Cell Reports, Volume *24*

Supplemental Information

Accommodation of Helical Imperfections in *Rhodobacter sphaeroides* Argonaute Ternary Complexes with Guide RNA and Target DNA

Yiwei Liu, Daria Esyunina, Ivan Olovnikov, Marianna Teplova, Andrey Kulbachinskiy, Alexei A. Aravin, and Dinshaw J. Patel

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Protein Expression and Purification

6xHis-tagged full-length *R. sphaeroides* Ago (RsAgo) gene was inserted into pET-30a expression vector (Invitrogen). The plasmid was transformed into *E. coli* strain BL21 (DE3). The cells were grown in Luria-Bertrani (LB) medium supplemented with 50 mg/mL kanamycin and harvested after 16 h incubation with 0.1 mM IPTG at 16 °C. The cells were lysed in 20 mM Tris-HCl buffer, 0.5 M NaCl, 0.1 mM DTT, pH 7.5. The protein was first purified using HisTrap HP column (GE Healthcare). The protein was then diluted to 0.1 M NaCl concentration and purified with tandem HiTrap Q HP and HiTrap SP HP columns. The protein was finally purified by gel filtration column Superdex 200 (GE Healthcare). It was concentrated to 50 mg/ml in 20 mM Tris-HCl pH 7.5, 0.2 M NaCl and 0.1 mM DTT. The mutant variants of RsAgo analyzed in biochemical experiments were obtained by site-directed mutagenesis in vector pET-30a and purified using HisTrap HP and Heparin Sepharose (GE Healthcare) columns.

Crystallization and Data Collection

RNA oligonucleotides were purchased from Thermo Scientific and DNA oligonucleotides were purchased from Invitrogen. All crystals were grown by the hanging-drop vapor diffusion method. Native or selenomethionine (SeMet)-labelled RsAgo was mixed with 5'-phosphorylated 18-base RNA and 24-base DNA at 1:1.1:1.1 molar ratios and protein concentration of 0.1-0.15 mM. Crystals of the ternary RsAgo - 18-mer gRNA - 24-mer tDNA complex were grown from 40% MPD, 30 mM magnesium acetate, 50 mM sodium cacodylate pH 5.7 at 20 ^{o}C. Protein-to-mother liquor ratio in the crystallization drop was 1:1. Crystals of the non-canonical pair and bulge containing sequences were grown under similar conditions.

Diffraction data were collected at 100 K at beam line NE-CAT 24-IDC at the Advanced Photon Source (APS), Argonne National Laboratory. All data were processed with the HKL2000 suite (Otwinowski and Minor, 1997). Data processing statistics are summarized in Table S1-S4.

Structure Determination and Refinement

The structure of the RsAgo - 18-mer gRNA - 24-mer tDNA complex was solved by the multi-wavelength anomalous dispersion method using the SeMet-labeled protein. Phases were calculated and the initial model comprising two copies of the complex in the crystallographic asymmetric unit was built with Phenix suite (Adams et al., 2010). The model was improved by several cycles of manual rebuilding with program COOT (Emsley and Cowtan, 2004; Winn et al., 2011) and was refined using the native dataset by phenix.refine (Table S1). The structures of all other complexes were solved by molecular replacement with the program Molrep, using the structure of the RsAgo - 18-mer gRNA - 24-mer tDNA complex as a search model and refined with Refmac5 of CCP4 suite or phenix.refine of Phenix suite. NCS restrains were not used in refinement (Tables S2-S4).

Analysis of RsAgo interactions with gRNA and tDNA

All RNA and DNA oligonucleotides used for analysis of RsAgo-nucleic acid interactions are shown in Figure S4. Determination of apparent K_d values was performed using nitrocellulose filtration method in binding buffer containing 40 mM Tris-HCl, pH 7.9, 40 mM KCl, 200 mM NaCl, 2 mM $MgCl_2$ and 50 μg/ml BSA. Nitrocellulose membrane filters (0.2 μm Amersham Protran, GE Healthcare) were presoaked in 0.4 M KOH for 10 minutes with gentle agitation, washed 3 times with 100 ml of distilled water and stored in the binding buffer at +4°C for up to 2 days. Nylon membrane filters (Hybond N+, GE Healthcare) were presoaked in the binding buffer for 10 minutes before the experiment. Guide RNA was phosphorylated with T4-PNK (New England Biolabs, NEB) and γ-[P^{32}]-ATP for radioactive labeling or with ATP for non-radioactive labeling and stored at -20°C. Target DNA was labeled using γ-[P^{32}]-ATP and T4-PNK, purified on Sephadex G-25 column (Roche) and stored at -20°C. For analysis of gRNA binding, 5'-P^{32}-labeled gRNA (0.1 nM) was incubated with increasing amounts of RsAgo concentrations for 20 minutes at 30 °C, followed by nitrocellulose filtration. For analysis of tDNA binding, the binary complex was first obtained by incubation of gRNA (5 μM) with RsAgo (10 μM) for 20 minutes at 30°C. Labeled tDNA (0.05 nM) was then mixed with increasing concentrations of the binary complex (0, 0.1, 0.3, 0.5, 1, 2, 5, 10, 30, 100 nM), and the samples were incubated for additional 20 minutes at 30°C. The filtration module was prepared by placing Bio-Rad Bio-Dot filter paper, nylon membrane and nitrocellulose membrane into the Bio-Dot Microfiltration Apparatus (Bio-Rad), and the wells were washed 3 times with 0.5 ml of the binding buffer (without BSA). The samples were applied at mild vacuum, and the wells were washed 3 times with the binding buffer. The membranes were dried at 50°C for 20 minutes and analyzed by phosphorimaging (Typhoon 9500, GE Healthcare). The data were analyzed using ImageQuant, Microsoft Excel and Graphit software and fitted to the hyperbolic equation:

