

# The crystal structure of the Argonaute2 PAZ domain reveals an RNA binding motif in RNAi effector complexes

Ji-Joon Song<sup>1–3</sup>, Jidong Liu<sup>1,3</sup>, Niraj H Tolia<sup>1,2</sup>, Jonathan Schneiderman<sup>1</sup>, Stephanie K Smith<sup>1,2</sup>, Robert A Martienssen<sup>1</sup>, Gregory J Hannon<sup>1</sup> & Leemor Joshua-Tor<sup>1,2</sup>

RISC, the RNA-induced silencing complex, uses short interfering RNAs (siRNAs) or micro RNAs (miRNAs) to select its targets in a sequence-dependent manner. Key RISC components are Argonaute proteins, which contain two characteristic domains, PAZ and PIWI. PAZ is highly conserved and is found only in Argonaute proteins and Dicer. We have solved the crystal structure of the PAZ domain of *Drosophila* Argonaute2. The PAZ domain contains a variant of the OB fold, a module that often binds single-stranded nucleic acids. PAZ domains show low-affinity nucleic acid binding, probably interacting with the 3' ends of single-stranded regions of RNA. PAZ can bind the characteristic two-base 3' overhangs of siRNAs, indicating that although PAZ may not be a primary nucleic acid binding site in Dicer or RISC, it may contribute to the specific and productive incorporation of siRNAs and miRNAs into the RNAi pathway.

The presence of double-stranded RNA (dsRNA) in a eukaryotic cell can trigger an evolutionarily conserved gene silencing mechanism known as RNA interference or RNAi<sup>1</sup>. This pathway generally responds to exogenously introduced dsRNAs by targeting homologous, endogenous RNAs for sequence-specific degradation<sup>2</sup>. There are also numerous endogenous noncoding RNAs known as miRNAs, and at least some of these enter either the same or related silencing pathways, regulating gene expression at the level of either mRNA stability or protein synthesis<sup>3</sup>. In many systems, the RNAi machinery has been linked to the metabolism of repetitive genomic sequences. For example, in *Schizosaccharomyces pombe*, an RNAi-related machinery is essential for regulating repetitive elements at the centromeres and mating type loci through effects on heterochromatin formation<sup>4,5</sup>.

A close relationship between these seemingly distinct processes is suggested by their sharing not only a common structure of the proposed silencing trigger, namely dsRNA, but also by the presence of certain common features that indicate related biochemical mechanisms. siRNAs are a central element of RNAi-related pathways. These comprise a dsRNA of 19–24 nucleotides in which the double helix is bounded by two-nucleotide 3' overhangs<sup>6</sup>. Terminal phosphate groups at the 5' ends of each strand have also been implicated as essential features for the incorporation of such small RNAs into the silencing pathway<sup>7</sup>. siRNAs are produced through the action of a specific RNase III family nuclease known as Dicer<sup>8</sup>. This enzyme cleaves either dsRNAs or short hairpin miRNA precursors to generate the active forms of these silencing triggers<sup>1</sup>. In the current working model of the RNAi mechanism, each of these structures enters an effector complex (or a series of related complexes) known as RISC,

which contains as a signature component a member of the Argonaute family of proteins<sup>9,10</sup>.

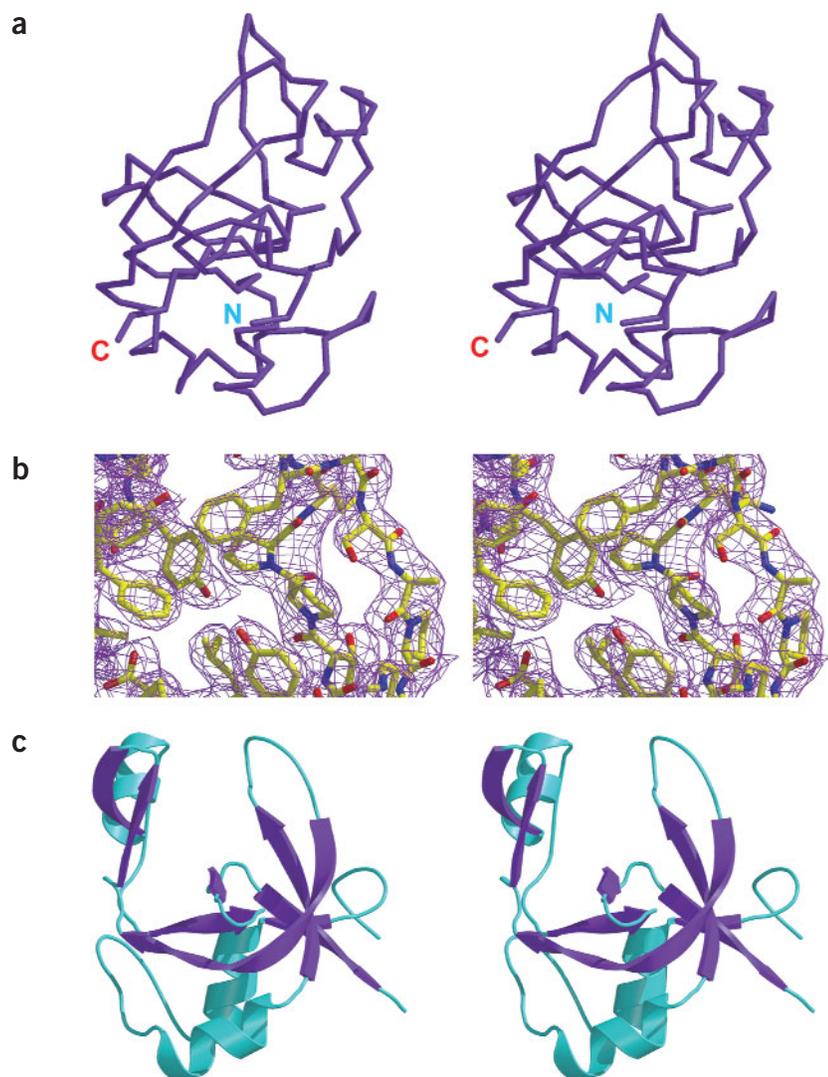
Argonaute proteins are evolutionarily conserved and are composed of two distinguishing domains, PAZ and PIWI<sup>11,12</sup>. Whereas the PIWI domain is restricted to Argonautes, the PAZ domain is found also in Dicer family proteins<sup>8</sup>. To better understand the biochemical function of Argonaute proteins, we have determined the crystal structure of the PAZ domain of *Drosophila melanogaster* Argonaute2 (Ago2-PAZ) and, based on this structural information, suggest a model for how this domain functions in Dicer and RISC.

