nature portfolio

https://doi.org/10.1038/s41586-021-04225-4

Supplementary information

Time-resolved structural analysis of an RNA-cleaving DNA catalyst

In the format provided by the authors and unedited

Supplementary Figures

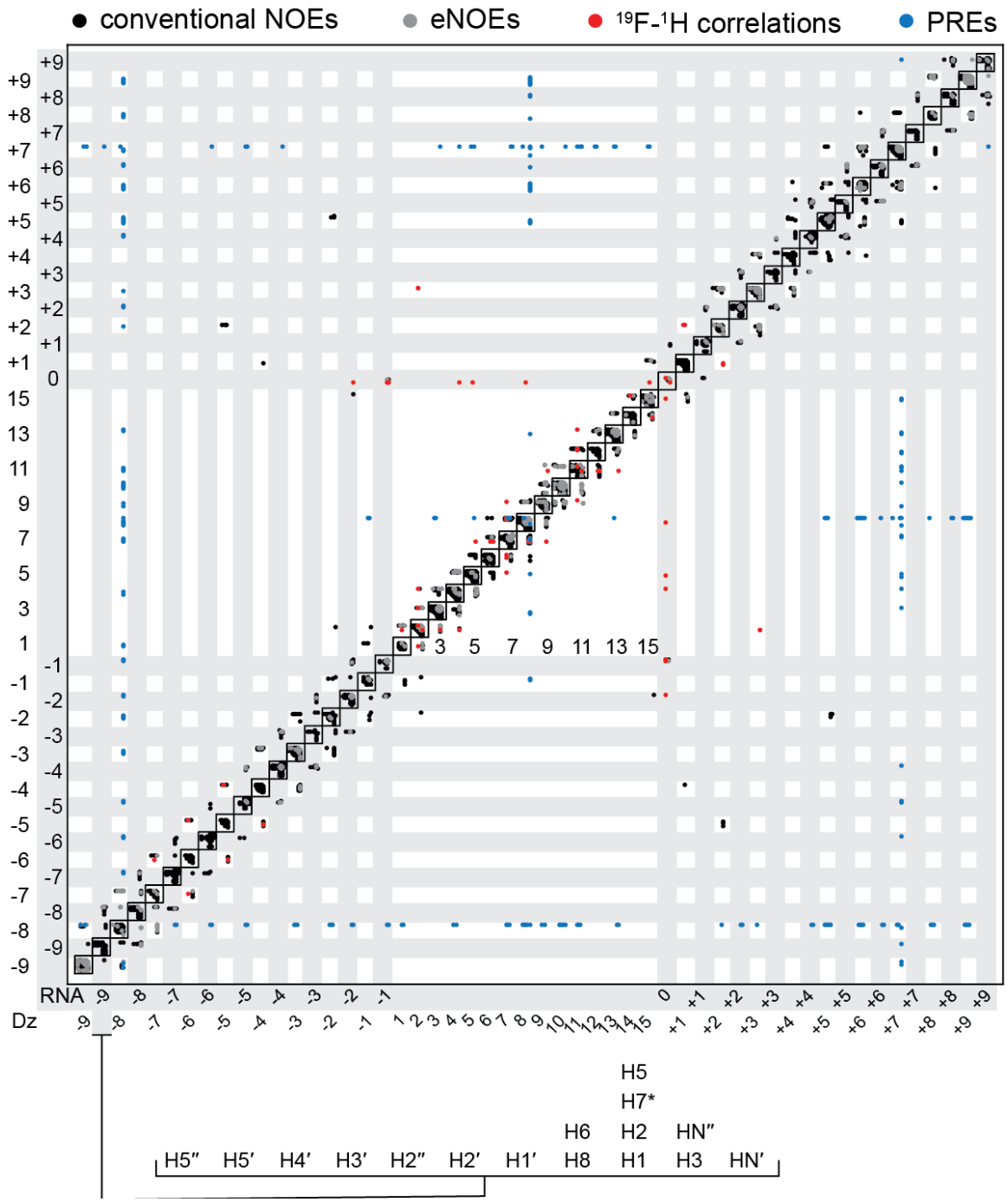

Fig. S1. Atom-resolved contact plot. All intra- and inter-nucleotide distance restraints obtained via, regular/exact NOE (black/grey, respectively), PREs (blue), and ^{19}F-STD (red) experiments are shown. The order of protons within each nucleotide is shown below.

Fig. S2. Stereoview of the structure of precatalytic Dz^{5C}–$RNA^{2'F}$ complex shown in Fig. 1g. Dz is shown in dark red, RNA is shown in grey. The phosphate of the scissile bond is shown as a golden sphere. Base-paired nucleotides are shown as straight sticks, unpaired nucleobases are shown as rings.

