Novel modes of protein–RNA recognition in the RNAi pathway
Andreas Lingel and Michael Sattler

Gene silencing mediated by RNA interference (RNAi) depends on short interfering RNAs (siRNAs) and micro RNAs (miRNAs). These RNAs have unique features, namely a defined size of 19–21 base pairs, and characteristic two-nucleotide single-stranded 3’ overhangs and 5’ monophosphate groups. These molecular features of siRNAs and miRNAs are produced by RNase III enzymes, which are a hallmark of gene silencing induced by double-stranded RNA. Recent structural studies of components of the RNAi pathway, including PAZ, Piwi and RNase III domains, as well as full-length Argonaute and viral p19 proteins, have revealed distinct and novel modes of sequence-independent recognition of the characteristic features of siRNAs and miRNAs in the RNAi pathway.

Addresses
EMBL Heidelberg, Meyerhofstrasse 1, 69117 Heidelberg, Germany
Corresponding author: Sattler, Michael (sattler@embl.de)

This review comes from a themed issue on Protein–nucleic acid interactions
Edited by Aneel K Aggarwal and Jennifer A Doudna
Available online 26th January 2005
0959-440X/5 – see front matter © 2005 Elsevier Ltd. All rights reserved.
DOI 10.1016/j.sbi.2005.01.010

Abbreviations
ds double-stranded
dsRBD dsRNA-binding domain
$K_d$ dissociation constant
miRNA micro RNA
OB-fold oligonucleotide/oligosaccharide-binding fold
PAZ Piwi, Argonaute, Zwille
PTGS post-transcriptional gene silencing
RISC RNA-induced silencing complex
RNAi RNA interference
siRNA short interfering RNA

Introduction
RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) is an evolutionarily conserved cellular response to the presence of double-stranded (ds) RNA [1,2,3,4,5]. RNAi is thought to originate from an ancient endogenous defense mechanism against viral and other heterologous dsRNAs [6,7]. Gene silencing induced by dsRNA is widely used as a tool to study gene function in higher eukaryotes [8]. In the cell, the dsRNA is first cleaved by the RNase III enzyme Dicer into short interfering RNA duplexes (siRNAs) of 21–23 nucleotides (nt) with 3’ overhangs of 2 nt (Figure 1a). In a subsequent step, one strand of the siRNA is incorporated into a multimeric RNA-induced silencing complex (RISC), where it guides the selection of a complementary mRNA. The evolutionarily conserved family of Argonaute proteins are essential components of the RISC. An endonuclease in the RISC effector complex (termed Slicer) cleaves the mRNA in the region that has sequence complementarity to the siRNA guide. Subsequently, the mRNA initially cleaved by Slicer is degraded by exonucleases and thus silenced.

A related pathway of PTGS involves endogenously expressed non-coding RNAs of 21–25 nt called micro RNAs (miRNAs). These RNAs are derived from long primary transcripts termed pri-miRNAs, which are cleaved by the RNase III enzyme Drosha into precursor miRNA (pre-miRNA) hairpins of about 70 nt with a 3’ overhang of 2 nt (Figure 1a) [9]. Double-stranded miRNAs are excised from pre-miRNAs by Dicer. As products of RNase III cleavage, the resulting double-stranded miRNAs have 2-nt 3’ overhangs and 5’ monophosphate groups, similar to siRNAs. In contrast to siRNAs, which guide endonuclease cleavage, the single-stranded miRNA guide shows imperfect sequence complementarity to a target mRNA (usually in 3’ untranslated regions) and leads to the inhibition of translation. However, siRNAs with imperfect sequence complementarity have also been shown to effect translational inhibition and, conversely, miRNAs with full sequence complementarity can induce mRNA degradation (reviewed in [1,2]). Thus, it is thought that the siRNA and miRNA pathways lead to the assembly of related effector complexes and that the mode of gene silencing (mRNA degradation or inhibition of translation) depends on the degree of sequence complementarity between the siRNA or miRNA guide and the mRNA.

