Supporting information for

"Roles of long-range tertiary interactions in limiting dynamics of the *Tetrahymena* group I ribozyme"

Xuesong Shi[†], Namita Bisaria[†], Tara L. Benz-Moy[‡], Steve Bonilla[§], Dmitri S. Pavlichin[‖], Daniel Herschlag[†,‡,*]

[†]Department of Biochemistry, [‡]Department of Chemistry, [§]Department of Chemical Engineering, [‖]Department of Physics, Stanford University, Stanford, California 94305, United States

SI Methods

Single molecule FRET

Docking rate and equilibrium constants (k_{dock}, k_{undock}, and K_{dock}) were measured using single molecule FRET (smFRET) of individual *Tetrahymena* ribozyme molecules. These molecules were modified (L-16 *Sca*I version) containing a 3′ extension for annealing with a 3′ Cy5-labelled and 5′ biotin-labeled oligonucleotide with the sequence used previously (5′-biotin-ACCAAAAUCAACCUAAAACUUACACA-Cy5-3′).[1] Molecules were prepared by first folding the ribozyme at 50 °C for 30 min in 50 mM Na-MOPS, pH 7.0, and 10 $MgCl_2$, and then annealing the ribozyme with the 5′ biotin-3′Cy5 DNA oligonucleotide and the substrate 5′-r(CCCUC)dUr(AAACC)-Cy3 for 10 min at room temperature. The sample was then diluted to a concentration of ~75 pM, and attached to the surface of quartz slides for imaging in a total internal reflection microscope, as described in ref 1. Data were taken in 50 mM Na-MOPS, pH 7.0, with 10 mM $MgCl_2$, unless noted otherwise, and with an oxygen scavenging system of 44 mM glucose, ~1 mM Trolox, and small amounts of glucose oxidase and catalase. Image data were taken over a range of frame rates (5-20 ms) at a signal to noise ratio of 2-3, and the average dye lifetime was 30-40 s. The FRET traces of individual molecules displayed transitions between two FRET states: a high FRET state of ~0.95, corresponding to the docked states, and a low FRET state of ~0.4 corresponding to the undocked state.[1-2] Traces were accepted if there was single step photo-bleaching, as expected for a single fluorophore, with a corresponding increase in donor fluorescence or decrease in acceptor fluorescence, and if there was a sufficiently high signal to noise ratio (SNR), as determined visually.

Rate and equilibrium constants for docking were determined by analyzing FRET traces with the SMART[3] analysis package using a hidden Markov Modeling (HMM) based algorithm and fitting to a two-state model. The data were also fit to a three-state model and found to fit better to the two-state

model than to a three-state model according to the Bayesian information criterion (BIC) [3] (data analysis not shown).

Reported errors were determined by 95% confidence intervals of the mean from bootstrapping the parameters from individual molecules. Heterogeneity of docking behavior was previously observed and was described in terms of an 'heterogeneity parameter' H, the standard deviation (s.d.) of the ΔG_{dock} distribution.[2, 4] Values obtained herein are reported in Table S2, and the value for the wild type ribozyme is similar to the previous reported value (H = 0.67 and 0.70 respectively).

Table S1 Function effects (as fold change) for mutants compared to WT ribozyme (data from ref 5)

Tertiary contact mutated	$K_{dock,rel}$[a]	$K_{d,UCG,rel}$[b]	Coupling[c]	Chemical step[d]
none (WT)	(1)	(1)	(1)	(1)
P14 (L2)	5.4	0.7	3	≤1.8
P14 (L5c)	4.4	0.8	2	≤1.8
P13 (L2.1)	1.5	2.0	1.5	1.1
P13 (L9.1)	1.5	2.0	1.1	1.1
MC/MCR (ARB)	20.8	1.4	4	≤1.7
TL/TLR (L5b)	0.7	1.4	1	1.0
TL/TLR (J6a/b)	1.0	0.9	1.4	0.9
L9/P5	1.3	11.6	≤1.6	2.3

[a]The docking equilibrium constants for P1 forming tertiary interactions within the core of the ribozyme were determined from pulse-chase assays described in previously published results.[5] These values are the fold decrease in the overall docking of P1 for the mutant relative to that of WT so that $K_{dock,rel}=K^{WT}/K^{mutant}$ in the absence of guanosine. Reaction conditions: 30 °C, 50 mM Na-MOPS, pH 6.9, 10 mM $MgCl_2$.

[b]The UCG binding equilibrium constants presented here are for the dissociation of UCG from the (E•S) complex in the absence of P1 forming tertiary interactions with the ribozyme core except in the case of the L9/P5 mutant; the L9/P5 mutant was obtained with the P1 interactions formed for both the mutant and the WT. These values are the fold increase in the dissociation equilibrium of UCG for the mutant relative to that of WT so that $K_{d,UCG,\ rel} = K^{mutant}/K^{WT}$ in the absence of guanosine. Reaction conditions: 30 °C, 50 mM Na-MOPS, pH 6.9, 10 mM $MgCl_2$.

[c]In the WT ribozyme, guanosine binding and the oligonucleotide substrate binding are coupled, so that once guanosine is bound, the oligonucleotide substrate binds more tightly and vice versa. The degree that the binding of these substrates are energetically linked is referred to as the coupling constant. These values are the fold decrease in the coupling constant the mutant relative to that of WT so that $Coupling = K^{WT}/K^{mutant}$. Reaction conditions: 30 °C, 50 mM Na-MOPS, pH 6.9, 10 mM $MgCl_2$.

[d]The first-order rate constant for the transformation of (E•S•UCG) → P was determined with saturating UCG with respect to E and an oligonucleotide substrate that forms tertiary interactions with the ribozyme core in the WT ribozyme. Mutants that dock poorly appear as limits because S may not be fully docked for them under these conditions. These values are the fold decrease in the chemical step relative to that of WT so that Chemical Step = k^{WT}/k^{mutant}. Reaction conditions: 30 °C, 50 mM Na-MOPS, pH 6.9, 10 mM $MgCl_2$.