$$B=B_{max}\times C/(C+K_d),$$

where B is the fraction of bound DNA at a given RsAgo concentration (calculated as a ratio of radiolabeled DNA bound to the nitrocellulose membrane to the sum of DNA bound to the nitrocellulose and nylon membranes minus background values), B_{max} is the maximal fraction of bound DNA at saturating RsAgo concentrations, C is concentration of RsAgo and K_d is the apparent dissociation constant.

For analysis of binary and ternary complexes by native gel electrophoresis the samples with either labeled gRNA or labeled tDNA were prepared as described above and loaded onto 10% polyacrylamide gel in 1^x TBE. The gel was run at constant voltage (200 V) for 50 minutes in ice-cold bath. For experiments with preformed gRNA-tDNA duplex (Figure 5A), gRNA (250 nM) and tDNA (1 nM) were first annealed to each other by heating the sample to 65°C and slowly cooling it down to 30°C, followed by the addition of RsAgo.

SUPPLEMENTAL REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

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

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235-242.

Table S1. Related to Figure 1. X-ray Statistics of a Ternary RsAgo Complex Containing Guide RNA Paired with Target DNA

	Ternary Complex 18-mer Guide RNA – 24-mer Target DNA		
	Se-Met		**Native**
	Peak	**Remote**	
Beam line	APS-ID24C	APS-ID24C	APS-ID24C
Wavelength	0.9792	1.0000	0.9792
Space group	$P2_1$	$P2_1$	$P2_1$
Unit cell			
a, b, c (Å)	68.1, 118.5, 118.3	68.0, 118.4, 118.7	68.2, 119.2, 117.7
α, β, γ (^{o})	90.0, 94.7, 90.0	90.0, 94.8, 90.0	90.0, 95.6, 90.0
Resolution (Å)	100-2.50 (2.59-2.50)[a]	100-2.40 (2.49-2.50)[a]	50-2.10 (2.18-2.10)[a]
R_{merge}	0.132 (0.676)	0.114 (0.589)	0.057 (0.365)
I/ σ (I)	16.3 (1.8)	18.3 (2.2)	27.1 (2.6)
Completeness (%)	99.3 (98.1)	99.4 (99.3)	93.5 (83.6)
Redundancy	6.9 (4.2)	7.0 (4.9)	3.4 (2.9)
Number of unique reflections	63868	71757	101956
R_{work}/R_{free} (%)			18.2/23.5
Number of non-H atoms			
RNA/DNA			1566
Protein			11603
Water			720
Ligands(molecule)			115
Average B factors (Å^2)			
RNA/DNA			35.25
Protein			36.59
Water			37.00
Ligands			68.69
R.m.s. deviations			
Bond lengths (Å)			0.008
Bond angles (^{o})			1.200

Table S2. Related to Figure 6. X-ray Statistics of RsAgo Ternary Complexes Containing A•A Non-canonical Pairs

Crystal	**Mismatch A3 · A3'**	**Mismatch A8 · A8'**
Beam line	APS-ID24C	APS-ID24C
Wavelength	0.9792	0.9792
Space group	$P2_1$	$P2_1$
Unit cell		
a, b, c (Å)	68.2, 118.9, 118.6	68.2, 120.0, 118.7
α, β, γ (°)	90.0, 95.7, 90.0	90.0, 95.5, 90.0
Resolution (Å)	70-1.81 (1.87-1.81)[a]	70-1.85 (1.92-1.85)[a]
R_{merge}	0.060 (0.743)	0.058 (0.602)
I/ σ (I)	15.8 (1.9)	12.8 (1.8)
$CC_{1/2}$	0.993 (0.748)	0.989 (0.658)
Completeness (%)	93.8 (91.3)	98.4 (98.4)
Redundancy	4.3 (4.1)	3.0 (2.8)
Number of unique reflections	160184	159065
R_{work}/R_{free} (%)	19.8/24.0	21.0/25.0
Number of non-H atoms		
RNA/DNA	1589	1632
Protein	11691	11478
Water	848	666
Ligands(molecule)	26	4
Average B factors (Å^2)		
RNA/DNA	41.42	44.15
Protein	41.68	48.33
Water	43.69	46.18
Ligands	60.67	36.27
R.m.s. deviations		
Bond lengths (Å)	0.009	0.011
Bond angles (°)	1.416	1.539

Table S3. Related to Figure 6. X-ray Statistics of RsAgo Ternary Complexes Containing G•A Non-canonical Pairs

