Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity

Ji-Joon Song, *et al.*

Science **305**, 1434 (2004);
DOI: 10.1126/science.1102514

Crystal Structure of Argonaute and Its Implications for RISC Slicer Activity

Ji-Joon Song,[1,2] Stephanie K. Smith,[2] Gregory J. Hannon,[1] Leemor Joshua-Tor[1,2]*

Argonaute proteins and small interfering RNAs (siRNAs) are the known signature components of the RNA interference effector complex RNA-induced silencing complex (RISC). However, the identity of "Slicer," the enzyme that cleaves the messenger RNA (mRNA) as directed by the siRNA, has not been resolved. Here, we report the crystal structure of the Argonaute protein from *Pyrococcus furiosus* at 2.25 angstrom resolution. The structure reveals a crescent-shaped base made up of the amino-terminal, middle, and PIWI domains. The Piwi Argonaute Zwille (PAZ) domain is held above the base by a "stalk"-like region. The PIWI domain (named for the protein piwi) is similar to ribonuclease H, with a conserved active site aspartate-aspartate-glutamate motif, strongly implicating Argonaute as "Slicer." The architecture of the molecule and the placement of the PAZ and PIWI domains define a groove for substrate binding and suggest a mechanism for siRNA-guided mRNA cleavage.

RNA interference (RNAi) is triggered by the presence of double-stranded RNA (dsRNA) (*1*). A ribonuclease (RNase) III family enzyme, Dicer, initiates silencing by releasing ~20 base duplexes, with two-nucleotide 3′ overhangs called siRNAs (*2*, *3*). The RNAi pathway also mediates the function of endogenous, noncoding regulatory RNAs called microRNAs (miRNAs) [reviewed in (*4*)]. Both miRNAs and siRNAs guide substrate selection by similar if not identical effector complexes called RISC (*4*). These contain single-stranded versions of the small RNA and additional protein components (*5*–*7*). Of those, the signature element, which virtually defines a RISC, is a member of the Argonaute family of proteins (*8*).

Argonaute proteins are defined by the presence of PAZ and PIWI domains (*9*). Recent structural and biochemical analyses of the PAZ domain have begun to reveal Argonaute as the protein that interacts directly with the small RNA in RISC (*10*–*14*). The PAZ domain forms a deviant oligonucleotide/oligosaccharide-binding (OB) fold containing a central cleft lined with conserved aromatic residues that bind specifically to single-stranded 3′ ends (*10*, *12*). This was confirmed by subsequent structural studies of PAZ complexed with nucleic acids (*13*, *14*). On the basis of these studies, we first proposed a model in which the PAZ domain interacts with the 3′ ends of siRNAs in the two proteins containing this domain, Dicer and Argonaute (*10*). In RISC, the Argonaute PAZ domain would hold the 3′ end of the single-stranded siRNA, perhaps orienting recognition and cleavage of mRNA substrates. However, the nuclease responsible for cleavage, dubbed "Slicer," has so far escaped identification.

In an effort to deepen our understanding of the role of Argonaute proteins in RNAi, we have conducted structural studies of a full-length Argonaute protein from *P. furiosus*.

Overall architecture. The structure of the full-length Argonaute from the archaebacterium *P. furiosus* (PfAgo) was determined by x-ray crystallography to 2.25 Å resolution (table S1). The N-terminal, middle, and PIWI domains form a crescent-shaped base, with the PIWI domain at the center of the crescent. The region following the N-terminal domain forms a "stalk" that holds the PAZ domain above the crescent and an interdomain connector cradles the four domains of the molecule (Fig. 1). This architecture forms a groove at the center of the crescent and the PAZ domain closes off the top of this groove.

The N-terminal domain consists of a long strand at the bottom of the crescent, followed by a region of a small four-stranded β sheet, three α helices, and a β hairpin, which then extends to the three-stranded antiparallel β sheet stalk. The PAZ domain (residues 152 to 275) is a globular domain that adopts an OB-like β barrel fold with an attachment of two α helices on one side of the barrel and a cleft in between. This cleft is angled toward the crescent. The middle domain (residues 362 to 544) is an α/β open sheet domain composed of a central three-stranded parallel β sheet surrounded by α helices. This domain is similar to the glucose-galactose-arabinose-ribose–binding protein family and is most similar to Lac repressor (*15*). The middle domain also has a small three-stranded β sheet on the outer surface of the crescent, connecting it to the rest of the molecule.

