Supporting information for **"A role for a single-stranded junction in RNA binding and specificity by the *Tetrahymena* group I ribozyme"**

Xuesong Shi, Sergey V Solomatin, and Daniel Herschlag

Ribozyme preparation

L-16 *Sca*I ribozyme was prepared by *in vitro* transcription followed by PAGE purification, as previously reported[1]. Mutant ribozymes were transcribed from DNA templates generated with DNA primers containing the corresponding mutations. The L-16 *Sca*I ribozyme for fluorescence polarization anisotropy (FPA) measurements has a 5'-internal guide sequence (IGS) of 5'-G_{17}GACAG_{22}GAGGG-3' and the L-16 *Sca*I ribozyme for single molecule FRET and activity measurements have 5'-IGS sequence of 5'-G_{17}GUUUG_{22}GAGGG-3' and a 3'-tether for surface immobilization (see below).[1-3] The 5'-ACA sequence in the IGS of the ribozymes for FPA measurements is designed to minimize sequence dependent quenching of 6-MI on the complementary substrate strand.[3,4]

Fluorescence polarization anisotropy (FPA) measurement

An open complex fluorescence substrate 5'-r(CCC$_m$UCC U**F**U CC)-3', where m is 2'-methoxy substitution and **F** is 6-methyl isoxanthopterin (6-MI), was obtained from Fidelity Systems (Gaithersburg, MD) and used for all FPA measurements. The FPA ribozyme-substrate complex was prepared by refolding the ribozyme at 50 °C for 30 min with 10 mM $MgCl_2$ and subsequently incubating with the HPLC-purified substrate for 20 min at room temperature. The sample was then filter exchanged three

times with experimental buffer of 50 mM Na•MOPS, pH 7.0, and 10 mM Mg^{2+} using a 50 kDa centrifugal filter (Millipore, MA) to remove unbound fluorescent substrate. All anisotropy measurements were carried out at 25 °C using a Fluorolog-3 spectrometer following a previously reported procedure[3]. The reported errors are standard deviations from 3-4 repeated measurements on two different days with independently prepared samples.

Single molecule FRET

Docking constants, K_{dock}, of individual molecules were measured by single molecule FRET experiments as follows. L-16T2 version of the *Tetrahymena* ribozyme (L-16 *Sca*I version extended on the 3'-end by 26 nucleotides: *ACCAAAAUCAACCUAAAACUUACACA, the T2 extension*)[2] was pre-folded for 30 min at 50 °C, bound with the substrate (5'-*CCCUCdUAAACC-Cy3*), and attached to the surface of quartz slides for imaging in a total internal reflection microscope, as described.[5] All experiments were performed in 50 mM Na•MES, pH 7.0, with 10 mM $MgCl_2$, 25 mM NaCl, and an oxygen scavenging system (44 mM glucose, 1 mM Trolox, and trace amounts of glucose oxidase and catalase). FRET data were collected at a 5 ms frame rate and a signal-to-noise ratio of 2. The average photobleaching lifetime of the dyes is 12 s. FRET traces of individual molecules of the ribozyme/substrate complex displayed fluctuations between two states (FRET = 0.95, corresponding to the docked state, and FRET = 0.4, corresponding to the undocked state).[2] Rate and equilibrium constants were obtained by analyzing FRET traces using a Hidden Markov-based algorithm, as described.[5] Error analysis was performed by bootstrapping the data from individual molecules.

Ribozyme activity

Ribozyme activity was measured with 2 mM guanosine (G), using 5′-^{32}P radiolabeled open complex substrate, *-1r,dSA$_5$, 5′-CCCUCrUAAAAA-3′,* under single turnover condition with L-16 *Sca*I ribozyme saturating with respect to oligonculeotide substrate (>0.5 µM ribozyme and ~0.5 nM substrate, see below). A guanosine concentration of 2 mM is saturating for the WT ribozyme.[6] The reaction conditions were 50 mM Na•MOPS, pH 7.0 and 10 mM $MgCl_2$ at 25 °C. The ribozymes were prefolded at 50 °C for 30 min and equilibrated at 25 °C with G for 5 min before addition of the 5′-^{32}P radiolabeled substrate to initiate the reaction. Time points were taken by transferring 2 µl of the reaction mixture into 4 µl of stop/gel loading solution including 50 mM EDTA and 90% formamide, as described previously.[7]

The reaction followed under these conditions is the pseudo-first order reaction: $(E\bullet{*}S\bullet G)_o \rightarrow (E\bullet{*}P)$, with the full reaction being:

$$(E\bullet{*}S\bullet G)_o \overset{K_{dock,G}}{\rightleftarrows} (E\bullet{*}S\bullet G)_c \overset{k_c}{\rightarrow} (E\bullet{*}P) \qquad k_{obs} = K_{dock,G} \times k_c \quad (K_{dock,G} << 1)$$

First-order reaction rate constants, k_{obs}, were obtained by non-linear least square fitting of data for disappearance of *S. For slow reactions that did not go to completion, values of k_{obs} were fit to the initial rate and end points of 95% were used. Errors were determined as the standard deviation from 2-4 repeated measurements.

The disassociation constant for -1r,dSA$_5$ binding to L-21 *Sca*I IGS was reported to be 85 nM at 30 °C.[8] Under similar salt conditions, the disassociation constant of -1r,dSA$_5$ binding to L-16 *Sca*I IGS (with 5′-extension of UUUGG to the L-21) at 25 °C is expected to be much lower than 85 nM due to the three extra AU base pairs between A$_5$ of -1r,dSA$_5$ and the UUUGG extension.[9] This substrate is expected to be fully bound at the start of the reaction because the substrate binding rate under these conditions

is much faster than the reaction (k_{on} = 9x10^7 M^{-1} min^{-1}).[10] Saturation was confirmed by demonstrating that the reaction rate constant was independent of ribozyme concentration (0.5 – 4.5 μM; data not shown).

