Supporting Information

Eiler et al. 10.1073/pnas.1414571111

SI Materials and Methods

The substrate strands were chemically synthesized with a 2′-deoxy substitution at the scissile phosphate. To make the RNA enzymes strands for the twister constructs, we first generated double-stranded DNA (dsDNA) templates by PCR according to the manufacturer's recommendations. The twister A template strand used was 5′-taatacgactcactataGGCAATAAAGCGGTTACAAGCCCGCAAAAATAGCAGAGTAATGTCGCG-3′ with forward primer 5′-taatacgactcactataGGCAATAAAGCGGTTACAAGCC-3′ and reverse primer 5′-CGCGACATTACTCTGCTATTTTTGCGGGC-3′ (the T7 RNA polymerase promoter sequence is in lowercase, and the coding sequence is capitalized). The twister B template strand used was 5′-taatacgactcactatagTAAATGATATAGCCGGTCCCAAGCCCGGAAAAAGGAGGAGGGTATA-3′ with forward primer 5′-taatacgactcactatagTAAATGATATAGCCGGTCCCAAGCC-3′ and reverse primer 5′-TATACCCTCCTCCTTTTTCCGGGCTTGG-3′ (the T7 RNA polymerase promoter sequence is in lowercase, and the coding sequence is capitalized). The dsDNA PCR product was separated from other reaction components by ethanol precipitation. After redissolving the dsDNA, we performed in vitro transcription reactions as described (1). RNA transcripts were purified over a G25 column to desalt and remove nucleotides and then separated over a Bio-Rad prep cell. Fractions having the correct RNA transcription product were analyzed using denaturing polyacrylamide gel electrophoresis (8 M urea, 1× TBE), pooled, and precipitated by ethanol. Pelleted RNA samples were washed by 70% ethanol, dissolved in water, and stored at −80 °C until use in either buffer A [30 mM Hepes (pH 7.5), 20 mM $MgCl_2$, and 100 mM KCl] or buffer B [10 mM Tris·HCl (pH 7.5), 200 mM NaCl, and 1 mM EDTA NaOH (pH 8.0)]. The substrate strand for twister A used was 5′-CGCGGCAU$\underline{U}$AAUGCAGCUUUAUUGCC-3′, and the substrate strand for twister B used was 5′-UAUG$\underline{U}$AACUCCGCCUAUGUCAUCUUA-3′ (nucleotide with the $\underline{2}$′-deoxy substitution is underlined).

Structure Determination

Initial structure determination of twister A derived from the environmental sequence in the $P6_122$ space group was carried out independently from two heavy-atom (HA) derivatives: (*i*) Cs derivative under a pseudo-MAD option, and (*ii*) Sm under the SIRAS option using the program suite ShelX (2). The pseudo–two-wavelength MAD procedure included two Cs datasets, both of which were collected at the Cs absorption edge wavelength but were soaked for 3 h with different concentrations of Cs_2SO_4 (20 and 100 mM) in the cryogenic freezing solutions. The derivative dataset with a lower concentration of Cs served as a pseudoinflection point dataset. There were three Cs binding sites corresponding to one site per molecule in the asymmetric unit but with different occupancies between the two Cs datasets (Fig. S1*A*). The Sm derivative was prepared by soaking a crystal with 96 mM samarium acetate in the cryogenic freezing solutions for 3 h and had six binding sites with two sites per molecule (Fig. S1*B*). After the two HA solutions were placed into a common crystallographic origin, phases were combined using the CCP4 program Sigma-A (Fig. S1*C*) (3). These phases were treated as external phases for rerefinement of the HA substructures from 21 Sm datasets and 8 Cs datasets (Table S3). The highest effective anomalous resolutions were 6.08 and 6.36 Å for the Sm and Cs derivatives, respectively (Table S3). When external phases from the anomalous signals were calculated and combined from the rerefined HA substructures (4), the solvent boundary for each of the three molecules was well defined. Density modification with threefold, noncrystallographic symmetry (NCS) averaging was carried out using the CCP4 program DMmulti, which greatly improved the experimental maps (Fig. S1*D*) (3). When phases from the two datasets of the most isomorphous SIR pairs for each of the Cs and Sm derivatives were added to all anomalous phases, the quality of experimental map was further improved, leading to the complete interpretation of the twister A structure.