The PIWI domain, which is at the C terminus of Argonaute (residues 545 to 770), is the most surprising portion of the structure, as we describe below. It sits in the middle of the

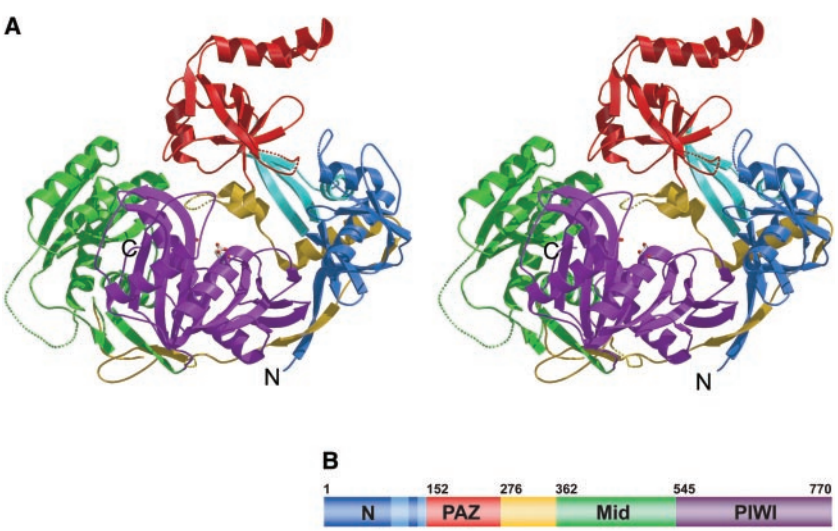

Fig. 1. Crystal structure of *P. furiosus* Argonaute. (**A**) Stereoview ribbon representation of Argonaute showing the N-terminal domain (blue), the "stalk" (light blue), the PAZ domain (red), the middle domain (green), the PIWI domain (purple), and the interdomain connector (yellow). Active site residues are drawn in stick representation. Disordered loops are drawn as dotted lines. (**B**) Schematic diagram of the domain borders.

[1]Watson School of Biological Sciences, [2]Keck Structural Biology Laboratory, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA.

*To whom correspondence should be addressed. E-mail: leemor@cshl.edu

crescent and below the PAZ domain. The crystal structure reveals the presence of a prominent central five-stranded β sheet flanked on both sides by α helices. A smaller β sheet extends from the central β sheet and attaches PIWI to the N-terminal domain and to portions of the interdomain connector.

The PAZ domain. The PAZ domain superimposes well with other PAZ domains that have known structures (*10*–*12*, *14*), although the attachment in the archael protein has two α helices rather than an α helix and a β hairpin (Fig. 2A). Other differences lie in loop regions. The root mean square deviation (RMSD) between human Argonaute-1 (hAgo1)–PAZ (*14*) and the PAZ domain in this structure is about 1.4 Å (for 53 Cα's). Despite close structural similarities, primary sequence comparisons failed to reveal a PAZ domain in PfAgo (fig. S1), whereas the presence and location of the PIWI domain was easily detected in Basic Local Alignment Search Tool (BLAST) searches.

