THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 282, NO. 18, pp. 13824–13832, May 4, 2007

Structure of *Aquifex aeolicus* Argonaute Highlights Conformational Flexibility of the PAZ Domain as a Potential Regulator of RNA-induced Silencing Complex Function*[S]

Received for publication, September 6, 2006, and in revised form, November 21, 2006 Published, JBC Papers in Press, November 27, 2006, DOI 10.1074/jbc.M608619200

Umar Jan Rashid[‡], Dirk Paterok[§], Alexander Koglin[‡], Holger Gohlke[¶], Jacob Piehler[§], and Julian C.-H. Chen[‡1]

From the Institutes of [‡]Biophysical Chemistry, [§]Biochemistry, and [¶]Cell Biology and Neuroscience, J. W. Goethe University Frankfurt, Max-von-Laue-Strasse 9, D-60438 Frankfurt, Germany

Gene silencing mediated by RNA interference requires the sequence-specific recognition of target mRNA by the endonuclease Argonaute, the primary enzymatic component of the RNA-induced silencing complex. We report the crystal structure of *Aquifex aeolicus* Argonaute, refined at 3.2 Å resolution. Relative to recent Argonaute structures, a 24° reorientation of the PAZ domain in our structure opens a basic cleft between the N-terminal and PAZ domains, exposing the guide strand binding pocket of PAZ. This rearrangement leads to a branched, Y-shaped system of grooves that extends through the molecule and merges in a central channel containing the catalytic residues. A 5.5-ns molecular dynamics simulation of Argonaute shows a strong tendency of the PAZ and N-terminal domains to be mobile. Binding of single-stranded DNA to Argonaute monitored by total internal reflection fluorescence spectroscopy shows biphasic kinetics, also indicative of domain rearrangement upon DNA binding. Conformational rearrangement of the PAZ domain may therefore be critical for the catalytic cycle of Argonaute and the RNA-induced silencing complex.

RNA interference, an ancient mechanism of gene silencing mediated by small RNAs, is widely distributed throughout most eukaryotes (1, 2). Potent silencing of genes is achieved through the processing of double-stranded RNA (dsRNA)[2] into small interfering RNAs (siRNA) by the enzyme Dicer. The ~22-nt-long siRNAs contain a signature 5′-phosphate group and a 2-nt-long 3′-overhang (3). The siRNA is then loaded into the RNA-induced silencing complex (RISC), of which Argonaute is the primary catalytic component (4, 5). Energetic asymmetry of the siRNA ends allows for its directional loading into RISC (6–8). Argonaute cleaves the passenger strand of the siRNA, leaving the guide strand of the siRNA bound to RISC (9–11). This single-stranded guide strand siRNA bound to Argonaute is required for the sequence-dependent recognition of mRNA by RISC. Recombinantly expressed human Argonaute 2 in complex with single-stranded siRNA forms a minimal RISC, capable of mRNA recognition and cleavage (12). Recent crystal structures of Argonaute apoenzymes show the architecture of Argonaute to be a multidomain protein composed of N-terminal, PAZ, MID, and PIWI domains (13, 14).

The catalytic cycle of RISC requires the binding of a number of different nucleotide structures to Argonaute, and we expect Argonaute to undergo a number of conformational changes during the cycle of mRNA recognition by RISC (15, 16). Nevertheless, it remains unclear how the multidomain arrangement of Argonaute recognizes and distinguishes between single-stranded and double-stranded oligonucleotides that correspond to the Dicer-processed siRNA product, guide strand siRNA, and the guide strand/mRNA duplex.

We report the crystal structure of *Aquifex aeolicus* Argonaute (AaAgo), crystallized in the presence of a single-stranded siRNA and refined using data to 3.2 Å resolution. We propose a model for Argonaute activity in the context of the RISC complex. We also report a kinetic and 5.5-ns molecular dynamics characterization of Argonaute, which highlights biphasic binding kinetics with ssDNA, and a tendency for pronounced motion of the PAZ and N-terminal domains. Together, these studies highlight a functional role for domain flexibility, especially of the PAZ, in Argonaute activity that may be crucial for the function of the RISC.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—The full-length Argonaute gene from *Aquifex aeolicus* was amplified by PCR from genomic DNA and cloned into a modified pET16b vector (Novagen) containing an N-terminal His_7 tag. AaAgo was overexpressed in *Escherichia coli* host cell strain BL21 (DE3) pLysS (Stratagene) and induced by 1 mM isopropyl-1-thio-β-D-galactopyranoside (A_{600} ~ 0.8) at 37 °C for 3 h. Cells were resuspended and sonicated in 10 mM Hepes, 300 mM NaCl, pH 7.4. Non-thermostable proteins were removed by a heating step of 72 °C for 30 min. Precipitated proteins were removed by centrifugation at 20,000 rpm, and the supernatant was applied to an immobilized metal affinity column (nickel-nitrilotriacetic acid-agarose; Qiagen), with the partially purified target protein eluting at 300 mM imidazole. The N-terminal His_7 tag was

* This work was funded by the Deutsche Forschungsgemeinschaft SFB 579 and the Hessisches Ministerium für Wissenschaft und Kultur. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 2NUB) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Figs. S1 and S2 and supplemental data.

[1] To whom correspondence should be addressed. Tel.: 49-69-798-29641/30; Fax: 49-69-798-29632; E-mail: chen@chemie.uni-frankfurt.de.

[2] The abbreviations used are: dsRNA, double-stranded RNA; ssRNA, single-stranded RNA; RISC, RNA-induced silencing complex; AaAgo, *Aquifex aeolicus* Argonaute; ssDNA, single-stranded DNA.

removed by addition of TEV protease overnight at 4 °C. The target was further purified by gel filtration on a Superdex 200 column (Amersham Biosciences), equilibrated with 10 mM Hepes, 300 mM NaCl, pH 7.4. For binding experiments, the protein was purified by affinity chromatography and gel filtration in 20 mM Hepes, 150 mM KCl, pH 7.4.

Crystallization, Data Collection, and Refinement—siRNA nucleotides (5′-PO_4 GGG AAU GCG ACU CAU GAA GCU-3′) were purchased from Ambion and solubilized in 10 mM Hepes, pH 7.4. An equimolar ratio of protein at 10 mg/ml and single-stranded siRNA was mixed together and incubated at 37 °C for 15 min prior to crystallization trials. Crystals were grown by the hanging drop method at 22 °C, with 3.0 M sodium formate, 100 mM Hepes, pH 7.4, as the precipitant. Crystals grew within 2 weeks to a maximum size of ~200 × 200 × 30 μm. For cryoprotection, crystals were transferred into mother liquor containing 25% glycerol (v/v) and immediately frozen in liquid nitrogen. Data were collected at beamline ID23–1 at European Synchrotron Radiation Facility in 1° oscillations. Diffraction was observed to a useable resolution of 3.2 Å and indexed in tetragonal space group I422, with one molecule/asymmetric unit. Data were processed using MOSFLM and SCALA from the CCP4 suite (17). Although the data were weak, with reflections at an $I/\sigma(I)$ of 2 at 3.6 Å resolution, a plot of the SigmaA *versus* resolution (supplemental Fig. S1) showed that the weak data were still usable, and thus the refinement utilized all data to 3.2 Å resolution (18, 19).