The structure of twister B from *O. sativa* in $P2_1$ space group was determined by a combination of molecular replacement (MR) and SIR methods. Using a partially interpreted twister A as a search model, we found four copies of twister in the asymmetric unit, although these models could not be refined and model-calculated phases failed to solve any HA substructures from data collected from crystals soaked with iridium hexamine chloride (Ir) or cobalt hexamine chloride (Co). These solutions revealed a recurring dimeric interface involving the five adenosines where the optional P5 stem would be (Fig. S4 *A* and *C*). When such a dimer derived from the twister A structure was used as a MR search model, the peak of the cross-rotation function increased to 8.31 σ, although the peak of the translation function was only 3.36 σ for the first of the two dimers in the asymmetric unit. This solution was able to be refined with a cross-validated R_{free} of 51.33% and 48.59% after rigid-body and restrained refinement at 3.1-Å resolution, respectively. The calculated phases from this model were able to solve the Ir derivative using a Ir minus Co isomorphous difference Fourier method (Ir − Co). There were a total of 28 Ir binding sites with 7 per molecule. Using the calculated phases as external phases, we refined the Ir HA substructure against the Ir − Co isomorphous signals and external phases, followed by fourfold NCS averaging using DMmulti (3). These maps were of sufficient quality so that we could easily distinguish purine from pyrimidine bases in many locations (Fig. S6). Interestingly, anomalous signals in this low-symmetry space group $P2_1$ appeared to be severely corrupted by unknown sources, even though the apparent anomalous resolution in this $P2_1$ crystal form was much higher than that in the $P6_122$ space group. An estimation for the anomalous resolution of the Ir derivative was 5.4 Å using the anomalous ratio criteria. We failed to find initial Ir positions using anomalous-difference Fourier methods with phases calculated from initial MR solution, although we did find them using phases calculated from the final refined coordinates. When phases calculated from anomalous signals were added to those calculated using SIR pairs, the quality of the resulting experimental maps was degraded substantially and the maps were uninterpretable.

1. Pata JD, King BR, Steitz TA (2002) Assembly, purification and crystallization of an active HIV-1 reverse transcriptase initiation complex. *Nucleic Acids Res* 30(22):4855–4863.
2. Sheldrick GM (2008) A short history of SHELX. *Acta Crystallogr A* 64(Pt 1):112–122.
3. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.
4. Wang J, Li Y, Modis Y (2014) Exploiting subtle structural differences in heavy-atom derivatives for experimental phasing. *Acta Crystallogr D Biol Crystallogr* 70(Pt 7): 1873–1883.