Importantly, conserved aromatic residues that bind the two-nucleotide 3′ overhang of an siRNA (*10*, *13*, *14*) are all present in PfAgo (Fig. 2B). Curiously, in some cases, these side chains occupy similar positions in space, although they are anchored to positions on the peptide backbone differing from those in the eukaryotic proteins. Specifically, Y^{212}, Y^{216}, H^{217}, and Y^{190} of PfAgo are equivalent to Y^{309}, Y^{314}, H^{269}, and Y^{277}, respectively, of hAgo1 (*16*), which bind the oxygens of the phosphate that links the two bases in the overhang. Residue Y^{190} of PfAgo superimposes perfectly on hAgo1-Y^{277}, which binds the 2′ hydroxyl of the penultimate nucleotide. Residues L^{263} and I^{261} can assume the role of L^{337} and T^{335}, which anchor the sugar ring of the terminal residue through van der Waals interactions. An aromatic residue, F^{292} in hAgo1 stacks against the terminal nucleotide. This position is occupied by another aromatic, W^{213}, in PfAgo. Finally, R^{220} in our structure is positioned similarly to K^{313} that contacts the penultimate nucleotide. Residues that bind other regions of the RNA include K^{191} (R^{278} in hAgo1) and Y^{259} (K^{333} in hAgo1) to bind phosphates. Additional PAZ residues, such as K^{252}, K^{248}, Q^{276}, and N^{176} are probably also used to bind that siRNA strand. We therefore reason that the PAZ domain in PfAgo binds RNA 3′ ends, as do PAZ domains of fly and human Argonautes.

PIWI is an RNase H domain. The PIWI domain core has a tertiary structure belonging to the RNase H family of enzymes. This fold is also characteristic of other enzymes with nuclease or polynucleotidyl transferase activities, such as human immunodeficiency virus and avian sarcoma virus integrases (*17*, *18*), RuvC (a Holliday junction endonuclease) (*19*), and transposases such as Mu (*20*) and Tn5 (*21*). The closest matches, however, are

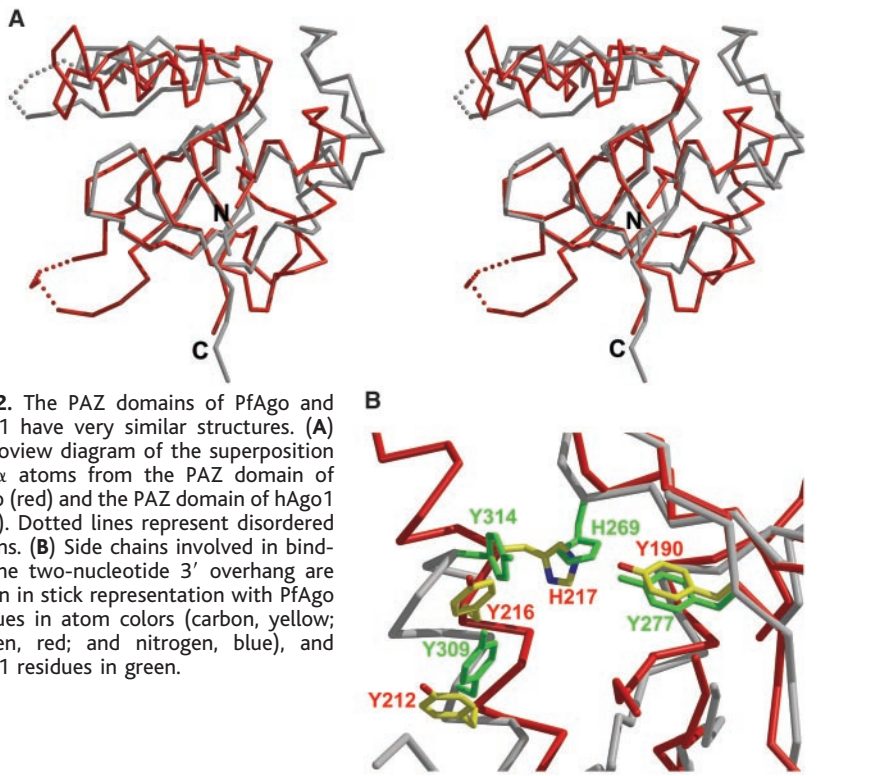

Fig. 2. The PAZ domains of PfAgo and hAgo1 have very similar structures. (**A**) Stereoview diagram of the superposition of Cα atoms from the PAZ domain of PfAgo (red) and the PAZ domain of hAgo1 (gray). Dotted lines represent disordered regions. (**B**) Side chains involved in binding the two-nucleotide 3′ overhang are shown in stick representation with PfAgo residues in atom colors (carbon, yellow; oxygen, red; and nitrogen, blue), and hAgo1 residues in green.

