Review

Structural Foundations of RNA Silencing by Argonaute

Jessica Sheu-Gruttadauria and Ian J. MacRae

Department of Integrative Structural and Computational Biology, *The Scripps Research Institute, La Jolla, CA 92037, USA*

Correspondence to Ian J. MacRae: *macrae@scripps.edu*
http://dx.doi.org/10.1016/j.jmb.2017.07.018
Edited by Helene Hodak

Abstract

Nearly every cell in the human body contains a set of programmable gene-silencing proteins named Argonaute. Argonaute proteins mediate gene regulation by small RNAs and thereby contribute to cellular homeostasis during diverse physiological process, such as stem cell maintenance, fertilization, and heart development. Over the last decade, remarkable progress has been made toward understanding Argonaute proteins, small RNAs, and their roles in eukaryotic biology. Here, we review current understanding of Argonaute proteins from a structural prospective and discuss unanswered questions surrounding this fascinating class of enzymes.

Introduction

Since their discovery over 25 years ago, small RNAs [21–30 nucleotides (nt)] have emerged as major regulators of gene expression in nearly all aspects of animal biology. The widespread utility of small RNAs in biology arises from the remarkable molecular assemblies in which small RNAs function: the RNA-induced silencing complexes (RISCs) [1]. RISC combines the natural capacity of nucleic acids to store and convey complex sequence information with the ability of proteins to fold into highly functional and dynamic molecular machines. The result is a family of fully programmable and highly efficient molecular complexes capable of identifying and silencing essentially any gene.

The functional core of RISC is composed of a small RNA loaded into a member of the Argonaute protein family [2–5]. Argonaute uses the sequence information encoded in the small RNA as a guide to identify complementary RNAs targeted for silencing. RISC is versatile in silencing diverse genes because Argonaute can be loaded with a small guide RNA of essentially any sequence. RISC is also remarkably fast and fastidious in finding its targets, with the ability to identify targets at rates approaching the limit of diffusion while avoiding the off-targets that are highly abundant in the cellular milieu [6]. Silencing mechanisms used by RISC are diverse, as a variety of distinct silencing pathways have evolved around the central ability of Argonaute proteins to rapidly and efficiently identify target RNAs.

The past 10 years has seen major advances in structural and mechanistic understanding of the Argonaute proteins. Great strides have been made toward understanding how different forms of Argonaute are programmed through the assembly of RISC, how Argonaute reshapes its RNA guide to enable target searches, how targets recognized by RISC are silenced, and how RISC itself is regulated. In this review, we reflect upon these advances from a structural perspective and look forward to the major open questions and challenges ahead.

Classes of small RNAs and their biogenesis

microRNAs

microRNAs (miRNAs), which are typically 21–23 nt in length, regulate gene expression in eukaryotes and thereby participate in diverse physiological processes, including epithelial regeneration [7], cardiac development and function [8], ovulation [9],

and neuronal plasticity [10]. In 1993, the first miRNA, *lin-4*, was discovered in *Caenorhabditis elegans* [11,12]. Subsequently, over 2000 putative miRNAs have been identified in humans. The most evolutionarily conserved of these can be grouped into 87 miRNA families, which have predicted target sites in over half the protein-coding genes in the human transcriptome [13–15]. It is therefore unsurprising that dysregulation of miRNA silencing has been implicated in several pathologies, including diabetes, cardiovascular disease, and various forms of cancer [16].

miRNAs are transcribed endogenously by RNA polymerase II as long primary miRNA transcripts that contain a stem–loop structure that houses the mature miRNA [17–19]. The Microprocessor, composed of Drosha, an RNase III enzyme, and DGCR8 (DiGeorge syndrome chromosomal region 8, Pasha in flies), a double-stranded RNA-binding protein that facilitates primary miRNA transcript recognition, removes the precursor miRNA (pre-miRNA) hairpin from this stem–loop [20–25]. The hairpin is subsequently exported from the nucleus by Exportin-5 [26–28]. This pre-miRNA duplex is then processed by the cytoplasmic RNAse III enzyme Dicer, producing an RNA duplex about 22 nt long with 5′ phosphates and 2-nt 3′ overhangs on both ends [29–31]. This duplex is loaded into Argonaute [29,32–34] and one strand is preferentially retained, functioning thereafter as the guide strand for targeting.