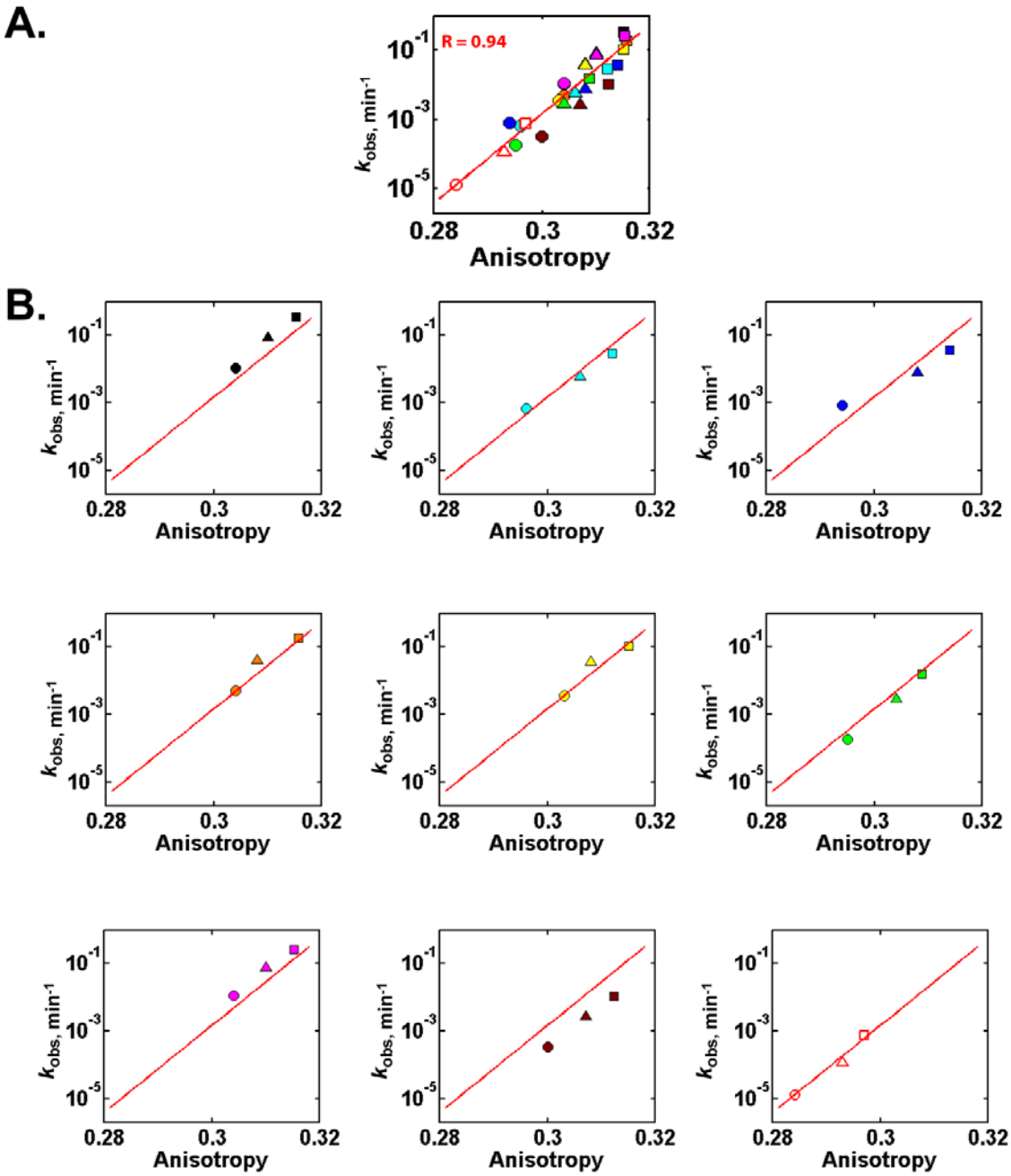

Figure S1. Correlation between ribozyme activity and P1 anisotropy. Activity and open complex P1 anisotropy of the following ribozymes: wild type (black) and mutant (MC/MCR green; P14 (L2), cyan; P14 (L5C) blue; P13 (L2.1), yellow; P13 (L9.1), orange; TL/TLR, magenta; L9/P5, brown; P14(L5C)&MC/MCR double mutant, red unfilled). The reaction rate, k_{obs}, was obtained under varying concentrations of $MgCl_2$ (10 mM circle, 30 mM triangle, 100 mM square) together (**A**, same as Figure 2C) or individually (**B**). The correlation line in **A** is kept in **B**. See the legend of Figure 2 for details about the reaction and the FPA measurements.

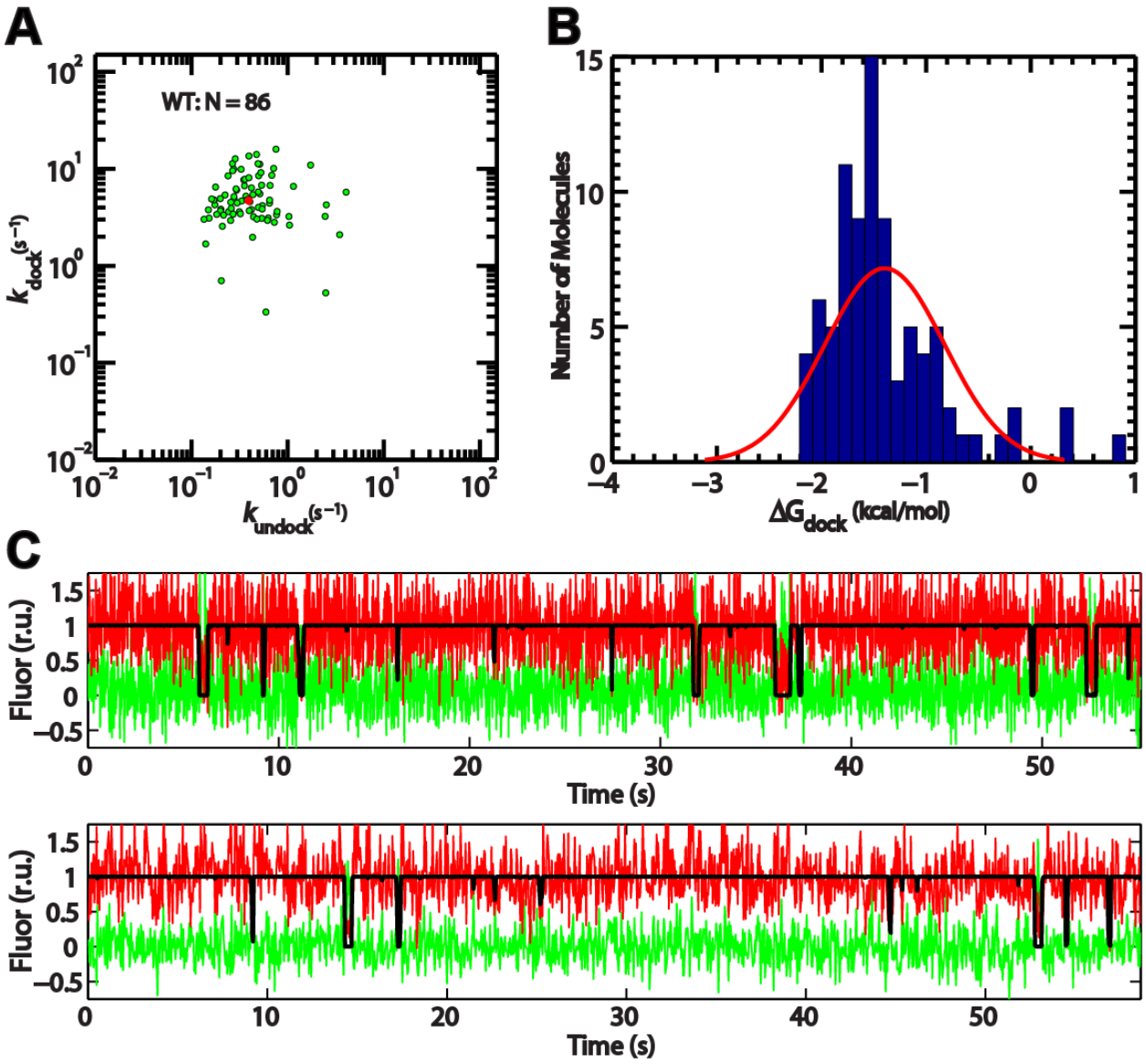

Figure S2. smFRET study of the P1 docking kinetics of the WT *Tetrahymena* ribozyme. (A) The docking, k_{dock}, and undocking, k_{undock}, rate constants for individual molecules (green) and the median (red). **(B)** Histogram of the distribution of docking free energy, with the red line indicating a fit to a single Gaussian distribution. **(C)** Typical FRET traces. The black line in each trace represents the probability of the high FRET state determined by a hidden Markov model fit to data.[3]

