et al. 2004). The 5′ end of the guide lies within a binding site in the mid domain that was identified previously through structures of a PIWI-domain protein with RNA, though the trajectory of the strand differs in the two complexes (Ma et al. 2005; Parker et al. 2005).

The existence of a 5′ end-binding pocket within the Ago protein itself is consistent with the many indications that different family members have different 5′ end preferences. In *Arabidopsis*, the 12 Ago proteins each show a different spectrum of preferences for the 5′ nucleotide of the bound small RNA (Mi et al. 2008; Montgomery et al. 2008). In fact, individual species can be redirected from one family member to another simply by changing the terminal base. In worms, different Ago proteins likely alter this pocket to engage small RNAs with mono versus triphosphate termini, an accommodation that is also likely made in certain fungal Argonautes. Some Piwi proteins bind almost exclusively to small RNAs beginning with U (Aravin et al. 2007), though it is presently unclear whether this preference is established during biogenesis or small RNA selection by members of this clade. Though the *pyrococcus*, *thermophilus*, and *aquifex* structures form a basis from which to build hypotheses regarding small RNA selection by Ago proteins, what is clearly required as a future step is to define binding pockets within family members whose natural small RNA preferences are clear.

A second correlate between structural, biochemical, and genetic studies extends to the way in which the seed is used for target selection. There appears to be no pressure on natural targets to pair with the first base of the small RNA (Lewis et al. 2003). This is in accord with its limited availability within the structure (Wang et al. 2008b). Moreover, the structure reveals that the bases of the first and second nucleotides are destacked and that the edges of the bases for nucleotides 2–6 are exposed for base pairing. The middle part of the guide (nucleotides 12–17) is disordered in the structure lacking substrate. However, one can see an arginine intercalating between nucleotides 10 and 11 resulting in a catalytically incompetent conformation.

The exploitation of active site mutants of the *thermophilus* enzyme, substituting one of the active site aspartates with asparagine, enabled the examination of structures with target RNAs (Wang et al. 2009). In a complex containing a quite short, 12-mer target, the guide remained anchored at both ends as in the binary enzyme (Fig. 3). The 5′-phosphate was bound to several conserved residues and a Mg++ ion, underscoring the importance of the
presence of this phosphate for selectivity in guide binding and for establishing fidelity in cleavage sites selection (Ma et al. 2005; Rivas et al. 2005). The first guide-strand base remained destacked from the next as was shown previously. The DNA-RNA duplex that follows at positions 2 through 12 are reported to resemble an A-form conformation and the scissile bond is placed at the active site of the enzyme in a helical and catalytically competent conformation in contrast to their orthogonal arrangement in the binary complex. The following nucleotides of the guide strand are presumably unstructured, because they are not observed. The final two nucleotides again appear anchored within the PAZ domain. The placement of the guide is a bit different in the ternary structure, but most notable is the substantial movement of the PAZ domain with respect to the binary complex. One must keep in mind that the PAZ domain is involved in significant crystal contacts that might affect these conformational changes, however, it does illustrate the range of possible motions it could sample. A different active site mutant with a longer 15-mer target RNA resulted in a smaller movement of the PAZ domain. However, in this case, the guide strand appears to have exited the PAZ domain to pair with the RNA and its 3’ end is now disordered. It is worth noting that guide-strand exiting the PAZ during the catalytic cycle had been previously predicted by careful kinetic analysis of Drosophila RISC (Schwarz et al. 2004). The 3’ end pairing between the passenger and guide is accompanied by a widening of the channel between the PIWI and N-domains. An even longer RNA target, 19-nucleotides in length (Fig. 4), helps localize the 16th nucleotide of the target RNA and in this case the N-domain appears to block simple propagation of the DNA-RNA duplex, though additional conformational changes in the protein might occur with an even longer target. Throughout the seed pairing, Argonaute contacts only the guide strand, though contacts to the target are observed around nucleotides 10–13.