Measurement of docking equilibria ($K_{dock,P}$)

The docking equilibria of product, $K_{dock,P}$, of wild type and mutant ribozymes were measured using a pulse-chase native gel binding assay as described previously. [11] Specifically, 150 nM of L-16 or L-21 *Sca*I ribozyme were first prefolded as described in the activity assay above. After annealing the prefolded ribozyme with sub-saturating amount of 5'-^{32}P radiolabeled product*, -1d,rP, 5'-CCCUCdU-3'* (for measure $k_{off,P}$ in the closed complex) or *-3m,-1d, rP, 5'-CCC$_m$UCdU-3'* (for measure $k_{off,P}$ in the open complex; m represents a 2'-methoxy modification which destabilize docking) at 25 °C for 20 minutes, excess non-radiolabeled -1d,rP was added to a final concentration of 9 μM to start the pulse chase reaction. The reaction conditions were 50 mM Na•MOPS, pH 7.0 and 10 mM $MgCl_2$ at 25 °C. Time points were taken by transferring 2 μl of the reaction mixture onto a running native gel, which separates the bound and unbound radiolabeled -1d,rP. The dock equilibrium $K_{dock,P}$ equals to $k_{off,P}$(open)/ $k_{off,P}$(closed) when $K_{dock,P}$ >> 1 (Table S4 and S5).[7]

Table S1. Anisotropy, K_{dock}, and overall activity k_{obs} of wild type and mutant *Tetrahymena* ribozymes

Ribozyme				
J1/2 sequence, 5′	P2 mutation	Anisotropy	K_{dock}	k_{obs} (min^{-1})
AAA*	N/A	0.303 ± 0.002	28 ± 5	0.43 ± 0.02
AAU	N/A	0.290 ± 0.001	12 ± 3	0.77 ± 0.08
UAA	N/A	0.299 ± 0.002	17 ± 2	0.38 ± 0.07
UAU	N/A	0.288 ± 0.003	30 ± 2	0.23 ± 0.04
AUA	N/A	0.296 ± 0.001	0.6 ± 0.1	0.0050 ± 0.0003
AUU	N/A	0.288 ± 0.001	0.6 ± 0.1	0.0018 ± 0.0004
UUA	N/A	0.287 ± 0.002	0.5 ± 0.1	0.010 ± 0.0014
UUU	N/A	0.284 ± 0.001	1.0 ± 0.2	0.0038 ± 0.0004
AGA	N/A	–	–	0.0049 ± 0.0010
ACA	N/A	–	–	0.0015 ± 0.0003
AAA	A31•U56 to C31•G56	–	–	0.0099 ± 0.0002
AUA	A31•U56 to C31•G56	–	–	0.0020 ± 0.0001
AAA	G32•C55 to U32•A55	–	–	0.52 ± 0.07
AUA	G32•C55 to U32•A55	–	–	0.0021 ± 0.0004

*Wild type J1/2 sequence

Table S2. Covariation of A29 and the A31•U56 base pair.[12]

Fraction (f) of residue 31 as A*			Information content (I) of residue 31**			Mutual information (MI) between residues***	
$f_{A29=A}$	$f_{A29\neq A}$	Δf	$I_{A29=A}$	$I_{A29\neq A}$	ΔI	29 and 31	29 and 56
0.91	0.39	0.52	0.54	1.58	-1.04	0.29	0.19

* When residue 31 is an A, it is paired with a U (residue 56) 99.8% of the time.

$f_{A29=A}$ refers to the fraction of IC1 group I intron sequences with A at position 31 when residue 29 is an A, and $f_{A29\neq A}$ refers to the fraction of sequences with A at position 31 when residue 29 is not an A.

$\Delta f = f_{A29=A} - f_{A29\neq A}$

** Information content[13] is defined as $I = 2 + \text{sum}(f_i * \log_2 f_i)$, where f_i is the fraction of i = A/U/G/C. $\Delta I = I_{A29=A} - I_{A29\neq A}$.

***Mutual information (MI)[14] between residue a and b is defined as $MI_{a,b} = \sum_{i,j} f_{a,b}(ij) \log_2 \frac{f_{a,b}(ij)}{f_a(i) f_b(j)}$, where $f_a(i)$, $f_b(j)$ and $f_{a,b}(ij)$ are the fraction of residue a being I = A/U/C/G, residue b being j=A/U/C/G and the pair of residues ab being ij, respectively. For comparison, the background MI of between residue 29 and 91 residues other than residue 31 and 56 is 0.06 ± 0.06.

The MI between residue 29 and the base pair 31/56 is 0.35 (not shown) and is two standard deviations higher than MI = 0.19 ± 0.07 for residue 29 and base pairs in P2 and P2.1 other than base pair 31/56 (also not shown).

