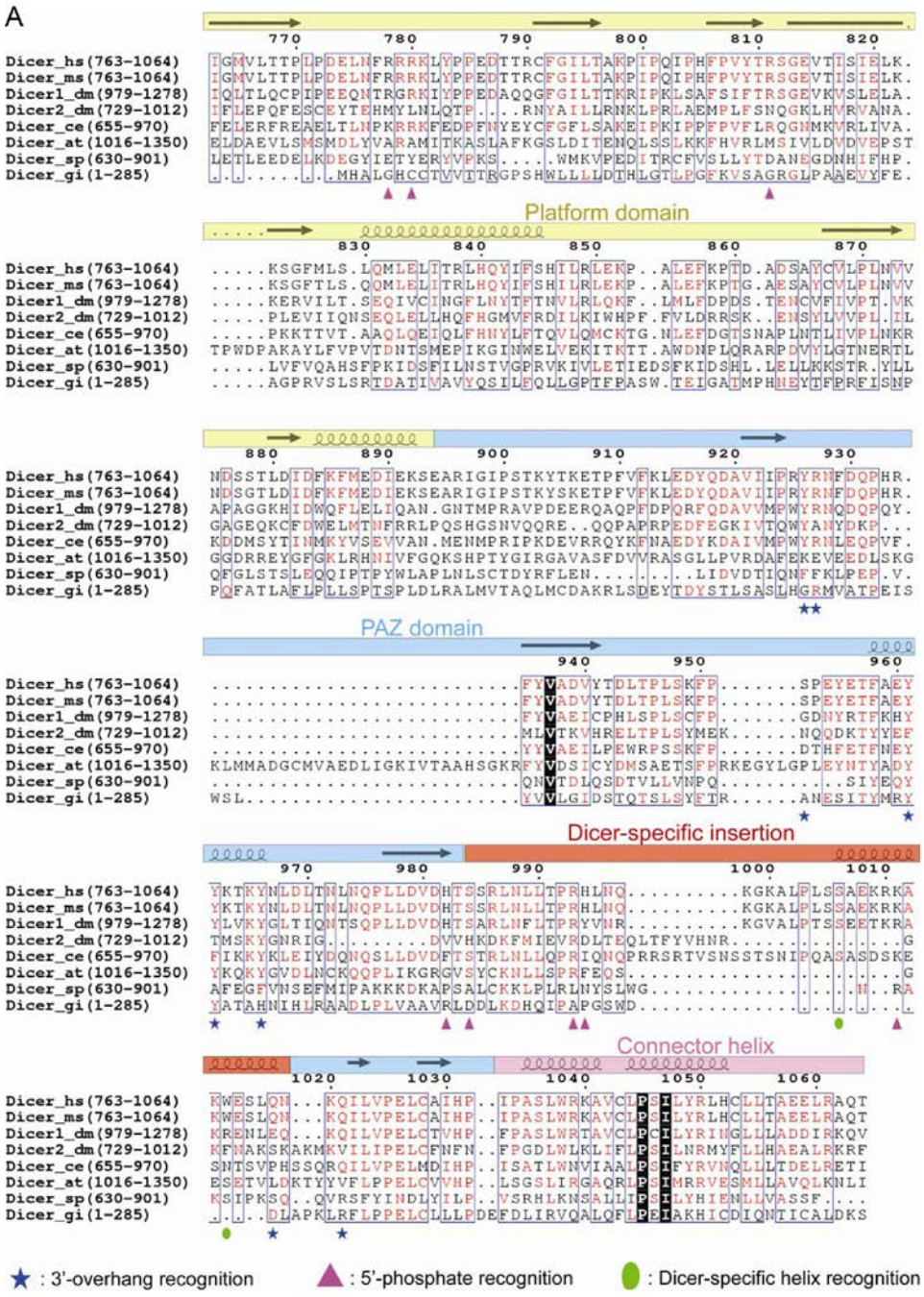

B

```
                    *         *         *         *         *
hAgo1-PAZ    225:  --AQPVIEFXCEVLDIRNIDEQPKPLTDSQRVRFTKEIKGLKVEVTHCGQ-
hDicer-PAZ   887:  MEDIEKSEARIGIPSTKYTKETP-------FVFKLEDYQDAVIIPRYRNFD

                    *         *         *         *         *
hAgo1-PAZ    273:  XKRKYRVCNVTR-RPASHQTFFLQVECTVAQYFKQKYNLQLKYPHLPCLQV
hDicer-PAZ   931:  QPHRFYVADVYTDLTPLS-KFESPEYETFAEYYKTKYNLDLTNLNQELLGV

                    *         *         *         *         *
hAgo1-PAZ    329:  G----------------------QEQK-HTYLPLEVCNIVAGQR
hDicer-PAZ   981:  DHTSSRLNLLTPLPLSSAEKRKAKWESLQNKQILVPELCAIHP---
```

Figure S1. **Species-specific Sequences of Dicer PAZ Domain Flanking Elements and Sequence Comparison Between PAZ Domains of Human Ago1 and Dicer.**

(A)**,** Comparison of Dicer PAZ domain and flanking element sequences. The sequences from top to bottom are for human, mouse, *D. melanogaster* (Dicer1 and Dicer2), *C. elegans, A. thaliana, S. pombe* and *G. intestinalis.* Amino acids involved in the recognition of 3'-overhang and 5'-phosphate are shown below the aligned sequence. Amino acids present in Dicer-specific helix which interact with siRNA are also indicated. (B), Structure-based sequence alignment of PAZ domains from human Ago1 (PDB number:1SI3) and human Dicer showing the region of Dicer specific insertion (984-1020). Similar and conserved residues are shaded in pink and blue, respectively.