Crystal	**Mismatch A8 · G8'**	**Mismatch G8 · A8'**
Beam line	APS-ID24C	APS-ID24C
Wavelength	0.9792	0.9792
Space group	$P2_1$	$P2_1$
Unit cell		
a, b, c (Å)	68.1, 119.1, 118.1	68.0, 118.3, 117.8
α, β, γ (°)	90.0, 95.6, 90.0	90.0, 95.5, 90.0
Resolution (Å)	70-2.00 (2.07-2.00)[a]	50-2.60 (2.69-2.60)[a]
R_{merge}	0.067 (0.591)	0.109 (0.777)
I/ σ (I)	14.9 (1.8)	11.1 (2.3)
$CC_{1/2}$	0.991 (0.663)	0.990 (0.722)
Completeness (%)	97.4 (90.4)	98.1 (97.8)
Redundancy	3.8 (2.6)	3.5 (3.4)
Number of unique reflections	123233	55738
R_{work}/R_{free} (%)	19.5/24.2	21.2/25.3
Number of non-H atoms		
RNA/DNA	1591	1570
Protein	11478	11603
Water	354	139
Ligands(molecule)	30	2
Average B factors (Å^2)		
RNA/DNA	54.28	49.58
Protein	56.32	54.10
Water	51.90	42.83
Ligands	62.46	37.21
R.m.s. deviations		
Bond lengths (Å)	0.013	0.005
Bond angles (°)	1.613	0.972

Table S4. Related to Figure 7. X-ray Statistics of RsAgo Ternary Complexes Containing A-A and T-T Bulges Within the Seed Segment of the Target Strand

Crystal	**Bulge A-A**	**Bulge T-T**
Beam line	APS-ID24C	APS-ID24C
Wavelength	0.9792	0.9792
Space group	$P2_1$	$P2_1$
Unit cell		
a, b, c (Å)	68.4, 119.0, 118.8	68.0, 118.3, 117.7
α, β, γ (°)	90.0, 95.5, 90.0	90.0, 95.5, 90.0
Resolution (Å)	50-2.25 (2.33-2.25)[a]	50-2.15 (2.23-2.15)[a]
R_{merge}	0.09 (0.65)	0.095 (0.637)
I/ σ (I)	11.8 (2.4)	12.7 (2.2)
$CC_{1/2}$	0.989 (0.754)	0.992 (0.732)
Completeness (%)	99.7 (100.0)	99.9 (100.0)
Redundancy	3.7 (3.6)	3.8 (3.8)
Number of unique reflections	88612	100379
R_{work}/R_{free} (%)	19.8/23.9	20.1/24.0
Number of non-H atoms		
RNA/DNA	1622	1580
Protein	11600	11618
Water	477	750
Ligands(molecule)	46	47
Average B factors (Å^2)		
RNA/DNA	51.20	45.2
Protein	51.9	44.7
Water	50.2	45.4
Ligands	60.9	61.8
R.m.s. deviations		
Bond lengths (Å)	0.012	0.008
Bond angles (°)	1.3	0.9
[a] Highest resolution shell (in Å) shown in parentheses.		