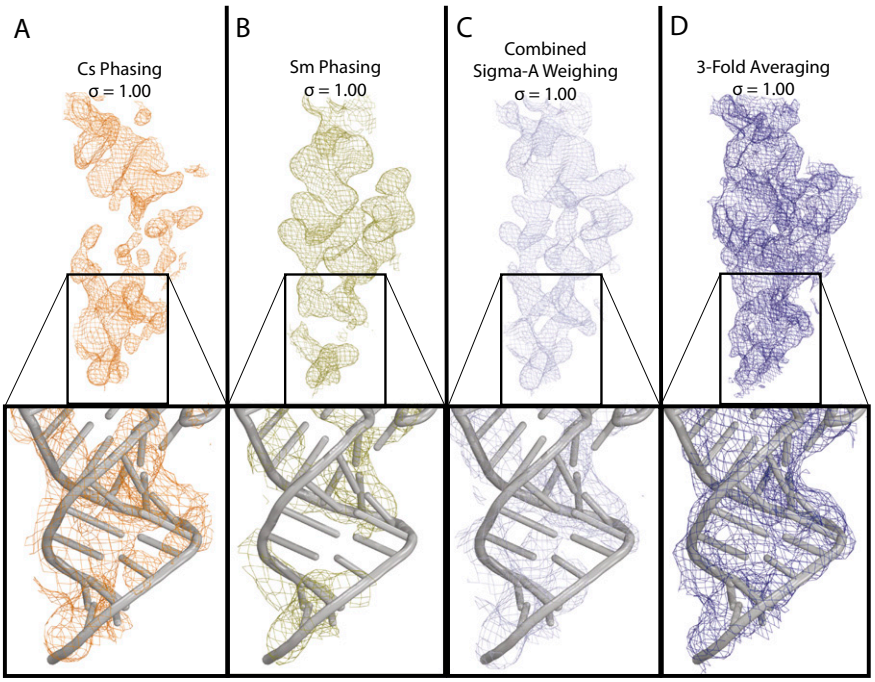

Fig. S1. Experimental electron density maps for twister structure derived from the environmental sequence. (*A*) The electron density map colored orange was generated from two-wavelength MAD phasing based on three Cs sites. (*B*) The electron density map colored green was calculated from SIRAS phasing of six Sm sites. (*C*) A combined Sigma-A weighted map colored light blue was obtained from phases derived from the Cs and Sm sites. (*D*) The electron density map colored dark blue was generated by DMmulti using threefold NCS averaging. The *Insets* show the improvement in the electron density around the P1 stem for each map.

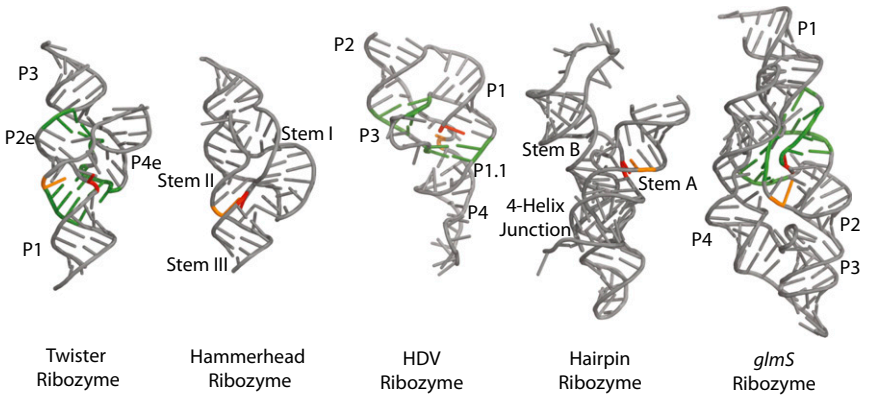

Fig. S2. Comparison of the crystal structures among small self-cleaving ribozymes. Crystal structures of members of the small self-cleaving ribozyme family including twister (PDB ID code 4QJH), hammerhead (PDB ID code 3ZD5), HDV (PDB ID code 1DRZ), hairpin (PDB ID code 1M5K), and *glmS* (PDB ID code 2GCS) ribozymes. Major stems and helices are labeled. The nucleotide 5′ to the cleavage site is colored red, and the nucleotide that deprotonates its 2′-hydroxyl is colored orange. Pseudoknots in twister, HDV, and *glmS* are green.