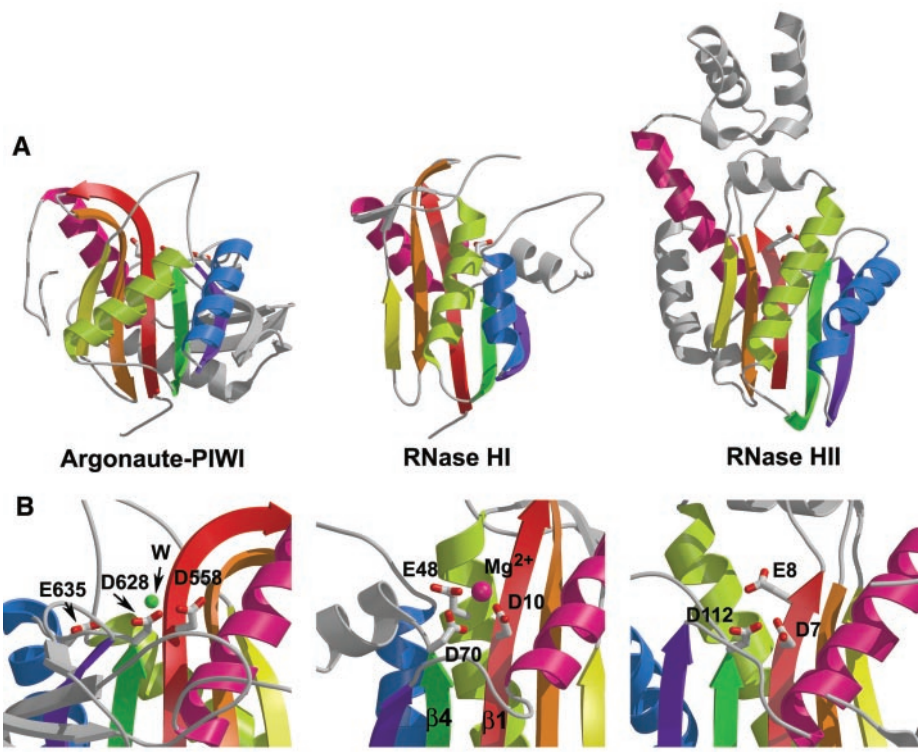

Fig. 3. PIWI is an RNase H domain. (**A**) Ribbon diagrams of the PIWI domain, *Escherichia coli* RNase HI and *Methanococcus jannaschii* RNase HII. The three structures are shown in a similar view with the secondary structure elements of the canonical RNase H fold in color. Active site residues are shown in stick representation. (**B**) This view of the active site residues is rotated ~180° about the *y* axis compared with the view in (A). The Mg^{2+} ion in RNase HI is shown as a pink sphere. A strong difference electron density (>4.5σ) found in the active site of PIWI that was assigned as a water molecule is shown as a green sphere. Secondary structural elements of the RNase H fold are colored from red to pink (red, orange, yellow, green, blue, purple, pink) as ordered in the protein sequence.

with RNase HII (*22*) and RNase H1 (*23*). The domains are topologically identical: The RMSDs between RNase HII and PfAgo (for 74 Cα's) and between RNase HI and PfAgo (for 66 Cα's) are both 1.9 Å (Fig. 3A). RNase H fold proteins all have a five-stranded mixed β sheet surrounded by helices. PIWI has an insertion between the last strand and the last helix of the RNase H fold that links it to the rest of the protein.