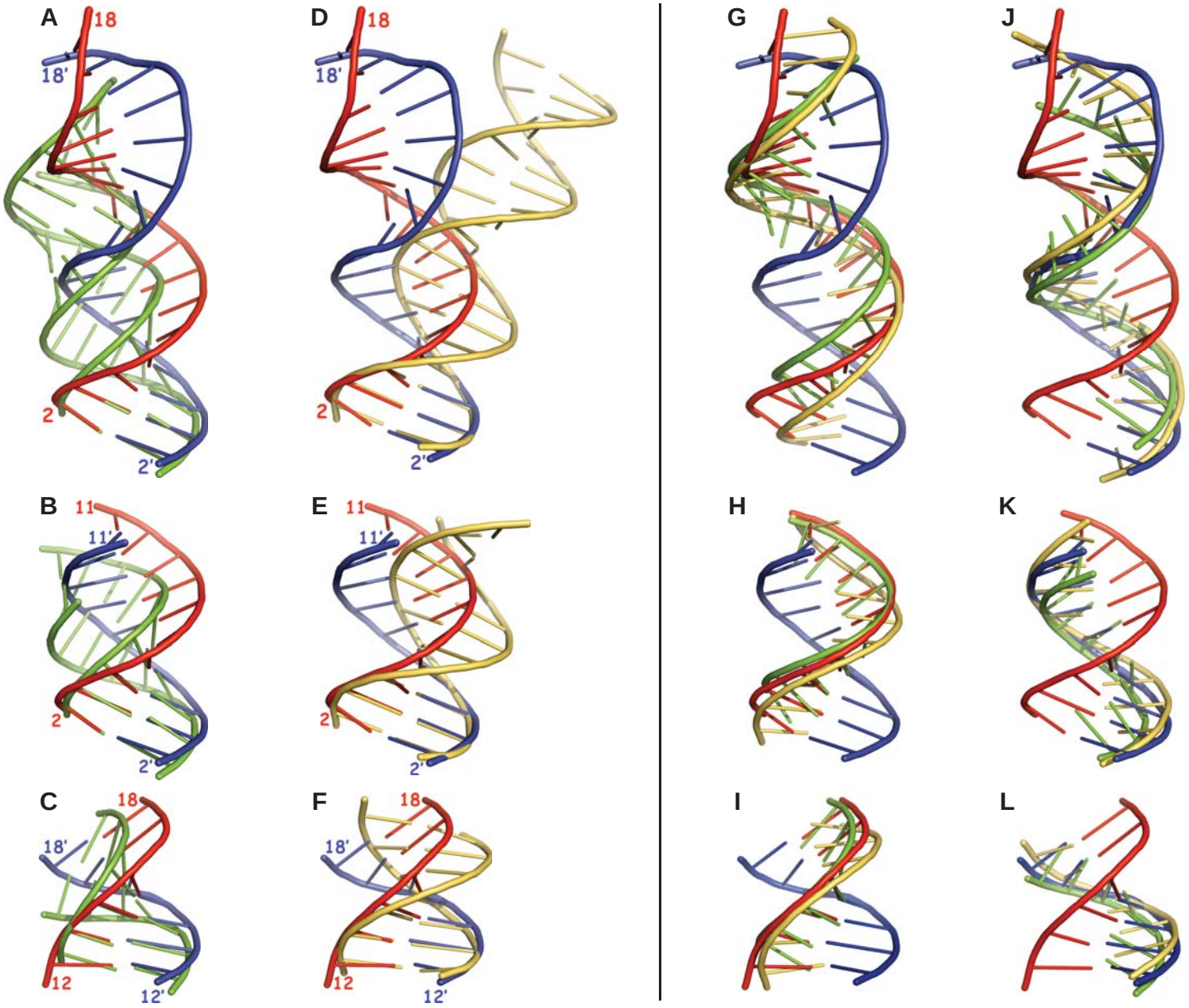

Figure S1. Related to Figure 1.

Panels A to F. Comparative Overlay of gRNA - tDNA Duplex in RsAgo Ternary Complex with Canonical A-form and B-form DNA.

(A-C) Comparative overlay of gRNA (in red) - tDNA (in blue) duplex in RsAgo ternary complex with canonical A-form DNA (in green) spanning 2-2' to 18-18' bps (panel A), 2-2' to 11-11' bps (panel B) and 12-12' to 18-18' (panel C).

(D-F) Comparative overlay of gRNA (in red) - tDNA (in blue) duplex in RsAgo ternary complex with canonical B-form DNA (in yellow) spanning 2-2' to 18-18' bps (panel D), 2-2' to 11-11' bps (panel E) and 12-12' to 18-18' (panel F).

Panels G to L. Comparative Overlay of gRNA and tDNA Strands in RsAgo Ternary Complex with Canonical A-form and B-form DNA.

(G-I) Comparative overlay of gRNA strand (in red) in RsAgo ternary complex with canonical A-form DNA strand (in green) or canonical B-form DNA (in yellow) spanning 2-2' to 18-18' bps (panel G), 2-2' to 11-11' bps (panel H) and 12-12' to 18-18'(panel I).

(J-L) Comparative overlay of tDNA strand (in blue) in RsAgo ternary complex with canonical A-form DNA strand (in green) or canonical B-form DNA strand (in yellow) spanning 2-2' to 18-18' bps (panel J), 2-2' to 11-11' bps (panel K) and 12-12' to 18-18' (panel L).