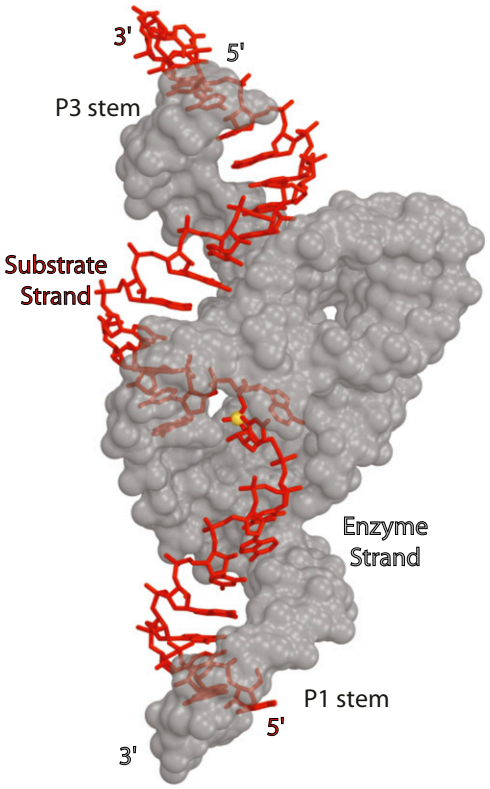

Fig. S3. The substrate strand folds into an A-form helix. The substrate strand (red sticks) adopts an A-form helix throughout the structure except for two active-site nucleotides. The enzyme strand (gray spheres) makes several complex interactions to form a rigid active site. The scissile phosphate is shown as a yellow sphere.

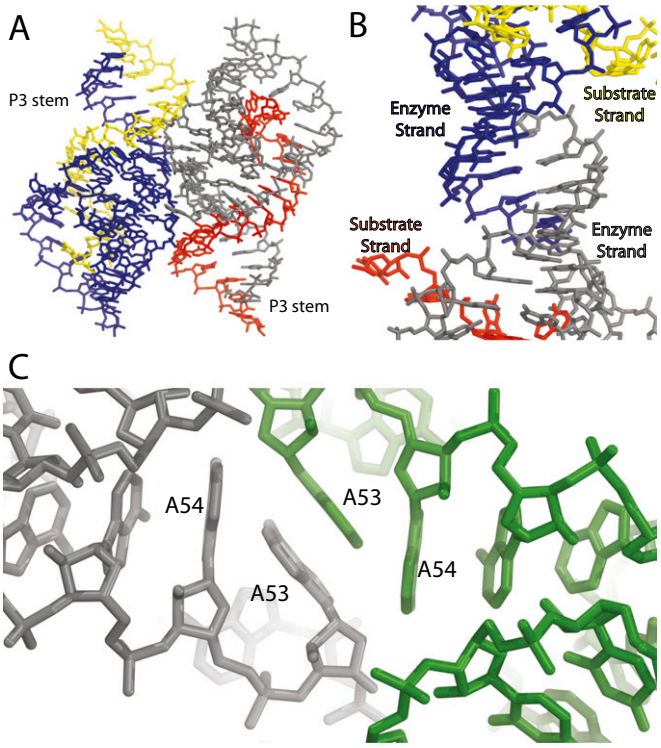

Fig. S4. Dimer of the *O. sativa* twister used to for molecular replacement, intermolecular dimer crystal artifact, and dimer interface of the environmental sequence twister. (*A*) Dimer of *O. sativa* twister used for molecular replacement where the P1 stem was deleted. Substrate strands are shown in yellow and red, and the enzyme strands are shown in blue and gray, respectively. (*B*) The intermolecular dimer involving the enzyme strands in twister B are yellow and gray. Displaced substrate strands are yellow and red. Note that the twister B molecules in *A* are the not the same as in *B* even though the coloring is the same. (*C*) The diametric interface formed by the five adenosines that replaced the optional P5 stem in the environmental sequence twister where the crystallographic copy is shown in green and its symmetry mate is colored gray.