The structure was solved by molecular replacement, using 1YVU as a search model. Although an unambiguous solution was initially obtained by the CCP4 program MOLREP (17), the density for the PAZ domain was poor. A second run using 1YVU excluding the PAZ domain was done in MOLREP, while placement of the PAZ domain was facilitated by EPMR (20). The model was refined by a combination of rigid body, positional, simulated annealing, and grouped and individual B-factor refinement in CNS (21) with manual rebuilding in Coot (22). B-factor refinement, though it potentially introduces a significantly higher number of parameters, is restrained in CNS and is recommended by the authors of the program, even at moderate resolution (18). Tests of grouped and individual B-factor refinement favored individual B-factor refinement in the early stages and grouped B-factor refinement in the later stages. Data collection and refinement statistics are summarized in Table 1. The choice of refinement protocols, especially in regard to B-factor refinement, is described in greater detail in the supplemental data.

Molecular Dynamics—The final crystallographic model was used for molecular dynamics simulations. First, loops of missing residues in the model were built using MODELLER (23). A 5.5-ns molecular dynamics simulation was then performed using the AMBER 8.0 (24) suite of programs together with the *parm99* force field and a modification for backbone torsion angles (25). The starting structure was placed into an orthorhombic periodic box of TIP3P water molecules (26) such that the distance between the edges of the water box and the closest protein atom was at least 10 Å. The particle mesh Ewald method (27) was used to treat long-range electrostatic interactions with a direct space non-bonded cutoff of 9 Å. Bond lengths involving bonds to hydrogen atoms were constrained using SHAKE (28). The time step for the simulation was 2 fs. After minimization, the system was gradually heated to 300 K over a period of 50 ps, after which the solvent density was adjusted during another 50 ps. Harmonic restraints applied during these steps to all solute atoms were then reduced over 50 ps. After an additional 350 ps of simulation at 300 K, the following 5.0 ns were used for analysis, considering snapshots obtained every 1 ps.

TABLE 1
Data collection and refinement statistics

Source	ESRF ID23–1
Wavelength (Å)	0.9795
Resolution (Å)[a]	3.2
Space group	I422
Unit cell (Å)	a = 125.0, c = 253.5
R_{sym}[b]	0.135 (0.797)
Redundancy independent R_{merge}[b]	0.145 (0.852)
Completeness (%)[b]	100 (100)
Measured reflections[b]	134396 (19238)
Independent reflections[b]	16989 (2443)
Redundancy[b]	7.9 (7.9)
$\langle I/\sigma(I)\rangle$[b]	4.5 (1.0)
Overall (Å^2)	65
PAZ domain	101
N-terminal / PIWI / Mid	61
R_{free} / R_{cryst} (%)	31.0 / 23.2
Root mean square deviation bond angles (°)	1.47
Root mean square deviation bond lengths (Å)	0.009
Ramachandran allowed / generous / disallowed (%)	97.0 / 2.0 / 1.0

[a] The threshold for an I / $\sigma(I)$ of 2 occurs at 3.6 Å resolution.
[b] Values in parentheses refer to values in the highest resolution shell, 3.37-3.2 Å.

Structural changes during the simulation were analyzed in terms of root mean square deviations of C_α atoms from the starting structure and variations in the C_α-C_α distances between particular residues at the top of the M groove. To analyze dynamic properties of the molecule, equal time cross-correlations of atomic fluctuations $C(i,j)$ between atoms i and j are obtained (29). For completely correlated motions, $C(i,j) = 1$, and for completely anticorrelated motions, $C(i,j) = -1$. Finally, a principal component analysis of C_α atom motions in Cartesian space was performed, yielding collective variables (also termed "modes") that describe the essential dynamics of the system (30). The modes most involved in the observed conformational change Δr between 1YVU and the structure determined here were then obtained by calculating the overlap between each mode and Δr according to Ref. 31. An overlap of one indicates that the directions of both kinds of collective C_α displacements are identical. In all cases, only those C_α atoms not contained in the N-terminal or PAZ domain were used for least-squares fitting to avoid the introduction of spurious correlations (32, 33).

Binding Kinetics—Binding of fluorescently labeled DNA to immobilized AaAgo was detected in real time by total internal reflection fluorescence spectroscopy. This technique was used in combination with reflectance interference for mass-sensitive quantification of the amount of immobilized AaAgo (34). Site-specific immobilization of the protein onto the sensor chip through its His tag was achieved by using multivalent chelator heads coupled to a polyethylene glycol polymer brush, which were loaded with Ni^{2+} ions (35).

100 nM of 21-nt-long ssDNA (5′-CGT ACG CGG AAT ACT TCG ATT-3′), 15-nt-long ssDNA (5′-CGT ACG CGG AAT

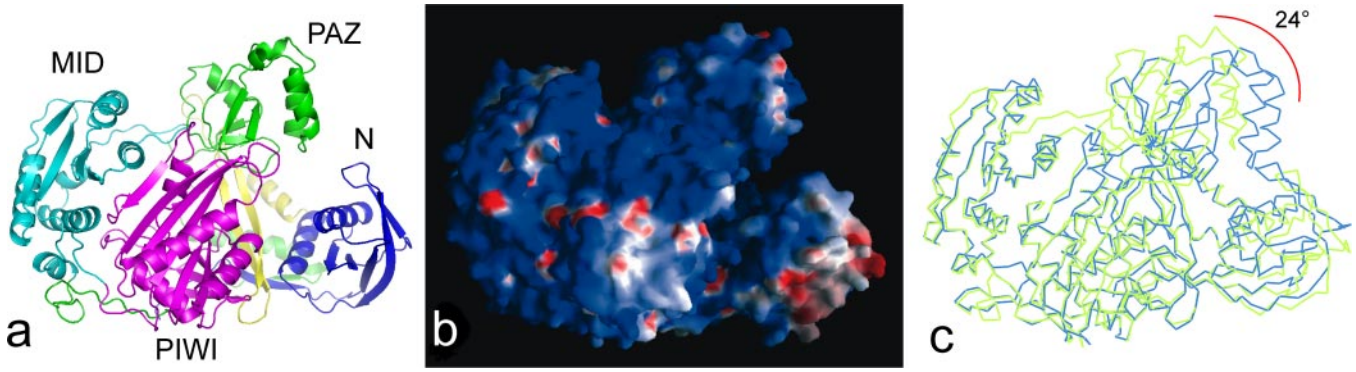

FIGURE 1. **Overall architecture of Argonaute.** *a*, structure of AaAgo, with the N-terminal (*blue*), PAZ (*green*), MID (*cyan*), and PIWI (*pink*) domains highlighted. Linker regions L1 and L2 are indicated in *yellow* and *dark green*, respectively. *b*, electrostatic surface representation of AaAgo, from −10/kT (*red*) to +10/kT (*blue*), highlighting the network of positively charged grooves. *c*, C_α trace showing 24° shift of PAZ between the structure presented here (*light green*) and 1YVU (*blue*).