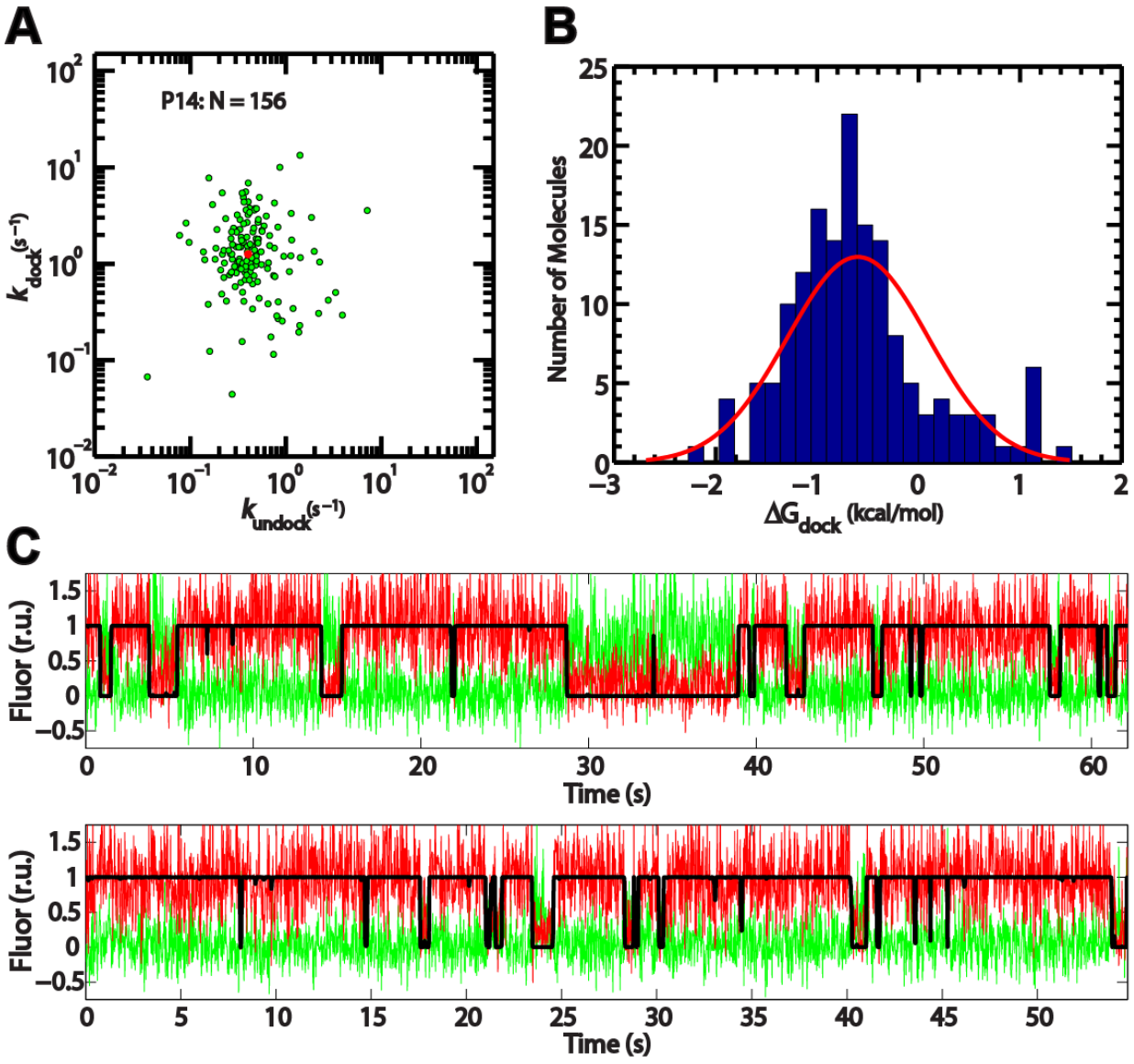

Figure S3. smFRET study of the P1 docking kinetics of the P14 mutant of the *Tetrahymena* ribozyme. (A) The docking, k_{dock}, and undocking, k_{undock}, rate constants for individual molecules (green) and the median (red). **(B)** Histogram of the distribution of docking free energy, with the red line indicating the fit to a single Gaussian distribution. **(C)** Typical FRET traces. The black line in each trace represents the probability of the high FRET state determined by a hidden Markov model fit to data.[3] The P14 construct has the L5C mutation (Figure 1A).

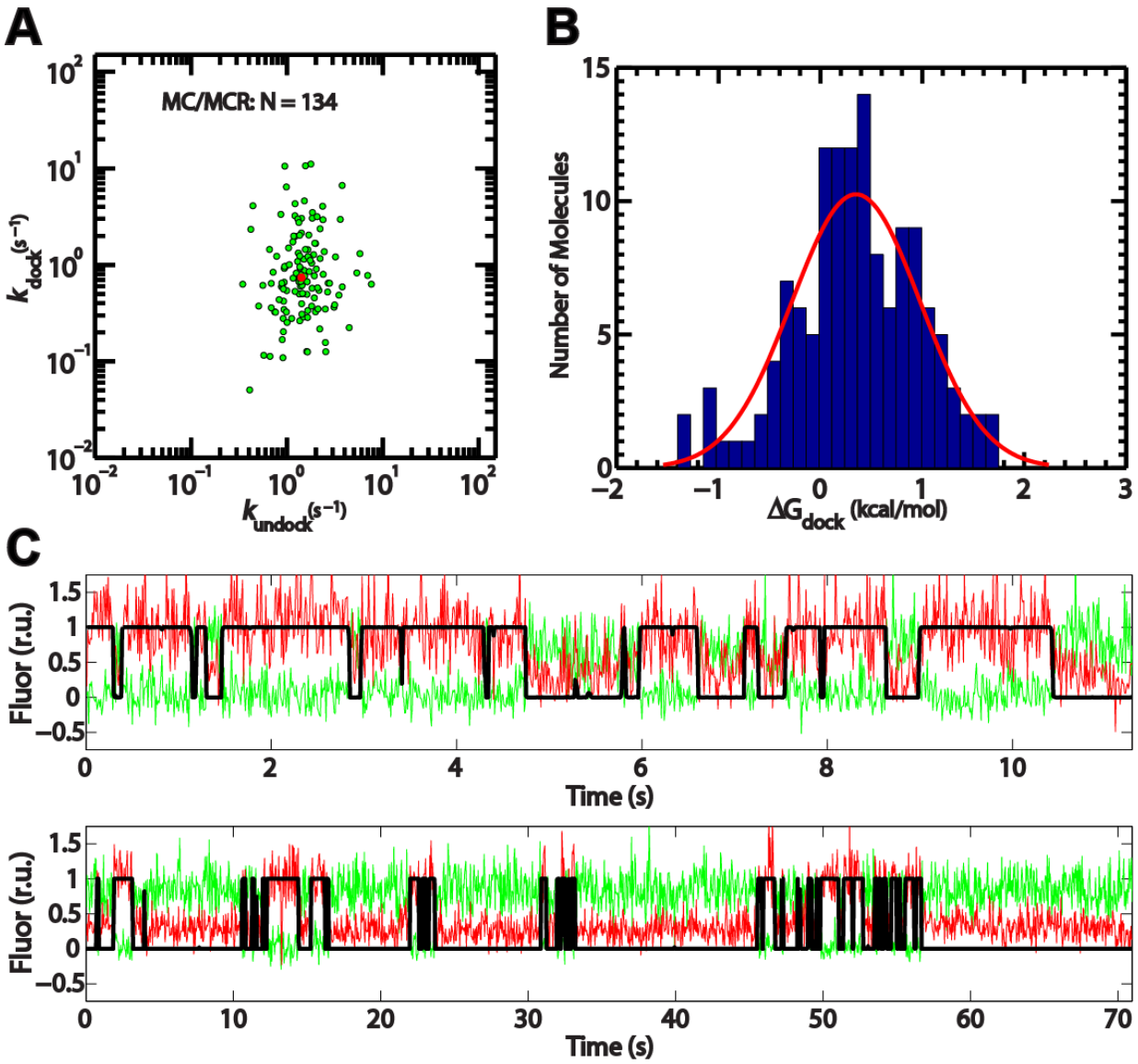

Figure S4. smFRET study of the P1 docking kinetics of the MC/MCR mutant of the *Tetrahymena* ribozyme. (A) The docking, k_{dock}, and undocking, k_{undock}, rates for individual molecules (green) and the median (red). **(B)** Histogram of the distribution of docking free energy, with the red line indicating the fit to a single Gaussian distribution. **(C)** Typical FRET traces. The black line represents the probability of the high FRET state determined by a hidden Markov model fit to data.[3]