Related to Figure 1.

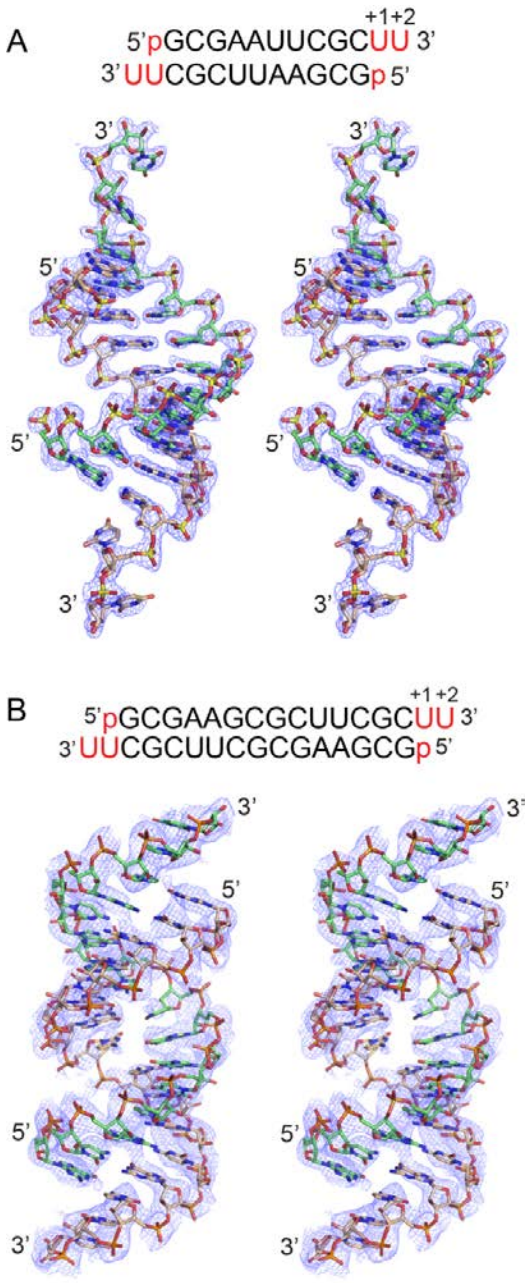

Figure S2. **Omit Maps for Bound siRNA 12-mer and 16-mer containing 5'phosphates and UU-overhangs at 3'-ends in their complexes with hDicer PAZ.**

(A) Stereo view of omit map (light blue mesh) contoured at 1.2 σ for the bound siRNA in the hDicer PAZ cassette-siRNA 12mer complex (1.95 Å resolution).

(B) Stereo view of omit map (light blue mesh) contoured at 1.0 σ for the bound siRNA in the hDicer PAZ cassette -siRNA 16mer complex (3.40 Å resolution).

Related to Figure 1.