The reloading of an RNA target into the complex during the catalytic cycle is a problem worth examining. In some of the aforementioned structures, the DNA guide appears to be blocking easy exit of the RNA. It appears to be nestled well between the guide and the protein. Would two turns of a duplex need to be unwound and rewound for each cycle to allow entrance of a new target RNA, or does the protein “help” distort the duplex so that there is no hindrance for exit and entrance of successive RNA target strands to be cleaved? This question is clearly relevant for achieving multiple catalytic cycles by each RISC but also likely for the conversion of RISC loaded with dsRNA guides to those complexes primed with ssRNA and ready for action.

5 SMALL RNA GUIDED CLEAVAGE

Though Argonaute proteins can impact their targets in many ways, one of the most critical and conserved is endonucleolytic cleavage (Song et al. 2004). The substrate specificity of Argonaute is determined by the sequence of the bound guide RNA, as the target is recognized via hybridization to the guide strand. As described earlier, Argonaute is an RNase H-like enzyme creating a 5’ product with a 3-OH and a 3’ product carrying a 5’ phosphate, with the DNA strand generally being replaced by a guide RNA. The RNase H protein family consists of well-characterized enzymes such as retroviral integrases and transposases (Nowotny 2009). Escherichia coli RNase H1 catalyzes a single reaction resulting in substrate cleavage. Integrase and transposases catalyze two consecutive reactions, donor-end processing and nucleotidyl transfer, resulting in strand
transfer. The nucleophile in these reactions is either a water molecule or a nucleotide 3′-OH. They involve a two-metal ion catalysis mechanism, with one metal activating the nucleophile and the second stabilizing the intermediate (Nowotny et al. 2005). Two metals have been observed in crystal structures of substrate-bound complexes of Tn5 transposase and both human and B. halodurans RNase H1 (Lovell et al. 2002; Nowotny et al. 2005; Nowotny et al. 2007). So far, the presence of at least one metal ion is clear in some of the Argonaute structures (Rivas et al. 2005; Yuan et al. 2005; Wang et al. 2008a; Wang et al. 2009). A second metal ion was reported in one of the ternary complexes, largely based on possible coordination with neighboring oxygen atoms (Wang et al. 2009).

A combination of structural and mutational analyses identified the two invariant aspartates characteristic of RNase H enzymes as well as a third residue that in most cases is a histidine rather than another carboxylate, critical for catalysis (Fig. 2) (Joshua-Tor 2006). The presence of all three residues is not sufficient to predict whether a particular Argonaute has catalytic activity. For example, the four human Argonaute clade proteins, HsAgo1, HsAgo2, HsAgo3, and HsAgo4 have been tested for activity, with HsAgo2 alone having slicer activity, even though HsAgo3 has an intact catalytic triad (Liu et al. 2004; Meister et al. 2004). The remaining Piwi clade proteins also conserve catalytic potential, and are presumed active based upon their conserved mechanism of transposon recognition and silencing (Faehnle and Joshua-Tor 2007).

The presence of the Argonaute protein and the guide RNA in a minimal recombinant RISC, is sufficient for target cleavage, but does not show ATP-stimulated product release and turns over rather slowly. However, other factors likely also contribute to the function of a complete RISC. In one report, a complex containing Argonaute, Dicer, and human immunodeficiency virus-1 trans-activating response protein...
(TAR) RNA-binding protein (TRBP) is capable of guide-strand loading and multiple rounds of target cleavage, stimulated by nucleotides (Gregory et al. 2005). Another report indicated that Dicer dissociates once Ago2 is loaded with the guide RNA (Maniataki and Mourelatos 2005). More recent studies support the dissociation of both Dicer and TRBP from Ago2 after loading, and these appear to not participate in target cleavage (MacRae et al. 2008).

Many questions still remain to be answered to fully understand target cleavage by the Argonautes. How is multiple turnover achieved? Does PAZ domain movement assist in substrate engagement and placement, and in product release? What is essential to constitute an active Slicer in addition to the active site requirements?