ACT-3′), and 9-nt-long (5′-CGT ACG CGG-3′), each 5′-phosphorylated and 3′-labeled with Oregon Green 488, were used for all binding experiments. Purified His-tagged AaAgo in 20 mM Hepes, 150 mM KCl, pH 7.4, was immobilized onto the sensor chip at various concentrations between 1 and 50 nM. 10 mM $MgCl_2$ was included in the buffer. The fluorescent label was excited with an argon ion laser at 488 nm, at ~100 μW. All experiments were performed at 25 °C with a DNA injection of 36 s, followed by a 200-s wash period. Immobilized protein was washed off the sensor surface by addition of 500 mM imidazole. Kinetic constants were calculated using BIAevaluation 3.1 software.

RESULTS

Structure of AaAgo—AaAgo is a roughly elliptical molecule measuring ~55 × 70 × 90 Å in size. Nearly all of the amino acids are defined in the electron density, with the exception of residues 176–177, 245–247, and 262–271. The domains of AaAgo are divided into three lobes, comprising the N-terminal, PAZ, and the Mid/PIWI domains. The N-terminal domain is a mixed α-β fold (Fig. 1*a*, *blue*), resembling that of other DNA/RNA-binding proteins, as revealed by a DALI search for structural neighbors. A linker region, designated as L1 (Fig. 1*a*, *green*), lies at the junction between the three lobes, forming a base to two clefts defined by the N-terminal and PIWI domains and the N-terminal and PAZ domains.

The PAZ domain (Fig. 1*a*, *green*) shares an overall architecture similar to that of other PAZ domain structures, divided into characteristic α-helical and β-sheet subdomains. The PAZ domain recognizes the two-nucleotide 3′-overhang of the guide strand siRNA. X-ray and NMR structures of PAZ-RNA complexes show the nucleotide end sandwiched between α-helical and β-sheet subdomains of PAZ (36, 37). Although PAZ domains share little sequence homology, the binding cleft for the 3′-end of the oligonucleotide is hydrophobic and is mediated by conserved aromatic residues. In AaAgo, these residues are Trp-226 and Phe-209, within the α-helical region of PAZ. A third conserved residue, Arg-195, is involved in recognition and orientation of the guide strand phosphodiester backbone.

A second linker region, L2 (Fig. 1*a*, *dark green*), connects the PAZ domain to the lobe formed by the MID and PIWI domains (Fig. 1*a*, *cyan* and *pink*, respectively). The binding pocket for the 5′-phosphate of the guide strand lies at the interface between the MID and PIWI domains. This positively charged binding pocket is highly conserved among Argonaute and PIWI proteins. Residues Tyr-439, Lys-443, Arg-444, and Lys-480, contributed by the MID domain, and Arg-681 and the C terminus of the protein (Leu-706), contributed by the PIWI domain, line this pocket in AaAgo. The PIWI domain has an RNase H fold, similar to other endonucleases such as transposases and viral integrases (38, 39) and as noted in structures of other Argonaute species and PIWI proteins. Cleavage of the target mRNA strand likely proceeds through a similar mechanism as other endonucleases, mediated by divalent cations such as Mg^{2+} or Mn^{2+} (12).

RNA Binding Sites—The architecture of this three-lobed structure results in a Y-shaped system of positively charged grooves that runs throughout the protein (Fig. 1*b*). The "stem" of the groove is flanked by the PAZ and MID/PIWI domains and contains the catalytic amino acids. Because of the location of the active site and the positively charged amino acids lining the stem, the dsRNA intermediate, composed of the guide strand siRNA and the target mRNA, likely binds here (13). Crystal structures of PIWI bound to dsRNA strongly support this model (37, 40). In our structure, the groove stem measures ~17–19 Å in diameter. Notably, the side chains of Arg-184 and Arg-601 approach within 4.5 Å of one another, thus requiring a conformational change or side chain rearrangement to properly accommodate the dsRNA substrate. The PAZ lies in an orientation that encloses part of the groove, forming a short tunnel.

During the course of this work, a crystal structure of AaAgo was reported (14), crystallized in a different space group from ours (henceforth referred to as 1YVU). The differences between the structures are most pronounced in the relative orientations of the PAZ domain. This movement arises from a gradual shift of the L1 β-sheet connecting the N-terminal and PAZ domains. Analysis of the conformational changes between our structure and 1YVU using DynDom (42) shows a ~24° shift in the position of the PAZ in our structure, away from the N-terminal domain (Fig. 1*c*). Consequently, this domain shift in our structure exposes the binding site for the 3′-end of the guide strand siRNA. The orientation of the PAZ domain observed in our