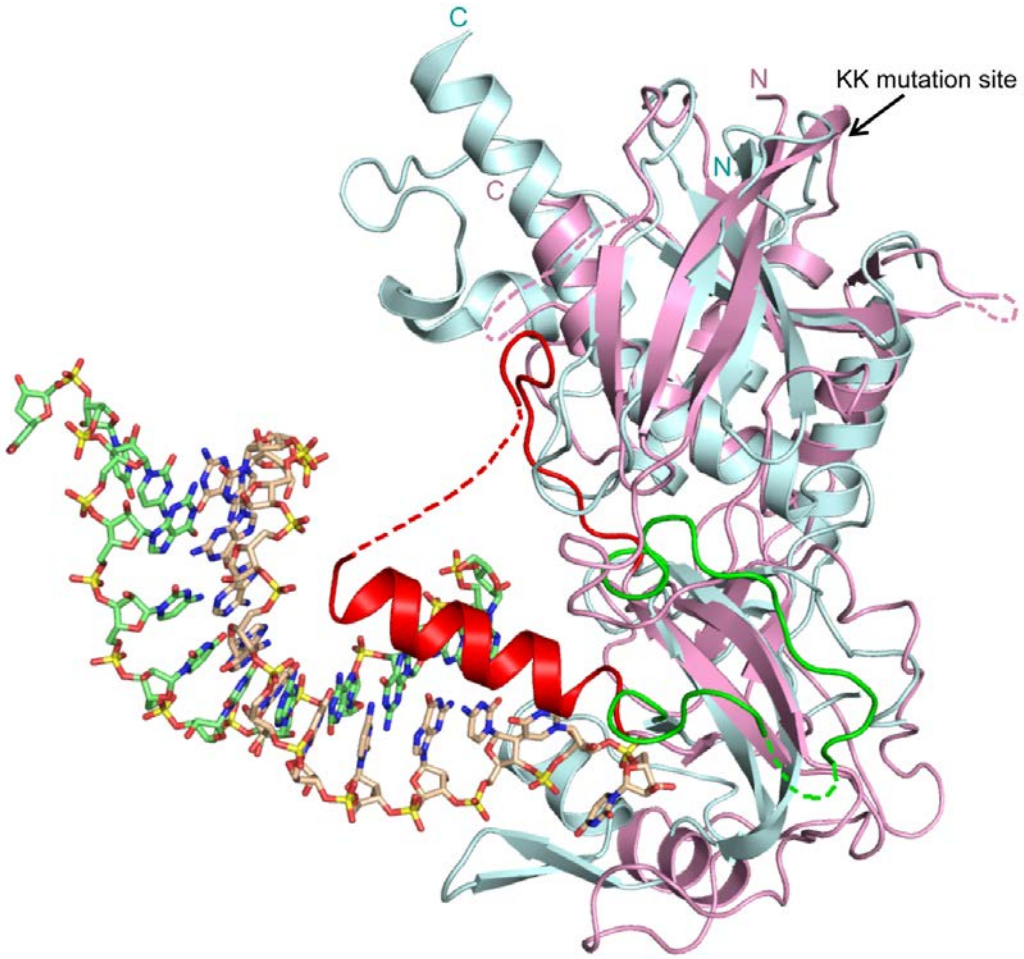

Figure S3. **Superposition of Platform-PAZ-connector Helix Segments of hDicer-RNA Complex and *G. intestinalis* Dicer in Free State (PDB number: 2FEL).**

Superposition of the complex of hDicer platform-PAZ-connector helix (in magenta) with bound RNA duplex containing 2-nt overhangs at 3'-end with that of *G. intestinalis* Dicer platform-PAZ-connector helix (in blue) in the free state. Note that the hDicer-specific insert is shown in red and its counterpart in *G. intestinalis* Dicer is shown in green. Position of two surface lysines which were mutated to alanine in hDicer PAZ cassette for obtaining better diffraction is shown.

Related to Figures 1 and 2.

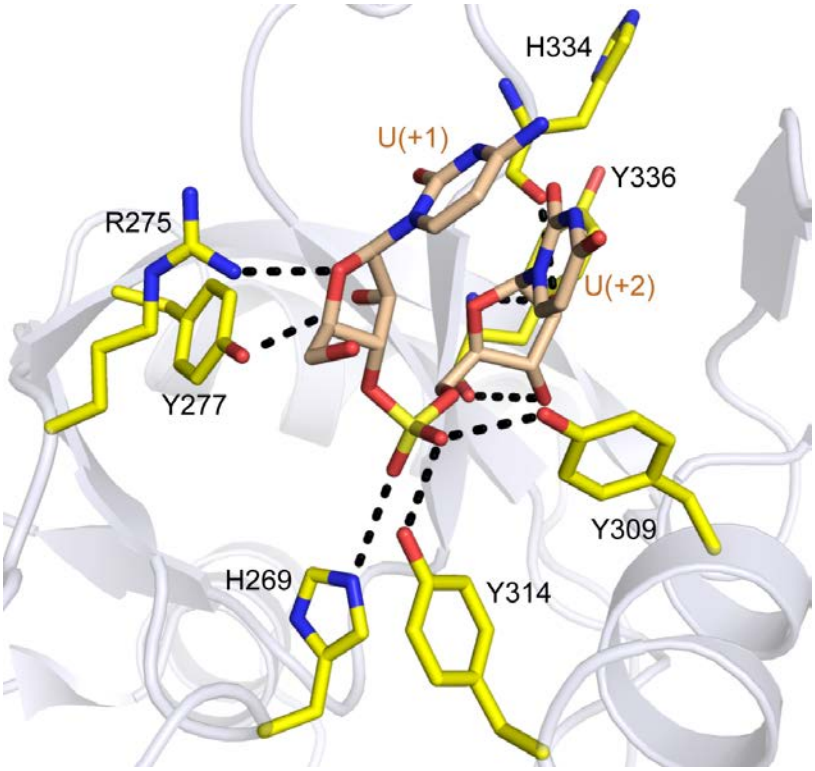

Figure S4. Structure of 2-nt Overhang Binding Pocket in the hAgo PAZ-siRNA Complex (PDB number: 1SI3).

Intermolecular hydrogen bonds involving the backbone phosphate and 2'-OH groups of the 2-nt overhang at the 3'-end and residues lining the 3'-pocket in the hAgo PAZ-siRNA complex. The overhang residues are labelled as U(+1) and U(+2)

Related to Figure 2.