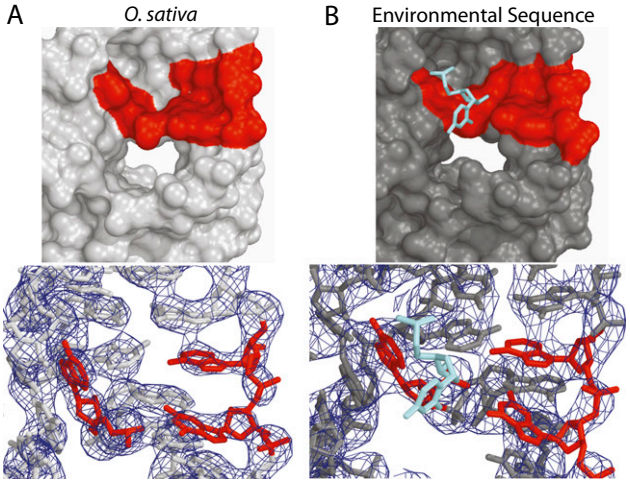

Fig. S5. Comparison of the active-site regions between twister A and B. Surface representation and electron density of active sites from (*A*) *O. sativa* and (*B*) the environmental sequence twisters. Nucleotides not proposed to be involved in catalysis or active-site formation are colored light gray in *O. sativa* and dark gray in the environmental sequence with nucleotides proposed to be involved in catalysis or active site formation are colored red in both. The ordered nucleotide 5′ to the cleavage site and scissile phosphate are colored cyan and represented as sticks in *B*. Electron density maps are contoured at 1.3 and 2.1 σ for *O. sativa* and the environmental sequence, respectively. The two crystal structures show a rigid active site that is preformed regardless whether the nucleotide 5′ to the cleavage site and scissile phosphate is well ordered or not. The rmsd of the active site region is 1.56 Å (chains C and D for twister A and chains E and F of twister B).

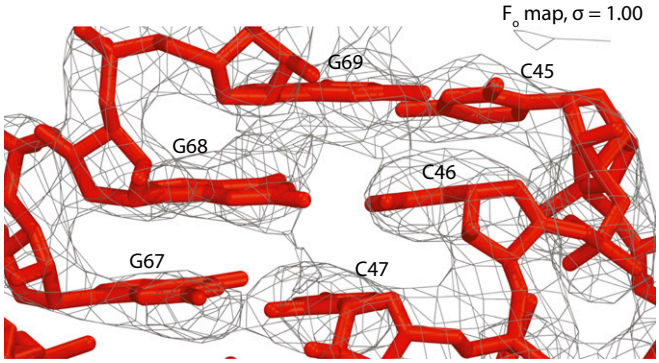

Fig. S6. Electron density maps of the *O. sativa* twister calculated from phases derived from the Ir-Co SIR method followed by fourfold NCS averaging. The calculated phases from the molecular placement model were used to solve the HA substructure of the iridium hexamine (Ir) derivative using an Ir minus a cobalt hexamine (Co) isomorphous difference Fourier method. Starting with the calculated phases as external phases, the Ir HA substructure was refined against the Co dataset as the native and calculated experimental phases. Subsequently, fourfold NCS averaging was performed using DMmulti to generate the electron density map that distinguished purines from pyrimidines. Nucleotides in the PK2 region are shown.

Table S1. Data collection, phasing, and refinement statistics for twister from an environmental sequence (MAD/SIRAS)

	Native	100 mM Cs soak	20 mM Cs soak	Native	96 mM Sm soak
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Cell dimensions					
a, *b*, *c*, Å	228.77, 228.77, 101.06	227.25, 227.25, 101.37	226.93, 226.93, 101.32	227.21, 227.21, 101.13	226.84, 226.84, 101.64
α, β, γ, °	90.0, 120.0, 90.0	90.0, 120.0, 90.0	90.0, 120.0, 90.0	90.0, 120.0, 90.0	90.0, 120.0, 90.0
		Peak	Inflection		Peak
Wavelength	1.1051	1.8456	1.8456	1.8445	1.8445
Resolution, Å	4.04–50.0 (4.04–4.14)	4.72–50.0 (4.72–4.84)	4.47–50.0 (4.47–4.59)	4.58–50.0 (4.58–4.70)	4.71–50.0 (4.71–4.83)
R_{sym} or R_{merge}	8.0 (515.6)	9.3 (506.5)	9.1 (380.4)	7.1 (406.0)	9.2 (580.7)
$I/\sigma I$	23.49 (1.0)	15.54 (1.0)	15.78 (1.0)	19.0 (1.0)	15.67 (1.0)
Completeness, %	99.9 (100.0)	99.9 (100.0)	99.9 (99.9)	99.9 (100.0)	99.9 (99.9)
Redundancy	38.2 (44.4)	20.5 (19.2)	20.4 (21.1)	20.6 (21.5)	20.5 (19.2)
Refinement					
Resolution, Å	3.88–50.0				
No. reflections	14,918				
R_{work}/R_{free}	19.7/24.4				
No. atoms	4,774				
RNA	4,764				
Ligand/ion	4				
Water	6				
B factors					
RNA	165.2				
Ligand/ion	81.3				
Water	102.8				
Rms deviations					
Bond lengths, Å	0.005				
Bond angles, °	1.393				