All of these enzymes contain three highly conserved catalytic carboxylates, which comprise the DDE motif (*24*). Two of these side chains are always located on the first strand, β1, which is the central strand of the β sheet, and at the C terminus of the fourth strand, β4, of the RNase H fold, which is adjacent to β1. The position of the third carboxylate varies. Notably, two aspartate residues in PIWI were located at the same positions as the invariant carboxylates of the RNase H fold (Fig. 3B). These are D^{558}, located on the first β strand, and D^{628}, located at the end of the fourth strand of the PIWI domain. The only requirement for the third variable carboxylate is a reasonable spatial position at the active site. E^{635} is in close proximity to the two aspartates and we suggest that this glutamate serves as the third active-site residue. This residue is positioned on the second helix of the RNase H fold of PIWI (the blue helix in Fig. 3). These three residues are almost completely invariant in the 136 Argonaute protein sequences examined (fig. S2). Interestingly, an arginine, R^{627}, is also positioned at the center of the active site, as in the case of the IS4 family of transposases such as Tn5, which appear to have a DDRE motif (*21*). The active site is thus positioned in a cleft in the middle of the crescent in the groove below the PAZ domain.

Ago as Slicer. The observation that the PIWI domain in Argonaute is an RNase H domain immediately implicated Argonaute as Slicer, the enzyme in RISC that cleaves the mRNA. RNase H enzymes cleave single-stranded RNA "guided" by the DNA strand in an RNA/DNA hybrid. Similarly, Argonautes might specialize in RNA cleavage, guided by the siRNA strand in a dsRNA substrate. Moreover, RNase H enzymes produce products with 3′-OH and 5′ phosphate groups (*25*), in agreement with the products of mRNA cleavage by RISC (*26*, *27*). A dependence on Mg^{2+} for activity is another hallmark of RNase H enzymes, a requirement that RISC shares (*27*). The PAZ domain, recognizing the 3′ ends of siRNAs, and the PIWI domain, now shown to be an RNase H domain, combine the necessary features of the slicing component of the RNAi machinery. Therefore, Argonaute, the signature component of RISC, appears to be Slicer itself.

A model for siRNA-guided mRNA cleavage. The overall structure of Argonaute defines a distinct groove through the protein, which has a claw shape and bends between the PAZ and N-terminal domains. A notable feature of the structure is evident when the electrostatic potential is mapped on the surface of the protein. As shown in Fig. 4A, the surface of this inner groove is lined with positive charges suitable for interaction with the negatively charged phosphate backbone and with the 2′ hydroxyl moieties of RNA, implicating the groove for substrate binding.

To examine possible substrate binding modes for Argonaute, we superimposed the PAZ domains from PfAgo and hAgo1 (*14*) and examined the position of the RNA in the hAgo1 complex with respect to PfAgo. The strand that interacts with the PAZ cleft is the siRNA guide (Fig. 4B). Leaving the cleft, the RNA tracks the top of the PAZ β barrel allowing similar, if not identical, interactions with the PfAgo PAZ, as observed in the hAgo1 PAZ-RNA complex. A long loop present in the PfAgo PAZ domain would probably move up slightly to accommodate the siRNA. The other strand would represent the mRNA substrate and would enter the binding groove with its 5′ end between the PAZ and N-terminal domains, with the latter acting as an "mRNA grip." Another extension of the groove lies between the N-terminal and the PIWI domains and could accommodate a single-stranded nucleic acid. However, the function of this feature is presently difficult to predict.

Extending the dsRNA further into the molecule along the binding groove by model building positions the mRNA above the active site located in the PIWI domain, nine nucleotides from the 5′ side end of the double-stranded region. On the basis of this model, the scissile bond falls between nucleotides 11 and 12, counting from the 3′ end of the guide. This precisely coincides with the demonstrated cleavage of mRNAs by RISC, 10 nucleotides from the 5′ end of a 21-nucleotide siRNA. The remainder of the RNA would continue along the binding groove (Fig. 4C). The length of the groove appears sufficient to accommodate the entire siRNA guide, with the 5′ end of the guide probably interacting with the other side of the groove to sense the siRNA 5′ phosphate. Additionally, from studies of other RNase H enzymes, we expect Argonaute to sense the minor groove width of the dsRNA, which differs from that of dsDNA and from an RNA/DNA hybrid. Such a hypothesis is in accord with the inability of RISC to cut DNA substrates (*26*). Although