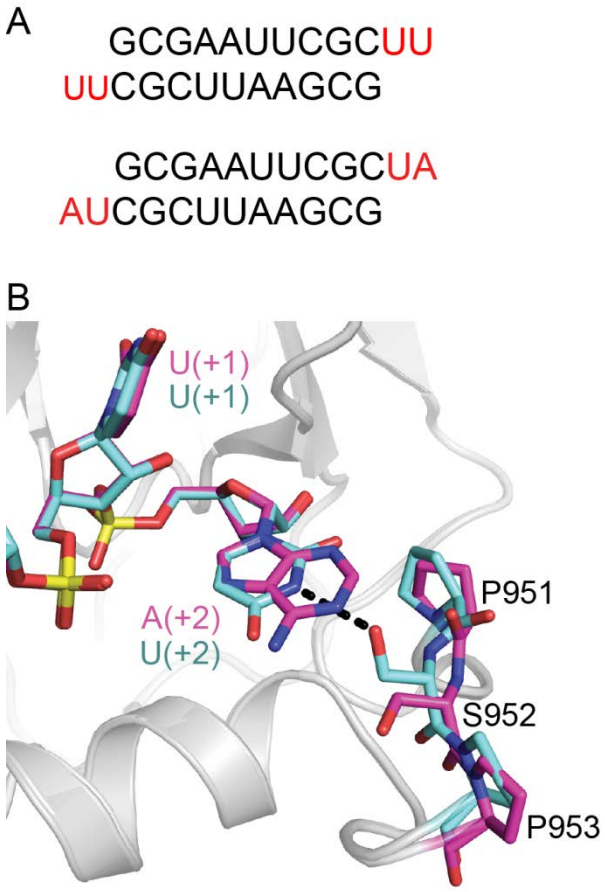

Figure S5. Plasticity of the 2-nt Overhang-binding Pocket and Structural Details of hDicer PAZ-RNA Complex Containing 2-nt UU and UA Overhangs.

(A) RNA sequences used to generate 2-nt UU and UA overhangs at 3'-end for complex formation with hDicer PAZ cassette. These sequences lack a 5'phosphate.

(B) Comparison of the overhang bases at the (+1) and (+2) positions and the Pro951-Ser952-Pro953 segment of the 3'-pocket, when U(+2) (protein and RNA in cyan) is replaced by A(+2) (protein and RNA in magenta).

Related to Figure 2.

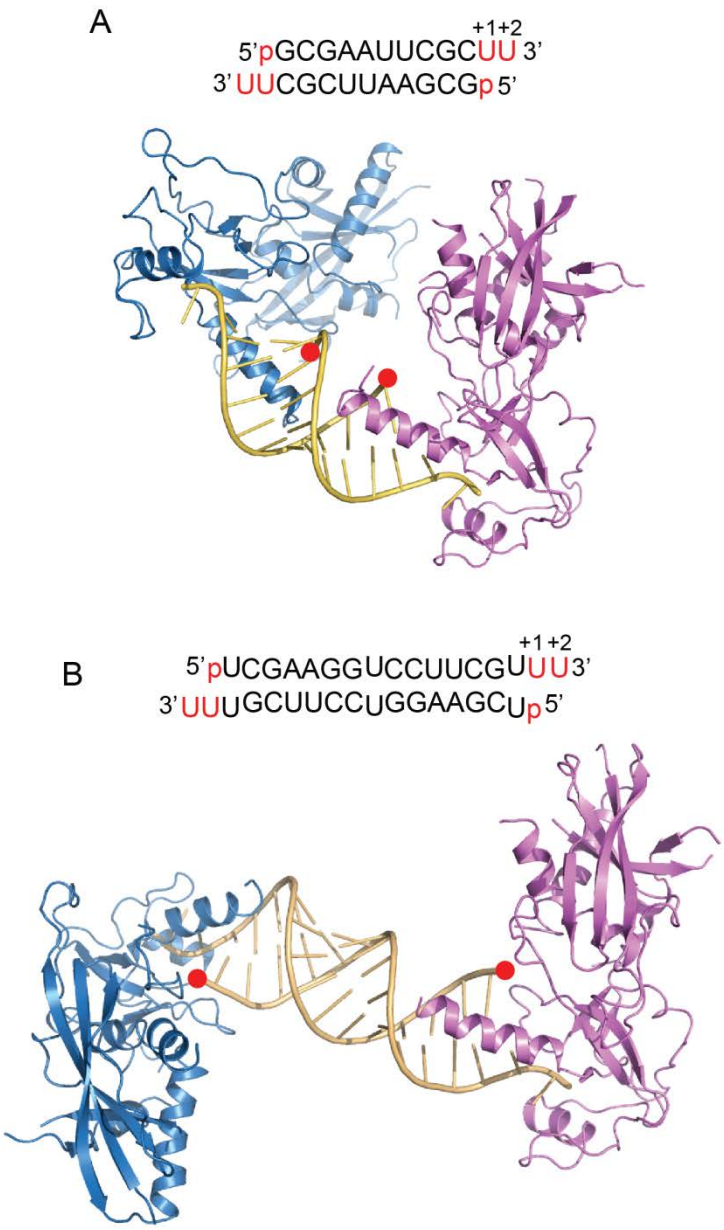

Figure S6. Crystal Packing Interaction for the RNA in hDicer PAZ cassette Complexes Obtained with (A) 12-mer and (B) 17-mer siRNAs.

Each end of siRNA is bound by one hDicer PAZ cassette. In the hDicer PAZ cassette complexes, RNA is positioned away the protein surface so that 5'phosphate is far from the 5'-pocket whereas the 2-nt overhang at 3'-end is tightly anchored in the 3'-pocket. Position of 5'-phosphate is highlighted by a red ball.

Related to Figure 1 and 2.