Table S2. Data collection and refinement statistics for *Orzyza sativa* twister (molecular replacement by 4QJH)

	$Co(NH_3)_6Cl_3$ soak
Data collection	
Space group	$P2_1$
Cell dimensions	
a, *b*, *c*, Å	70.91, 53.07, 106.75
α, β, γ, °	90.0, 106.86, 90.0
Resolution, Å	3.18–50.0 (3.18–3.27)
R_{sym} or R_{merge}	11.1 (108.8)
$I/\sigma I$	7.01 (1.0)
Completeness, %	97.9 (97.7)
Redundancy	3.7 (3.8)
Refinement	
Resolution, Å	50–3.1
No. reflections	11,370
R_{work}/R_{free}	21.3/28.8
No. atoms	6,201
RNA	6,183
Ligand/ion	4
Water	14
B factors	
RNA	155.7
Ligand/ion	74.0
Water	48.8
Rms deviations	
Bond lengths, Å	0.007
Bond angles, °	1.43

Table S3. Statistics of Sm and Cs derivative data for phasing twister A

Datasets	Overall resolution, Å	R_{meas}, %	Anomalous resolution, Å	
			$CC_{1/2}$	Anomalous signal
Sm31_7_peak	4.71	9.5	6.35	6.08
Sm31_7_infl	5.20	7.6	7.80	7.70
Sm31_7_hrem	5.90	6.8	8.80	8.34
Sm23_13_peak	4.83	7.5	6.50	6.50
Sm23_13_infl	4.78	11.1	7.60	7.13
Sm23_13_hrem	4.72	9.9	7.46	6.90
Sm30_16_peak	5.02	11.2	8.10	7.50
Sm30_16_infl	5.02	10.6	8.50	8.00
Sm30_16_hrem	5.31	8.3	9.30	8.40
Sm30_8_peak	6.62	9.7	8.55	8.55
Sm30_8_infl	6.97	12.4	10.39	10.39
Sm30_8_hrem	7.98	12.9	11.97	11.97
Sm31_4	4.74	10.8	7.80	6.90
Sm23_2A	5.24	9.4	8.00	7.60
Sm23_2B	5.36	11.2	8.48	8.20
Sm30_4	5.38	12.7	6.40	6.40
Sm23_16	5.58	13.9	6.50	6.50
Sm30_11	6.20	12.6	8.00	8.00
Sm23_8	6.40	16.6	9.00	9.00
Sm23_15	6.94	19.9	8.00	7.90
Sm30_14	7.00	8.2	9.00	9.00
Cs31_6	4.72	9.6	6.36	6.36
Cs31_3	4.47	9.3	6.66	6.40
Cs23_12	4.50	7.3	7.40	7.50
Cs23_10	4.84	10.3	6.00	7.22
Cs30_3	4.85	11.2	9.70	7.67
Cs30_7	4.85	9.1	8.20	7.23
Cs30_15	4.85	9.1	8.20	7.67
Cs23_6	4.96	9.7	9.50	9.06

Corresponding wavelengths for Sm peak, inflection (infl), and high-energy remote (hrem) are A, B, and C Å, respectively. All Sm single-wavelength datasets were collected at Sm peak wavelength. All Cs single-wavelength datasets were collected at Cs peak wavelength of X Å. In the highest resolution shells, the overall $CC_{1/2}$ values varied from 35% to 54% for all datasets. Estimated effective anomalous resolutions using anomalous $CC_{1/2}$ (>50%) and anomalous signals (>1.15) were included.