Small interfering RNAs

Small interfering RNAs (siRNAs) are typically 21 nt in length and generated by the cleavage of long double-stranded RNAs by Dicer [35]. siRNAs silence extensively complementary target RNAs by endonucleolytic cleavage via a process called RNA interference (RNAi). RNAi was first identified in *C. elegans* [36], although it has since been demonstrated to be essential for viral defense and protection of the genome from transposable elements in many eukaryotic organisms [37–42]. In mammals, Ago2 is the only Argonaute capable of catalyzing siRNA-directed cleavage of target RNAs [3,4]. This activity appears to be necessary during mammalian development, as Ago2-deficient mice are embryonic lethal [3]. Moreover, knock-in with a slicer-deficient Ago2 results in death shortly after birth. This observation revealed that miR-451, involved in erythropoiesis, requires endonucleolytic cleavage by Ago2 for maturation [43,44]. In addition, siRNA-mediated cleavage has also proven to be a powerful tool for exogenous gene manipulation [45].

Piwi-interacting RNAs

Piwi-interacting RNAs (piRNAs) are the largest and most diverse class of small RNAs in metazoans [46]. Ranging in size from 24 to 31 nt in length, piRNAs are generated independently of the enzyme Dicer and are associated with members of a germline-specific subclade of the Argonaute family named Piwi [47–51]. piRNAs are initially transcribed as long precursors from defined genetic regions termed piRNA clusters [52,53]. Precursors are fragmented, loaded into Piwi proteins, and then trimmed from the 3′ end to their mature length [54–58]. In addition, secondary piRNAs are generated from the 5′ end of cleaved piRNA targets, leading to amplification and sharpening of the cellular piRNA pool in a process termed the ping-pong pathway [52,59]. Piwi proteins and piRNAs play an essential role in maintaining the integrity of the germline genome by repressing transposons via transcriptional or posttranscriptional mechanisms. A role for piRNAs controlling protein-coding genes early in development is emerging as well [60,61].

The architecture of Argonaute

Small RNA classes possess distinct base-pairing requirements for target recognition as well as unique mechanisms of silencing [62]. These differences are imparted by the structure and dynamics of the Argonaute proteins associated with each small RNA class. Despite their diversity of function, however, all Argonaute proteins share a common global architecture composed of two lobes formed by four globular and two linker domains that together form a central RNA-binding cleft [63–67] (Fig. 1A, B). The N and PAZ (Piwi–Argonaute–Zwille) domains comprise the N-terminal lobe, while the C-terminal lobe contains the MID (middle) and PIWI (P-body-induced wimpy testes) domains. The two lobes are bridged by the L1 and L2 linker domains, which combined form the central cleft that cradles the guide and target nucleic acid molecules [68]. The slicer active site, which is formed by an ribonuclease H (RNase H)-like fold located in the PIWI domain, resides in the center of the central cleft [66] (Fig. 1B).

The guide RNA is held in an extended conformation through the central cleft and is contacted by all domains of the protein [69] (Fig. 1B). In keeping with the notion that Argonaute can bind guide RNAs of any sequence, almost all interactions with the protein are mediated through the RNA sugar-phosphate backbone. The 5′-phosphate, which is critical for licensing small RNA entry into the RNA silencing pathway, is buried within a hydrophilic pocket at the interface of the MID and PIWI domains (Fig. 1C). The 5′-nt of the guide binds to the MID domain in a manner that renders it unavailable for pairing to complementary target RNAs [70,71]. This finding is in accordance with observations that the 5′-nt does not contribute to target recognition [72]. However, functional analysis demonstrated that human