A third but less well characterized dsRNA-induced pathway mediates chromatin-based gene silencing at the level of transcription in plants, yeast and metazoa [10,11]. Finally, viruses have developed mechanisms to counteract RNAi by providing proteins that sequester siRNAs targeted against viral RNAs [6,12,13].

The Dicer and Argonaute protein families constitute crucial components of the RNAi pathway. Members of the Dicer family are multidomain proteins of about 200 kDa that comprise two RNase III domains, a PAZ domain [named after the proteins Piwi, Argonaute and Zwille (also known as Pinhead)], a dsRNA-binding
domain (dsRBD), a domain of unknown function (DUF283) and a helicase domain (Figure 1b) [14,15**]. The Argonaute family comprises highly basic proteins of about 100 kDa that are characterized by a central PAZ domain and a C-terminal Piwi domain [16] (Figure 1b). The genomes of higher eukaryotic organisms encode about 100 kDa that are characterized by a central PAZ domain and a C-terminal Piwi domain [16] (Figure 1b). The PAZ domain is found exclusively in the Argonaute and Dicer protein families [16,17], and was initially hypothesized to mediate protein–protein interactions because Argonaute and Dicer proteins co-immunoprecipitate. Recent three-dimensional structures of the Drosophila Ago1 and Ago2 PAZ domains have revealed a central five-stranded β barrel flanked by two α helices, and a conserved inserted module of about 35 residues comprising a β hairpin and α helix (Figure 2) [18**–20**]. Unexpectedly, these PAZ domains have been found to bind nucleic acids in vitro [18**–20**]. ssRNAs and dsRNAs with single-stranded 3’ overhangs bind strongly, whereas blunt-ended dsRNA molecules show a reduced affinity for PAZ domains [19**,20**].

Nucleic acid 3’-end recognition by the PAZ domain

The PAZ domain is found exclusively in the Argonaute and Dicer protein families [16,17], and was initially hypothesized to mediate protein–protein interactions because Argonaute and Dicer proteins co-immunoprecipitate. Recent three-dimensional structures of the Drosophila Ago1 and Ago2 PAZ domains have revealed a central five-stranded β barrel flanked by two α helices, and a conserved inserted module of about 35 residues comprising a β hairpin and α helix (Figure 2) [18**–20**]. Unexpectedly, these PAZ domains have been found to bind nucleic acids in vitro [18**–20**]. ssRNAs and dsRNAs with single-stranded 3’ overhangs bind strongly, whereas blunt-ended dsRNA molecules show a reduced affinity for PAZ domains [19**,20**].

Using NMR chemical shift perturbation and mutational analysis, the nucleic-acid-binding site of the PAZ domain has been mapped to a hydrophobic cleft between the central β barrel and the conserved β hairpin/α helix module [18**–20**]. The conservation of residues in this region suggests that it performs a similar function in all PAZ domains. Indeed, conserved nucleic acid binding has been observed for PAZ domains from the Drosophila Ago1 (DmAgo1), Ago2 (DmAgo2), Piwi...
and Aubergine proteins, and the human Ago1 \((Hs\text{Ago1};\) also known as eIF2C1), Ago2 \((Hs\text{Ago2};\) also known as eIF2C2) and Piwi proteins \([18^-21^-];\) A Lingel and E Izaurralde, unpublished).

The molecular basis of nucleic acid recognition has been revealed by the crystal structure of the \(Hs\text{Ago1 PAZ domain}\) in complex with an siRNA mimic \([21^-];\) and by NMR solution structures of the \(Dm\text{Ago2 PAZ domain}\).
bound to single-stranded RNA and DNA oligonucleotides (Figure 2a,b) [22•]. In all structures, the 3′ end of the nucleic acid is bound in the hydrophobic cleft of the PAZ domain without sequence specificity. Recognition of the 3′-terminal nucleotide involves a stacking interaction with a conserved aromatic residue (Phe292 and Phe72 in HsaGao1 and DmaGao2, respectively) of the β hairpin/α helix module (Figure 2a,b). The phosphate groups of the two 3′-terminal nucleotides interact with tyrosine and histidine residues in the HsaGao1•RNA (Tyr309, Tyr314, His269 and Tyr277) and DmaGao2•RNA (Tyr44) complexes. It is conceivable that some of these residues (such as Tyr314 and His269, which are not conserved in DmaGao2) contribute to the selectivity of most PAZ domains for RNA over DNA. This may also be linked to the different conformations of the sugar–phosphate backbone in the complexes of DmaGao2 with RNA and DNA [22•].