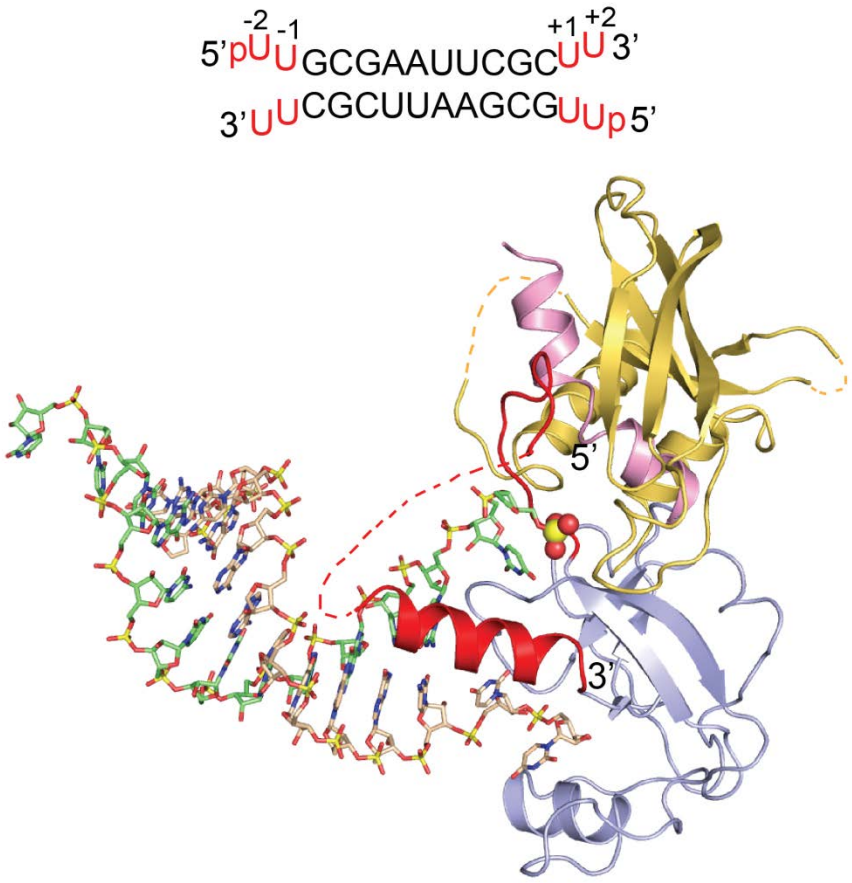

Figure S7. Structure of hDicer PAZ Cassette Bound to the 14-mer siRNA

Structure of hDicer PAZ cassette bound to a 14-mer siRNA. The platform, PAZ and connector helix are colored in yellow, blue and pink, respectively. The 'Dicer PAZ-insertion element' is colored in red and composed of a disordered segment followed by the 'hDicer-specific helix'. The 5'-phosphate is shown in a space-filling representation bound in the phosphate pocket and UU-3' overhangs bounds in the 3'-pocket.

Related to Figure 5.

Table S1. Data Collection, phasing and Refinement Statistics for hDicer PAZ cassette-siRNA 16-mer Complex. Values in parentheses are for highest-resolution shell.

	5'p/UU-3' **16-mer**
Data Collection	
Wavelength	0.9792 (Se-Peak)
Space group	$I4_122$
Unit cell a, b, c (Å) α, β, γ (°)	 174.46, 174.46, 74.76 90.0, 90.0, 90.0
Resolution (Å)	20.0-3.4 (3.52-3.40)
Total reflections	55186
Unique reflections	8201
$I/\sigma I$	30.4 (4.3)
Completeness (%)	99.4 (98.5)
Redundancy	6.8 (6.7)
R_{merge}	6.9 (38.2)
Phasing	
Anomalous completeness	99.2 (97.6)
Anomalous redundancy	3.7 (3.5)
Refinement	
R_{work} / R_{free}	27.7/31.7
No. of non-H atoms Protein RNA	 1774 340
Wilson B factor (Å^2)	107.4
Average B-factors (Å^2) Protein RNA	 134.1 188.1
RMSD Bond lengths (Å) Bond angles (°)	 0.006 1.191

Related to Figure 6.

Table S2. Data Collection and Refinement Statistics for hDicer PAZ Cassette-RNA Complexes. Values in parentheses are for highest-resolution shell.

	5'-p/UU-3' 12-mer	5'-pU/UU-3' 13-mer	5'-pUU/UU-3' 14-mer	5'-pUUU/UU-3' 15-mer	5'-p/UU-3' 17-mer
Data collection					
Space group	C2	C2	C2	C2	$C222_1$
Unit cell (Å) (°) a, b, c α β γ	 108.71 84.62 51.36 90.0 113.33 90.0	 109.32 83.85 51.09 90.0 113.19 90.0	 118.33 84.48 51.07 90.0 109.01 90.0	 117.04 84.13 51.50 90.0 109.67 90.0	 99.25 104.19 368.64 90.0 90.0 90.0
Resolution* (Å)	50.0-1.95 (2.02-1.95)	30.0-2.6 (2.69-2.60)	30.0-2.55 (2.64-2.55)	50.0-2.55 (2.64-2.55)	50.0-3.1 (3.21-3.10)
Total reflections	1061211	283246	227956	470104	3811656
Unique reflections	30619	13427	15324	15388	35057
$I/\sigma I$	39.9 (3.1)	23.0 (1.8)	28.1 (2.9)	27.7 (2.5)	28.9 (3.7)
Completeness (%)	93.9 (70.2)	99.4 (96.1)	94.9 (99.9)	99.5 (96.7)	91.6 (87.8)
Redundancy	8.8 (4.5)	4.6 (3.7)	3.0 (3.3)	4.0 (3.9)	13.1 (13.4)
R_{merge}	7.1 (28.4)	10.7 (69.5)	7.1 (50.6)	7.3 (44.9)	12.0 (65.0)
Refinement					
R_{work} / R_{free}	17.7/20.4	19.6/28.4	21.1/27.4	19.3/24.5	22.3/27.0
No. of atoms Protein RNA Water	 4625 383 254	 2128 274 14	 2176 285 10	 2215 308 16	 8282 1428 0
Average B-factor (Å^2) Protein RNA Water	 41.4 42.5 46.7	 75.4 102.5 67.5	 78.9 117.2 61.8	 80.0 113.3 58.9	 70.0 48.8
RMSD Bond lengths (Å) Bond angles (°)	 0.005 0.960	 0.007 1.015	 0.006 0.978	 0.007 1.089	 0.006 1.165

Related to Figures 1, 2 and 5.