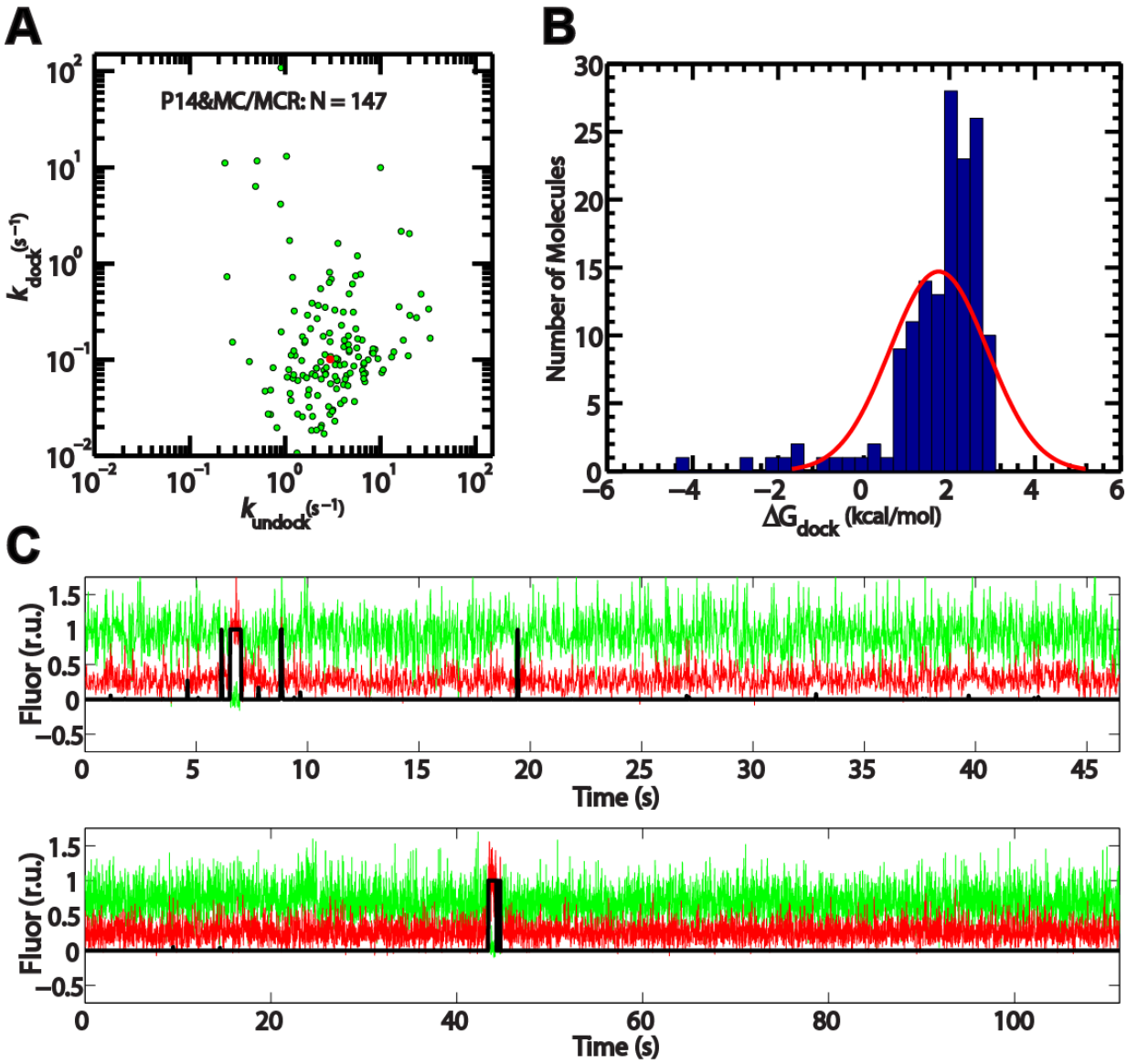

Figure S5. smFRET study of the P1 docking kinetics of the P14&MC/MCR double mutant of the *Tetrahymena* ribozyme. (A) The docking, k_{dock}, and undocking, k_{undock}, rate constants for individual molecules (green) and the median (red). **(B)** Histogram of the distribution of docking free energy, with the red line indicating the fit to a single Gaussian distribution. **(C)** Typical FRET traces. The black line represents the probability of the high FRET state determined by a hidden Markov model fit to data.[3] The P14&MC/MCR construct has the L5C mutation (Figure 1A).

The P14&MC/MCR double mutant smFRET data included a significant number of traces that had no docking transitions, as the docking rate is slow and comparable to the trace length. To obtain an accurate determination of the docking rate constants under conditions in which a significant fraction of the molecules exhibit no transitions, we developed an algorithm to include these traces in the determination of the docking rate constant through inclusion of the time spent in this undocked state in the Markov model (see SI Appendix for details). Exclusion of these traces (i.e. the traces without transitions), gives a four-fold increase in the docking rate constant but does not change the conclusions of this work.

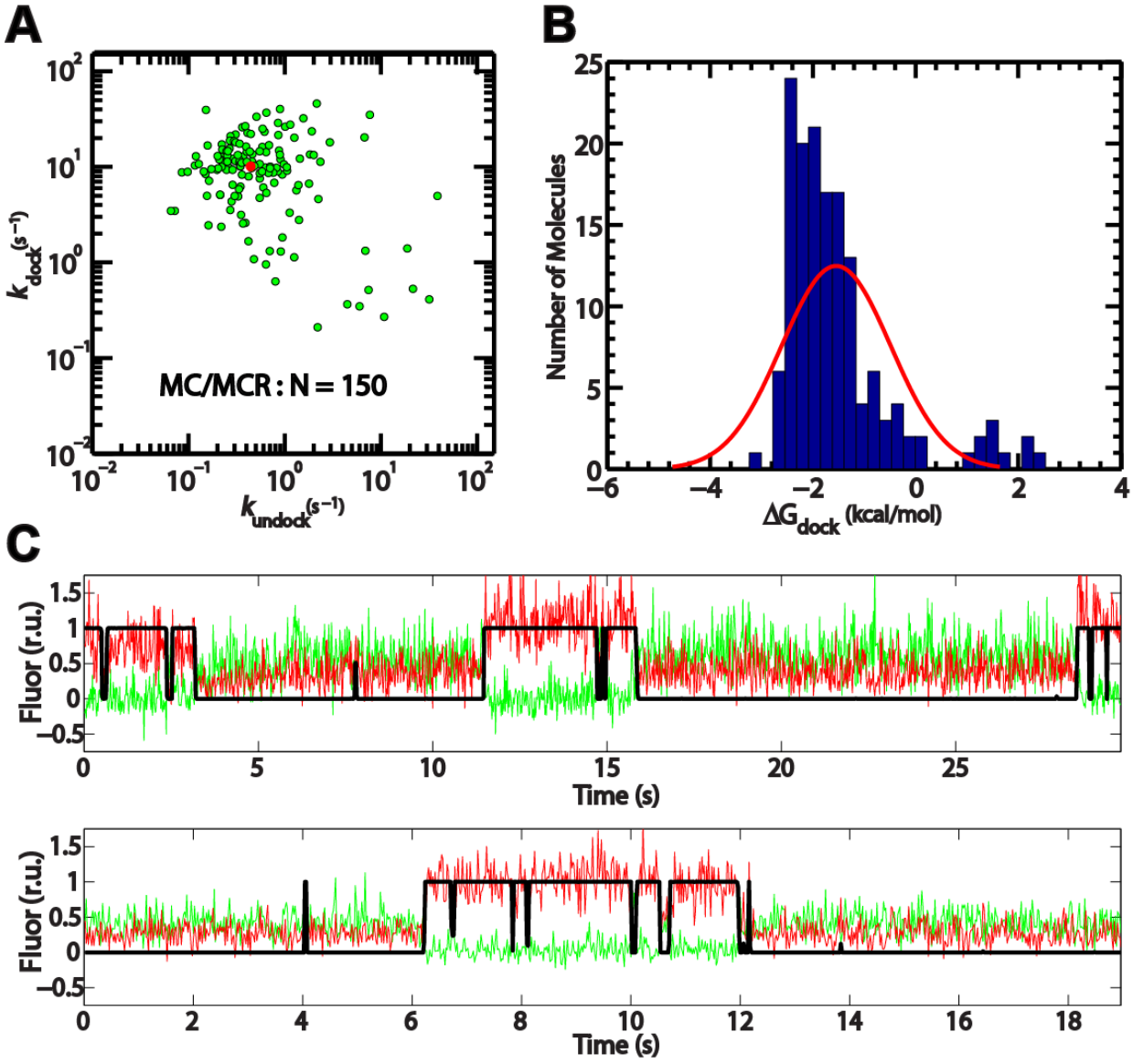

Figure S6. smFRET study of the P1 docking kinetics of the MC/MCR mutant in the *Tetrahymena* ribozyme in 100 mM Mg^{2+}. (A) The docking, k_{dock}, and undocking, k_{undock}, rate constants for individual molecules (green) and the median (red). **(B)** Histogram of the distribution of docking free energy, with the red line indicating the fit to a single Gaussian. **(C)** Typical FRET traces. The black line represents the probability of the high FRET state determined by a hidden Markov model fit to data.[3]