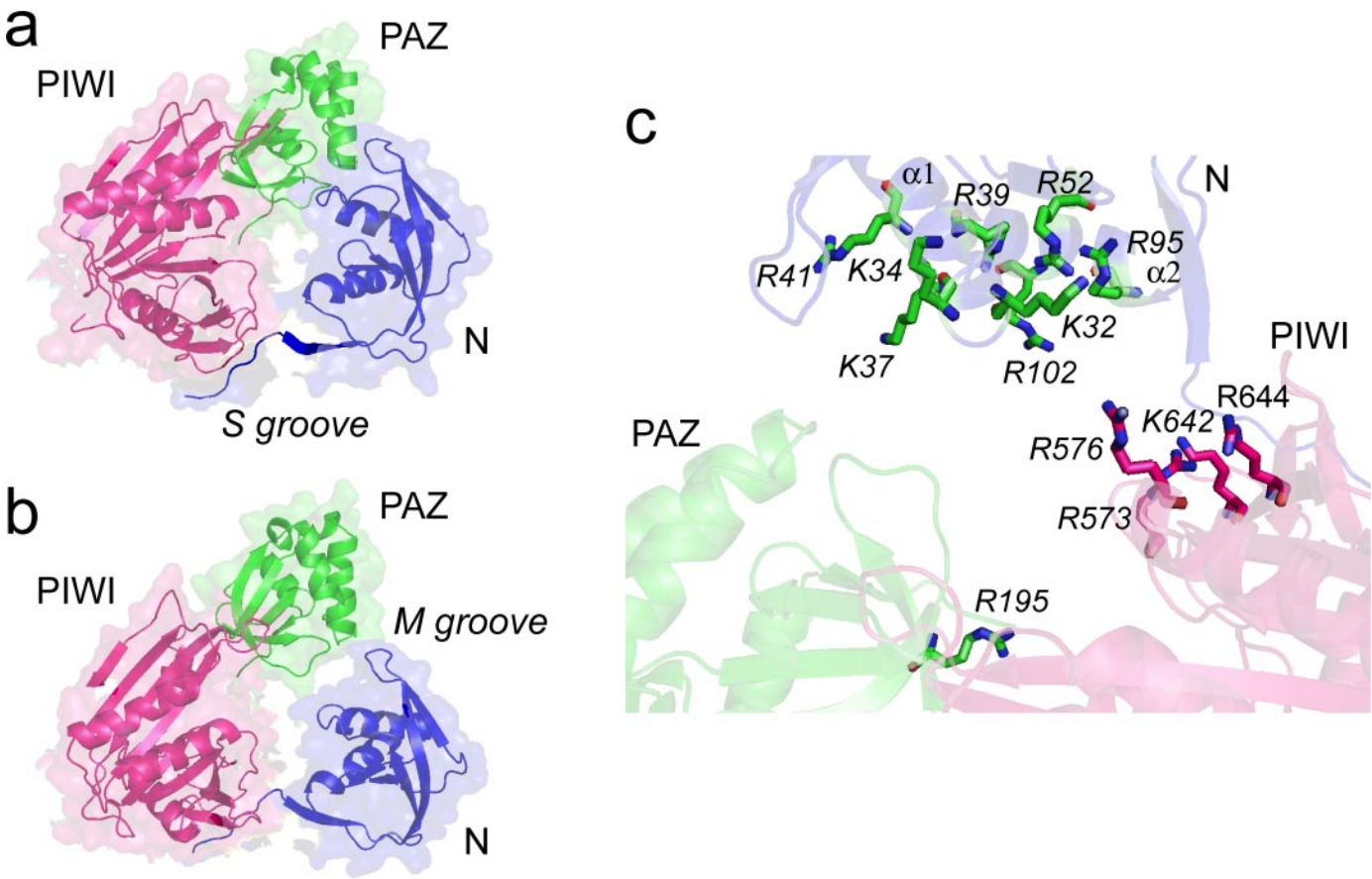

FIGURE 2. **Detail of M and S grooves.** *a*, S groove, between the N-terminal and PIWI domains. *b*, M groove, between the PAZ and N-terminal domains. *c*, overall view of M and S grooves showing basic amino acids contributed from α-helices α1 and α2 in the N-terminal domain and Arg-573, Arg-576, Lys-642, and Arg-644 from the PIWI domain. Conserved residue Arg-195 of the PAZ domain, which recognizes the phosphodiester backbone of the nucleic acid, is *highlighted*.

structure resembles the conformation of the *Pyrococcus furiosus* Argonaute apoenzyme, which also adopts a three-lobed structure with prominent grooves between the N-terminal domain and the PAZ and PIWI domains (13). The relatively high average B-factors of the PAZ domain, 101 Å^2 compared with 61 Å^2 for the remainder of the molecule, further support a tendency of the domain to be flexible. As the B-factors in the PAZ domain are relatively high, with an average amplitude of thermal vibration $u = 1.13$ Å ($B = 8\pi^2 <u^2>$), we calculated a composite omit map to investigate how well the structure of the PAZ domain is defined by the density map. The map was created by omitting a fixed 5% volume of the model in the asymmetric unit and subjecting the remainder to simulated annealing refinement. The reassembled map shows clear density for the PAZ domain (supplemental Fig. S2), demonstrating that the position of the PAZ domain is well defined in our structure. We also note that in the crystal the PAZ domain is relatively unhindered by crystal contacts, which may be the source of the higher B-factors. In contrast, the PAZ domain of 1YVU is involved in extensive crystal contacts with the N-terminal, MID, and PIWI domains. As the conformational differences in the two AaAgo structures alternatively expose or occlude the entrance to the nucleotide binding site in PAZ, this implies a functional flexibility of the PAZ domain in RISC function. An analysis of the surface accessibility of the two AaAgo structures shows that Lys-37 and Trp-40 in the α1 of the N-terminal domain are exposed in our structure relative to 1YVU (91.1 and 110.6 Å^2 in our structure *versus* 68.7 and 64.6 Å^2, respectively, in 1YVU). Likewise, residues Glu-215, Arg-216, and Thr-218, lining the entrance to the nucleotide binding pocket in PAZ, are also comparatively exposed in our structure (131.9, 152.9, and 115.7 Å^2 in our structure *versus* 109.7, 59.9, and 9.8 Å^2, respectively, in 1YVU).

One branch of the Y-shaped groove is lined by the N-terminal and PIWI domains (Fig. 2*a*, *S groove*) and measures ~17–19 Å in diameter. The size of the groove is very similar in 1YVU as well as *P. furiosus* Argonaute. The S groove is lined with basic amino acids, which may serve to recognize the backbone of dsRNA. The stretch of sequence corresponding to the PIWI box, residues 615–657, a highly conserved element present in eukaryotic Argonaute proteins, lies at one end of the S groove. Although the corresponding amino acids of the PIWI box in human Ago2 (residues 730–772) are not conserved, the region is solvent-exposed and is structurally localized at the base of the groove defined by the PIWI and N-terminal domains. A direct interaction between Dicer and Argonaute was observed, localized to an interaction between the PIWI box and the second RNase III domain of Dicer (43, 44), while subsequent studies have shown that Argonaute is responsible for the cleavage of the sense strand of Dicer-processed siRNA (9, 10). As such, these results constrain the location of Dicer within the RISC complex and suggest that the S groove may act as a channel for