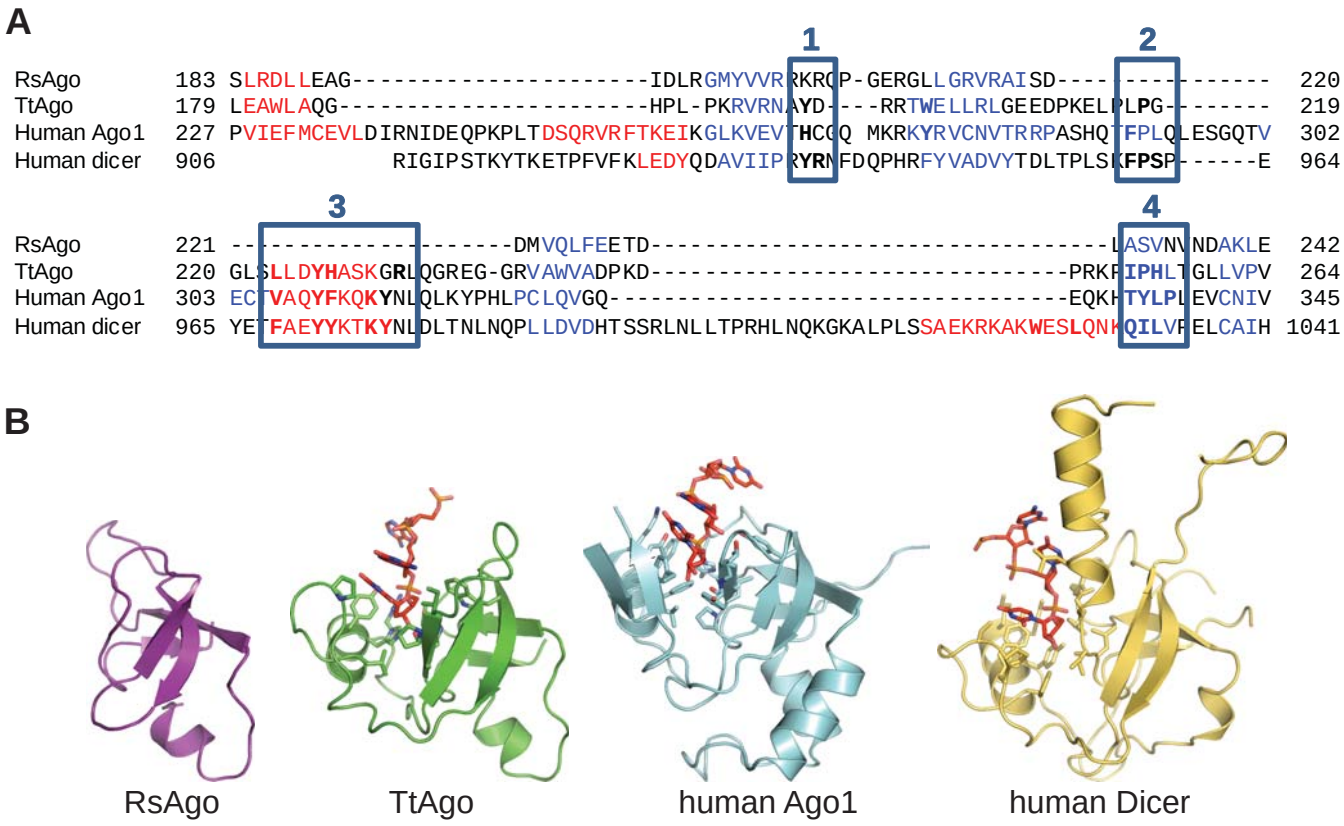

Figure S2. Related to Figure 1. Recognition Features of the PAZ Pocket in the Structure of RsAgo Bound to gRNA and tDNA.

(A) Sequence alignments of the PAZ domains of *R. sphaeroides* RsAgo, *T. thermophilus* TtAgo, human Ago1 and human Dicer.

(B) Structures of RsAgo, TtAgo (PDB: 3HO1), human Ago1 (PDB: 4KXT) and human Dicer (PDB: 4NGB) with bound nucleic acid, where observed, in orange stick representation.

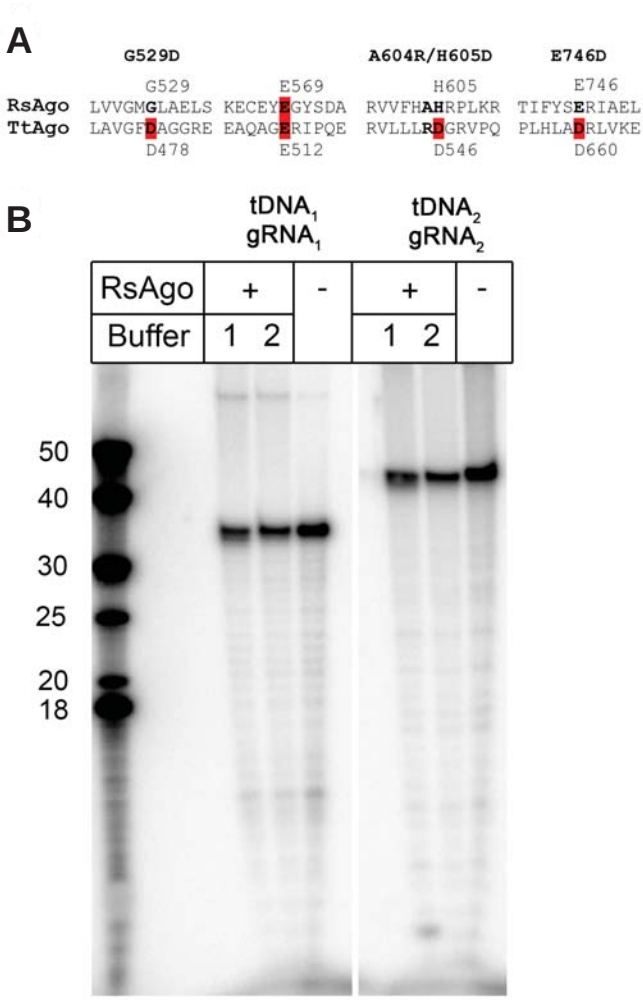

Figure S3. Related to Figure 2. Analysis of DNA cleavage by RsAgo with substitutions of the catalytic residues.
(A) Sequence alignments of active sites in the PIWI domain in RsAgo (top) and TtAgo (bottom). Shown are protein segments surrounding the catalytic tetrad residues. Amino acid substitutions in the mutant variant of RsAgo were obtained by site-directed mutagenesis and are shown above the alignment. In addition to the G529D, H605D and E746D substitutions in the catalytic tetrad, we also introduced an A604R substitution present in TtAgo.
(B) Analysis of DNA cleavage by the mutant RsAgo. The gRNA and tDNA sequences used in the cleavage experiments are shown in Figure S5. RsAgo (300 nM) was incubated with phosphorylated gRNA (100 nM) for 20 min at 30 °C. The 5'-P^{32}-labeled tDNA was added to 100 nM and the samples were incubated at 30 °C for 20 hours. The reactions were performed either in buffer 1 (40 mM Tris-HCl pH 8.0, 100 mM NaCl, 40 mM KCl, 10 mM $MgCl_2$, 2 mM $MnCl_2$, 1 mM DTT) or buffer 2 (40 mM potassium phosphate pH 7.4, 200 mM NaCl, 50 mM KCl, 10 mM $MgCl_2$ 10 mM, 2 mM $MnCl_2$, 1 mM DTT). Nucleic acids were separated by 15% urea PAGE (19:1) and analyzed by phosphorimaging. The first lane contains DNA length markers.