Supplementary Tables

Table S1: NMR and refinement statistics for nucleic acids

NMR distance and dihedral constraints	
Distance restraints	
Total NOE	
Conventional NOE	1008
Intra-residue	691
Inter-residue	317
Sequential ($\|i-j\|=1$)	274
Nonsequential ($\|i-j\|>1$)	43
Exact NOE	108
Intra-residue	68
Inter-residue	40
Hydrogen bonds	143
Paramagnetic Relaxation Enhancement	126
^{19}F-derived distances	44
Sequential ($\|i-j\|=1$)	21
Nonsequential ($\|i-j\|>1$)	23
Total dihedral angle restraints	208
Base pair	43
Sugar pucker	60
Backbone	105
Based on A-form geometry	105

Structure statistics	**Dz^{5C}–RNA$^{2'F}$ before refinement (20 structures)**	**Cluster I i.e., 7PDU.pdb (7 structures)**	**Cluster I after refinement (10 structures)**
Violations (mean and s.d.)			
Distance constraints (Å)	0.27 (0.13)	0.26 (0.13)	0.19 (0.01)
Dihedral angle constraints (º)	0.76 (0.06)	0.70 (0.03)	0.88 (0.07)
Max. dihedral angle violation (º)	17.56	8.51	13.94
Max. distance constraint violation (Å)	6.34	5.25	2.36
Deviations from idealized geometry			
Bond lengths (Å)	0.37	0.35	0.46
Bond angles (º)	1.10	0.99	1.19
Average pairwise r.m.s. deviation** (Å)			
All nucleotides	8.90	5.37	3.50
Binding arms	8.10	2.59	3.33
Loop region	12.84	6.18	4.64

**Pairwise r.m.s. deviation was calculated among indicated structures.

Table S2: Dihedral angular restraints for Dz^{5C}–$RNA^{2'F}$

	Involved atoms	Restraint	Conformation	Restrained nucleotides
Backbone dihedrals				
α	O3′-P-O5′-C5′	-60° ±40°	A from	
β	P-O5′-C5′-C4′	-180° ±50°	A from	DNA: -9 to -5 / +5 to +9
γ	O5′-C5′-O4′-C3′	60° ±30°	A from	RNA: -9 to -5 / +5 to +9
δ	O3′-P-O5′-C5′	-60° ±40°	A from	
ε	C4′-C3′-O3′-P	-160° ±50°	A from	
ζ	C3′-O3′-P-O5′	-70° ±50°	A from	
Sugar puckering				
ν_1	O4′-C1′-C2′-C3′	-20° ±10°	3′-*endo*	DNA: -9 to -5 / +5 to +9
ν_2	C1′-C2′-C3′-C4′	35° ±5°	3′-*endo*	RNA: -9 to -5 / +5 to +9
ν_3/ δ	C5′-C4′-C3′-O3′	80° ±20°	3′-*endo*	
ν_1	O4′-C1′-C2′-C3′	35° ±5°	2′-*endo*	
ν_2	C1′-C2′-C3′-C4′	-35° ±5°	2′-*endo*	No restraint
ν_3/ δ	C5′-C4′-C3′-O3′	145° ±20°	2′-*endo*	
Nucleobase orientation				
χ	O4′-C1′-N1-C2 O4′-C1′-N9-C4	180° ±90°	*anti*	DNA: -9 to 3 / 6,7,9 / 10 to +9 RNA: -9 to -3 / +3 to +9
χ	O4′-C1′-N1-C2 O4′-C1′-N9-C4	0° ±90°	*syn*	DNA: 4,5,8,11 RNA: -2 to 0

Supplementary Discussion

Details on EPR data fitting

The EPR binding isotherms were fitted with a non-cooperative and two cooperative models. The non-cooperative model (eq. 1) did not fit the data. A cooperative binding model (eq. 2) with two binding sites and one K_D of 362 ± 83 µM fits the data in the range $[Mn^{2+}_{free}]$ = 0 – 0.5 mM but not beyond this concentration range. The best fit of the data and covering the whole concentration range was obtained assuming cooperative binding and three different K_Ds (eq. 3). According to this model, the lowest K_D of 173 ± 109 µM is associated with two binding sites. One can therefore calculate from this macroscopic K_D a microscopic K_D for the two equivalent sites of K_D(macroscopic)/2 = K_D(microscopic) = 87 µM [26]. The other macroscopic K_Ds are all in the mM or even M range and are thus interpreted as being associated with unspecific binding.

$$\frac{[\mathrm{Mn^{2+}}]_{\mathrm{bound}}}{[\mathrm{oligo}]} = \frac{n[\mathrm{Mn^{2+}}]_{\mathrm{free,\ oligo}}}{K_\mathrm{D} + [\mathrm{Mn^{2+}}]_{\mathrm{free,\ oligo}}} \quad (1)$$

$$\frac{[\mathrm{Mn^{2+}}]_{\mathrm{bound}}}{[\mathrm{oligo}]} = \left(\frac{[\mathrm{Mn^{2+}}]_{\mathrm{free,\ oligo}}}{K_\mathrm{D} + [\mathrm{Mn^{2+}}]_{\mathrm{free,\ oligo}}}\right)^n \quad (2)$$

$$\frac{[\mathrm{Mn^{2+}}]_{\mathrm{bound}}}{[\mathrm{oligo}]} = \sum_{i=1}^{i=3}\left(\frac{[\mathrm{Mn^{2+}}]_{\mathrm{free,\ oligo}}}{K_{\mathrm{D}i} + [\mathrm{Mn^{2+}}]_{\mathrm{free,\ oligo}}}\right)^{n_i} \quad (3)$$