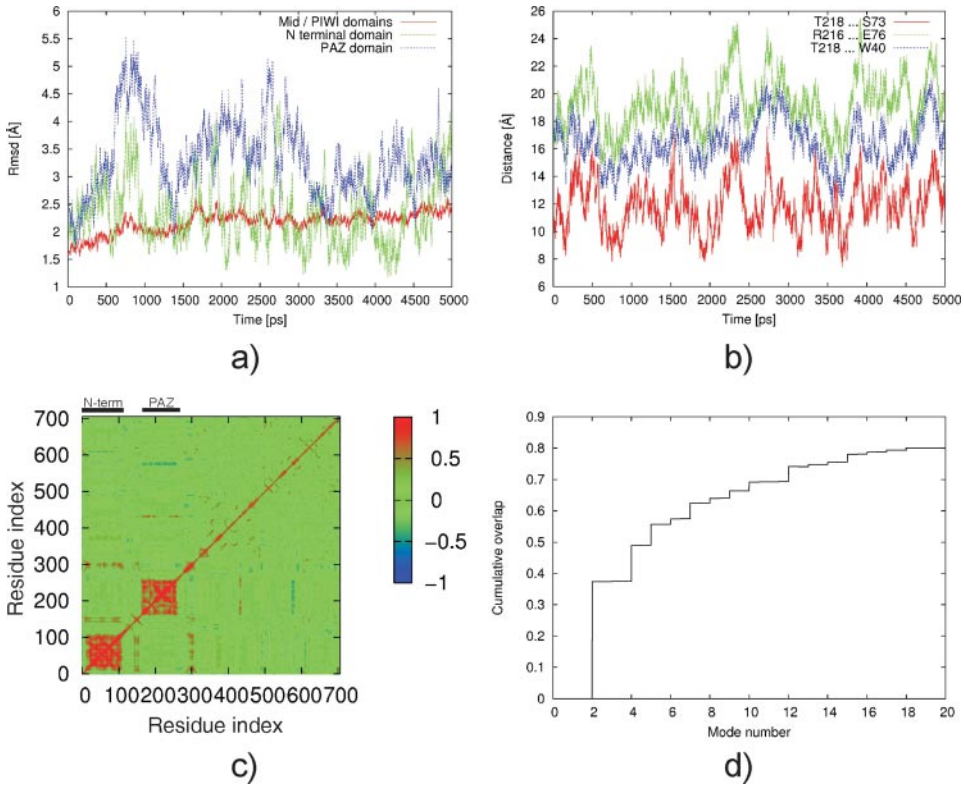

FIGURE 3. **Periodic motions of the PAZ and N-terminal domains.** *a*, time series of root mean square deviations of C_α atoms of the MID/PIWI domains (*red*), the N-terminal domain (*green*), and the PAZ domain (*blue*) from the starting structure over 5.0 ns of molecular dynamics simulation. The equilibration phase is not included. *b*, time series of C_α-C_α distances between residues at the tips of the PAZ and N-terminal domain: Thr-218-Ser-73 (*red*), Arg-216-Glu-76 (*green*), Thr-218-Trp-40 (*blue*). *c*, cross-correlation map of C_α atomic fluctuations of AaAgo. The two axes refer to residue indices. Positive correlations are indicated in *red*, negative ones in *blue*. Positions of the N-terminal domain and the PAZ domain are indicated by *bold lines*. *d*, the cumulative summation of the overlap (29) between the first 20 modes determined by principal component analysis of the 5.0-ns MD trajectory and the vector related to the conformational change between our structure and 1YVU as a function of the mode number. Only C_α atoms were considered.

passage of the Dicer-cleaved siRNA product into the active site of Argonaute.

The second branch of the Y-shaped groove is bordered by the N-terminal and PAZ domains (Fig. 2*b*, *M groove*), and measures ~17 Å in diameter. The closest interactions between the N-terminal and PAZ domains in our structure occur between the side chains of residues Ser-73 and Thr-218 (7.3 Å), which lie at the top of the groove. Importantly, in our structure, the 3′-nucleotide binding pocket in PAZ is exposed to the M groove. In contrast, 1YVU shows a closed conformation, characterized by a salt bridge between the side chains of residues Arg-216 and Glu-76 (2.9 Å), as well as a hydrogen bond between the indole nitrogen of Trp-40 and the carbonyl oxygen of Thr-218 (3.2 Å). Our structure shows no interactions between these residues, with a Arg-216/Glu-76 distance of 17.8 Å and a Trp-40/Thr-218 distance of 10.2 Å. Notably, although the size of the M groove shows significant variation between our structure and 1YVU, the size of the S groove is the same.

A sequence alignment of residues lining the S groove between AaAgo and human Ago1 and Ago2 reveals a number of conserved residues, including Lys-34, Arg-52, Arg-95, from the N-terminal domain, and Arg-573, Arg-576, Lys-642, and Arg-644 from the PIWI domain, which constitute the majority of the basic amino acids in the S groove (Fig. 2*c*).

Helices α1 and α2 act as a putative RNA recognition face of the N-terminal domain, spanning both the M and S grooves (Fig. 2*c*). Although AaAgo was crystallized in the presence of a single-stranded siRNA, only disordered, unassignable electron density was observed in the 5′-nucleotide binding pocket, with no discernible density observed in the 3′-nucleotide binding site in PAZ. This may be explained by the preference of AaAgo for ssDNA (K_d ~ 0.01 μM) relative to ssRNA (K_d ~ 0.97 μM) as measured by filter binding assays (14).

Molecular Dynamics—To further investigate the potential flexibility of the PAZ domain within Argonaute, we performed a 5.5-ns molecular dynamics simulation. Visual inspection of snapshots along the trajectory reveals a tendency for the PAZ domain, and to a lesser extent, the N-terminal domain, to change conformations, leading to a widening and narrowing of the M groove. This is corroborated by a time series of root mean square deviation of C_α atoms from the experimental structure (Fig. 3*a*). Whereas the MID/PIWI domains only show moderate deviations that remain below 2.5 Å, the N-terminal and PAZ domains show major conformational changes of up to 4.3 and 5.5 Å, respectively. Importantly, there was a tendency for both the PAZ and N-terminal domains to swing back and forth during the course of the simulation, with two and a half cycles observed for the PAZ and at least four cycles for the N-terminal domain (Fig. 3*a*). These *periodic* motions, together with the large conformational changes, show an inherent mobility of these two regions rather than a relaxation of the experimental structure during the simulation.

The motions lead to considerable changes of the distances between both domains, as depicted in Fig. 3*b* for those three pairs of residues at the "tips" that show closest interactions in our structure (Thr-218-Ser-73) or 1YVU (Arg-216-Glu-76; Thr-218-Trp-40), respectively. Changes in the C_α-C_α distances of up to 10 Å can be observed, which relate to changes in the diameter of the M groove of more than 5 Å.

Orientational correlations between motions of C_α atoms are displayed color-coded in Fig. 3*c*. The map shows strongly correlated motions within the PAZ and N-terminal domains, indicative of a rigid body type of motion of the domains.