## RESULTS

### Overall structure of Ago2-PAZ

The crystal structure of an MPB-PAZ fusion protein was determined at a resolution of 2.8 Å. The three MBP portions and the three PAZ domains are very similar, with pairwise r.m.s. deviations of 0.04–0.05 Å over C $\alpha$  atoms of the MBPs and 0.60–1.05 Å over C $\alpha$  atoms between the PAZ domains. However, the PAZ domains are not oriented similarly with respect to the MBPs in the three molecules, giving pairwise superposition r.m.s. deviations of 2.16–4.22 Å for the fusion proteins. The crystal lattice is stabilized by nine Ni<sup>2+</sup> metal ions (see Supplementary Fig. 1 online). The PAZ domain was found to have a novel fold, based on analysis using the DALI server, and is composed of two subdomains that are oriented to form a cleft (Fig. 1a–c). One subdomain has a five-stranded open  $\beta$ -barrel with two helices on one end of the barrel and an additional strand attached to the outer part of the barrel. This subdomain is highly reminiscent of an OB fold<sup>13</sup>, albeit with altered topology (Fig. 2 and Supplementary Fig. 2 online).

<sup>1</sup>Cold Spring Harbor Laboratory, Watson School of Biological Sciences and <sup>2</sup>Keck Structural Biology Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. <sup>3</sup>These authors contributed equally to this work. Correspondence should be addressed to G.J.H. (hannon@cshl.edu) or L.J. (leemor@cshl.edu).



**Figure 1** Crystal structure of Ago2-PAZ. (a) Stereo diagram of the C $\alpha$  trace of Ago2-PAZ. This figure and **Figures 2a**, **4c** and **5a** were prepared with BobScript<sup>28</sup>, MolScript<sup>29</sup>, and Raster3D<sup>30,31</sup>. (b) A portion of the final  $2F_o - F_c$  electron density map around the invariant aromatic residues at the back of the cleft (see below), contoured at  $1\sigma$ , superimposed on the refined model, drawn in stick representation. (c) Stereo ribbon diagram of the Ago2-PAZ domain with  $\beta$ -strands in purple and  $\alpha$ -helices in blue. The intersubdomain cleft is apparent in this view shown in a view  $-90^\circ$  from **a**.

The second subdomain is composed of a  $\beta$ -hairpin followed by an  $\alpha$ -helix.

### The PAZ domain binds RNA

Several lines of evidence suggested that PAZ provides a nucleic acid binding site. First, the PAZ domain is similar to the OB fold. The largest superfamily of OB-fold proteins contains those involved in single-stranded (ss) DNA or RNA binding. These proteins are involved in processes such as replication (for example, RPA), transcription (Rho termination factor), translation (aspartyl-tRNA synthetase) and telomere maintenance (*Saccharomyces cerevisiae* Cdc13 and *Oxytricha nova* telomere end-binding protein, TEBP)<sup>14</sup>. There are several other OB-fold superfamilies including the staphylococcal nucleases, bacterial enterotoxins and pyrophosphatases (SCOP database)<sup>15</sup>.

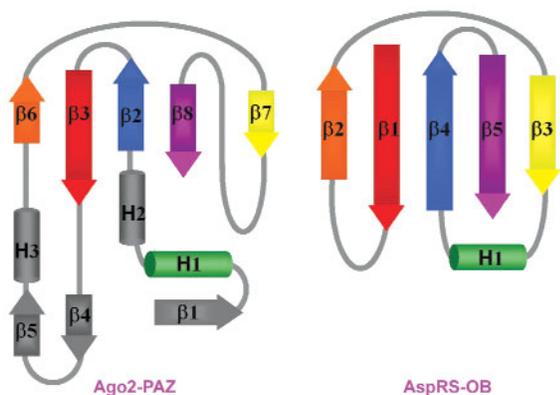
Second, in the purification of the Ago2-PAZ from *Escherichia coli*, this protein fragment coeluted from an affinity selection step with

heterogeneous RNAs. Third, a series of *in vitro* experiments suggested a direct interaction between Argonaute proteins and siRNAs. In these studies, we carried out UV-crosslinking experiments with cell extracts and <sup>32</sup>P-labeled siRNAs. Among only a few proteins that became linked to siRNAs was a band of  $\sim 100$  kDa that was identified by immunoprecipitation of both native and epitope-tagged species as an Argonaute protein (both human Ago1 and Ago2; J.L., J.S. and G.J.H., unpublished data). This interaction could have occurred with any single domain or with multiple domains within these proteins.

In light of these observations, we considered that the PAZ domain might act as a nucleic acid binding motif in Dicer and Argonaute proteins. To address this question, we examined interactions between single-stranded and double-stranded siRNAs and a GST-PAZ fusion protein by coprecipitation. Although nucleic acids could be detected in bound complexes, high-affinity binding was not seen with any substrate. Based on binding and competition experiments, we estimate the  $K_d$  to be  $>10\ \mu\text{M}$ . By comparison, CDC13, which uses multiple distinct domains including an OB fold to bind to its targets, does so with an affinity of 3 pM (ref. 16). We also examined nucleic acid binding with the PAZ domain of *Drosophila* Argonaute1. This protein fragment coprecipitated with nucleic acids, and bound with an affinity higher than that of the Ago2 PAZ domain (data not shown). Superficially, these experiments suggest that the PAZ domain can interact with nucleic acids. However, our data indicate that PAZ is not a high-affinity nucleic acid binding module.