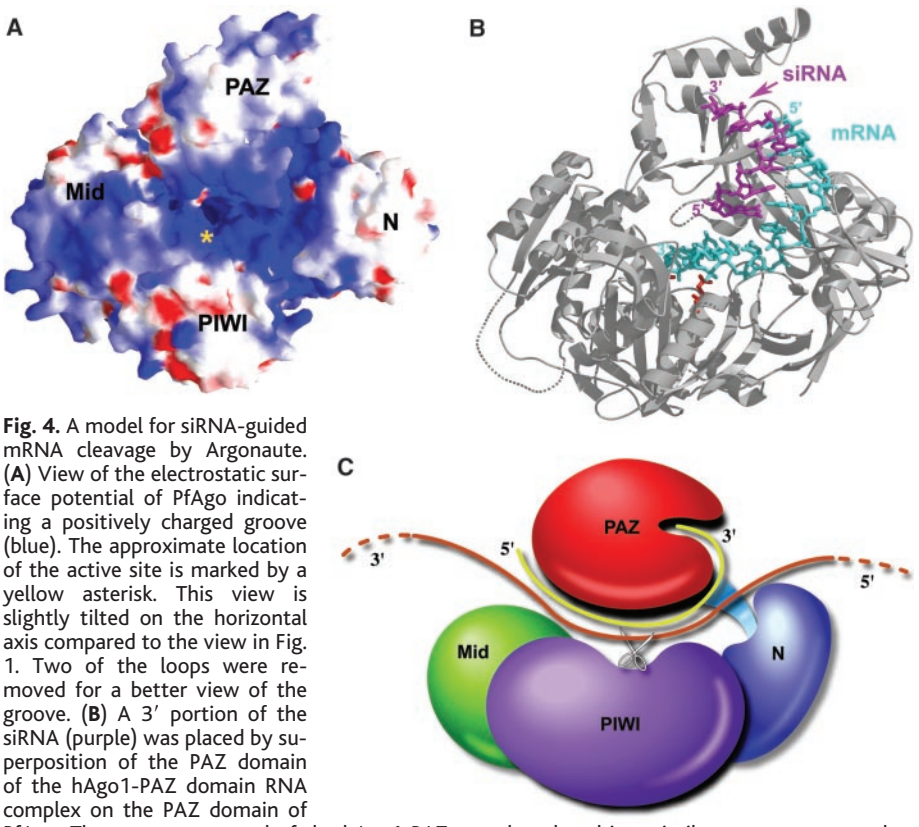

Fig. 4. A model for siRNA-guided mRNA cleavage by Argonaute. (**A**) View of the electrostatic surface potential of PfAgo indicating a positively charged groove (blue). The approximate location of the active site is marked by a yellow asterisk. This view is slightly tilted on the horizontal axis compared to the view in Fig. 1. Two of the loops were removed for a better view of the groove. (**B**) A 3′ portion of the siRNA (purple) was placed by superposition of the PAZ domain of the hAgo1-PAZ domain RNA complex on the PAZ domain of PfAgo. The passenger strand of the hAgo1-PAZ complex placed in a similar manner was used to model the mRNA strand (light blue) by extending the RNA two nucleotides at the 5′ end, and from the middle of that strand along the binding groove toward the active site in PIWI. The phosphate between nucleotides 11 and 12 from the 5′ end of the mRNA falls near the active site residues (red). The view is similar to the view in Fig. 1. (**C**) Schematic depiction of the model for siRNA-guided mRNA cleavage. The domains are colored as in Fig. 1. The siRNA (yellow) binds with its 3′ end in the PAZ cleft and the 5′ is predicted to bind near the other end of the cleft. The mRNA (brown) comes in between the N-terminal and PAZ domains and out between the PAZ and middle domain. The active site in the PIWI domain (shown as scissors) cleaves the mRNA opposite the middle of the siRNA guide.

a single-stranded RNA should bind fairly readily, opening the claw of the molecule might assist binding the mRNA, after which Argonaute might close on the double-stranded substrate. A possible hinge region exists in the interdomain connector at residues 317 to 320. This hinge could lift the PAZ away from the crescent base, perhaps allowing the RISC loading complex to assist in assembling an active complex (*28*, *29*).