Finally, a principal component analysis of the motions of the C_α atoms was computed yielding collective variables (or modes) for the description of the essential molecular dynamics

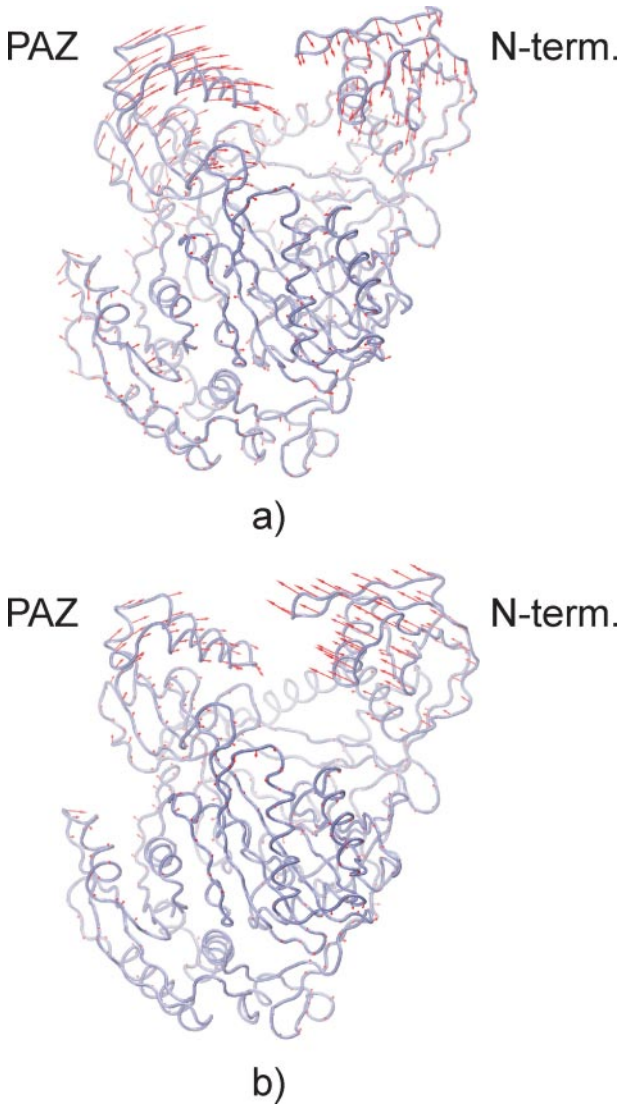

FIGURE 4. **Correlated motions of the PAZ and N-terminal domains.** Displacement along mode directions determined by principal component analysis of the 5.0-ns molecular dynamics trajectory considering only C_α atoms. The two modes most involved in the conformational change between 1YVU and the structure determined in this study are shown: mode number 2, overlap = 0.37 (*a*) and mode number 4, overlap = 0.34 (*b*). The amplitudes of the motions were scaled for best graphical representation. The indicated motions led to a narrowing of the M groove bordered by the N-terminal and PAZ domains in both cases.

(30). These analyses often show that protein motion is dominated by a relatively small number of low-frequency modes (45). One interpretation for this is that protein structures have evolved such that biologically relevant motions near the folded state predominantly occur along the directions of lowest-energy modes (41). Here, mode number 2 (Fig. 4*a*) and mode number 4 (Fig. 4*b*) showed the largest involvement in the observed conformational change between 1YVU and the structure determined in this study, with both modes explaining close to 50% of the experimentally observed motion (Fig. 3*d*). Taken together with the notion of periodic motions discussed above, this finding is encouraging as it demonstrates that the motions observed for the structure determined here may lead to conformational transitions that resemble those of the 1YVU structure. As a consequence, closed conformations of Argonaute may already exist in the ensemble of the unbound protein.

Kinetic Behavior of ssDNA Binding to Argonaute—To further examine the interaction between ssDNA and Argonaute, we performed kinetic binding experiments using immobilized AaAgo. We used surface-sensitive fluorescence detection by total internal reflection fluorescence spectroscopy. The fluorescently labeled ssDNA is excited by the evanescent wave emanating from the surface. The interaction of 3′-labeled 21-nt ssDNA containing a 5′-phosphate with immobilized AaAgo shows biphasic association and dissociation kinetics (Fig. 5*a*). To exclude that the signal was due to ssDNA binding to the surface, a control experiment with 100 nM ssDNA in the absence of immobilized protein was run, showing minimal fluorescence signal (Fig. 5*a*). The interaction kinetics was independent of the immobilized protein concentration (0.06, 0.12, and 0.23 ng/mm^2, corresponding to 2, 5, and 10 nM protein injected onto the chip) (Fig. 5*b*). The slow dissociation phase could be fit with a single exponential decay, with a mean dissociation rate constant k_d of 0.004 ± 0.002/s, corresponding to very slow dissociation of the ssDNA from Argonaute (Fig. 5*c*). No systematic dependence of the k_d on the surface concentration of AaAgo was observed, which excludes the possibility that rebinding or multivalent interaction of the ssDNA with adjacent immobilized protein molecules was the source of the slow dissociation kinetics.

Under simple binding conditions involving a protein and ligand, the K_d is the ratio of the association and dissociation rate constants (k_d/k_a). The kinetic behavior observed in the case of AaAgo and ssDNA precludes calculating a true K_d for the interaction. Nevertheless, a range of "K_d" values could be calculated based on the fitting of the two components of the association and dissociation curves. The K_d was estimated to be between ~2.9 nM (fast association, slow dissociation) and 270 nM (slow association, slow dissociation), in relative agreement with the K_d measured by Yuan *et al.* (14) by filter binding (10 nM). We note that the identical nucleotide sequence and buffer conditions were used for these kinetic experiments with immobilized AaAgo.

We then examined the kinetics of binding as a function of the length of ssDNA, using 9- and 15-mer ssDNA in addition to the 21-mer in the previous experiments. Surprisingly, the kinetic behavior is nearly identical regardless of the length of ssDNA used (Fig. 5*d*). One interpretation of this finding is that the "hotspot" of binding energy is localized to the 5′-region of the ssDNA.