Table S3. Length and sequence composition of J1/2 within the IC1 subclass of group I introns.[12]

J1/2 length (nt)	Occurrence (fraction)	Sequence conservation (from 3′ to 5′)*
0	204 (0.24)	N/A
1	206 (0.25)	$A_{0.43}$
2	48 (0.06)	$A_{0.73}\mathbf{A_{0.42}}$
3	53 (0.06)	$A_{0.74}A_{0.85}A_{0.51}$
4	301 (0.36)	$A_{0.61}\mathbf{A_{0.95}}A_{0.67}U_{0.35}$
5	12 (0.01)	$A_{0.58}\mathbf{A_{0.92}}A_{0.58}A_{0.67}G_{0.42}$
6	3 (<0.01)	AACAAA (x2), AAGAGC
7	1 (<0.01)	CCGACGU
8	4 (<0.01)	ACGAAAAA (x2), CCGAAGAC, CUAAUUGA
9	2 (<0.01)	AAGAGACGU, AAGACCCCG
11	1 (<0.01)	AAGUUCUUUUG

As the length of J1/2 varies amongst the IC1 subgroup, we separated these introns by J1/2 length and determined the conservation of J1/2 length and sequence for each length sequence.

* The most probable residue is shown at each position, with the fraction indicated by the subscript. Residues are colored according to their information content (ref 12, see also Figure 5) (purple: I > 1.5; blue: I = 1.5-1.5; green: I = 0.5-1.0; and red: I < 0.5). P2 is 3′ to J1/2 and reside 29, two residues upstream of the start of P2, is bolded in the first six rows. There is a preponderance of A residues two residues upstream of the start of P2.

Table S4. P1 docking of L-16 *Sca*I *Tetrahymena* ribozymes with J1/2, P2 and J2/3 mutations.

Ribozyme					
J1/2 sequence, 5′ to 3′	P2 mutation (Base pair 31•56)	$k_{off,P}$ (closed) (min^{-1})	$K_{dock,P}$(rel), fold down from WT**	$\frac{K_{dock,P}(A29)}{K_{dock,P}(U29)}$	Coupling***
AAA*	A•U*	0.00036	1	88	1
AUA	A•U	0.032	88		
AAA*	C•G	0.016	44	6	15
AUA	C•G	0.095	260		
AAA*	U•A	0.00070	2	155	0.6
AUA	U•A	0.11	310		
AAA*	G•C	0.0011	3	83	1
AUA	G•C	0.089	250		
AGA	U•A	0.14	380		
AGA	G•C	0.035	97	N/A	ND
AGA	C•G	0.12	330		
ACA	U•A	0.14	380		
ACA	G•C	0.12	330	N/A	ND
ACA	C•G	0.23	640		
	J2/3 mutation (residue 95)				
AAA*	A*	0.00036	1	88	1
AUA	A*	0.032	88		
AAA*	G	0.0044	12	28	3
AUA	G	0.12	330		

*Wild type sequence

** $K_{dock,p}$(rel) = $k_{off,p}$ (closed, mutant)/ $k_{off,p}$ (closed, WT), based on prior kinetic studies.[7,15] $k_{off,p}$ (open) = 0.29 min^{-1} (measured using open complex product -3m,-1d,rP* and a mutant ribozyme with J1/2=AUA) is assumed to be the same for wild type and mutant ribozymes, and represents an upper limit for the observed k_{off}.

***Coupling is the effect of A29 interaction, $K_{dock,P}$(A29)/$K_{dock,P}$(U29) without additional mutation outside of J1/2, divided by $K_{dock,P}$(A29)/$K_{dock,P}$(U29) in the presence of additional mutation in P2 or J2/3.

Table S5. P1 docking of the L-21 *Sca*I *Tetrahymena* ribozymes with J1/2, J7/3 and J8/7 mutations.

Ribozyme					
J1/2 sequence, 5′ to 3′	J7/3 and J8/7 (residues 269 & 304)	$k_{off,P}$ (closed) (min^{-1})	$K_{dock,P}$(rel), fold down from WT**	$\frac{K_{dock,P}(A29)}{K_{dock,P}(U29)}$	Coupling***
AAA*	A269*	0.00042	1	96	1
AUA	A269*	0.035	96		
AAA*	C269	0.00038	0.9	74	1
AUA	C269	0.028	67		
AAA*	U269	0.00035	0.8	91	1
AUA	U269	0.032	76		
AAA*	G269	0.00031	0.7	74	1
AUA	G269	0.023	55		
AAA*	A304*	0.00045	1.1	53	2
AUA	A304*	0.024	57		
AAA*	U304	0.0015	3.6	52	2
AUA	U304	0.079	190		
AAA*	G304	0.0011	2.6	59	2
AUA	G304	0.065	150		
AAA*	G269 and G304	0.00027	0.6	81	1
AUA	G269 and G304	0.022	52		

*Wild type sequence

** $K_{dock,p}$(rel) = $k_{off,p}$ (closed, mutant)/ $k_{off,p}$ (closed, WT), based on prior kinetic studies.[7,15]

***Coupling is the effect of A29 interaction, $K_{dock,P}(A29)/K_{dock,P}(U29)$ without additional mutation outside of J1/2, divided by $K_{dock,P}(A29)/K_{dock,P}(U29)$ in the presence of additional mutation in J7/3 or J8/7.

Table S6. Heterogeneity in smFRET measurement of P1 docking of *Tetrahymena* ribozymes

J1/2 sequence, 5′	Heterogeneity**, kcal/mol
AAA*	0.7 ± 0.1
AAU	0.7 ± 0.1
AUA	0.6 ± 0.2
UAA	0.7 ± 0.1
AUU	0.6 ± 0.1
UAU	0.6 ± 0.1
UUA	0.8 ± 0.1
UUU	0.5 ± 0.1