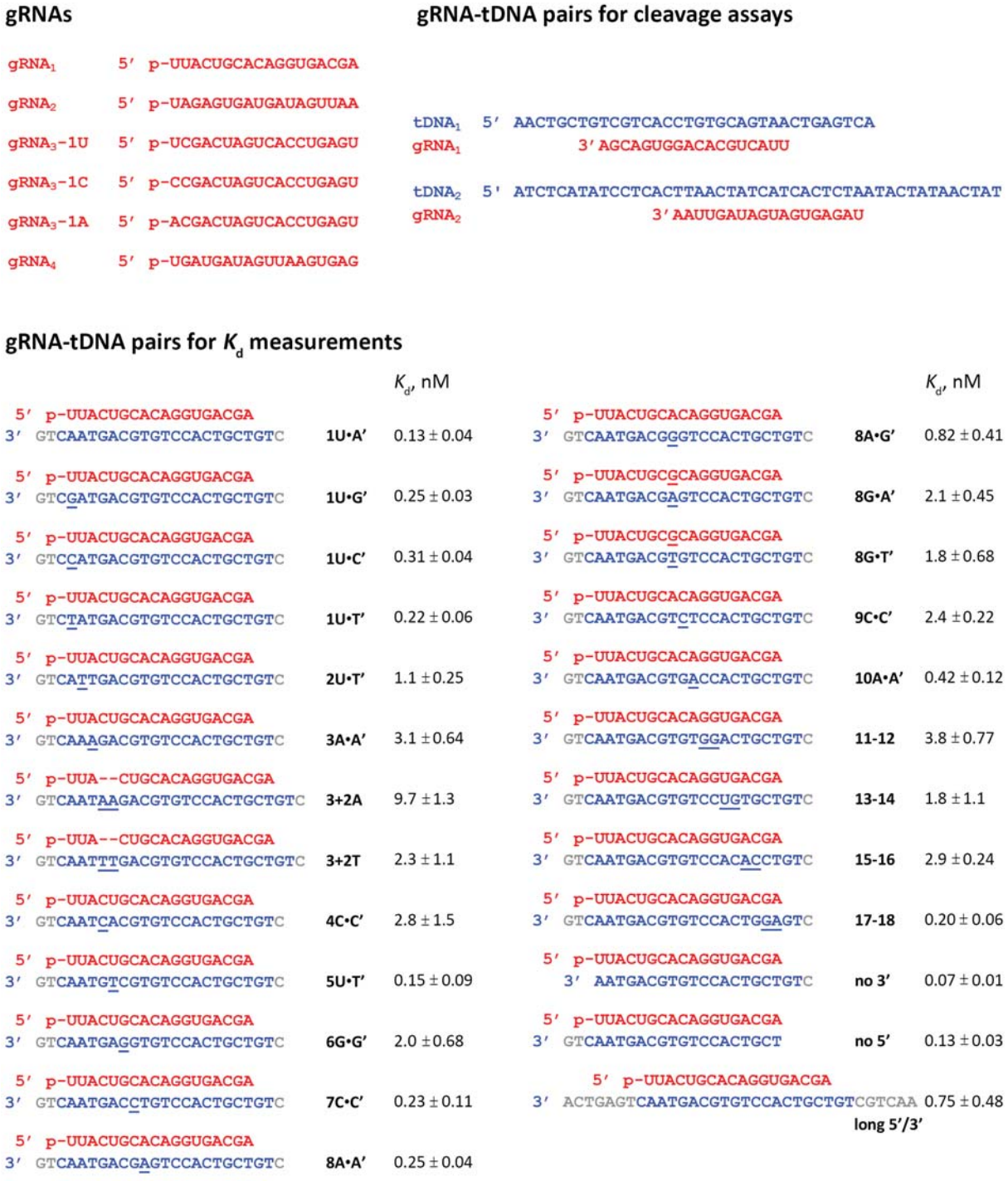

Figure S4. Related to Figures 3 and 4. Structures of gRNAs and gRNA-tDNA pairs used in the binding and cleavage assays. The gRNA-tDNA variants with mismatches and bulges used for K_d measurements are shown at the bottom part of the figure. Nucleotides substituted in each gRNAt-DNA pair are underlined. Apparent K_d values for tDNA binding by the preformed RsAgo-gRNA complex are shown on the right (correspond to Figure 4).