Table S3. Data Collection and Refinement Statistics for hDicer PAZ Cassette-RNA Complexes. Values in parentheses are for highest-resolution shell.

	5'/UU-3' **12-mer**	5'/UA-3' **12-mer**	5'-A/UU-3' **13-mer**
Data collection			
Space group	C2	C2	I222
Unit cell a, b, c (Å) α, β, γ (°)	 109.45, 84.68, 51.63 90.0, 113.30, 90.0	 109.08, 84.48, 51.55 90.0, 113.49, 90.0	 85.90, 97.24, 106.77 90.0, 90.0, 90.0
Resolution (Å)	50.0-2.25 (2.33-2.25)	50.0-2.1 (2.18-2.10)	50.0-2.6 (2.69-2.60)
Total reflections	513964	416840	359768
Unique reflections	20624	24912	14311
$I/\sigma I$	25.2 (8.3)	25.1 (3.1)	30.9 (2.4)
Completeness (%)	99.8 (100.0)	95.2 (75.8)	99.9 (100.0)
Redundancy	6.9 (6.7)	4.9 (3.1)	9.7 (9.0)
R_{merge}	9.9 (40.4)	6.5 (21.4)	8.8 (72.0)
Refinement			
R_{work} / R_{free}	18.2/22.0	17.2/20.2	18.6/23.9
No. of non-H atoms Protein RNA Water	 2234 250 194	 2284 252 227	 2157 270 26
Average B-factors (Å^2) Protein RNA Water	 37.4 47.2 42.9	 31.8 42.1 40.8	 50.0 91.9 55.0
RMSD Bond lengths (Å) Bond angles (°)	 0.000 1.131	 0.008 1.121	 0.008 1.094

Related to Figures 4 and 5.

Table S4. Crystallization Conditions in Which Crystals for Different Complexes Were Obtained.

RNA used in the study	Crystallization condition
Wild type human Dicer 'platform-PAZ-connector helix' cassette	
5'-p/UU-3' (16-mer)	0.2 M KCl, 0.1 M Mg acetate, 0.05 M Na-Cacodylate, pH 6.5, 12 % PEG8000
5'-p/UU-3' (17-mer)	0.08M HEPES sodium pH 7.5, 8% 2-Propanol, 16% PEG 4000
Human Dicer 'platform-PAZ-connector helix' cassette (K832A/K833A)	
5'-p/UU-3' (12-mer)	0.2 M Li_2SO_4, 0.1 M HEPES, pH 7-7.5, 16% PEG3350
5'/UU-3' (12mer)	0.8 M Na/K tartrate, 0.1 M Na citrate, pH 6.3
5'/UA-3' (12-mer)	0.2 M Na/K tartrate, 0.1 M bis-Tris-propane, pH 7.0, 18-22% PEG3350
5'-A/UU-3' (13-mer)	0.2 M NaCl, 0.1 M Na/K phosphate, pH 6.0-6.4, 10% PEG8000
5'-pU/UU-3' (13-mer)	0.2 M NaCl, 0.1 M Na citrate, pH 6.5, 18% PEG4000
5'-pUU/UU-3' (14-mer)	0.2 M Na formate, 0.1 M BICINE, pH 8.1-8.5, 16-19% PEGMME5000
5'-pUUU/UU-3' (15-mer)	0.2 M NaCl, 0.05 M Tris-HCl, pH 8.3-8.8, 12-15% PEG20000

Related to Figures 1, 2, 4, 5 and 6.

Table S5. RNA Sequences used for Crystallization Experiments and Biotinylated RNA Sequences used for SPR Binding Experiments.

5'-p/UU-3' (12-mer)	pGCGAAUUCGCUU UUCGCUUAAGCGp
5'/UU-3' (12-mer)	GCGAAUUCGCUU UUCGCUUAAGCG
5'/UA-3' (12-mer)	GCGAAUUCGCUA AUCGCUUAAGCG
5'-pU/UU-3' (13-mer)	^{pU}GCGAAUUCGC^{UU} $_{UU}$CGCUUAAGCG$_{Up}$
5'-A/UU-3' (13-mer)	A GCGAAUUCGC^{UU} $_{UU}$CGCUUAAGCG A
5'-pUU/UU-3' (14-mer)	^{pUU}GCGAAUUCGC^{UU} $_{UU}$CGCUUAAGCG$_{UUp}$
5'-pUUU/UU-3' (15-mer)	^{pUUU}GCGAAUUCGC^{UU} $_{UU}$CGCUUAAGCG$_{UUUp}$
5'p/UU-3' (16-mer)	pGCGUUGGCCAACGCUU UUCGCAACCGGUUGCGp
5'-p/UU-3' (17-mer)	pUCGAAGGUCCUUCGUUU UUUGCUUCCUGGAAGCUp
5'-p/UU-3' (21-mer)	5'- pUCGAAGUAUUCCGCGUACGUU 3'- Biotin - UUAGCUUCAUAAGGCGCAUGCp
5'-pUU/UU-3' (21-mer)	5'- pUCGAAGUAUUCCGCGUACG^{UU} 3'- Biotin - UUAGCUUCAUAAGGCGCAUGC$_{UUp}$
5'p/UU-3' (21-mer)	5'pUCGAAGUAUUCCGCGUACU^{UU} 3'- Biotin-UUAGCUUCAUAAGGCGCAUGUp