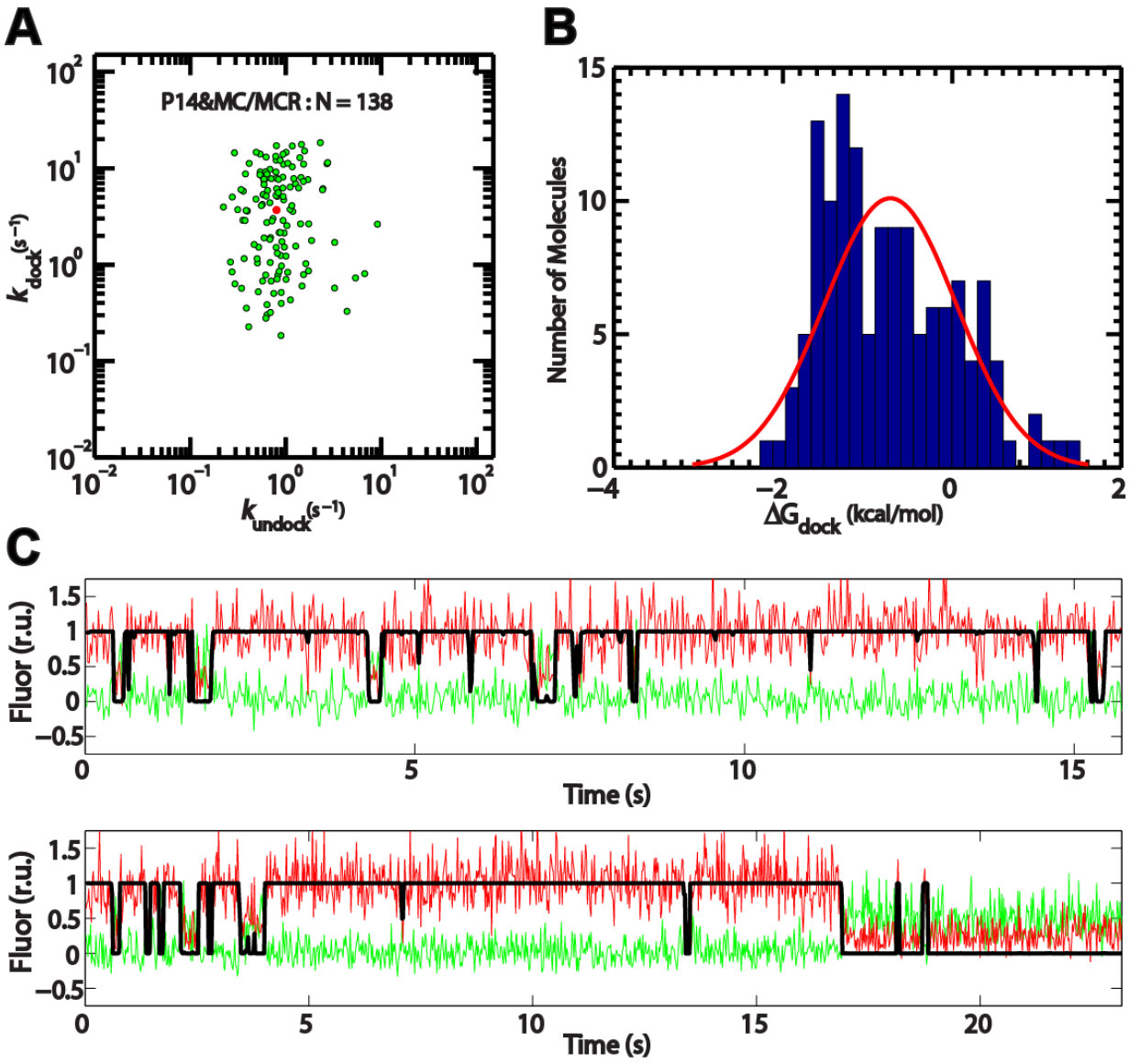

Figure S7. smFRET study of the P1 docking kinetics of the P14 &MC/MCR double mutant in the *Tetrahymena* ribozyme in 100 mM Mg^{2+}. (A) The docking, k_{dock}, and undocking, k_{undock}, rates for individual molecules (green) and the median (red). **(B)** Histogram of the distribution of docking free energy, with the red line indicating the fit to a single Gaussian. **(C)** Typical FRET traces. The black line represents the probability of the high FRET state determined by a hidden Markov model fit to data.[3] The P14&MC/MCR construct has the L5C mutation (Figure 1A).

Table S2: P1 docking parameters of the L-16 *Sca*I *Tetrahymena* ribozymes obtained from smFRET

Mutations in Ribozyme	k_{dock} (s^{-1})	k_{undock} (s^{-1})	K_{dock}	ΔG (kcal/mole)	Heterogeneity*	SNR Cy3	SNR Cy5	Number of molecules	Average lifetime traces (s)
WT	4.6 ± 0.65	0.42 ± 0.1	11.0 ± 2.3 (5.0)	-1.4 ± 0.14	0.67	2.3 ± 0.9	2.3 ± 0.9	86	26
P14	1.2 ± 0.18	0.44 ± 0.05	2.8 ± 0.5 (2.8)	-0.60 ± 0.1	0.65	1.9 ± 0.7	2.1 ± 0.6	156	63
MC/MCR	0.80 ± 0.14	1.5 ± 0.14	0.55 ± 0.12 (0.52)	0.35 ± 0.11	0.63	3.5 ± 1.4	3.0 ± 1.0	134	37
P14& MC/MCR	0.04 ± 0.01 *0.14 ± 0.04*[#]	2.8 ± 0.66	0.014 ± 0.007 (0.022) *0.05 ± 0.02*[#] *(0.08)*[#]	1.75 ± 0.21	0.88	3.0 ± 1.9	3.5 ± 1.5	147	39

WT is wild type. The P14 and P14&MC/MCR constructs have the L5C mutation (Figure 1A). Mean and bootstrapped 95% confidence interval of mean are given. K_{dock} is calculated as k_{dock}/k_{undock} and as the ratio of total time spend in the high FRET state to the low FRET state (in parentheses). SNR is the signal to noise ratio. Traces used in the analysis had SNR > 1 which avoids fitting noise instead of FRET transitions.

Values obtained not including traces without transitions (in italics; see SI Appendix).

* Represents the standard deviation of the ΔG distribution.

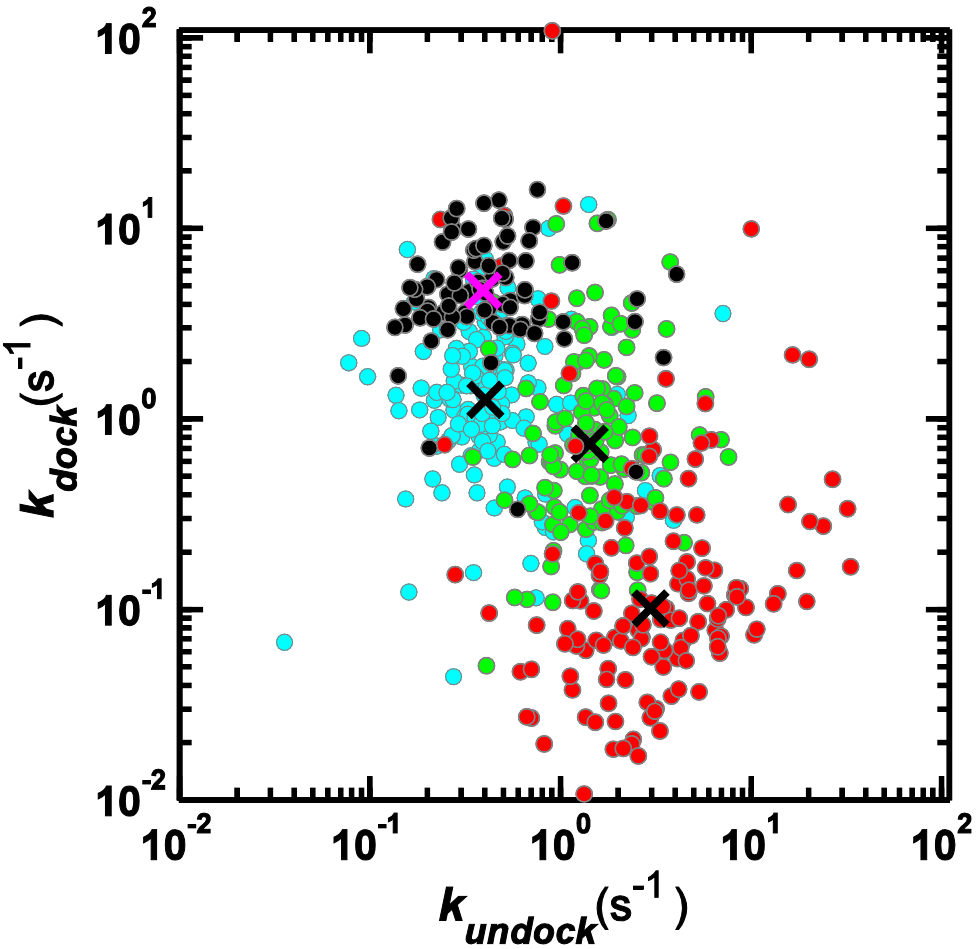

Figure S8. Combined smFRET P1 docking kinetics data. smFRET data for WT (black), P14 (cyan), MC/MCR (green) and P14&MC/MCR (red) ribozymes. The P14 and P14&MC/MCR constructs have the L5C mutation (Figure 1A). The mean values of each ribozyme clusters are labeled with a black (mutants) or a magenta (WT) cross.