With the hope of covalently trapping an interaction between the recombinant PAZ domain and an RNA species, we examined the ability of UV light to crosslink the Ago2-PAZ to single- and double-stranded siRNAs. As we chose to induce crosslinks by irradiation at 254 nm, we expected uridine and deoxythymidine residues to present the most reactive species<sup>17</sup>. For this reason, we used a number of siRNAs, each with a different sequence, so that U or dT residues might be present at different relative positions within the siRNAs. Two siRNAs in a panel of five gave rise to crosslinked species (Fig. 3a). We noted that both of these siRNAs shared the presence of 3' overhangs comprising two dT residues. One of the two siRNAs with the dTdT overhang was also represented in the panel by another siRNA with an identical double-stranded region but with terminal overhangs that matched the sequence of its luciferase target gene. This siRNA did not crosslink efficiently, indicating that the presence of the dTdT overhangs was important to our obtaining an adduct. Similar results were obtained with single-stranded siRNAs; species with photoreactive residues (namely dTdT) at their 3' ends crosslinked with the greatest efficiency (not shown).

The observation of a crosslink only when two potentially photoreactive residues are present at siRNA termini suggested a direct



**Figure 2** Ago2-PAZ has a deviant OB fold. A topology diagram for Ago2-PAZ on the right and aspartyl-tRNA synthetase<sup>32</sup> on the left depicting the differences in the topology of the  $\beta$ -barrels.

interaction between the PAZ domain and the 3' ends of ssRNAs, including the characteristic ends of siRNAs. However, the aforementioned data was also consistent with several other possible interpretations. First, the PAZ domain could bind preferentially to DNA, thus giving the greatest degree of crosslinking to those sequences with two terminal deoxythymidines. However, unlike ssRNA, ssDNA could not be recovered in a GST pulldown assay (not shown). Second, the PAZ domain could bind to single-stranded regions in a sequence-specific fashion, preferring dTdT among the sequences with which it was presented. Third, the PAZ domain could interact with any single-stranded or double-stranded region with crosslinking occurring solely because of the greater photoreactivity of dTdT bases as compared with standard RNA bases<sup>17</sup>.

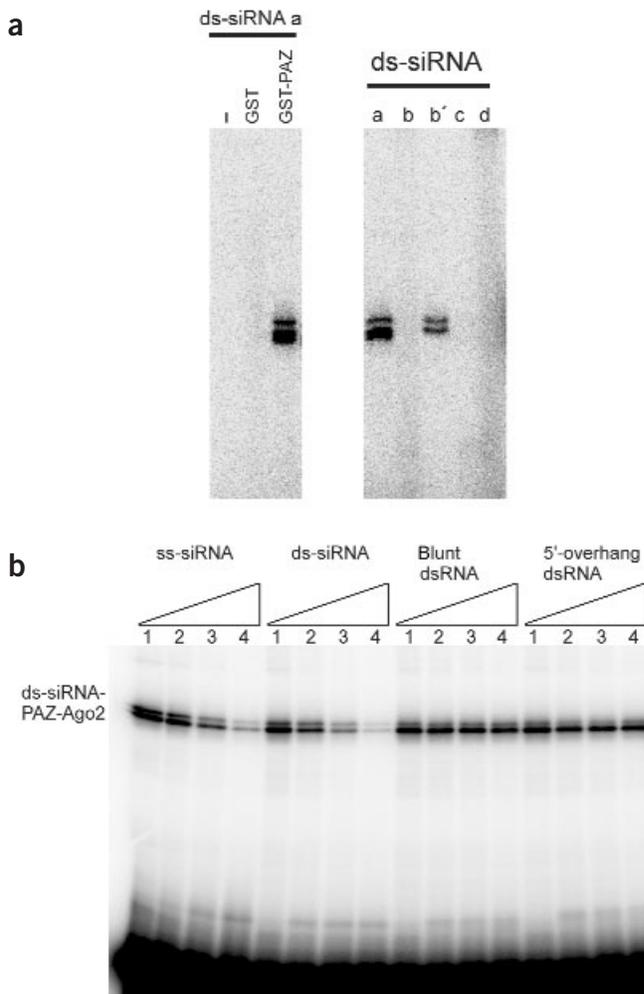
To distinguish among these four possibilities, we used a crosslinking competition assay with unlabeled nucleic acids. We found that single-stranded siRNAs or double-stranded siRNAs were equally capable of competing for crosslinking between PAZ and the siRNA containing a dTdT 3' overhang (Fig. 3b). Blunt-ended dsRNAs were unable to compete for crosslinking, even when present at a more than ten-fold greater concentration than that at which we saw competition with single-stranded and double-stranded siRNAs. Finally, double-stranded siRNAs with two-nucleotide 5' overhangs were also unable to compete for the crosslink (Fig. 3b). Together, these data indicate that the PAZ domain can bind to the 3' end of ssRNAs, including the characteristic 3' overhangs that are produced on siRNAs and pre-miRNAs by cleavage with Dicer or Drosha.