Related to Figures 1, 2, 4, 5 and 6.

SUPPLEMENTAL EXPERIMENTAL PROCEEDURES

Gene Cloning and Protein Preparation

The human Dicer "platform-PAZ-connector helix" cassette (residues 755-1,055) was generated using a standard PCR based cloning strategy. It was cloned into a modified pET-28b vector (EMD Biosciences) with a His_6-SUMO tag at the N-terminus. All clones were overexpressed in *E. coli* strain BL21(DE3) cells in LB medium. The cells were induced by 1 mM IPTG when OD_{600} reached a value of 0.6 and were cultured overnight at 20 C in shaker. The cell extract was prepared using a French press. The proteins were first purified from the soluble fraction by a nickel-chelating affinity column, followed by SUMO protease Ulp1 treatment overnight at 4ºC to cleave the His_6-SUMO tag, which was later removed via a second nickel-chelating column. Proteins were further purified by gel filtration chromatography using Hiload 26/60 Superdex-75 preparative grade column (GE Healthcare) in buffer containing 20 mM Tris-HCl, pH 8.0 and 100 mM NaCl. Protein purity was assessed by SDS-PAGE. The peak fractions containing pure protein were pooled and concentrated using 30 KDa cut-off centrifugal concentrator (Vivaspin). The purified protein was immediately frozen in liquid nitrogen and stored at -80 ºC. All mutants discussed in this study were generated by a two-step PCR-based overlap extension method and confirmed by DNA sequencing. For phasing, L-selenomethionine (Se-Met) derivative protein was produced using the feedback-inhibition of methionine synthesis pathway in minimum media (M9) and purified in the same way as native protein. 1 mM DTT was maintained during protein purification to prevent oxidation of selenium atoms.

RNA Preparation

All the 5'-phosphorylated RNAs were purchased from Dharmacon. The non-phosphorylated RNAs were chemically synthesized on a 3400 DNA synthesizer (Applied Biosystems), deprotected and purified by denaturing polyacrylamide gel electrophoresis. The self-complementary RNAs were finally dissolved in DEPC-treated water and annealed by heating to 95 ºC for 1 min followed by temperature reduction in a 37 ºC water bath over a period of one hour.

Crystallization and Data Collection

The complex crystals were grown using the hanging-drop vapor-diffusion method by mixing the protein-RNA complex (approx. 10 mg/ml in 100 mM NaCl, 20 mM Tris-HCl, pH 8.0) with equal volume of reservoir solution (Table S4). The diffraction data were collected using synchrotron

radiation at beamlines ID24-C and ID24-E at Advanced Photon Source and X29 at National Synchrotron Light Source. All measured diffraction spots were indexed, integrated, and scaled using the HKL2000 package (Otwinowski and Minor, 1997).

Structural Determination

The crystal structure of human Dicer PAZ cassette (residues 755-1055) in complex with self-complementary 16-mer siRNA containing a 5'-phosphate and a 2-nt overhang at 3'-end was solved by SAD method using data collected at Se-peak wavelength to a resolution of 3.4 Å in space group $I4_122$. The position of selenium atoms were located by the program SHELXD (Schneider and Sheldrick, 2002), which was further refined by the AUTOSHARP algorithm (Bricogne et al., 2003). Initial phases were further improved by density modification using SOLOMON (Abrahams and Leslie, 1996) as implemented in the SHARP module. The initial experimental map showed interpretable electron density in which protein and RNA could be built manually. Further model improvement was carried out with alternate rounds of refinement using Phenix.refine (Afonine et al., 2005) and model building via COOT (Emsley and Cowtan, 2004). In this structure, we observed relatively poor electron density for segments of the platform domain compared to the PAZ domain and bound RNA. Because of limited resolution and relatively poor electron density, it was difficult to assign amino acids unambiguously for the platform domain. Careful analysis of the data suggested no sign of twinning. Instead, poor crystal packing interactions for the platform domain and overall high B-factor of the data set, as well as anisotropic diffraction, appear to be responsible for the poor/noisy electron density for segments of the platform domain. Later, using the high resolution structure of hDicer PAZ cassette-siRNA 12-mer complex, amino acid sequence assignment was completed and the structure was further refined using Phenix.refine. During the refinement, we used protein atoms from the high-resolution structure of hDicer PAZ cassette as a reference structure to generate a set of dihedral restraints to each matching dihedral in the working model. The structure refinement was completed with cycles of minimization, group B-factor refinement along with TLS parameters, leading to a final structure with R_{work} of 27.7% and R_{free} of 31.7 %. The final map (2Fo–Fc) showed reasonably good quality electron density for the final model except for some segments in the platform domain, which has relatively poor density. The final model is of good stereochemical quality and Ramachandran plot of the main-chain torsion angles showed 89% of the residues were found in the most allowed regions. The data collection and refinement statistics for this structure of the complex are listed in Table S1.