Table S3. The effect of Mg^{2+} on the P1 docking kinetics for the MC/MCR mutant and the P14&MC/MCR double mutant ribozymes

	Mg^{2+} (mM)		
MC/MCR	10	100	Fold change*
k_{dock} (s^{-1})	0.80 ± 0.13	8.0 ± 1.3	10
k_{undock} (s^{-1})	1.45 ± 0.14	0.56 ± 0.12	0.39
K_{dock}	0.55 ± 0.12 (0.52)	14 ± 3 (7.8)	25
	Mg^{2+} (mM)		
P14&MC/MCR	10	100	Fold change*
k_{dock} (s^{-1})	0.04 ± 0.01	2.2 ± 1.1	16
k_{undock} (s^{-1})	2.8 ± 0.64	0.83 ± 0.10	0.30
K_{dock}	0.014 ± 0.007 (0.022) *0.05 ± 0.02*** (*0.08*)**	2.7 ± 1.4 (2.0)	54

The docking kinetics were determined by smFRET (see SI Methods for detail).
K_{dock} is calculated as k_{dock}/k_{undock} and as the ratio of total time spend in the high FRET state to the low FRET state (in parentheses).
The P14&MC/MCR constructs have the L5C mutation (Figure 1A).
*The change in value going from 10 to 100 mM, the value at 100 mM divided by the value at 10 mM.
**not including traces without transitions (italic).

Table S4. Comparison of K_{dock} values obtained from smFRET in this work and from prior bulk assays[5]

	K_{dock}	
Mutant ribozyme	smFRET	Bulk gel-shift assay[5]
WT	11 ± 2	10 ± 2.9
P14	2.8 ± 0.5	2.2 ± 0.7
MC/MCR	0.55 ± 0.12	0.47 ± 0.14

The K_{dock} values obtained from smFRET measurements herein (see SI methods for detail) and from bulk methods as reported in ref 5. Conditions: 50 mM Na-MOPS, pH 7.0, and 10 mM $MgCl_2$ at 23 °C (smFRET) or 30 °C (bulk). The P14 constructs have the L5C mutation (Figure 1A).

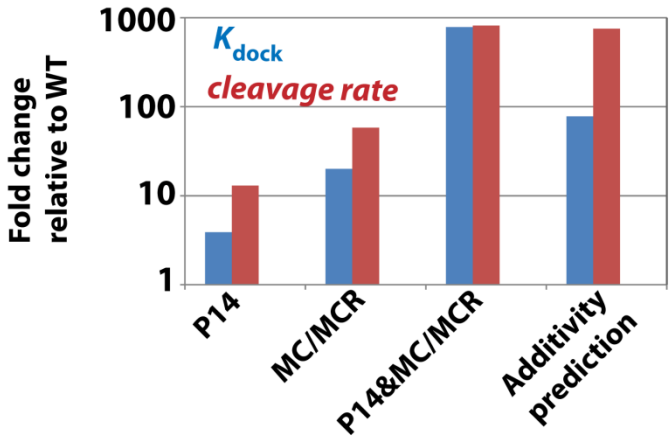

Figure S9. Comparison of the fold change in docking constant and reaction rate relative to WT for single and double mutants. The additivity prediction column is the predicted fold change for the double mutant if the effects from P14 and MC/MCR mutations are additive. K_{dock} was determined from smFRET under conditions of 50 mM Na-MOPS, pH 7.0, and 10 mM $MgCl_2$ at 23 °C (Table S2). The relative cleavage rate is for the reaction $(E{\cdot}S)_{open}$ + UCG → product, measured at 10 μM UCG (subsaturating) using the oligonucleotide substrate, d(CCCUC)rUA_5, which binds primarily in the open complex,[6] and is under the same solution conditions as the smFRET experiment except at 30 °C (also see Figure 2A). The P14 constructs have the L5C mutation (Figure 1A).

The larger effects observed in the cleavage assay than the K_{dock} effect for the single mutants may arise from small effects of steps other than docking.[5] The same effect is observed for the double mutant so that there may be no effects on other steps for the double mutant or there may be different behaviors for the different ribozyme forms used in the smFRET versus activity assays [see SI Method, small differences in docking have been observed with different constructs (refs 5,7 and unpub. results)]; and/or the differences may be the result of measurement error, which likely is greater for the docking measurements especially for the double mutant which has a very small docking equilibrium. The observed difference for the double mutant in smFRET docking versus the cleavage activity does not affect the conclusions herein.

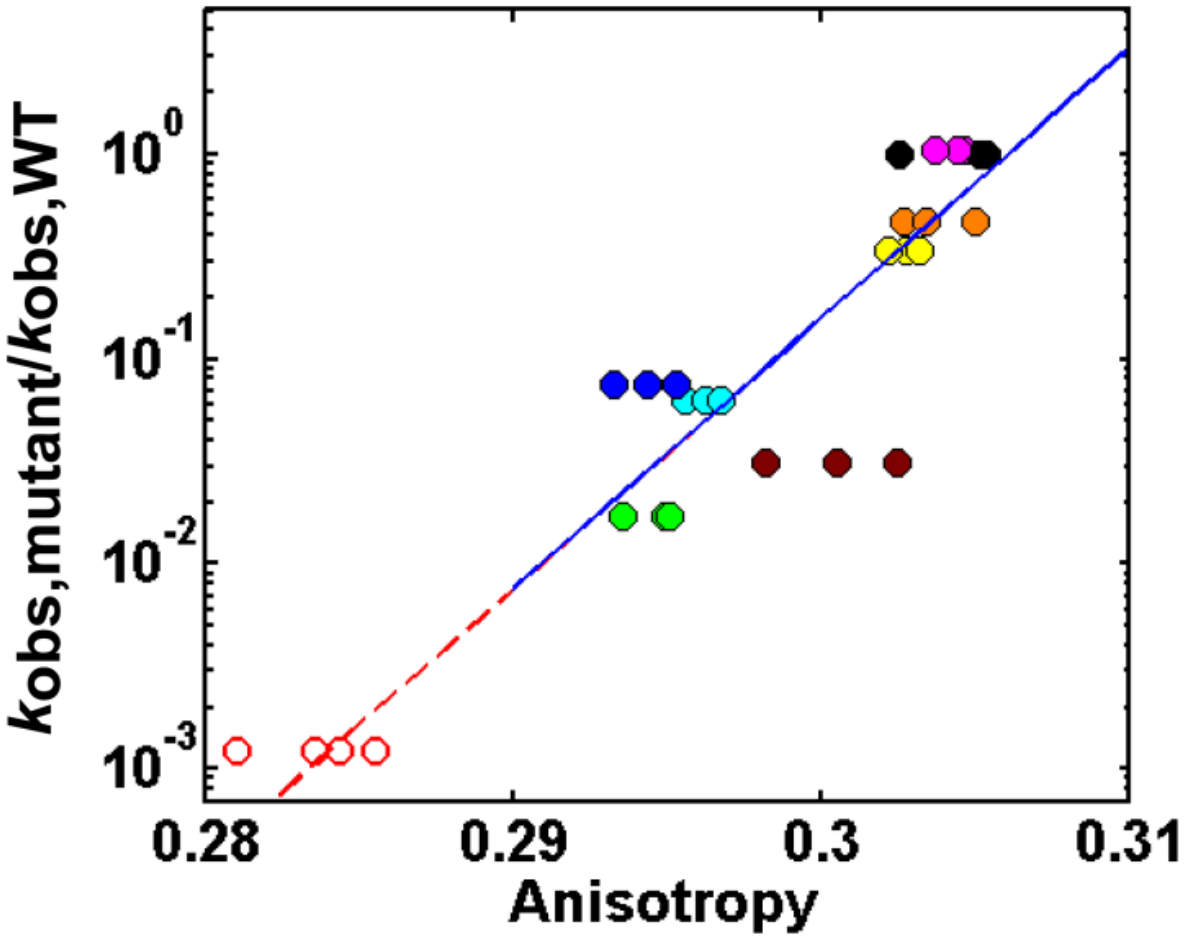

Figure S10. Dispersion in fluorescence polarization anisotropy (FPA) data. To illustrate the dispersion in the primary FPA data, Figure 2A is replotted by replacing the average FPA for wild type (black) and mutant (MC/MCR, green; P14 (L2), cyan; P14 (L5C), blue; P13 (L2.1), yellow; P13 (L9.1), orange; TL/TLR, magenta; L9/P5, brown; MC/MCR&P14 (L5C) double mutant, red unfilled) ribozymes with primary FPA values from multiple independent measurements. Experimental condition: 50 mM NaMOPS, pH 7.0, 30 °C and 10 mM $MgCl_2$.