*Wild type J1/2 sequence

**Standard deviation of the distribution of free energies of docking, obtained as described in ref 16.

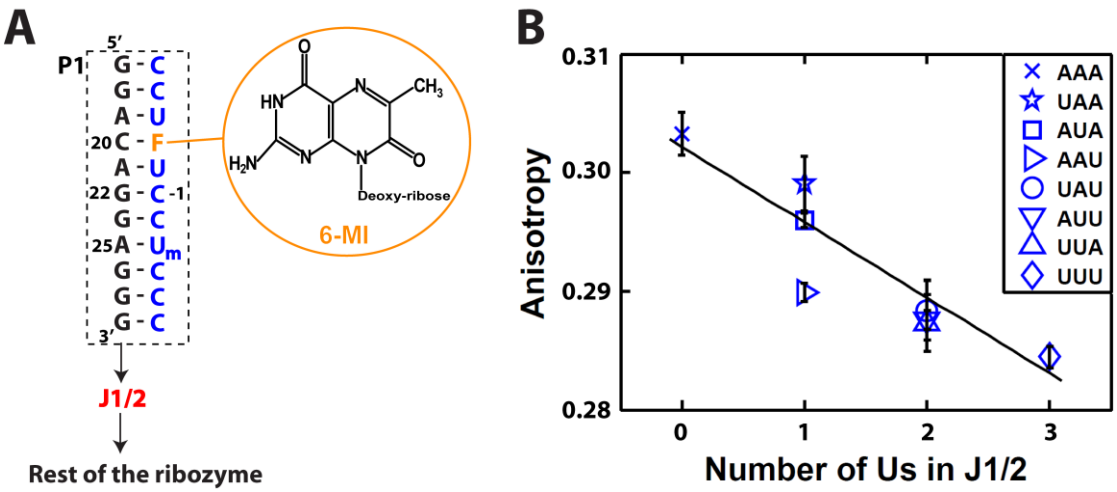
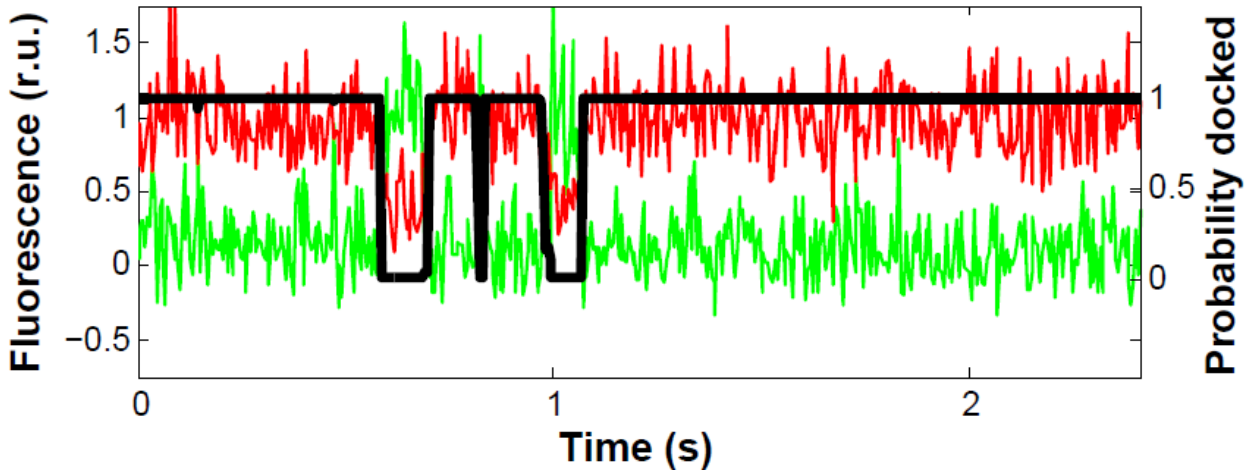


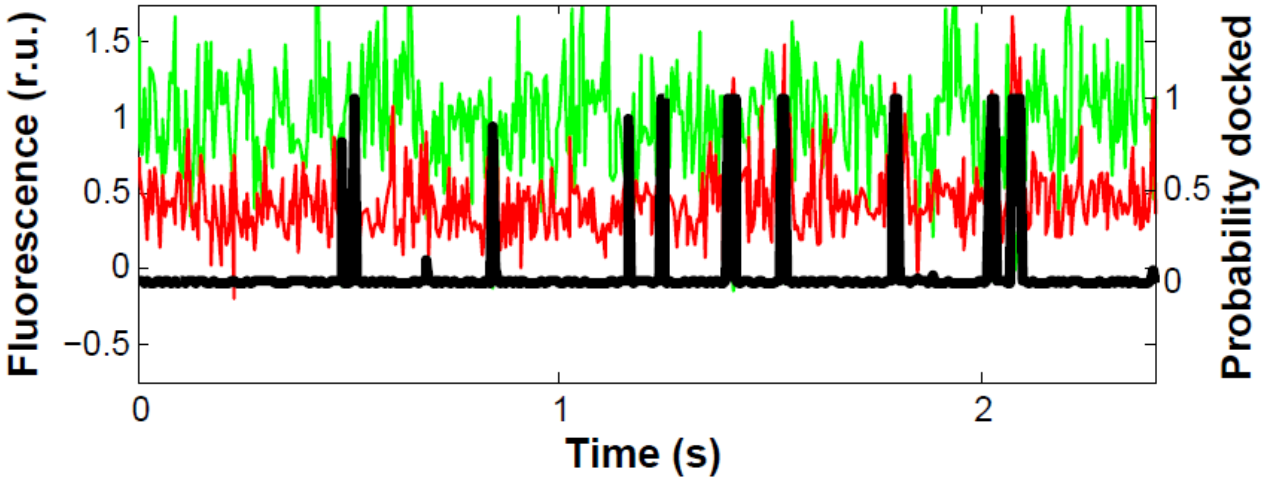
Figure S1. Anisotropy of the L-16 *ScaI* ribozyme with fluorescently labeled P1 duplex. (A) 6-MI-labeled P1 duplex. The P1 duplex was labeled with 6-MI by using an oligonucleotide substrate (blue) containing 6-MI (orange),[4] 5′-CCC$_m$UCCUFUCC, where F = 6-MI and m denotes a 2′-OCH$_3$ substitution. This substrate favors the formation of the open complex, with C(-1) and m(-3) destabilizing docking.[17] The 5′ extension sequence of P1 is altered from that of the natural sequence in order to optimize fluorescence properties of 6-MI. Potential perturbing effect of this sequence difference on docking is not an issue because in fluorescence anisotropy experiments we measured the P1 mobility only in the open complex where the helix is undocked and not interacting with the ribzoyme core. (B) The dependence of P1 anisotropy in the open complex on the number of A to U substitution in J1/2 (Table S1; J1/2 sequences are shown 5′ to 3′). Conditions: 50 mM Na•MOPS, pH 7.0, 10 mM $MgCl_2$, 25 °C.