For improving diffraction quality of crystals, we made three different double mutants (in which surface residues such as Lys or Glu were mutated to Ala) based on crystal contacts observed in the partially refined structure (see above), as well as predictions made by surface entropy reduction server (http://services.mbi.ucla.edu/SER/). One such mutant in which two surface lysines (Lys832 and Lys833) were mutated to Ala gave a crystal in complex with 12-mer siRNA having 5'-phosphate and 2-nt overhang at 3'-end which diffracted to a resolution of 2.25 Å. These two lysine residues are very far from the RNA binding site (Figure S3). This structure was solved by PHASER (McCoy et al., 2007) using protein atoms of partially refined low resolution structure of hDicer PAZ cassette-siRNA 16-mer complex as a search model. The initial model was further improved by manual and automated model building as well as refinement using the PHENIX program package (Adams et al., 2002; Afonine et al., 2005). At this stage, electron density corresponding to RNA was very clear. The model was further improved by refinement using rigid body, simulated annealing, positional and TLS refinement at different stages. Water molecules were included near the end of refinement. The majority of the model has a clear and well interpretable electron density map with the exception of a few solvent-exposed side chains, which were omitted in the final model.

The crystals of K832A/K833A mutant in complex with 5'-pU/UU-3' 13-mer RNA, 5'-pUU/UU-3' 14-mer RNA and 5'-pUUU/UU-3' 15-mer RNA were improved for better diffraction using repeated micro-seeding along with addition of 10% (v/v) detergent additive cyclohexylpropanoyl-N-hydroxyethylglucamide (C-HEGA®-9) into crystallization mixture drop. The crystals of wild-type hDicer PAZ cassette in complex with 17-mer siRNA having 5'-phosphate and 2-nt UU overhang at 3'-end were improved for better diffraction with addition of 10% (v/v) detergent additive cyclohexyl-ethyl-β-D-maltoside/2-cyclohexyl-1-ethyl-β-D-maltoside (CYMAL®-2).

All other crystal structures of hDicer PAZ cassette (wild-type or K832A/K833A) in complex with siRNAs having different overhangs at 5'- and 3'-ends (Table S5) discussed in this paper were solved by molecular replacement or by difference Fourier using protein atoms of the refined structure of human Dicer PAZ cassette (K832A/K833A) in complex with 12-mer siRNA having 5'-phosphate and 2-nt UU overhang at 3'-end. Model building and refinement were carried out using the protocol mentioned above. The correctness of stereochemistry of the models was verified using PROCHECK (Laskowski et al., 1993). All crystallographic data and refinement statistics are summarized in Tables S1, S2 and S3. Structural comparisons were performed using the SSM algorithm (Krissinel and Henrick, 2004) implemented in COOT. The figures were prepared using PyMOL (DeLano, 2008) and ESPript (Gouet et al., 2003).

Surface Plasmon Resonance (SPR) Binding Assay

All experiments were performed on BIAcore 2000 biosensor (BIAcore) at 25 ºC (Katsamba et al., 2002). The RNA sequences used in this assay were biotinylated at the 3'-end on one strand of the duplex RNA (Figure S7). All RNA oligomers were purchased from Dharmacon. The 3'-biotinylated RNA was immobilized on a Sensor Chip SA (GE Healthcare) after the surface had been conditioned with three injections of conditioning fluid (1 M NaCl, 50 mM NaOH). The low-density immobilization surface was achieved by generating a 25-35 response unit (RU). The duplex was formed by injecting 500 nM of complementary RNA in 1 M NaCl, followed by washing to remove all unbound RNAs. Wild-type or mutant proteins were diluted in running buffer HBS-EP (10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM EDTA, 0.005% Surfactant P20) at concentrations ranging from 1 nM to 10 µM. Proteins were then flown over the immobilized RNAs at a rate of 30 µl/min, followed by dissociation monitored by flowing the running buffer over a 3 minutes period. The chip was regenerated by injecting 20 µl of 2 M NaCl at 20 µl/min. Scrubber 2.0 (Center for Biomolecular Interaction Analysis, University of Utah) was used to calculate saturation fraction for use in the equilibrium K_d analysis.

SUPPLEMENTAL REFERENCES

Abrahams, J.P., and Leslie, A.G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallogr D Biol Crystallogr *52*, 30-42.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D *59*, 2023-2030.

DeLano, W.L. (2008). The PyMOL Molecular Graphics System. (Palo Alto, CA, USA, DeLano Scientific LLC).

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr *60*, 2126-2132.

Gouet, P., Robert, X., and Courcelle, E. (2003). ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. Nucleic Acids Res *31*, 3320-3323.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr D Biol Crystallogr *60*, 2256-2268.

Laskowski, R.A., Macarthur, M.W., Moss, D.S., and Thornton, J.M. (1993). Procheck - a Program to Check the Stereochemical Quality of Protein Structures. J Appl Crystallogr *26*, 283-291.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. Method Enzymol *276*, 307-326.

Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr *58*, 1772-1779.