Table S5. Effects of Peripheral Tertiary Contact Mutations on Ribozyme Catalysis

	Fold down relative to WT	
Ribozyme*	k_{obs}**	$k_2^{overall}$***
TL/TLR(L5B)	1.0 ± 0.3	0.5 ± 0.2
P13(L9.1)	2.1 ± 0.8	3.3 ± 0.4
P13(L2.1)	3.0 ± 1.0	3.6 ± 0.6
P14(L5C)	13 ± 4	18 ± 4
P14(L2)	16 ± 5	18 ± 4
L9/P5	32 ± 8	35 ± 7
MC/MCR	58 ± 21	66 ± 12

*Ribozymes are named by the long-range contact (Figure 1A) that is disrupted and the specific mutation site (in parenthesis; Figure 1A). For simplicity, mutants are referred to in the text by just the tertiary contact affected.

**Fold decrease relative to WT in the rate of the reaction: $(E.S)_{open}$ + UCG → P at 10 μM UCG; data are from ref 5.

***Fold decrease relative to WT in the second-order rate constant of the reaction: $(E.S)_{open}$ + UCG → P; data are from ref 5.

SI Text

As the P14&MC/MCR double mutant has a slow docking rate (at 10 mM Mg^{2+}) that is comparable to the trace length, it is uniquely expected to have a significant fraction of traces that exhibit no docking transitions. Indeed 303 out of 390 (78%) of the traces identified, as described in SI Methods, had no transitions. To accurately determine the overall docking rate constant, we included these traces in our calculation of k_{dock} and determined the overall k_{dock} to be 0.04 s^{-1} (see SI Appendix for detail), which is about three to four-fold slower than the k_{dock} determined using traces with transitions only (k_{dock} = 0.14 s^{-1}). Further evaluation of this model revealed additional complexities as described below, but does not affect the conclusions drawn from the main text.

A simple model is that the double mutant has a mean docking rate of 0.04 s^{-1} or a range of docking rates normally distributed around 0.04 s^{-1}, as the *Tetrahymena* ribozyme is known to have inherent conformational heterogeneity that leads to a broadened distribution of the docking free energies and rate constants.[1, 4] To test this simple two-state model, we simulated single-molecule data with the estimated rate constants, k_{dock} = 0.04 s^{-1} and k_{undock} = 2.8 s^{-1}, and with a broadened distribution of these rate constants with a range of heterogeneity factors (H = 0.5-1.4).[4] H-factor is the standard deviation in the distribution of docking free energy, with a non-zero component from measurement fitting uncertainties[1, 4] (about 0.5 for our system). These simulations yielded a prediction of 28-40% of traces with no transitions, approximately two-fold less than the observed percentage of 78% (Figure S11A). There are also qualitative differences in the distribution of the number of transitions per trace between the simulated and observed data (Figure S11A).

To account for these differences, more complex kinetic models were simulated and constrained by an overall $K_{dock,obs}$ = 0.014 to fit the observed data. The simplest model that had the best fit to the data was a three-state linear model ($U_1 \rightleftharpoons U_2 \rightleftharpoons D$) with a long-lived undocked state U1 (Figure S11B). This model yielded 67-75% of traces with no transitions and qualitative agreement in the distribution of the number of transitions per trace between the simulated and observed data (Figure S11B). For simplicity, the rate constants observed for traces with transitions (k_{dock} = 0.14 s^{-1}, k_{undock} = 2.8 s^{-1}) were used to model the transition between U_2 and F. When the $U_2 \rightleftharpoons D$ rate constants are increased or decrease by more than two-fold, the agreement worsened between the observed and simulated data. This observation suggests that the docking kinetics of $U_2 \rightleftharpoons D$ is well defined by the traces with transitions. This analysis suggests that the data for the P14&MC/MCR double mutant are better accounted for with a 3-state model than a 2-state model. Regardless, this complication does not change the conclusions of this work, as in either model the docking rate constants of the double mutant (U, U1 or U2, to D) are slower than either of the

single mutants. Further, the third state is not observed at 100mM Mg^{2+} and the data at this higher Mg^{2+} concentration lead to the same conclusions.

One potential model for U1, the long-lived undocked state, is a previously identified misfolded state,[8] which was shown to become significant with the elimination of the P5abc region (Figure 3),[9] the region that includes both of the mutations of the P14&MC/MCR double mutant. Further studies are needed to test this model and the origin of this ribozyme's heterogeneous behavior.

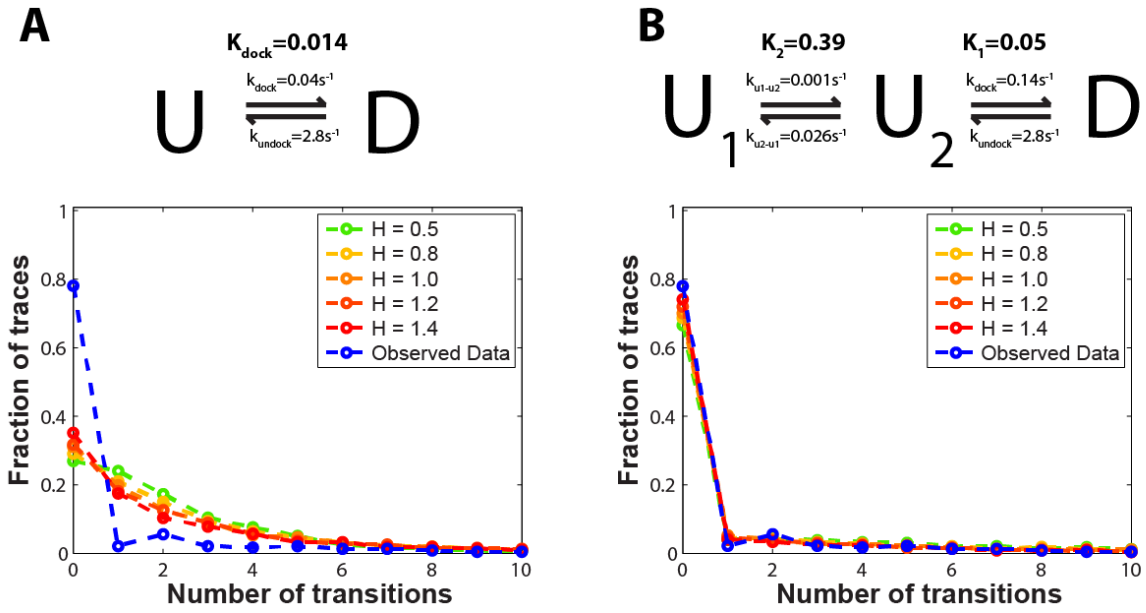

Figure S11: Comparison of simulated versus observed fraction of traces with different number of transitions per trace. Simulations were carried out with a two-state model (**A**) or a three-state model (**B**) with varying heterogeneities (H = 0.5-1.4, green to red).[4] The models and the mean rate constants used are depicted above the plots. See SI text for a more detailed description.

SI Appendix: Algorithm for including smFRET traces that have no transitions

Summary

It can be problematic to infer the transition rates and lifetimes for systems that have dynamics on a timescale comparable to or longer than the typical observation time. In these cases, many traces will exhibit no transitions during an observation period and are thus uninformative about some model parameters (e.g., the lifetimes of the unobserved states) but are informative about other parameters (e.g., the lifetimes of observed states). Excluding such traces while performing model inference can lead to significant bias in the inferred transition rates and state lifetimes, as one would be selecting a systematically biased set of molecules for analysis. An approach to avoid this bias is to use traces in which no transition is observed to correct the average rate and equilibrium constants obtained from the traces in which transitions are observed. Our implementation of this approach works in the high signal to noise ratio (SNR) regime, when transitions are easily resolvable.

Methods

Consider the problem of estimating the parameters of a system with dynamic time scales on the order of or longer than a typical observation time. In particular, say we have a molecule with a "high" and "low" FRET state that transitions between these states with rates $r_{\mathrm{H}\to\mathrm{L}}$ and $rL \to \mathrm{H}$. Equivalently, this dynamic model is determined by the mean lifetimes of the high and low states: t_{H} and t_{L}. It may happen that the lifetime of one of the states is comparable to or longer than the mean observation time (the "trace length") per molecule. In this case, a typical observation sequence ("trace") for this molecule will include at most a few total transition events between the high and low states, and with some non-trivial probability will include no state transitions at all.