The notion that RISC "Slicer" activity resides in Argonaute itself was tested in a mammalian system, by mutational analysis of hAgo2 (*30*). Conserved active site aspartates in hAgo2 were altered and the mutants lost nuclease activity but retained siRNA binding. This supports the model in which Argonaute itself functions as the Slicer enzyme in the RNAi pathway.

Many questions regarding the details of the mechanism for siRNA-guided mRNA cleavage remain. Several Argonaute protein family members appear to be inactive toward mRNA cleavage despite the presence of the catalytic residues. This situation might be analogous to the case of the Tn5 transposase and its inhibitor, which possess a catalytic domain with an RNase H–like fold. Tn5 inhibitor is a truncated version of the active Tn5 transposase and retains essential catalytic residues. However, there are major conformational differences between the two (*21*). Mutations have been introduced into a catalytically active Ago protein, hAgo2, in the vicinity of the active site, which change residues to corresponding residues in an inactive Ago, hAgo1. These inactivate Ago2 for cleavage, indicating that there are determinants for catalysis beyond simply the catalytic triad and that relatively minor alterations in the PIWI domain can have profound effects on its activity toward RNA substrates. In addition, interactions with other factors may be needed to create a fully active Slicer. The common fold in the catalytic domain of Argonaute family members and transposases and integrases is also intriguing given the relationship of RNAi with control of transposition. Notably, the identification of the catalytic center of RISC awaited a drive toward understanding RNAi at a structural level. Thus, it seems likely that, as in the present example, a full understanding of the underlying mechanism of RNAi will derive from a combination of detailed biochemical and structural studies of RISC.

References and Notes

1. A. Fire *et al*., *Nature* **391**, 806 (1998).
2. S. M. Elbashir, J. Martinez, A. Patkaniowska, W. Lendeckel, T. Tuschl, *EMBO J.* **20**, 6877 (2001).
3. E. Bernstein, A. A. Caudy, S. M. Hammond, G. J. Hannon, *Nature* **409**, 363 (2001).
4. D. P. Bartel, *Cell* **116**, 281 (2004).
5. A. Nykanen, B. Haley, P. D. Zamore, *Cell* **107**, 309 (2001).
6. D. S. Schwarz, G. Hutvagner, B. Haley, P. D. Zamore, *Mol. Cell* **10**, 537 (2002).
7. J. Martinez, A. Patkaniowska, H. Urlaub, R. Luhrmann, T. Tuschl, *Cell* **110**, 563 (2002).
8. S. M. Hammond, E. Bernstein, D. Beach, G. J. Hannon, *Nature* **404**, 293 (2000).
9. L. Cerutti, N. Mian, A. Bateman, *Trends Biochem. Sci.* **25**, 481 (2000).
10. J. J. Song *et al*., *Nature Struct. Biol.* **10**, 1026 (2003).
11. K. S. Yan *et al*., *Nature* **426**, 468 (2003).
12. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nature* **426**, 465 (2003).
13. A. Lingel, B. Simon, E. Izaurralde, M. Sattler, *Nature Struct. Mol. Biol.* **11**, 576 (2004).
14. J. B. Ma, K. Ye, D. J. Patel, *Nature* **429**, 318 (2004).
15. A. M. Friedman, T. O. Fischmann, T. A. Steitz, *Science* **268**, 1721 (1995).
16. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
17. F. Dyda *et al*., *Science* **266**, 1981 (1994).
18. J. Lubkowski *et al*., *Biochemistry* **38**, 13512 (1999).
19. M. Ariyoshi *et al*., *Cell* **78**, 1063 (1994).
20. P. Rice, K. Mizuuchi, *Cell* **82**, 209 (1995).
21. D. R. Davies, I. Y. Goryshin, W. S. Reznikoff, I. Rayment, *Science* **289**, 77 (2000).
22. L. Lai, H. Yokota, L. W. Hung, R. Kim, S. H. Kim, *Struct. Fold Des.* **8**, 897 (2000).
23. K. Katayanagi, M. Okumura, K. Morikawa, *Proteins* **17**, 337 (1993).
24. W. Yang, T. A. Steitz, *Structure* **3**, 131 (1995).
25. U. Wintersberger, *Pharmacol. Ther.* **48**, 259 (1990).
26. J. Martinez, T. Tuschl, *Genes Dev.* **18**, 975 (2004).
27. D. S. Schwarz, Y. Tomari, P. D. Zamore, *Curr. Biol.* **14**, 787 (2004).
28. Y. Tomari *et al*., *Cell* **116**, 831 (2004).
29. J. W. Pham, J. L. Pellino, Y. S. Lee, R. W. Carthew, E. J. Sontheimer, *Cell* **117**, 83 (2004).
30. J. Liu *et al*., *Science* **305**, 1437 (2004); published online 29 July 2004 (10.1126/science.1102513).