Very few contacts involve the sugar of the 3′-terminal nucleotide, indicating that the 3′ end of the nucleic acid is specified mainly by steric exclusion. In complexes of the DmaGao2 PAZ domain with single-stranded nucleic acids, the penultimate nucleotide interacts with conserved residues Tyr57 and Lys107. Notably, this binding pocket for the penultimate nucleotide interacts with conserved residues Tyr314 and His269, which are not conserved in DmaGao2 and HsaGao1. This observation indicates that there is conformational plasticity with respect to the location of the penultimate nucleotide. The mode of recognition of this nucleotide may depend on whether the upstream nucleic acid adopts a single-stranded or double-stranded conformation.

In both structures, the double-stranded stem of the different siRNA mimics extends along the β3 strand of the central β barrel of the PAZ domain [21••,22•]. The phosphodiester backbone of the RNA strand, which is bound via its 3′ end, mediates numerous contacts to positively charged arginine and lysine sidechains [21••]. These additional contacts depend on the presence of the dsRNA stem and contribute substantially to the binding affinity; extension of the 3′ single-stranded overhang from two to ten uridines reduces the interaction 50-fold (corresponding to a change in the dissociation constant, $K_d$, from ~2 nM to 100 nM). Notably, the complementary RNA strand, including its 5′ end, does not seem to be recognized by the PAZ domain [21••].

Taken together, these structural studies show that the PAZ domain binds the 3′ end of single-stranded nucleic acids by burying it in a hydrophobic cleft between the characteristic β hairpin/α helix module and a central β barrel. The mode of recognition of the two terminal nucleotides is unique and distinct from that used by other single-stranded nucleic acid binding domains [23]. The burial of the 3′ end by the monomeric PAZ domain also differs from 3′-end recognition by oligonucleotide/oligosaccharide-binding fold (OB-fold) domains [24,25], such as the recognition of telomeric DNA by the OB-fold of Schizosaccharomyces pombe Pot1 (Figure 2c) [26].

Intriguingly, in the crystal structure of a PAZ domain with an siRNA mimic [21••], the protein interacts almost exclusively with the 3′-anchored strand. This observation suggests that the PAZ domain might stay bound to this strand after the complementary strand is removed, marking it as the siRNA guide in an intermediate RISC loading complex (see Figure 1a).

### Piwi, RNase H and the structure of a full-length Argonaute protein

Recently, the crystal structure of a full-length Argonaute protein from the archaea Pyrococcus furiosus (P/Ago) has been solved at a resolution of 2.25 Å [27*] (see also Update). The structure comprises four domains, including a PAZ domain that had not been predicted from the amino acid sequence (Figure 2d). Notably, the β hairpin/α helix module found in the PAZ domains of eukaryotic Argonaute proteins (see above) is replaced by two helices in P/Ago (Figure 2d). Nevertheless, residues that are important for nucleic acid binding are found in similar locations, despite the partially different secondary structure of the insertion module.

The Piwi domain of Argonaute strongly resembles the fold of RNase H endonucleases, comprising a five-stranded mixed β sheet surrounded by α helices on both sides (Figure 3a,b). RNase H enzymes are known to cleave the RNA strand of an RNA–DNA hybrid [28]. Catalysis depends on divalent cations, and mechanisms based on one or two metal ions have been proposed [29–31]. The reaction mechanism of RNase H enzymes differs from that of most RNases but resembles that of DNases, producing 3′ hydroxyl and 5′ phosphate termini after cleavage [32]. Intriguingly, Slicer activity in the RISC has been shown to have a similar catalytic requirement for divalent cations (such as Mg$^{2+}$ or Mn$^{2+}$) and identical product chemistry (3′ hydroxyl and 5′ phosphate) [33,34].