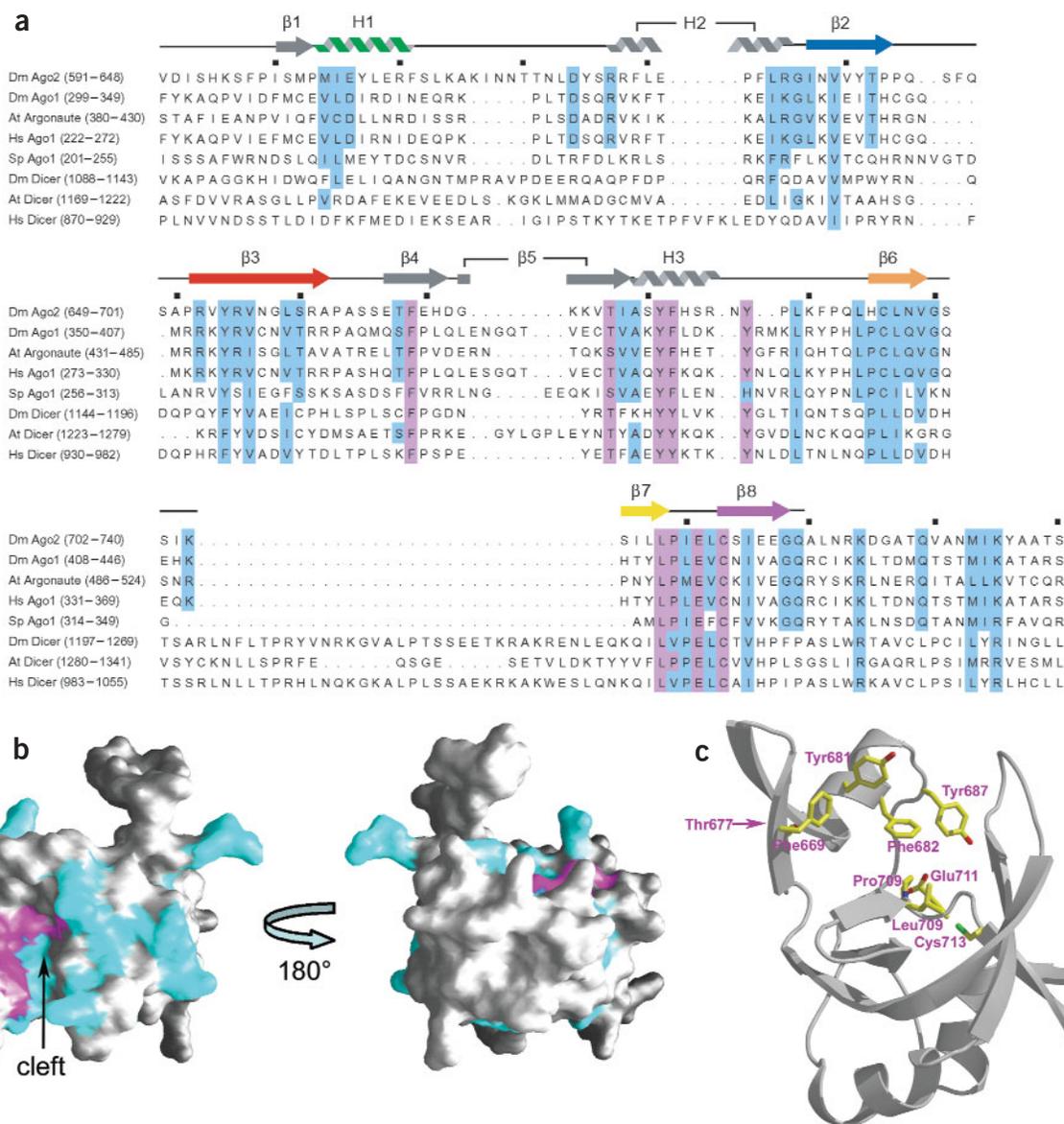
### The signature sequence motif of the PAZ domain

A sequence alignment of the PAZ domains of five Ago1 subfamily proteins and three Dicer family proteins<sup>12</sup> identified residues that are

**Figure 3** Binding properties of the PAZ domain. (a) UV crosslinking of siRNAs to the Ago2 PAZ. SiRNAs, as described below, were radiolabeled by phosphorylation with [ $\gamma$ -<sup>32</sup>P]ATP. Proteins and adducts were resolved by SDS-PAGE. Untagged PAZ or MBP-fused PAZ domain from Argonaute2 protein and GST-tagged PAZ domain from Argonaute1 protein showed similar crosslinking patterns. SiRNA a (see Methods for sequences) contains dTdT overhangs at both termini, siRNA b contains AC/UC overhangs, siRNA b' is identical to b except that it contains dTdT overhangs, siRNA c contains AG/UG overhangs and siRNA d contains GG/AA overhangs. (b) Competition experiments reveal preferential binding to the 3' ends of single-stranded regions. GST-PAZ (0.3  $\mu$ M) was incubated with siRNA b' (0.3  $\mu$ M) and increasing concentrations of the indicated competitors (lane 1 of each group, no competitor; lane 2, 1.5  $\mu$ M; lane 3, 6  $\mu$ M; lane 4, 30  $\mu$ M). Incubations and crosslinking were as described in Methods. The blunt dsRNA was 21 nt in length and the 5' overhang dsRNA is a siRNA-like structure with 2-nt overhangs on each 5' end.

conserved mainly among the Ago1 subfamily PAZ domains (Fig. 4a). These conserved residues were mapped onto the surface of the PAZ (Fig. 4b), and their side chains are shown in stick representation in Figure 4c. Two features are immediately apparent. First, the conserved residues are located on one side of the protein, leaving the other side completely variable. Second, apart from Thr677, the invariant residues are all located in the intersubdomain cleft of the molecule, forming two continuous regions. One, made up of aromatic residues, is on the inner surface of the  $\beta$ -hairpin of the second subdomain. The invariant threonine (Thr677) side chain in this region protrudes from strand  $\beta$ 5 and forms a hydrogen bond with the backbone amide of the N-terminal region of helix H3 (Ser680), stabilizing the fold of the second subdomain. The second region of invariant residues consists of hydrophobic residues, a cysteine and a glutamate, and is at the bottom of the intersubdomain cleft in the loop between strands  $\beta$ 7 and  $\beta$ 8 (Fig. 4c). This is the same surface where many OB-fold proteins bind single-stranded nucleic acids. The well ordered glutamate (Glu711) side chain forms two hydrogen bonds with backbone amide nitrogens of the second subdomain (Lys690 and Phe691), locking down the second subdomain. Of the





**Figure 4** Sequence conservation in the PAZ domain superfamily. **(a)** Sequence alignment of the PAZ domains from *D. melanogaster* Ago-2 (Dm Ago2), *D. melanogaster* Ago 1 (Dm Ago1), *A. thaliana* Argonaute (At Argonaute), *Homo sapiens* Argonaute1 (Hs Ago1), *S. pombe* Ago 1 (Sp Ago1), *D. melanogaster* Dicer (Dm Dicer), *A. thaliana* Carpel Factory (At Dicer) and *H. sapiens* Dicer (Hs Dicer). Conserved residues are blue and nearly invariant residues are purple (see text). These assignments were made based on inspection of an alignment of 35 PAZ domain sequences. The secondary structure elements are shown above the sequence and colored as in **Figure 2**. **(b)** Molecular surface representation of Ago2-PAZ showing the conserved residues in blue and invariant residues in purple. The invariant residues are concentrated in two areas in the intersubdomain cleft (see text). The surface shown on the right is devoid of conserved residues. This figure was prepared with GRASP<sup>33</sup>. **(c)** Ribbon diagram highlighting the invariant residues of the PAZ-domain superfamily. The side chains of these residues, shown in stick representation, lie in the intersubdomain cleft. Aromatic residues lie on the inner surface of the  $\beta$ -hairpin helix subdomain whereas hydrophobic residues and a conserved cysteine line the bottom of the cleft.