For concreteness, let's suppose that the total observation time per trace is T, the low FRET lifetime happens to be twice as large, $t_{\mathrm{L}} = 2T$, and the high FRET lifetime is much smaller than the total observation time, $t_{\mathrm{H}} \ll T$. Now on average 61% of all traces will exhibit no state transitions between the high and low state at all, such that they stay in the low FRET state for the entire observation time, so one may want to discard them as uninformative in estimating the transition rates or lifetimes. However, while an all-low FRET trace is uninformative about the high FRET lifetime t_{H} (and uninformative about the transition rate $r_{\mathrm{H}\to\mathrm{L}}$), it is informative about the low FRET lifetime (and about the transition rate $r_{\mathrm{L}\to\mathrm{H}}$), and its omission would bias the estimation of this transition rate.

Let's compare estimates for the model rates and lifetimes conditioned on including or not including the "all-low" traces. Denote by n_{ij} the number of times a transition from state i to state j was observed among all traces that include at least one low to high or high to low transition. These numbers form a matrix

$$N_{\text{only those traces that include a L to H or H to L transition}} = \begin{pmatrix} n_{\mathrm{LL}} & n_{\mathrm{LH}} \\ n_{\mathrm{HL}} & n_{\mathrm{HH}} \end{pmatrix}$$

corresponding to the inferred transition rate $r_{\mathrm{L}\to\mathrm{H}} = \frac{n_{\mathrm{LH}}}{n_{\mathrm{LH}}+n_{\mathrm{LL}}}$ and lifetime $t_{\mathrm{L}} = \frac{1}{r_{\mathrm{L}\to\mathrm{H}}} = \frac{n_{\mathrm{LH}}+n_{\mathrm{LL}}}{n_{\mathrm{LH}}}$. If we now include the "all-low" traces, we must add to the above matrix a term that counts the time spent in the low FRET state:

$$N_{\text{all traces}} = \begin{pmatrix} n_{\mathrm{LL}} & n_{\mathrm{LH}} \\ n_{\mathrm{HL}} & n_{\mathrm{HH}} \end{pmatrix} + \begin{pmatrix} n'_{\mathrm{LL}} & 0 \\ 0 & 0 \end{pmatrix}$$

where n'_{LL} denotes the total length of all the "all-low" FRET state traces. For a single "all-low" trace $n'_{\mathrm{LL}} = (\text{length}) - 1$, since we do not know what state preceded the first time point in a trace. For multiple "all-low" traces $n'_{\mathrm{LL}} = (\text{total length of traces}) - (\#\ \text{traces})$. The corresponding inferred transition rate and lifetime are now, respectively, $r'_{\mathrm{L}\to\mathrm{H}} = \frac{n_{\mathrm{LH}}}{n_{\mathrm{LH}}+n_{\mathrm{LL}}+n'_{\mathrm{LL}}}$, $t'_{\mathrm{L}} = \frac{1}{r'_{\mathrm{L}\to\mathrm{H}}} = \frac{n_{\mathrm{LH}}+n_{\mathrm{LL}}+n'_{\mathrm{LL}}}{n_{\mathrm{LH}}}$.

For the example given above, we have $n'_{\mathrm{LL}} \approx \frac{0.61}{0.39} n_{\mathrm{LL}} \approx 1.5 n_{\mathrm{LL}}$, so there are 1.5 times more low FRET data points that are excluded than are included when we discard the "all-low" FRET traces. We find that when we do not include the "all-low" traces, we underestimate the low FRET lifetime by a factor of about $(n_{LL}+n'_{LL})/n_{LL} \approx (1+1.5) / 1 = 2.5$. Thus, excluding traces that do not include every possible state transition can lead to significant bias when one of the state lifetimes is comparable to the observation. Table S6 below summarizes the inferred rate constants obtained by including or excluding the "all-low" FRET traces for the given parameter values and in symbolic form.

Table S6. The inferred rate constants obtained by including or excluding the "all-low" FRET traces

Fraction of traces that exhibit no transitions	61%	f
Inferred low FRET lifetime, excluding "all-low" traces	1	t_{L}
Inferred low FRET lifetime, including "all-low" traces	≈2.5	$t_{\mathrm{L}}\left(1+\frac{f}{1-f}\cdot\frac{n_{\mathrm{LL}}}{n_{\mathrm{LL}}+n_{\mathrm{LH}}}\right)\approx t_{\mathrm{L}}\cdot\frac{1}{1-f}$

Inferred low FRET state lifetimes excluding (middle row) and including (bottom row) the transitionless "all-low" traces for numeric parameters (middle column) and symbolic parameters (rightmost column). The symbols in the equation are defined [above] and the approximation sign indicates the rare transitions regime, $n_{\mathrm{LL}} \gg n_{\mathrm{LH}}$.

The analysis above relies on knowing the counts n_{ij} for each type of transition. In practice, the "with transitions" traces can be treated automatically by hidden Markov model-fitting software (e.g., SMART) to extract the estimated counts n_{LL}, n_{HH}, n_{LH} and $n_{HL,}$ whereas the "all-low" traces in principle directly provides n'_{LL}. With respect to real data used herein, only the P14&MC/MCR docking data obtained at 10 mM $MgCl_2$ had a docking rate constant slow enough to have an appreciable number of "all-low" traces, and required the consideration of n'_{LL} in the estimate of the docking rate constant. In practice, for this mutant molecule we observed that 22% of 390 traces have transitions, while 78% (303) have no transitions. For traces with transitions, 60% had a SNR > 2 and were treated by SMART and gave an estimated docking rate constant of 0.14 s^{-1}. For the "all-low" traces, we included the same fraction of molecules (182, corresponding to 60% of the 303 molecules with no transitions) in our analysis and we used this to calculate $n'_{LL.}$ The inclusion of n'_{LL} to the total low FRET lifetime, following the method described above, decreased the estimate of the docking rate constant by 3.5 fold to 0.04 s^{-1} (also see Table S2).

This method was tested and validated using simulated data in which rate constants were chosen such that a large fraction of all traces contained no transitions (k_{dock} = 0.010 s^{-1}, k_{undock} = 3.6 s^{-1}, N = 200). Using traces that were visually identified to contain transitions to estimate the docking rate constant lead to a 2.6 fold over-estimate. Including all traces, the docking rate constant was estimated to be 0.007 s^{-1}, within 40% of the actual rate constant.

In a typical smFRET experiment, in addition to fluorescent spots whose traces represent actual single molecules, there are intensities that correspond to fluorescent impurities, damaged molecules that only contain a single dye, and multiple molecules whose intensities that overlap. In practice, fluorescent spots

are scored as intact molecules if the trace exhibits a single-step photobleaching event in which both dyes change intensity or contain transitions in which the dye intensities exhibit anti-correlation in FRET.

For the experiments herein, the percent of fluorescent spots that corresponded to single, intact ribozyme molecules by the above criteria, was ~19% and ~23% for the P14&MC/MCR mutant at 10 and 100 mM $MgCl_2$, respectively, even though the lower $MgCl_2$ condition had many more traces with no transitions compared to the other. The similar number of estimated total molecules suggests that we are able to identify the subset of fluorescent spots that are molecules even when most molecules do not exhibit transitions.

REFERENCES

1. Solomatin, S. V.; Greenfeld, M.; Chu, S.; Herschlag, D. *Nature* **2010**, *463*, 681.
2. Shi, X.; Solomatin, S. V.; Herschlag, D. *J. Am. Chem. Soc.* **2012**, *134*, 1910.
3. Greenfeld, M.; Pavlichin, D. S.; Mabuchi, H.; Herschlag, D. *Plos One* **2012**, *7*, e30024.
4. Solomatin, S. V.; Greenfeld, M.; Herschlag, D. *Nat. Struct. Mol. Biol.* **2011**, *18*, 732.
5. Benz-Moy, T. L.; Herschlag, D. *Biochemistry* **2011**, *50*, 8733.
6. Herschlag, D.; Eckstein, F.; Cech, T. R. *Biochemistry* **1993**, *32*, 8299.
7. Bartley, L. E.; Zhuang, X.; Das, R.; Chu, S.; Herschlag, D. *J. Mol. Biol.* **2003**, *328*, 1011.
8. Russell, R.; Das, R.; Suh, H.; Traver, K. J.; Laederach, A.; Engelhardt, M. A.; Herschlag, D. *J. Mol. Biol.* **2006**, *363*, 531.
9. Johnson, T. H.; Tijerina, P.; Chadee, A. B.; Herschlag, D.; Russell, R. *P Natl Acad Sci USA* **2005**, *102*, 10176.