The RNase H active site comprises a so-called ‘DDE motif’, which consists of three acidic amino acids whose sidechain carboxylates are positioned to catalyze the cleavage reaction. In the Piwi domain of P/Ago, two of the potential catalytic residues (Asp558 and Asp628) are in positions that are structurally equivalent to those found in the RNase H fold (Figure 3a,b). The active site may be completed by a third carboxylate from a conserved
glutamate (Glu635 in PfAgo Piwi), even though this residue is located in a different sequence position in bacterial proteins (e.g. Glu48 in Escherichia coli RNase H1). A fourth residue, Arg627, is also positioned to contribute to the Piwi active site, in analogy to the active site of Tn5 transposase [35].

Strikingly, in the structure of full-length PfAgo [27**, the putative catalytic center is located at the end of a groove, about 30 Å away from the presumed binding site for the 3’ end of the nucleotide in the PAZ domain (Figure 3c). This distance corresponds to about 10 bp of an A-form RNA helix. It is conceivable that the PAZ domain could bind the 3’ end of the siRNA guide within a siRNA–mRNA duplex, thereby placing the mRNA above the Piwi active site. In such a model [27**], the scissile phosphodiester bond of the mRNA would be located near the center of the region spanned by the guiding siRNA, identical to the site of RISC-induced cleavage [36]. Further evidence suggesting that the Piwi domain has enzymatic activity has come from the mutation of predicted catalytic residues of the DDE motif in HsAgo2, which abolishes mRNA cleavage activity in human cells [37*]. Likewise, mutation of an arginine residue equivalent to the above-mentioned Arg627 in PfAgo eliminates mRNA cleavage activity in the Ago1 protein of trypanosomes [38]. Notably, mutation of two HsAgo2 residues to the corresponding amino acids in HsAgo1 (Gln633Arg and His634Pro) abolishes Slicer activity [37*]. These residues are located in a region similar to the ‘primer grip’ of the RNase H domain of HIV-1 reverse transcriptase, where they contribute to the catalytic activity [39].

Taken together, the structure of the full-length PfAgo protein [27**] and the biochemical studies [37*,40] indicate that an Argonaute protein is responsible for the Slicer activity of the RISC (see also Update). This finding is a significant step towards understanding the molecular mechanisms of RNAi. Clearly, the next challenge will be to reveal the structural details of Argonaute–RNA complexes at atomic resolution.

Dicer and RNase III enzymes

Dicer contains a dsRBD and a helicase domain, in addition to two RNase III domains, a PAZ domain and a DUF238 domain (Figure 1b). The recognition of dsRNA by members of the dsRBD family has been reviewed [41,42], and it is conceivable that similar modes of RNA recognition are used by the Dicer dsRBD. Generally, dsRBDs bind to the stem of an A-form RNA helix, accompanied by some widening of the major groove. Mismatches and bulges in the RNA can be tolerated, and may contribute to target selection by structure-specific rather than sequence-specific recognition, as has been recently observed for the dsRBD of the yeast RNase III Rntp1 [43]. The function of the RNA helicase

![Diagram](a) Crystal structure of RNase H1 from E. coli (PDB code 1RDD), showing sidechains of the catalytic DDE triad residues and a bound Mg^{2+} ion. Conserved and unconserved secondary structure elements are colored green and yellow, respectively. (b) The Piwi domain in the crystal structure of PfAgo protein (PDB code 1U04). Sidechains of putative residues of the catalytic DDE triad and Arg627 are shown. Structural elements are colored as in (a). (c) Crystal structure of full-length PfAgo. The PAZ domain is shown in blue, the Piwi domain is colored as in (a). Sidechains of residues implicated in endonucleolytic cleavage by Piwi (see text) and putative RNA binding by the PAZ domain are shown. Residues in the Piwi domain that might contribute to substrate binding and/or selectivity (Glu635 and Gln593, corresponding to Gln633 and His634 in HsAgo2, respectively; see text) are shown in orange.
domain in Dicer is unclear at present, but it presumably contributes to unwinding of the double-stranded siRNA precursor.