Figure S2: Raw smFRET traces for the *Tetrahymena* ribozymes with the wild type and mutant J1/2 analyzed with the SMART program.[5] The green and the red lines are fluorescent intensity traces for the donor dye and for the acceptor dye, respectively. For display purposes, fluorescent intensity scale was set between 0 and 1 by dividing by the intensity at each time point by the mean intensity in the high fluorescence state, as calculated by SMART ($I_{donor} = I_{donor,\ undocked}$; $I_{acceptor} = I_{acceptor,\ docked}$). The black line displays the SMART-calculated probability that the molecule is in the docked state. Only the first 500 frame for each trace are displayed for better visualization of fast docking transitions.

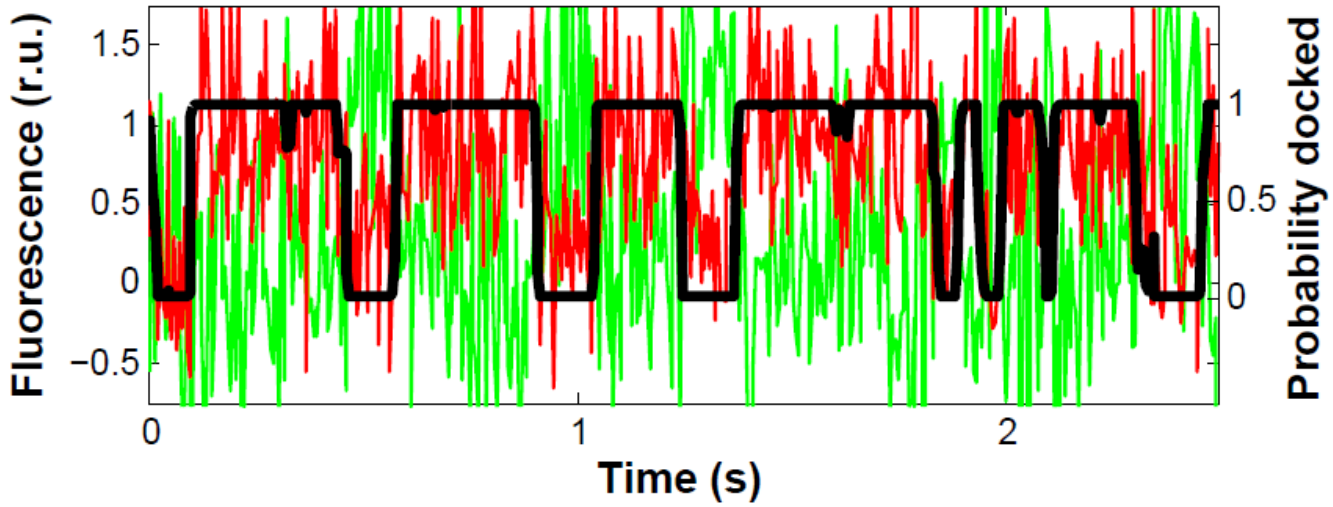
WT:



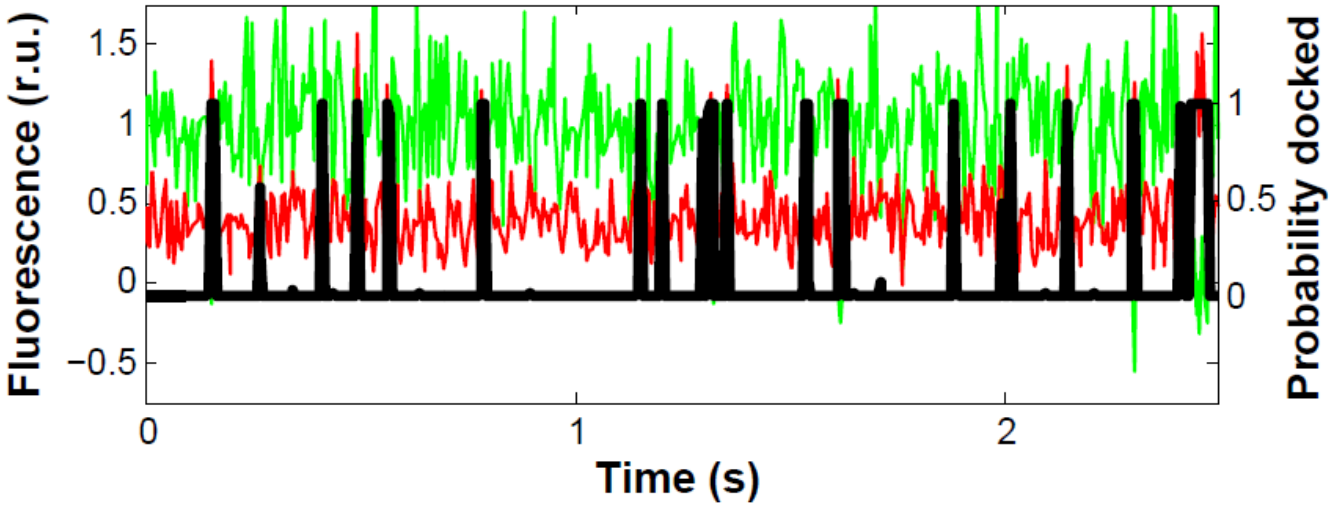
J1/2 = UUU:



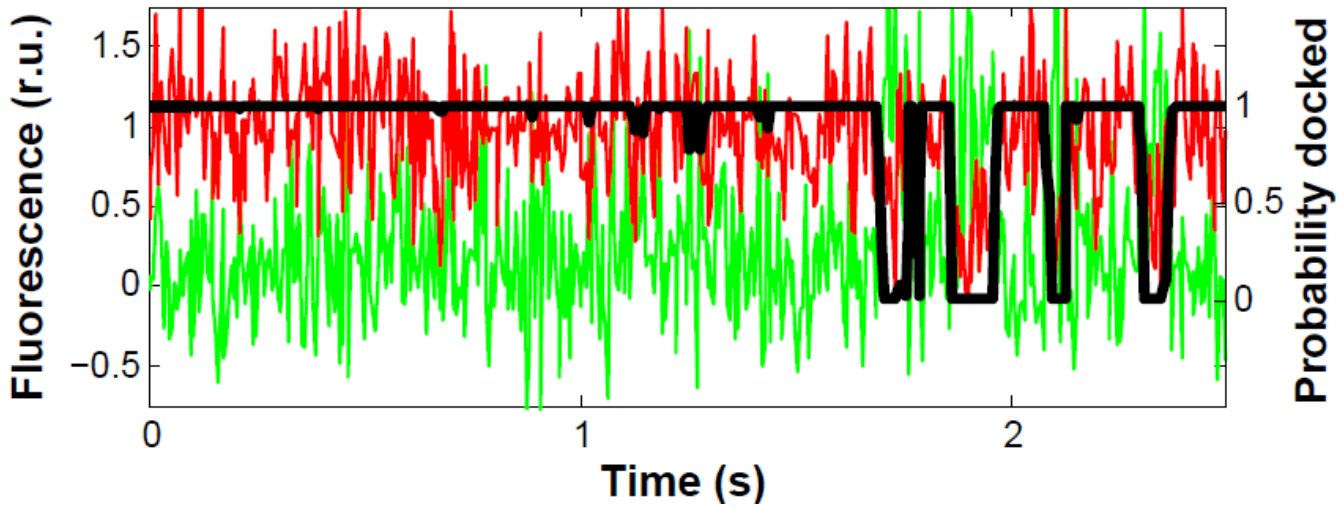
J1/2 = AAU:



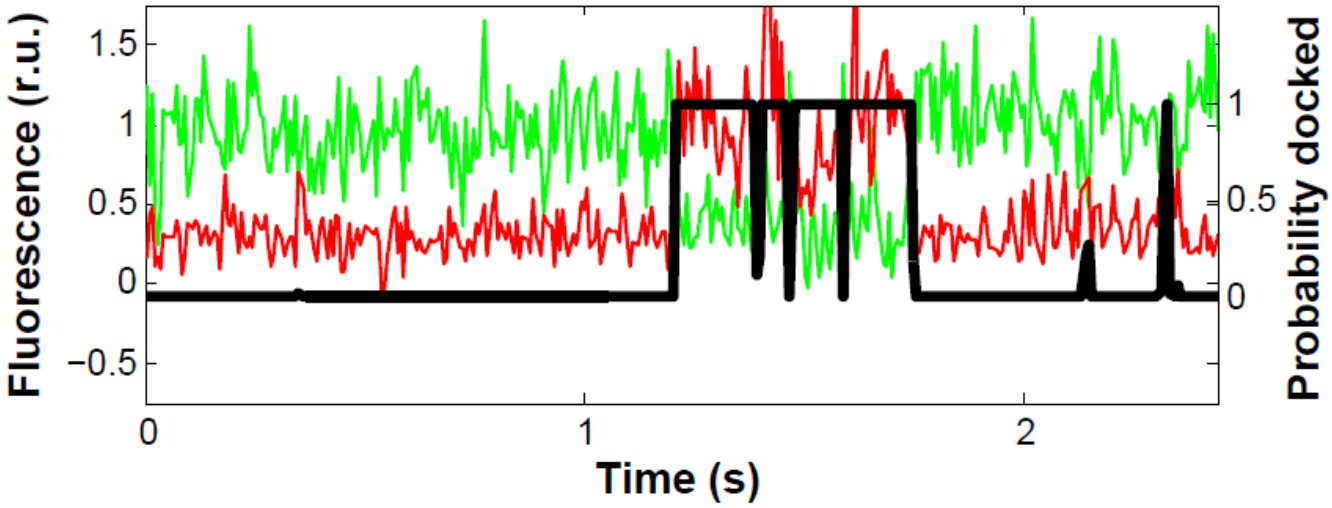
J1/2 = AUA:



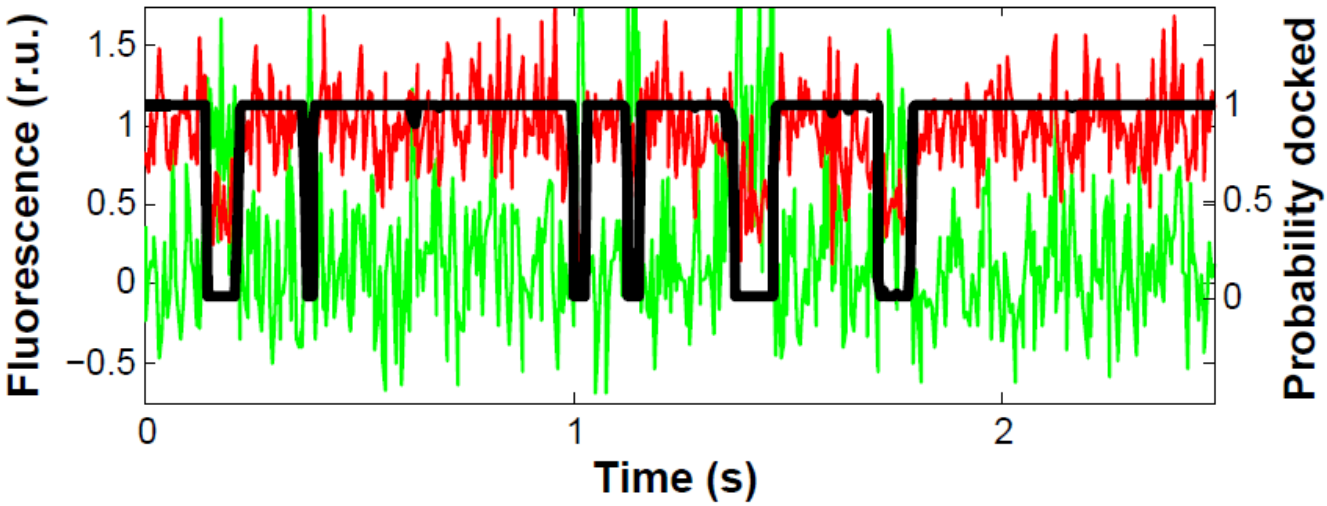
J1/2 = UAA:



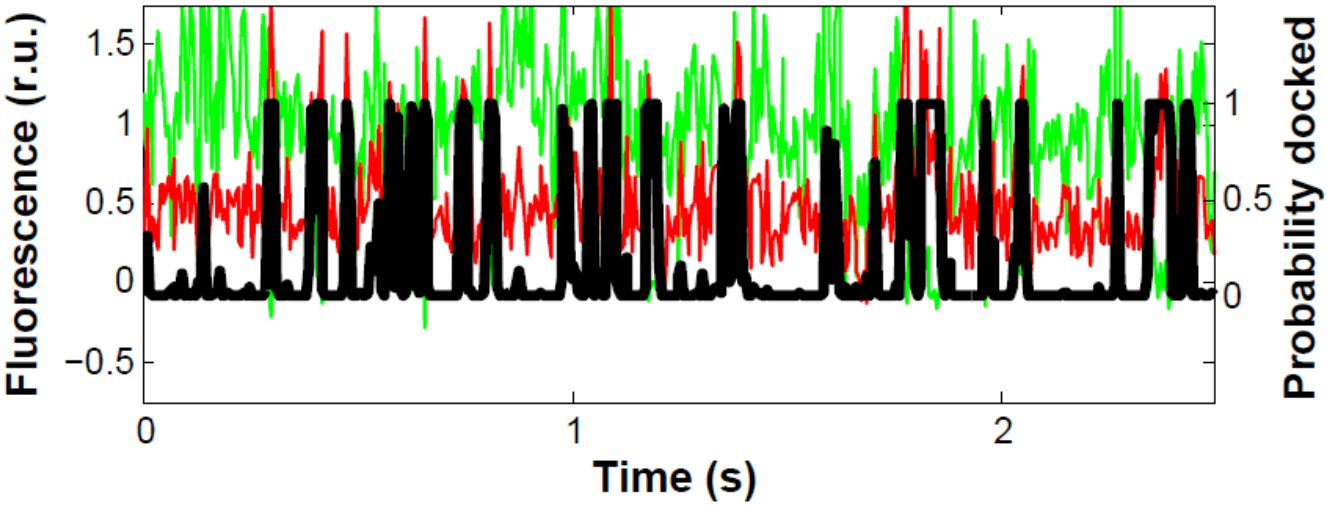
J1/2 = AUU:



J1/2 = UAU:



J1/2 = UUA :