6 COMPARATIVE STUDIES OF ARGONAUT FAMILY HINT AT ANCESTRAL FUNCTION

In essence, RISC is a programmable homology search engine, which can confer a variety of properties upon a target. Some of these depend upon the catalytic activity of Ago proteins, whereas others do not. This is reflected in a diversification of the Ago family into multiple clades and into groups of family members, which either retain or have lost catalytic potential. Given the staggering breadth of Ago functions, it seems difficult to infer the ancestral roles of the RNAi pathway. In both plants and animals, Argonautes are programmed by microRNAs to regulate gene expression, mainly at the posttranscriptional level (Bartel 2009). However, differences in microRNA biogenesis and effector mechanisms and the lack of any shared microRNAs between plant and animal kingdoms strongly suggests that gene regulatory functions of small RNAs may have evolved several times independently. Plants, animals, and fungi have a variety of small RNA-based pathways which impact chromatin structure and organization and which in some cases are essential for chromosome function or even control chromosome content (Mochizuki et al. 2002; Volpe et al. 2002; Mathieu and Bender 2004; Alexander et al. 2008). Although these pathways play different roles in different species, one shared feature is that all are driven by the recognition of repetitive sequences. In fact, recognition and control of repeated elements, transposons, and viruses—generally foreign nucleic acids—seems the one theme that unifies small RNA pathways wherever they are found.

The first hints of immune roles for RNAi came from plants, where canonical RNAi mechanisms—involving Dicer, siRNAs and Ago-clade proteins—proved a key mediator of antiviral defense (Hamilton and Baulcombe 1999; Mlotshwa et al. 2008). Here, simply the double-stranded nature of replication intermediates seemed sufficient to indicate that an RNA was “foreign” and to tag it for nucleolytic destruction. Similar observations in Drosophila and C. elegans extended the antiviral paradigm to animals (Lu et al. 2005; Wang et al. 2006). In mammals, where an adaptive immune system has evolved to fight viral infection, the roles of RNAi in antiviral immunity are less clear. However, mammals clearly use small RNA pathways to recognize and control endogenous nucleic acid parasites.

An animal-specific clade of Ago proteins, the Piwis, forms the basis of an elegant, RNA-based innate immune system that selectively silences transposons in the germ cells of animals (Aravin et al. 2007). This incorporates two key elements, a programmable evolutionary record of transposon exposure and control and an adaptive mechanism to optimize the response to ongoing transposon challenge.

Small RNAs that enter the piRNA pathway are produced by at least two distinct biogenesis mechanisms. So-called “primary piRNAs” are generated from specific loci within animal genomes, called piRNA clusters. These are often highly enriched for fragmented, diverged transposon remnants, and in some cases represent the most transposon-rich areas of the host genome. These loci appear to be transcribed as large single-stranded precursors, which are parsed into individual small RNAs to create the equivalent of an immune repertoire. In order to gain control over a new element, it must be incorporated into the piRNA pathway by integration into a piRNA cluster, and some clusters are known to act as hotspots for insertion of the elements that they control.

Once made, primary piRNA/Piwi protein complexes can recognize targets much as in the canonical RNAi pathway, catalyzing small RNA-dependent cleavage, in this case of transposon mRNAs. Here, the canonical and piRNA pathways diverge in the sense that target cleavage by Piwi can create the 5’ end of a new small RNA (Brennecke et al. 2007; Gunawardane et al. 2007). This target-derived RNA can catalyze the production of further small RNAs through the same mechanism. If target RNAs can recognize antisense transposon content synthesized as part of the precursors generated from piRNA clusters, then an amplification loop can be formed where interactions between the immune repertoire and their targets can increase populations of small RNAs to meet a particular transposon challenge. In this way, the adaptive amplification loop, called the ping-pong cycle, is analogous to the expansion of adaptive immune cells triggered by antigen-receptor interactions (Aravin et al. 2007).

piRNA pathways in animals clearly use post-transcriptional mechanisms as part of their regulatory arsenal. However, there is clear evidence that piRNA pathways direct DNA methylation of repeats in mammals (Carmell et al. 2007) and that posttranscriptional controls may also figure prominently in silencing of mobile elements in flies.
(Pal-Bhadra et al. 2004; Brower-Toland et al. 2007). In plants, a 24-nucleotide class of small RNAs directs the methylation of repeat elements via binding to an Ago-clade protein (Zilberman et al. 2003). These depend upon a Dicer for their production, implicating dsRNA precursors that are absent from the piRNA pathway. Unlike the piRNA pathway in animals where we are at least starting to gain glimpses into the mechanism of self-non-self recognition, there are no concrete hypotheses regarding how repeat elements are singled out for silencing in plants. Two distinct RNA polymerases, polIV and polV, appear essential for RNA-directed DNA methylation (Herr et al. 2005; Onodera et al. 2005) suggesting at least one construction for a reinforcing loop similar to the one in animals. If transcription by these polymerases triggers small RNA production, which in turn directs methylation, one could close a cycle if that same methylation also helped recruit the polymerase variants. The question, however, would be how the cycle would be initiated in the first place.