Insight into the molecular details of the RNase III activity of Dicer has been derived from the crystal structure of the endonuclease domain of the RNase III homolog from *Aquifex aeolicus* (Figure 4) [44]. This structure, together with biochemical studies, has established that the catalytically active enzyme in prokaryotes comprises a homodimer of two RNase III domains. The conservation of two clusters of carboxylic sidechains and one divalent metal ion per monomer has led to the proposal that each RNase III domain contains two catalytic sites. Because these two catalytic sites may form one processing center that cleaves two nearby phosphodiester bonds located on opposite strands of the dsRNA, it has been suggested that the RNase III homodimer contains two centers for processing dsRNA [44]. Notably, the Drosha and Dicer enzymes each contain two RNase III domains (RNase IIIa and RNase IIIb), which could dimerize to form an enzymatically active unit that resembles the RNase III homodimer of *A. aeolicus*.

In a recent study, Filipowicz and co-workers [15**] have shown that the RNase IIIa and RNase IIIb domains of human Dicer constitute only a single center for processing dsRNA. This single processing center comprises two catalytic sites (one from each subunit) and cleaves both strands of the dsRNA substrate at one location, leaving a 3' overhang of 2 nt (Figure 4). Filipowicz and co-workers suggest that the characteristic size (~20 bp) of the siRNAs resulting from Dicer cleavage may be linked to 3'-end recognition of the dsRNA substrate by the Dicer PAZ domain. 3’-end binding by the PAZ domain might place the processing center formed by the two RNase III domains at a location roughly 20 bp from the 3' end. This model is consistent with the known preference of Dicer for cleaving the termini of dsRNAs with 2-nt 3' overhangs [45]. The exact details of the catalytic mechanism must await, however, the structure of a Dicer–dsRNA complex.

**Interfering with RNAi: size-selective RNA binding by p19**

The p19 protein from tombusviruses suppresses RNAi in plants by sequestering siRNAs and preventing their incorporation into the RISC (Figure 1a) [12,13]. Crystal structures of a p19 homodimer bound to siRNAs [46**,**47**] reveal a C-terminal domain composed of a four-stranded β sheet flanked on one side by three α helices (Figure 5). Two N-terminal helices are docked onto the C-terminal domain by salt bridges. The secondary structure topology of the C-terminal domain corresponds to a circular permutation of the ribosomal L1 domain.
protein, which binds dsRNA as a monomer with its β-sheet surface [48]. Homodimeric p19 forms an extended concave antiparallel β-sheet surface, which mediates several interactions with the sugar–phosphate backbone of the double-stranded helical region of the siRNA (Figure 5).

Recognition of the siRNA stem by the β-sheet surface is distinct from the mode of binding of dsRBDs, in which mainly loop regions interact with the stem of the RNA duplex. The p19 homodimer binds the siRNA with very high affinity (K_d ≈ 0.17 nM) [46**]. Recognition of the siRNA is sequence nonspecific, employing hydrogen bonds, electrostatic contacts (involving phosphate and 2'-hydroxyl groups) and stacking interactions between aromatic residues and bases. Several conserved serine and threonine residues mediate key interactions with 2'-hydroxyl groups in the minor groove of the RNA duplex. These interactions presumably contribute to bending the axis of the dsRNA helix by roughly 40° towards the protein.