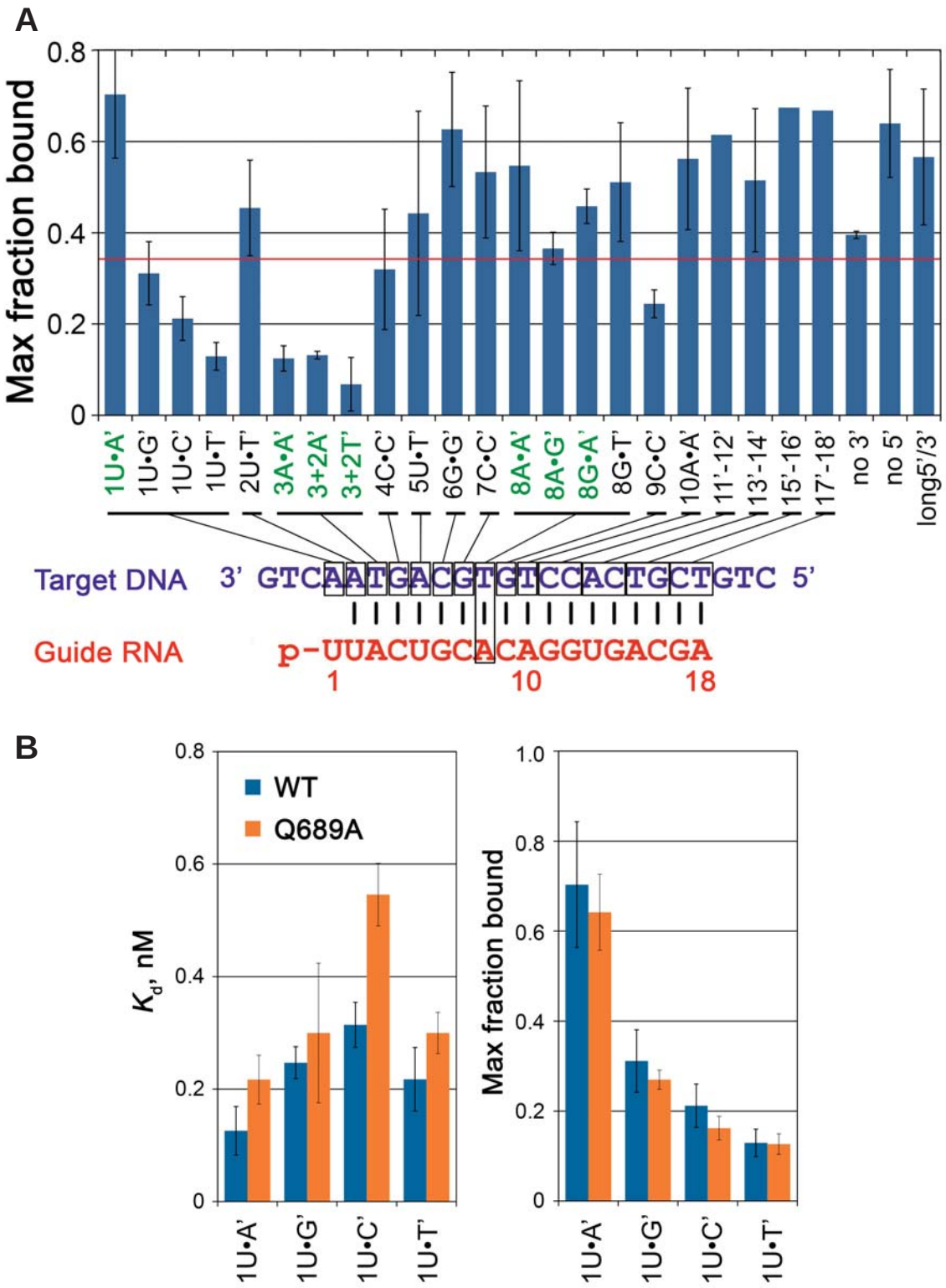

Figure S5. Related to Figures 3 and 4.

Panel A. The maximal fractions of bound tDNA observed at saturating gRNA-RsAgo concentrations.

The plot shows the maximal binding (B_{max}) of tDNA variants with mismatches and bulges as a fraction of total labeled tDNA in the sample. The structure of the gRNA-tDNA heteroduplex is shown below (see Figure 4 and Figure S4). The red line indicates the two-fold lower level of the binding relativeto fully-complementary tDNA (1U-A'). The data are means and standard deviations from 3-5 independent measurements.

Panel B. No effects of the Q689A substitution in RsAgo on the apparent K_d and maximal binding values for tDNA variants with different nucleotides at the 1' position.

Shown are means and standard deviations from 3-4 independent experiments.

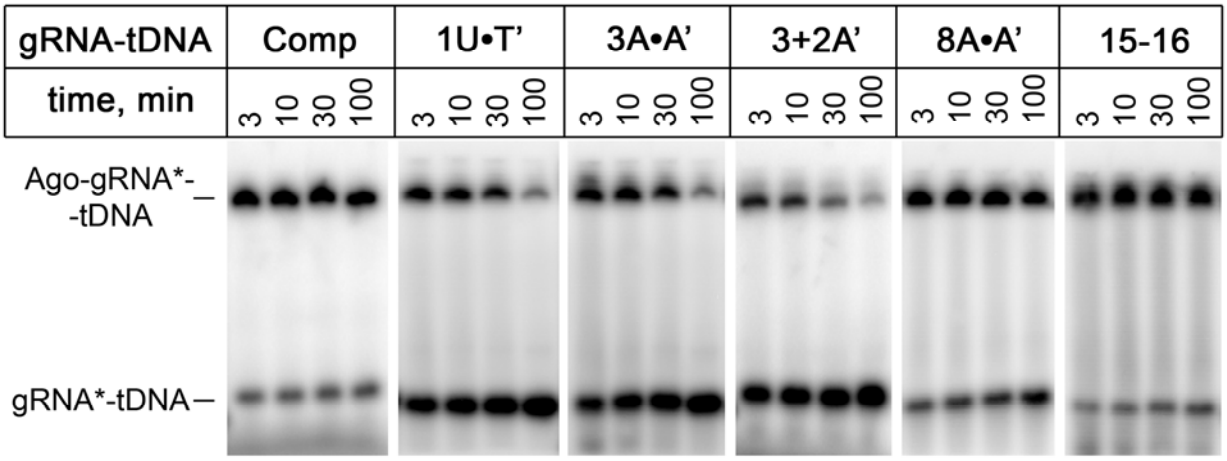

Figure S6. Related to Figure 5. Dissociation kinetics of tDNA variants with mismatches and bulges. tDNA (100 nM) was added to preformed binary complexes of RsAgo with gRNA1 (RsAgo 40 nM, labeled gRNA 10 nM) together with excess competitor gRNA4 (1000 nM) in standard binding buffer containing 40 mM Tris-HCl, pH 7.9, 40 mM KCl, 200 mM NaCl, 2 mM $MgCl_2$. The samples were incubated for indicated time intervals at 30 °C and analyzed by 10% native polyacrylamide gel electrophoresis. Positions of the ternary complex and free gRNA-tDNA duplex are shown; the labeled component (gRNA) is indicated with asterisks.