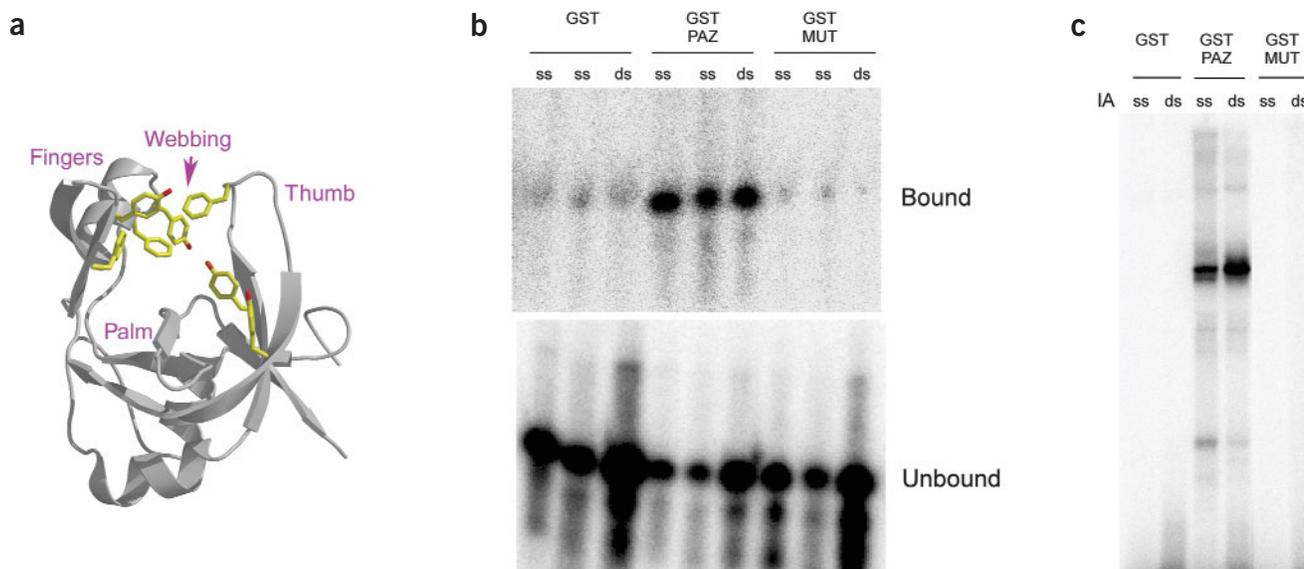
conserved residues, about half are buried, probably contributing to the fold of the PAZ domain, and half are exposed. However, notably, all of the invariant residues are exposed on the surface of the molecule. This strongly indicates that the intersubdomain cleft of the PAZ domain is involved in its biochemical function.

#### The intersubdomain cleft

The back side of the intersubdomain cleft is closed with side chains from both sides of the cleft and thus this region of the molecule resembles a right-handed baseball glove with the fingers made up of the

second subdomain, the thumb of portions of strands  $\beta 2$  and  $\beta 3$  and the loop between them, and the webbing between the thumb and the fingers of the side chains at the back of the cleft. The second region of invariant residues would thus be located on the palm of the glove. The cleft is lined with well ordered aromatic residues from the fingers, thumb and webbing that point toward the interior of the cleft (Figs. 5a and 1b). Also conserved are Lys704, located at the entrance of the cleft, and Arg652 on the thumb, which may contact phosphates.

Mapping the electrostatic potential on the surface of the protein does not yield an obvious positively charged surface for nucleic acid



**Figure 5** Aromatic residues lining the intersubdomain cleft are involved in RNA binding. **(a)** Ribbon diagram highlighting the aromatic residues lining the intersubdomain cleft. The structure resembles a right-handed baseball glove. Structural elements analogous to the fingers, palm, thumb and webbing are indicated. **(b)** GST-PAZ lacking conserved aromatic residues does not bind to single-stranded siRNAs. GST, wild-type or mutant GST-PAZ (as indicated) was used in GST pull-down experiments (see Methods) with single-stranded siRNA b'. Both the bound and unbound fractions are shown. The bound fraction is shown from a longer exposure, and quantification reveals that ~1% of the input RNA was recovered under the binding conditions used (see Methods). **(c)** GST-PAZ lacking conserved aromatic residues does not crosslink to siRNAs. GST, wild-type GST-PAZ or mutant GST-PAZ (see text) was incubated with siRNA b' as described. UV crosslinking (see Methods) reveals interaction with the wild type but no detectable interaction with the mutant protein.

binding. However, the stretch of invariant aromatic residues could interact with single-stranded nucleic acids via stacking interactions as is observed for many single-stranded nucleic acid-binding proteins<sup>18</sup>, including those containing OB-fold proteins<sup>14</sup>. Moreover, the intersubdomain cleft would need to open up considerably to accommodate a double-stranded A-type helix. It is well suited to bind an overhang of two nucleotides, both in its depth and width. Opening the cleft would require disruption of the contacts between the invariant Glu711 from the palm and the backbone of the fingers. Opening the cleft would require disruption of the contacts between the invariant glutamate Glu711 from the palm and the backbone of the fingers. In addition, the aromatic residues point inward (Fig. 5a) and as double-helical RNA would not fit in the cleft, it is unlikely that these residues contact a double helix sitting above the cleft.

To examine the role of the aromatic residues in the cleft in RNA binding, we mutated three of the invariant aromatic amino acids to alanine (Phe669, Tyr681 and Tyr687). These are all solvent exposed within the cleft and their side chains are not involved in any stabilizing interactions within the protein and therefore mutating them should not affect proper folding of the protein (Figs. 1b and 4c). Although we cannot completely rule out differences in stability between the mutant and the wild-type proteins, the mutant protein used in this study was well expressed, soluble and eluted with the same retention volume as the wild-type protein from a gel filtration column. The mutant was expressed as a GST fusion protein and examined for its ability to bind single-stranded RNAs and double-stranded siRNAs in both crosslinking and pull-down assays. Whereas the wild-type protein coprecipitated with both ssRNAs and siRNAs, the mutant protein was unable to bind either structure at a level above the background observed for GST alone (Fig. 5b). Similarly, the mutant was unable to crosslink to either single-stranded or double-stranded siRNAs (Fig. 5c). These data support the hypothesis that conserved aromatic residues within the cleft are involved in RNA binding.