DISCUSSION

Domain Motions in Argonaute—Assuming that a range of PAZ orientations is represented by our crystal structure and 1YVU, our molecular dynamics simulation demonstrates that the unliganded form of Argonaute can assume a number of conformations (Figs. 1*c*, 3, *a* and *b*, and 4). The molecular dynamics simulation thus lends further support to the notion of a flexible PAZ domain and, to a lesser extent, the N-terminal domain. Furthermore, this is evident from the different conformations of AaAgo seen in our structure and 1YVU, as well as the unliganded structure of *P. furiosus* Argonaute. Thus the

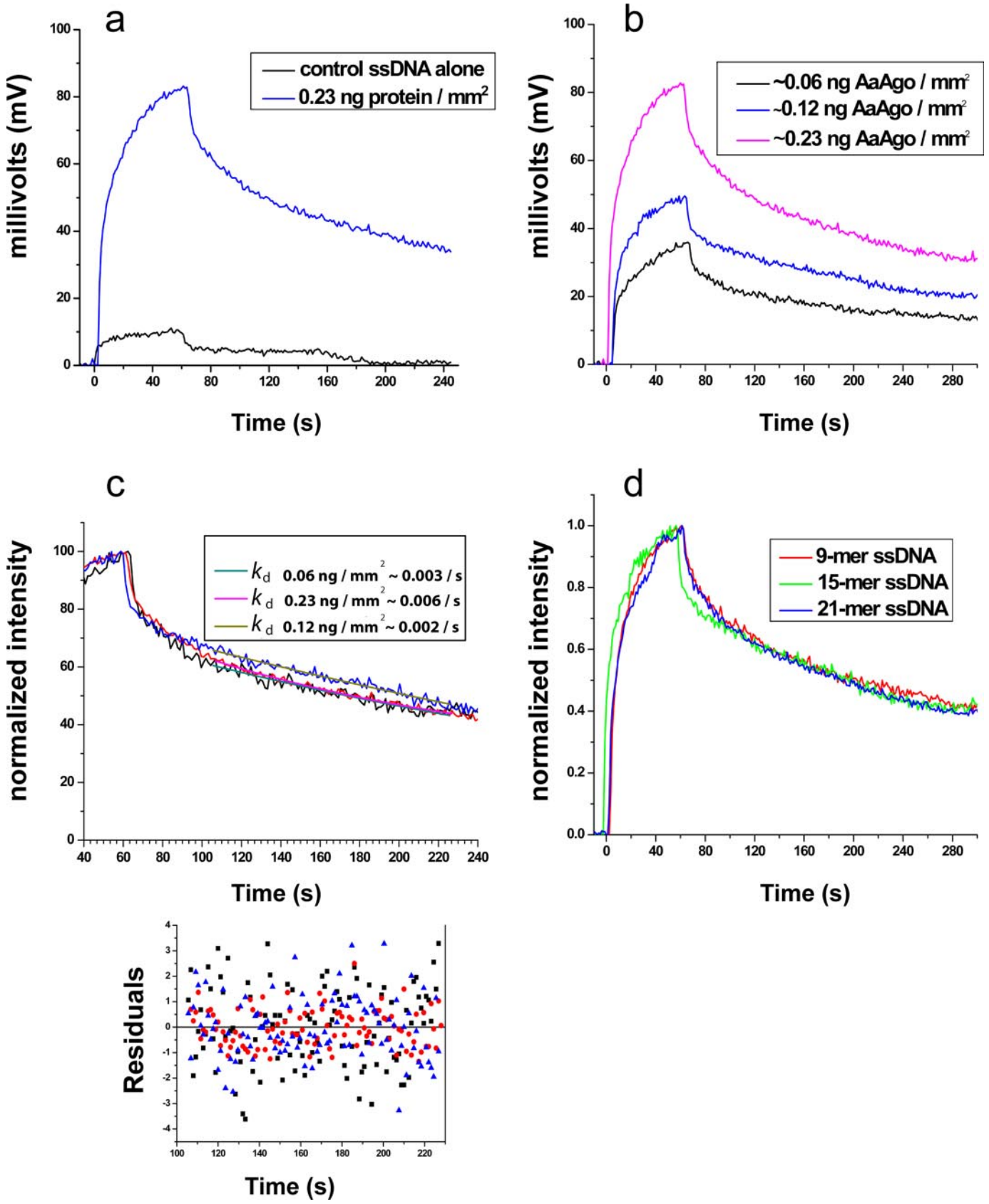

FIGURE 5. **Kinetics of binding of ssDNA to AaAgo.** *a*, sensorgram of kinetics of ssDNA binding to immobilized AaAgo. The surface concentration of AaAgo was estimated to be 0.23 ng/mm^2, corresponding to an injection of 10 nM AaAgo (*blue*). A control with 100 nM ssDNA shows minimal binding (*black*). *b*, sensorgram of protein concentration dependence of ssDNA binding, with 0.06 (*black*), 0.12 (*blue*), and 0.23 ng/mm^2 (*pink*) surface concentration of AaAgo. *c*, single exponential fit, with residuals, of the slow dissociation at three protein concentrations (*colors*). *d*, length dependence of kinetics of ssDNA binding to AaAgo, for 9- (*red*), 15- (*green*), and 21-mer (*blue*) oligonucleotides. Curves were normalized to the peak fluorescence intensity.

closed and open forms of Argonaute may represent low energy conformations that become fixed in specific orientations upon nucleotide binding.

The biphasic kinetics of ssDNA binding to AaAgo is further suggestive of a conformational rearrangement in the protein. We note that the fluorescence intensity observed is primarily due to the interaction of the ssDNA with the protein and not due to the ssDNA alone (Fig. 5*a*).

The total internal reflection fluorescence spectroscopy measurements show little difference in the association and dissociation kinetics of 9-, 15-, and 21-mer ssDNA nucleotides (Fig. 5*d*). We conclude the following from this observation. First, this suggests that the majority of the interaction energy is localized to the so-called seed region at the 5′-end and that the DNA beyond this point may be only loosely associated with Ago. Secondly, the kinetics of the ssDNA/AaAgo interaction are also independent of the length of ssDNA substrate used and the surface concentration of immobilized AaAgo. Together with the data from AaAgo crystal structures and the molecular dynamics simulation, the kinetic behavior points toward domain rearrangements during binding and dissociation, as reflected by the changes in DNA fluorescence (Fig. 5*a*). These data are fully consistent with the previously reported crystal structures of the *Archaeoglobus fulgidus* PIWI-dsRNA complex, which shows well ordered nucleotide density for the so-called "seed region" (37, 40).

Functional Implications of PAZ Flexibility—The structural, biochemical, and computational data presented here, in concert with the 1YVU and *P. furiosus* Argonaute structures, present an unprecedented opportunity to look at potential conformational changes in Argonaute during the cycle of mRNA recognition and cleavage by RISC. As the previous two structures of Argonaute have come from two different species with limited sequence identity (25%), the structure reported here allows for the first time the comparison of two Argonaute structures from the same species. In the context of the RISC, which has recently been identified as a complex composed of Dicer, Argonaute, and TRBP (11), our structure and biochemical data place constraints on the entrance and exit paths for the loaded Dicer-processed siRNA as well as the mRNA.

It is plausible that the initial loading of the Dicer-processed siRNA may proceed through the S groove prior to Argonaute cleavage of the passenger strand. The position of the PIWI box proximal to the S groove and the interaction between the PIWI box and the catalytic domains of Dicer support this hypothesis. The positively charged residues of the S groove may therefore contribute to the recognition of the siRNA in the "initial loading" complex.