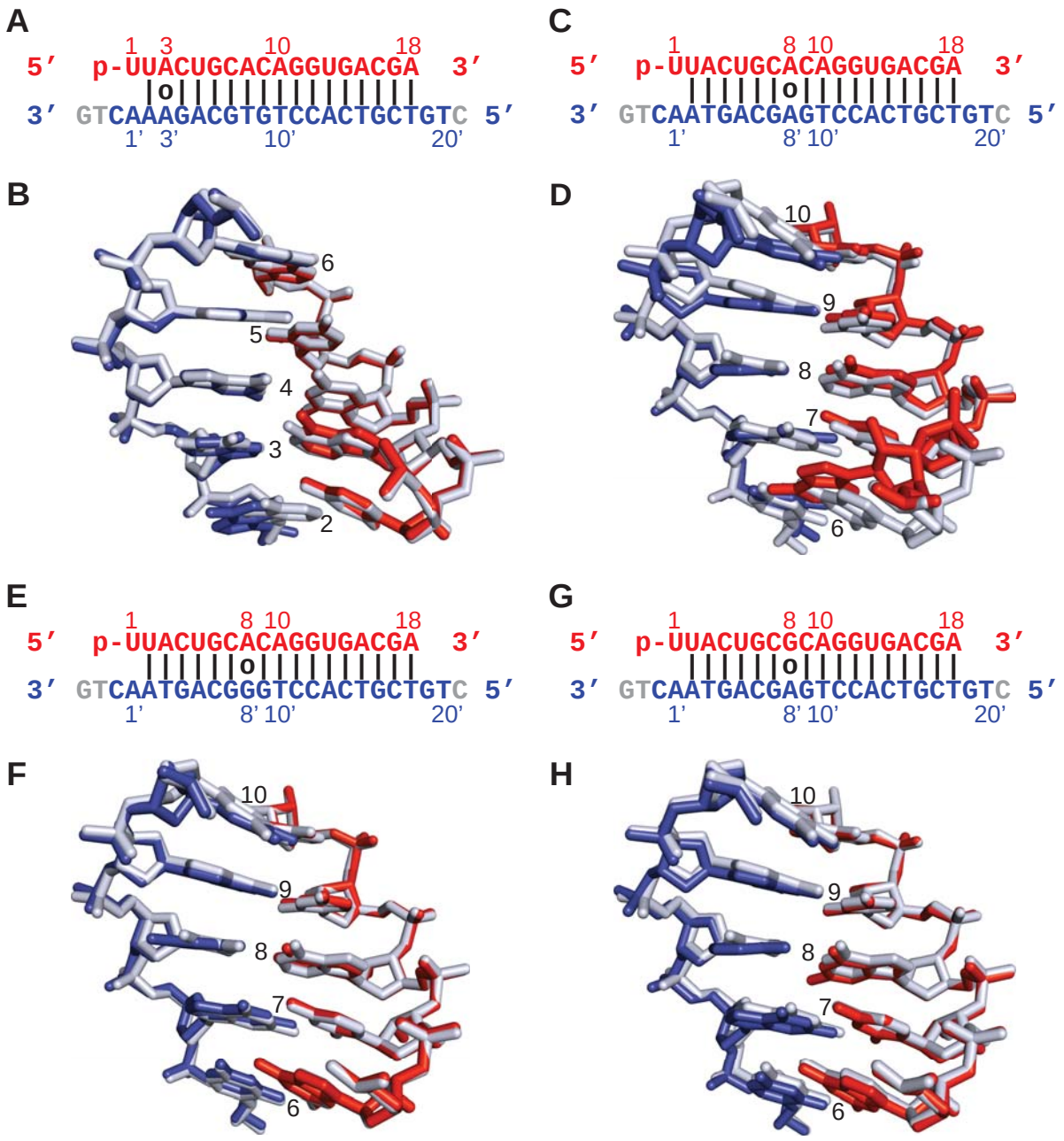

Figure S7. Related to Figure 6.

Panels A to D. Pairing Alignments of A•A Non-canonical Pairs Within the Seed Segment in the Complex of RsAgo with Guide RNA and Target DNA.

(A) Sequence alignment of the guide RNA - target DNA showing the position of the A3•A3’ non-canonical pair.

(B) Superposition of the RNA-DNA segment for the wt duplex (in silver) with the A3•A3’ non-canonical pair-containing duplex.

(C) Sequence alignment of the guide RNA - target DNA showing the position of the A8•A8’ non-canonical pair.

(D) Superposition of the RNA-DNA segment for the wt duplex (in silver) with the A8•A8’ non-canonical pair-containing duplex.

Panels E to H. Pairing Alignments of A•G and G•A Non-canonical Pairs Within the Seed Segment in the Complex of RsAgo with Guide RNA and Target DNA.

(E) Sequence alignment of the guide RNA - target DNA showing the position of the A8•G8’ non-canonical pair.

(F) Superposition of the RNA-DNA segment for the wt duplex (in silver) with the A8•G8’ non-canonical pair-containing duplex.

(G) Sequence alignment of the RNA guide - target DNA showing the position of the G8•A8’ non-canonical pair.

(H) Superposition of the RNA-DNA segment for the wt duplex (in silver) with the G8•A8’ non-canonical pair-containing duplex.