Size selectivity for an RNA helix of 19 bp is defined by symmetric interactions of the highly conserved residue Trp42 and the more variable residue Trp39 with the 5' and 3'-terminal bases, respectively, at both ends of the RNA duplex region of a 21-nt siRNA (Figure 5). Consistent with this, p19 does not interact with single-stranded nucleic acids or dsDNA, and has reduced binding affinity for siRNAs with longer or shorter double-stranded regions [46**,47**]. In one of the structures, the 2-nt single-stranded 3' overhang makes few interactions with p19 [46**], whereas in the other structure no electron density is observed for the two nucleotides at the 3' end [47**]. This suggests that recognition of the 2-nt overhang is not essential, a proposal that is consistent with the comparable binding affinities of p19 for a 19-bp RNA duplex with and without a 2-nt single-stranded 3' overhang [46**,47**]. By contrast, a hydrogen bond between the indole amide of the conserved Trp42 and the 5' phosphate seems to be important for binding, because the affinity of p19 for an siRNA lacking a 5' phosphate is reduced 23-fold [46**]. The major determinant of siRNA specificity is thus size-selective recognition of the 19 bp duplex — a binding mode that is unprecedented and provides a fascinating example of structure-specific recognition of the key mediators of RNAi.

**Conclusions**

Structural studies of the proteins involved in the RNAi pathway have revealed unique and novel ways of recognizing the characteristic molecular features of prototypic siRNAs, such as the size of a 19 bp RNA duplex (p19), a 2-nt 3' overhang (PAZ) and 5' phosphates (p19). Other components of the RNAi pathway may also recognize these features of siRNAs and miRNAs. For example, the export of pre-miRNAs by the nuclear export factor exportin-5 is facilitated by the presence of the 2-nt single-stranded 3' overhang [49–51]. These features are produced by RNase III enzymes, which are a hallmark of RNA-induced gene silencing. In general, the molecular recognition of siRNAs and miRNAs is not sequence
specific and, in most cases, the cognate proteins show a preference for RNA over DNA. Even though some of the protein folds involved are reminiscent of other RNA-binding motifs (e.g. PAZ and p19), they employ distinct and unique modes of RNA recognition.

The next challenges will be to understand the molecular basis of the different biochemical and biological activities of RISCs assembled on siRNAs and miRNAs. The relative thermodynamic stability of each side of the siRNA or miRNA duplex has been shown to define which strand is incorporated as a guide into the RISC [52,53]. This asymmetry may correlate with the susceptibility of the siRNA or miRNA duplex to unwinding by RNA helicases during assembly of the RISC. Furthermore, it is unclear how the degree of sequence complementarity between the siRNA or miRNA guide and the mRNA might define the mode of gene silencing. It is also unknown why only HsAgo2 shows endonuclease activity, even though the putative catalytic residues are conserved in all four human Argonaute homologs. Presumably, additional molecular interactions confer specificity and selectivity in different RISCs, but the functional and structural roles of additional RISC components are still poorly characterized. We expect that, in the future, a combination of structural and biochemical analyses will help to further clarify the mechanisms of RNAi.

**Update**

A recent publication by Parker et al. [55] describes the crystal structure of a Piwi protein from the archaea Archaeoglobus fulgidus (A/Piwi). The protein comprises two domains, which resemble the middle and Piwi domains of Pf/Ago; however, unlike Pf/Ago, it does not contain a PAZ domain. In addition to a putative RNase-H-like active site in the Piwi domain, the authors identify a conserved C-terminal region in A/Piwi, which is required for siRNA binding. The authors propose a model whereby the C-terminal region could be involved in recognition of the 5’ phosphate of the siRNA guide strand, thus positioning the target mRNA strand for cleavage above the putative active site of the Piwi domain.

**Acknowledgements**

We thank Elisa Izaurralde, Bernd Simon, Ana Messias and Leonic Unterholzner for discussions and comments on the manuscript, and Petra Riedinger for help with figures. We apologize to those colleagues whose work we could not discuss in detail owing to space limitations.

**References and recommended reading**

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest


The crystal structure of the PAZ domain of HsAgo1 bound to a small siRNA mimic shows that the characteristic 2 nt single-stranded overhang of siRNAs is recognized. Additional contacts involving the phosphates of the siRNA duplex region are required for high-affinity binding.