31. We thank E. Enemark for help with data collection, R. Martienssen and members of the Joshua-Tor laboratory for helpful discussions, and M. Becker (beamline X25) for support with data collection at the National Synchrotron Light Source (NSLS). The NSLS is supported by the U.S. Department of Energy, Division of Material Sciences and Division of Chemical Sciences. J.-J.S. is a Bristol-Myers Squibb Predoctoral Fellow. Coordinates have been deposited in the Protein Data Bank (accession code 1U04).

Supporting Online Material

www.sciencemag.org/cgi/content/full/1102514/DC1
Materials and Methods
Figs. S1 to S3
Table S1
References

8 July 2004; accepted 21 July 2004
Published online 29 July 2004;
10.1126/science.1102514
Include this information when citing this paper.

Argonaute2 Is the Catalytic Engine of Mammalian RNAi

Jidong Liu,[1]* Michelle A. Carmell,[1,2]* Fabiola V. Rivas,[1] Carolyn G. Marsden,[1] J. Michael Thomson,[3] Ji-Joon Song,[1] Scott M. Hammond,[3] Leemor Joshua-Tor,[1] Gregory J. Hannon[1]†

Gene silencing through RNA interference (RNAi) is carried out by RISC, the RNA-induced silencing complex. RISC contains two signature components, small interfering RNAs (siRNAs) and Argonaute family proteins. Here, we show that the multiple Argonaute proteins present in mammals are both biologically and biochemically distinct, with a single mammalian family member, Argonaute2, being responsible for messenger RNA cleavage activity. This protein is essential for mouse development, and cells lacking Argonaute2 are unable to mount an experimental response to siRNAs. Mutations within a cryptic ribonuclease H domain within Argonaute2, as identified by comparison with the structure of an archeal Argonaute protein, inactivate RISC. Thus, our evidence supports a model in which Argonaute contributes "Slicer" activity to RISC, providing the catalytic engine for RNAi.

The presence of double-stranded RNA (dsRNA) in most eukaryotic cells provokes a sequence-specific silencing response known as RNA interference (RNAi) (*1*, *2*). The dsRNA trigger of this process can be derived from exogenous sources or transcribed from endogenous noncoding RNA genes that produce microRNAs (miRNAs) (*1*, *3*).

RNAi begins with the conversion of dsRNA silencing triggers into small RNAs of ~21 to 26 nucleotides (nts) in length (*4*). This is accomplished by the processing of triggers by specialized ribonuclease III (RNase III)–family nucleases, Dicer and Drosha (*5*, *6*). Resulting small RNAs join an effector complex, known as RISC (RNA-induced silencing complex) (*7*). Silencing by RISC can occur through several mechanisms. In flies, plants, and fungi, dsRNAs can trigger chromatin remodeling and transcriptional gene silencing (*8*–*11*). RISC can also interfere with protein synthesis, and this is the predominant mechanism used by miRNAs in mammals (*12*, *13*). However, the best studied mode of

[1]Cold Spring Harbor Laboratory, Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA. [2]Program in Genetics, Stony Brook University, Stony Brook, NY 11794, USA. [3]Department of Cell and Developmental Biology, University of North Carolina, Chapel Hill, NC 27599, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: hannon@cshl.edu