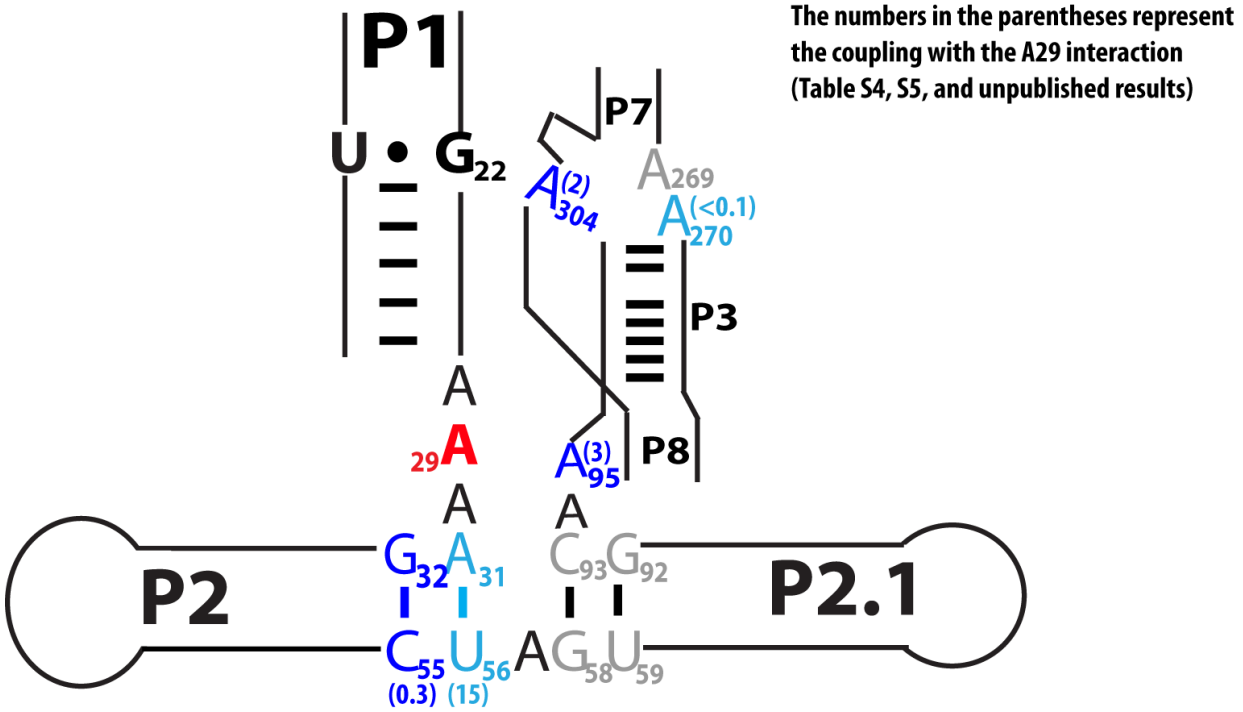


Figure S3. The J1/2 junction and its surroundings in the *Tetrahymena* group I ribozyme. P is for paired region and J is for junction. Residues that have been tested mutagenically for coupling with J1/2 residue A29 (red) are color-coded to represent the degree of coupling, which is given numerically in the parentheses: blue for 2-10 fold coupling, cyan for >10 fold coupling, and grey for no significant coupling (Tables S4, S5, and unpublished results). Synergistic versus anti-synergistic coupling is not distinguished - i.e., residues whose mutations enhance or diminish the A29 effect by the same about are shown with the same color but are represented inversely numerically. A269, A270, and A304 are core residues that may be involved in functionally important interactions.[18] The phosphate of A304 is also known to coordinate a functionally important Mg^{2+} cation.[19] A95 is a conserved residue and has been proposed to form a functionally important interaction with A57.[20]

References

(1) Karbstein, K.; Lee, J.; Herschlag, D. *Biochemistry* **2007**, *46*, 4861-4875.
(2) Solomatin, S. V.; Greenfeld, M.; Chu, S.; Herschlag, D. *Nature* **2010**, *463*, 681-684.
(3) Shi, X.; Mollova, E. T.; Pljevaljcic, G.; Millar, D. P.; Herschlag, D. *J. Am. Chem. Soc.* **2009**, *131*, 9571-9578.
(4) Hawkins, M. E.; Pfleiderer, W.; Balis, F. M.; Porter, D.; Knutson, J. R. *Anal. Biochem.* **1997**, *244*, 86-95.
(5) Greenfeld, M.; Pavlichin, D. S.; Mabuchi, H.; Herschlag, D. *Plos one, in press*.
(6) McConnell, T. S.; Cech, T. R. *Biochemistry* **1995**, *34*, 4056-4067.
(7) Benz-Moy, T. L.; Herschlag, D. *Biochemistry* **2011**, *DOI: 10.1021/bi2008245*.
(8) Knitt, D. S.; Narlikar, G. J.; Herschlag, D. *Biochemistry* **1994**, *33*, 13864-13879.
(9) Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. *J. Mol. Biol.* **1999**, *288*, 911-940.
(10) Herschlag, D.; Cech, T. R. *Biochemistry* **1990**, *29*, 10159-10171.
(11) Karbstein, K.; Carroll, K. S.; Herschlag, D. *Biochemistry* **2002**, *41*, 11171-11183.
(12) Zhou, Y.; Lu, C.; Wu, Q.-J.; Wang, Y.; Sun, Z.-T.; Deng, J.-C.; Zhang, Y. *Nucleic Acids Res.* **2008**, *36*, D31-D37.
(13) Schneider, T. D.; Stormo, G. D.; Gold, L.; Ehrenfeucht, A. *J. Mol. Biol.* **1986**, *188*, 415-431.
(14) Gutell, R. R.; Power, A.; Hertz, G. Z.; Putz, E. J.; Stormo, G. D. *Nucleic Acids Res.* **1992**, *20*, 5785-5795.
(15) Herschlag, D.; Cech, T. R., *Biochemistry* **1990**, *29*, 10159−10171
(16) Solomatin, S. V.; Greenfeld, M.; Herschlag, D. *Nat. Struct. Mol. Biol.* **2011**, *18*, 732-734
(17) Bartley, L. E.; Zhuang, X. W.; Das, R.; Chu, S.; Herschlag, D. *J. Mol. Biol.* **2003**, *328*, 1011-1026.
(18) Donnelly, L. O.; Szewczak, A. A.; Gutell, R. R.; Strobel S. A. *RNA* **1998**, *4*, 498-519.
(19) Forconi, M.; Lee J.; Lee J. K.; Piccirilli J. A.; Herschlag D. *Biochemistry* **2008**, *47*, 6883-6895.
(20) Downs, W. D.; Cech T. R. *Genes Dev.* **1994**, *8*, 1198-1211.