## DISCUSSION

The structure of the PAZ domain of *Drosophila* Argonaute2 forms a new fold, structurally related to the OB fold present in several families of nucleic acid-binding proteins. Our crosslinking experiments suggest a possible role for the PAZ domain in interaction with the 3' ends of ssRNAs. Notably, this is not the first time such a function has been attributed to OB-containing proteins. Cdc13 from *S. cerevisiae* and telomere end-binding protein from *O. nova* are OB fold-containing proteins that are used to bind specifically to the single-stranded 3'-overhang of telomeric DNA<sup>19</sup>. In a DNA complex of TEBP with DNA, which uses three OB domains to bind ssDNA, the 3' end is well buried in the complex<sup>14</sup>.

The ability of the PAZ domain to interact with the 3' end of ssRNAs suggests a model in which the PAZ domain in Dicer and Argonaute proteins might recognize the unique terminal structures of siRNAs and pre-miRNAs. This contrasts with a previous model proposed by Bateman and colleagues that the PAZ domain might represent a protein-protein interaction motif<sup>11</sup> and our suggestion, based on that model, that PAZ might mediate the interaction between Argonaute2 and Dicer<sup>9</sup>. With the relatively low affinity of the Ago2-PAZ for either single- or double-stranded siRNAs, the majority of the affinity for siRNAs within RISC must come from other Argonaute domains or from other protein components of the complex. Indeed, canonical OB folds are used for nucleic acid binding in a modular fashion<sup>14</sup>. Some, such as the telomeric protein TEBP and RPA use multiple OB-fold domains, whereas in others, the OB fold forms part of a multicomponent nucleic acid binding domain.

The presence of the PAZ domain in both Argonaute and Dicer families supports the model that PAZ domains might interact with the 3' overhangs of siRNAs. siRNA-like ends are one common structure that must be recognized both by Argonaute-containing RISC complexes and by Dicer. In the former case, siRNAs must be accurately

**Table 1** Data collection and refinement statistics

	MBP-PAZ
<b>Data collection</b>	
Space group	$P6_5$
Unit cell dimensions (Å)	
$a = b$	89.81
$c$	380.33
Resolution limits (Å) <sup>a</sup>	50.0–2.80 (2.90–2.80)
Reflections <sup>a</sup>	172,206 (3,383)
Unique reflections	41,500
Completeness (%) <sup>a</sup>	97.3 (78.9)
$R_{\text{sym}}^{\text{a,b}}$	0.105 (0.548)
$\langle I \rangle / \langle \sigma I \rangle^{\text{a}}$	12.74 (1.76)
<b>Refinement</b>	
Reflections used (all data)	50.0–2.80 Å
Number of atoms	
Protein	11,702
Sugar	69
$\text{Ni}^{2+}$	9
Water	57
Total	11,837
$R_{\text{work}}^{\text{c}}$ (Reflections)	0.2268 (37,408)
$R_{\text{free}}^{\text{c}}$ (Reflections)	0.2809 (1,986)
Ramachandran statistics	
Core (%)	87.8
Allowed (%)	12.2
R.m.s. deviations	
Bond length (Å)	0.0079
Bond angle (°)	1.32

<sup>a</sup>Values in parentheses are for the highest-resolution shell. <sup>b</sup> $R_{\text{sym}} = \sum (I_{hkl} - \langle I_{hkl} \rangle) / (\sum I_{hkl})$ , where  $\langle I_{hkl} \rangle$  is the mean intensity of all reflections equivalent to reflection  $hkl$  by symmetry. <sup>c</sup> $R_{\text{work}}$  ( $R_{\text{free}}$ ) =  $\sum |F_o - F_c| / \sum F_o$  where 5% of randomly selected data were used for  $R_{\text{free}}$ .

discriminated from other small RNAs in the cell that might be generated by random degradation. For RISC, this is probably accomplished by a series of coupled assembly steps that result in the generation of an active RISC complex<sup>20</sup>. Numerous experiments support the notion that this discrimination occurs at least in part through recognition of the characteristic terminal structures of siRNAs, including both 3' overhangs and 5' terminal phosphates<sup>6,7</sup>. As Argonaute components are central players in RISC, they are reasonable candidates for providing some component of such recognition.

RISC, in its active form, contains single-stranded siRNAs, and indeed these can be used to create active RISC under *in vitro* assembly conditions<sup>7,10</sup>. However, the timing of the events that give rise to an active RISC complex containing an unwound siRNA *in vivo* has not been clarified. *In vitro* data indicates that RISC assembly includes an ATP-dependent unwinding step that is proposed to occur within the context of a siRNP<sup>20</sup>. However, it is not known whether unwinding occurs before incorporation into RISC, concurrently with assembly or after binding of a double-stranded siRNA to RISC subunits. In any of these cases, the PAZ domain may interact with the 3' end of the siRNA, present either in a single- or double-stranded form.

Dicer proteins cleave dsRNAs into siRNAs<sup>8</sup>. Experiments with purified Dicer proteins suggest that this family cleaves its substrates after recognition of an end of the dsRNA<sup>21</sup>, although no preference for specific end structures was noted. We have found that purified Dicer cleaves with increased efficiency if blunt dsRNAs are first pretreated

with limiting amounts of *E. coli* RNase III (J. Silva and G.J.H., unpublished data). Such experiments suggest that beyond end recognition, Dicer recognizes specific terminal structures. Such a model is consistent with recent studies of miRNA maturation in which miRNA precursors are first cleaved in the nucleus by Drosha, an RNase III family nuclease, before export to the cytoplasm and subsequent processing by Dicer<sup>22</sup>. As Drosha is predicted to generate termini with two-nucleotide 3' overhangs, Dicer might use such a structure to promote recognition and maturation of this intermediate. Again, the PAZ domain may contribute to recognition of this signature RNA structural element.