Following cleavage of the passenger strand in the initial loading complex, the single-stranded guide strand-Argonaute complex is now in the catalytically active state. We have shown that unliganded Argonaute is flexible, most notably in the PAZ domain. Upon formation of the complex, the flexibility of the PAZ domain may be restricted. This may allow for recognition of the incoming target mRNA as well as discrimination between perfect and imperfect base pairing in the seed region. The mRNA strand may be directed to the seed region and the enzyme active site through two possible pathways, either along the M groove between the PAZ and N-terminal domains or along the S groove between the N-terminal and PIWI domains. This will have to be tested by further biochemical and structural studies.

The structure presented here, together with our molecular dynamics and kinetic studies, has highlighted a critical role for PAZ flexibility in the function of Argonaute. Future biochemical and structural studies will be necessary to further clarify the cycle of siRNA loading and mRNA recognition and cleavage by Argonaute.

Acknowledgments—We thank the staff at European Synchrotron Radiation Facility beamline ID23–1 for assistance with data collection and Volker Dötsch, Wesley McGinn-Straub, and Prakash Chandra for critical comments on the manuscript. We thank Sheryl Tsai and Janet Finer-Moore for helpful discussions about the structural refinement.

REFERENCES

1. Meister, G., and Tuschl, T. (2004) *Nature* **431,** 343–349
2. Hannon, G. J. (2002) *Nature* **418,** 244–251
3. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001) *Nature* **409,** 363–366
4. Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., Hammond, S. M., Joshua-Tor, L., and Hannon, G. J. (2004) *Science* **305,** 1437–1441
5. Meister, G., Landthaler, M., Patkaniowska, A., Dorsett, Y., Teng, G., and Tuschl, T. (2004) *Mol. Cell* **15,** 185–197
6. Khovorova, A., Reynolds, A., and Jayasena, S. D. (2003) *Cell* **115,** 209–216
7. Schwarz, D. S., Hutvanger, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003) *Cell* **115,** 199–208
8. Tomari, Y., Matranga, C., Haley, B., Martinez, N., and Zamore, P. D. (2004) *Science* **306,** 1377–1380
9. Rand, T. A., Petersen, S., Du, F., and Wang, X. (2005) *Cell* **123,** 621–629
10. Matranga, C., Tomari, Y., Shin, C., Bartel, D. P., and Zamore, P. D. (2005) *Cell* **123,** 607–620
11. Gregory, R. I., Chendrimada, T. P., Cooch, N., and Shiekhattar, R. (2005) *Cell* **123,** 631–640
12. Rivas, F. V., Tolia, N. H., Song, J. J., Aragon, J. P., Liu, J., Hannon, G. J., and Joshua-Tor, L. (2005) *Nat. Struct. Mol. Biol.* **12,** 340–349
13. Song, J. J., Smith, S. K., Hannon, G. J., and Joshua-Tor, L. (2004) *Science* **305,** 1434–1437
14. Yuan, Y. R., Pei, Y., Ma, J. B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H. Y., Dauter, Z., Tuschl, T., and Patel, D. J. (2005) *Mol. Cell* **19,** 405–419
15. Filipowicz, W. (2005) *Cell* **122,** 17–20
16. Tomari, Y., and Zamore, P. D. (2005) *Genes Dev.* **19,** 517–529
17. Bailey, S. (1994) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **50,** 760–763
18. De La Barre, B., and Brunger, A. T. (2006) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62,** 932–932
19. Ling, H., Boodhoo, A., Hazes, B., Cummings, M. D., Armstrong, G. D., Brunton, J. L., and Read, R. J. (1998) *Biochemistry* **37,** 1777–1788
20. Kissinger, C. R., Ghlhaar, D. K., and Fogel, D. B. (1999) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **55,** 484–491
21. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **54,** 905–921
22. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60,** 2126–2132
23. Marti-Renom, M. A., Stuart, A. C., Fiser, A., Sanchez, R., Melo, F., and Sali, A. (2000) *Annu. Rev. Biophys. Biomol. Struct.* **29,** 291–325
24. Case, D. A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., and Woods, R. J. (2005) *J. Comput. Chem.* **26,** 1668–1688

25. Simmerling, C., Strockbine, B., and Roitberg, A. E. (2002) *J. Am. Chem. Soc.* **124,** 11258–11259
26. Jorgensen, W. L., Chandrasekhar, J., Madura, J., and Klein, M. L. (1983) *J. Chem. Phys.* **79,** 926–935
27. Darden, T., York, D., and Pedersen, L. (1993) *J. Chem. Phys.* **98,** 10089–10092
28. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) *J. Comput. Phys.* **23,** 327–341
29. Ichiye, T., and Karplus, M. (1991) *Proteins* **11,** 205–217
30. Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) *Proteins* **17,** 412–425
31. Marques, O., and Sanejouand, Y. H. (1995) *Proteins* **23,** 557–560
32. Hunenberger, P. H., Mark, A. E., and van Gunsteren, W. F. (1995) *J. Mol. Biol.* **252,** 492–503
33. Karplus, M., and Ichiye, T. (1996) *J. Mol. Biol.* **263,** 120–122
34. Gavutis, M., Lata, S., Lamken, P., Muller, P., and Piehler, J. (2005) *Biophys. J.* **88,** 4289–4302
35. Lata, S., and Piehler, J. (2005) *Anal. Chem.* **77,** 1096–1105
36. Lingel, A., Simon, B., Izaurralde, E., and Sattler, M. (2004) *Nat. Struct. Mol Biol.* **11,** 576–577
37. Ma, J. B., Yuan, Y. R., Meister, G., Pei, Y., Tuschl, T., and Patel, D. J. (2005) *Nature* **434,** 666–670
38. Lai, L., Yokota, H., Hung, L. W., Kim, R., and Kim, S. H. (2000) *Structure* **8,** 897–904
39. Chen, J. C.-H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A. H., Leavitt, A. D., and Stroud, R. M. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 8233–8238
40. Parker, J. S., Roe, S. M., and Barford, D. (2005) *Nature* **434,** 663–666
41. Bahar, I., and Rader, A. J. (2005) *Curr. Opin. Struct. Biol.* **15,** 586–592
42. Hayward, S., and Berendsen, H. J. C. (1998) *Proteins* **30,** 144–154
43. Tahbaz, N., Kolb, F. A., Zhang, H., Jaronczyk, K., Filipowicz, W., and Hobman, T. C. (2004) *EMBO Rep.* **5,** 1–6
44. MacRae, I. J., Zhou, K., Li, F., Repic, A., Brooks, A. N., Cande, W. Z., Adams, P. D., and Doudna, J. A. (2006) *Science* **311,** 195–198
45. Hayward, S., and Go, N. (1995) *Annu. Rev. Phys. Chem.* **46,** 223–250