In ciliates, small RNAs are critical for repeat management through strategies that show some conceptual similarities to the piRNA pathway but also some remarkable differences (Mochizuki and Gorovsky 2004). *Tetrahymena thermophila* is unusual in many respects. Among these are that somatic and germ nuclei, the macronucleus and micronucleus respectively, are propagated within a single vegetatively growing cell. The nuclei differ not only in their transcriptional activity but also in their genomic content. During the sexual cycle when a new somatic nucleus is produced following fusion of the meiotic products of the micronucleus, a significant fraction of the genome is discarded through a process that is guided by small RNAs. The small RNA population is adjusted based on the content of the old macronucleus through mechanisms that are not precisely clear. However the result is that through a process of genome comparison, small RNA populations targeting repeats are enriched, and these direct the physical elimination of sequences from the developing somatic genome. Small RNAs that direct elimination reside in a Piwi family protein but are Dicer-dependent (Malone et al. 2005; Mochizuki and Gorovsky 2005), a strange convergence of pathways that seem distinct in other species.

Examples of small RNA and repeat management can also be drawn from the fungal kingdom. In fission yeast, small RNAs are essential for the functional integrity of centromeres (Folco et al. 2008; White and Allshire 2008). Their effect in this case is on the organization of the centromeric heterochromatin domain, which is in turn essential for proper segregation. In *S. pombe*, a series of centromere-associated repeats becomes periodically transcribed during the S-phase of the cell cycle (Chen et al. 2008; Kloc et al. 2008). These are targeted by the RNAi machinery, in this case via a pathway that requires Dicer, an RdRP, and an Ago-clade protein, resulting in the deposition of histone methyl marks characteristic of heterochromatin (White and Allshire 2008). Here, the conserved theme of targeting repeats seems to have been co-opted to serve a positive role in the host cell. A second example can be seen in the phenomenon of unpaired silencing (MSUD, meiotic silencing of unpaired DNA) in *Neurospora* (Shiu et al. 2001). During MSUD, sequences that differ between two homologs are selectively and heritably silenced during the sexual reproductive cycle. Such a comparison would presumably pick up new transposon insertions in addition to other types of copy number variation between synapsed chromosomes. Though no small RNAs driving this process have yet been identified, it does require the same types of RNAi-related components required for the creation of centromeric heterochromatin in *pombe* (Alexander et al. 2008).

Though RNAi-related pathways are not the universal mechanism by which hosts battle genomic parasites, the conceptual framework that emerged particularly by studies of the piRNA pathway also underlies phage resistance in both archaean and eubacteria. Here the CRISPR-CAS system uses small RNAs (crRNAs) to identify and cleave phage mRNAs or genomic DNAs (Jore et al. 2010; Karginov and Hannon 2010). There is also analogy in the mechanism by which invading elements are incorporated into a genetic memory of resistance. Fragments of invading phage become integrated into a locus (CRISPR), which serves as the source of phage-targeting crRNAs. Upon challenge with an unrecognized phage, the clusters can clearly evolve by the acquisition of new sequence information near the 5′ end of the CRISPR unit. Fragments can also be lost from this heritable resistance determinant in the absence of pressure to maintain them.

Thus, through all kingdoms of life, small RNAs underlie the ability of genomes to selectively propagate in the presence of endogenous and exogenous invaders that seek to hijack replicative potential. However, the flexible nature of small RNAs themselves and the protein complexes, which use these to guide target selection, has resulted in an explosion of regulatory function that now pervades almost every aspect of biology.

REFERENCES


L. Joshua-Tor and G.J. Hannon