Dicer proteins come in two flavors, one of which contains a recognizable PAZ domain and the other of which does not. Organisms in which miRNAs have been identified all contain at least one PAZ-containing Dicer, including *D. melanogaster*, *Caenorhabditis elegans*, mammals and plants. However, some of these organisms also have Dicer proteins that lack a PAZ motif, including *Arabidopsis thaliana*, rice and *D. melanogaster*. Furthermore, *Schizosaccharomyces pombe*, in which no miRNAs have been identified to date, contains only one Dicer protein that lacks a PAZ motif. Although no biochemical or genetic data have yet functionally compared these two Dicer subclasses, one hypothesis based on our working model is that the Dicers, which lack PAZ motifs, might not participate in miRNA maturation.

Overall, the studies presented here provide a first glimpse into the molecular structure of a piece of the RNAi machinery. In combination with a biochemical characterization of the PAZ domain, these studies allow the proposal of a working model for how this domain fits into the biology of dsRNA-dependent gene silencing. This hypothesis must be tested by further biochemical and genetic experiments examining the role of the PAZ domain both in the context of complete proteins and intact multiprotein RNP complexes.

## METHODS

**Protein expression and purification.** The *D. melanogaster* Ago2 protein was subjected to limited proteolytic digestion (J. Jenco and L.J., unpublished data) to generate the PAZ domain fragment (residues 591–726). Ago2-PAZ domain was cloned into modified pMAL\_c2x vector. PAZ domain was expressed as a MBP fusion protein (MBP-PAZ) in *E. coli*, and purified using amylose affinity chromatography, anion exchange and gel filtration. Purified MBP-PAZ was concentrated up to 15 mg ml<sup>-1</sup> in 100 mM Tris-HCl, pH 8.0, and 100 mM NaCl. For GST-PAZ fusion proteins, the Ago2-PAZ domain was cloned into pGEX4T-1 vector (Amersham Pharmacia) and was expressed in *E. coli*. GST-PAZ proteins were purified using GST resins and eluted with 10 mM glutathione after extensive washing with lysis buffer containing 1 M NaCl. Further purification was done using anion-exchange chromatography and gel filtration.

**Crystallization.** Crystals of MBP-PAZ were grown at 17 °C by the hanging-drop method, mixing 1 μl of the protein solution with 1 μl of a reservoir solution containing 15% (w/v) PEG-MME 2K, 10 mM NiCl<sub>2</sub>, 5% (v/v) ethylene glycol, 100 mM NaF, 30 mM meso-erythritol (meso-1,2,3,4-tetrahydroxybutane) and 100 mM Tris-HCl, pH 8.5.

**Diffraction data measurement and structure determination.** For cryoprotection crystals were soaked for 30 min in crystallization solutions containing increasing amounts of glycerol in 2.5% steps to a final glycerol concentration of 20% (v/v). After an incubation for 3 h at –20 °C, the crystals were mounted in a fiber loop and flash frozen in liquid nitrogen. Diffraction data were collected to a resolution of 2.8 Å under cryogenic conditions (100 K) at beamline X26C at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL, Upton, New York, USA). Data were processed with the HKL suite (<http://www.hkl-xray.com>). The crystals are hexagonal and in the  $P6_5$  space group. They contained three molecules in the asymmetric unit, which we

named molecules P, A and Z. Further details on data collection statistics are shown in Table 1.

The structure was solved by molecular replacement with the program AMoRe<sup>23</sup> using the coordinates of MBP (PDB entry 1ANF) without the sugar or solvent atoms as a search model and a resolution range of 15–3.5 Å. The three highest peaks from the MR solution gave rise to placement of the three MBP protein molecules. Model phases were then calculated and used for initial  $2F_o - F_c$  and  $F_o - F_c$  electron density maps. About 60% of the backbone of the PAZ domain was built using O<sup>24</sup> and the new model phases were then used as starting phases for density modification using DM in CCP4 (ref. 25), and Prime and Switch in RESOLVE<sup>26</sup>. Subsequent iterative model building and density modification yielded placement of ~90% of the backbone followed by placement of ~70% of the side chain atoms. Crystallographic refinement was carried out with CNS<sup>27</sup> by using separate noncrystallographic symmetry (NCS) restraints on the MBP and the PAZ portions of the fusion protein and excluding the linker region. Iterative model building and refinement resulted in further placement of main chain and side chain atoms, the maltose sugar molecules, metal ions and solvent molecules. Group temperature factors were also refined later in the refinement.

The C-terminal seven, nine and ten residues are missing from molecules Z, P and A, respectively, in the final model, as are residues in two short loops in the MBP portion of the molecule. Further details on the refinement are shown in Table 1. Residues were numbered with reference to the Argonaute2 sequence, PFAM accession code Q9VUQ6.

**RNA binding and UV crosslinking.** All binding reactions were carried out in Buffer F<sup>9</sup> for 30 min at 30 °C. After incubation of RNAs and wild-type or mutant GST-PAZ, complexes were either UV crosslinked or recovered on glutathione agarose. UV crosslinking was done using a Stratelinker (Stratagene) and samples were exposed to 254 nM UV light for 10 min at room temperature. siRNA (0.1 pmol) was mixed with 25 pmol GST-tagged PAZ and incubated in 10 ml for 30 min at 30 °C before irradiation with 254 nM UV light. For GST pulldowns, binding reactions were diluted into 300 µl buffer F, mixed with 10 µl of packed glutathione agarose, incubated for 1 h at 4 °C, and washed three times with a buffer containing 150 mM NaCl, 0.5% (v/v) NP-40, 2 mM MgCl<sub>2</sub>, 2 mM CaCl<sub>2</sub> and 10 mM Tris, pH 7.5. Sequences of siRNAs were as follows: a, CUCCAGC-GAGCUCAGAGUCdTdT/GACUCUGAGCUCGUGGAGdTdT; b, CUUAC-GCUGAGUACUUGAAC/UCGAAGUACUCAGCGUAAAGUC; b', as b with dTdT overhangs, c, GGCGCCGUCGGUGUGGGCAAG/UGCCACACCC-GACGGCGCCUG; d, GGAGUCGUGAAAAUCGCGG/GCGAUUUU-UCAGCGACUCCAA. For DNA binding experiments, the oligonucleotide sequence was AAGGTTGACAAGATGGCACAG.

**Mutagenesis.** Mutagenesis was done on the GST-PAZ fusion protein using outward PCR with TGCGTCCGCCTTCCACAGTCGCAACGCC-CAATTGAAGTT as the forward primer and ATTGTGACCTTCTCCCATCGTGTTCGGCGGTCTCACT as the reverse primer. This reaction created a silent MfeI site and changed Phe669, Tyr681 and Tyr687 to alanine (underlined). Once the PCR product was obtained, it was phosphorylated, circularized and transformed into *E. coli*. The presence of the mutation was detected by MfeI digestion. The mutant fusion protein was purified using the same purification scheme as that for the wild-type GST-PAZ.

**Coordinates.** The atomic coordinates and structure factors have been deposited with the Protein Data Bank (accession code 1R6Z).

*Note: Supplementary information is available on the Nature Structural Biology website.*

#### ACKNOWLEDGMENTS

We thank E. Enemark and M. Carmell for help with refinement and figures, A. Caudy for aid in characterizing the biochemical properties of GST-PAZ, M. Myers for mass spectrometry and A. Heroux (beamline X26C) for support with data collection at the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. The NSLS is supported by the US Department of Energy, Division of Materials Sciences and Division of Chemical Sciences. J.J.S. is a Bristol-Myers Squibb Predoctoral Fellow, N.H.T. is a Leslie C. Quick Jr. Predoctoral Fellow. This work was supported by the Watson School of Biological Sciences (L.J.) and the US National Institutes of Health (G.J.H.). G.J.H. is a Rita Allen Foundation

Scholar and is supported by an Innovator Award from the US Army Breast Cancer Research Fund.

#### COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 16 October; accepted 24 October 2003

Published online at <http://www.nature.com/naturestructuralbiology/>

- Hannon, G.J. RNA interference. *Nature* **418**, 244–251 (2002).
- Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
- Carrington, J.C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336–338 (2003).
- Volpe, T. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297**, 1833–1837 (2002).
- Hall, I.M. *et al.* Establishment and maintenance of a heterochromatin domain. *Science* **297**, 2232–2237 (2002).
- Elbashir, S.M., Martinez, J., Patkaniowska, A., Lendeckel, W. & Tuschl, T. Functional anatomy of siRNAs for mediating efficient RNAi in *Drosophila melanogaster* embryo lysate. *EMBO J.* **20**, 6877–6888 (2001).
- Schwarz, D.S., Hutvagner, G., Haley, B. & Zamore, P.D. Evidence that siRNAs function as guides, not primers, in the *Drosophila* and human RNAi pathways. *Mol. Cell* **10**, 537–548 (2002).
- Bernstein, E., Caudy, A.A., Hammond, S.M. & Hannon, G.J. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**, 363–366 (2001).
- Hammond, S.M., Boettcher, S., Caudy, A.A., Kobayashi, R. & Hannon, G.J. Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293**, 1146–1150 (2001).
- Martinez, J., Patkaniowska, A., Urlaub, H., Luhrmann, R. & Tuschl, T. Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* **110**, 563–574 (2002).
- Cerutti, L., Mian, N. & Bateman, A. Domains in gene silencing and cell differentiation proteins: the novel PAZ domain and redefinition of the Piwi domain. *Trends Biochem. Sci.* **25**, 481–482 (2000).
- Carmell, M.A., Xuan, Z., Zhang, M.Q. & Hannon, G.J. The Argonaute family: tentacles that reach into RNAi, developmental control, stem cell maintenance, and tumorigenesis. *Genes Dev.* **16**, 2733–2742 (2002).
- Murzin, A.G. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861–867 (1993).
- Theobald, D.L., Mitton-Fry, R.M. & Wuttke, D.S. Nucleic acid recognition by OB-fold proteins. *Ann. Rev. Biophys. Biomol. Struct.* **32**, 115–133 (2003).
- Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540 (1995).
- Anderson, E.M., Halsey, W.A. & Wuttke, D.S. Site-directed mutagenesis reveals the thermodynamic requirements for single-stranded DNA recognition by the telomere-binding protein Cdc13. *Biochemistry* **42**, 3751–3758 (2003).
- Hockensmith, J.W., Kubasek, W.L., Vorachek, W.R., Evertsz, E.M. & von Hippel, P.H. Laser cross-linking of protein–nucleic acid complexes. *Methods Enzymol.* **208**, 211–236 (1991).
- Anston, A. Single-stranded RNA binding proteins. *Curr. Opin. Struct. Biol.* **10**, 87–94 (2000).
- Theobald, D.L., Cervantes, R.B., Lundblad, V. & Wuttke, D.S. Homology among telomeric end-protection proteins. *Structure* **11**, 1049–1050 (2003).
- Nykanen, A., Haley, B. & Zamore, P.D. ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107**, 309–321 (2001).
- Zhang, H., Kolb, F.A., Brondani, V., Billy, E. & Filipowicz, W. Human Dicer preferentially cleaves dsRNAs at their termini without a requirement for ATP. *EMBO J.* **21**, 5875–5885 (2002).
- Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
- Navaza, J. & Saludjian, P. AMoRe: an automated molecular replacement program package. *Methods Enzymol.* **276**, 581–594 (1997).
- Jones, T.A. & Kjeldgaard, M. Electron-density map interpretation. *Methods Enzymol.* **277**, 173–208 (1997).
- Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Terwilliger, T.C. Maximum likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
- Brünger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Esnouf, R.M. An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph.* **15**, 132–134 (1997).
- Kraulis, P.J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950 (1991).
- Bacon, D.J. & Anderson, W.F. A fast algorithm for rendering space-filling molecule pictures. *J. Molec. Graph.* **6**, 219–220 (1988).
- Merritt, E.A. & Murphy, M.E.P. Raster3D version 2.0—a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
- Rees, B., Webster, G., Delarue, M., Boeglin, M.A. & Moras, D. Aspartyl tRNA-synthetase from *Escherichia coli*: flexibility and adaptability to the substrates. *J. Mol. Biol.* **299**, 1157–1164 (2000).
- Nicholls, A., Sharp, K.A. & Honig, B. Protein folding and association—insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).