Features of single-stranded Dz's and post-catalytic Dz–RNA complexes

Minimalistic fold of free Dz

NMR data analysis reveals that while for the Dz–RNA complex all resolved nucleotides of the Dz (with the exception of flipped out +1 position, *vide infra*) show clear cross peaks indicative of π-stacking to their neighboring nucleotides (Extended Data Fig. 9d, black), a considerably smaller fraction shows these contacts already in the free Dz (Extended Data Fig. 9d red). Interestingly, while in the free Dz only about 30% of the nucleotides of the arm region show cross peaks (6 out of 18), 60% of the nucleotides in the loop region (9 out of 15) show sequential cross peaks, suggesting that the 10-23 Dz's catalytic loop may have already some defined regions in the absence of RNA substrate. In particular, the nucleotides from 7–13 in the loop appear to have similar NOE patterns. Interestingly this region also shows relatively small chemical shift perturbations upon complex formation with its substrate (in particular C10 and C13, Extended Data Fig. 9b). Overall, the data point to a minimally structured central region of the catalytic loop that displays transiently and locally restricted features found for the same region after complex formation.

Post-catalytic states

The spectrum of the post-catalytic state reveals the occurrence of different populations in a stable equilibrium that is dominated by peaks characteristic for a still complexed state, as visible by nearly identical peak positions for several nucleotides in the binding arms. The occurrence of multiple peaks for several nucleotides is linked to the presence of multiple conformations.
At least 2-3 states with populations above 5% and peak positions indicative of complexed conformations can be detected. The relative populations of these post-catalytic-complexed states are strongly dependent on temperature (Extended Data Fig. 9n), revealing states that likely reflect more ordered configurations and states that are entropy-driven. In general, the relatively high concentration of Dz and RNA (200 µM) should increase the dissociation temperature of the two resulting RNA fragments from the post-catalytic complex. Under the applied conditions, it can be predicted that about 90% of the Dz molecules should still be in a complexed configuration with the RNA after cleavage (Extended Data Fig. 9q), which is in line with the occurrence of weak peaks representative of single-stranded RNA. Interestingly, for most affected nucleotides, the observed peak multiplicity affects the H6 proton frequency considerably stronger than the H5 frequency. Since, unlike the H5 proton, the H6 position is directly flanking the sugar-nucleobase bond (χ-angle), such behavior could be in line with an exchange of the *syn* and *anti*-configurations of the respective base.

Features of temperature-dependent NMR data

To investigate possible interactions in the different regions of the Dz^{5C}–$RNA^{2'F}$ complex, we followed the temperature-dependent variation of NMR signal intensities with nucleotide-specific resolution. This approach exploits the denaturing behavior of nucleic-acid polymers that can oppose the signal-intensity gain expected from increasing temperatures. As a reference, we measured the temperature profile for free Dz (Extended Data Fig. 3b. The observed continuous intensity increase is in line with a constantly increasing molecular tumbling of a single-stranded DNA that is either unstructured or does not change its structure within the measured temperature range. On the contrary, clearly different temperature profiles could be detected for (the same) nucleotides in the Dz^{5C}–$RNA^{2'F}$ complex (Extended Data Fig. 3c,d. Deviation from the constantly increasing temperature profile (dashed lines in Extended Data Fig. 3b-d) can directly report on the presence of restrained

conformations. The nucleotides of the binding arms with known Watson-Crick base pairing can serve as positive controls of the approach and indeed show strongly reduced signal intensities around the expected duplex melting temperature.

These data are in line with conformational exchange processes in the transition temperature range (e.g., association and dissociation into single-stranded RNA/DNA) that reduce the intensity of the NMR detectable signal albeit overall molecular tumbling should increase. Since the occurrence of such exchange processes requires the presence of alternating conformations, only one (if any) of the involved states can be fully disordered. Interestingly, the different regions of the Dz–RNA complex show different temperature profiles, suggesting that the level of restraining also differs considerably within the structure of the pre-catalytic state. In this respect, in particular the nucleotides surrounding the cleavage site closely follow the behavior of the binding arms indicating a well-restrained conformation in this region. While smaller deviations are also detected for nucleotides in the first half of the catalytic loop (including positions 5 and 7), no deviations from the single-stranded DNA are seen for the second half of the catalytic loop (including positions 8 and 13). Interestingly, when comparing our data to previously reported full mutation screens of the Dz[5A]'s loop region, a remarkable overlap of the less restrained nucleotides (positions 8–13) and a stretch of subsequent nucleotides whose mutations do not or only weakly affect Dz activity (positions 7–12) is apparent[3,16]. This strongly suggests that this part of the loop is amenable to conformational plasticity, which is not directly involved in the catalysis mechanism.

Details of time-resolved NMR analysis

To allow a more detailed peak shape analysis of the 2D real-time NMR data, the peak of interest needs to be sufficiently well separated from other signals in each relevant state occurring in the monitored transition. All assigned peaks matching this criterium were manually identified and selected for further analysis. Respective rectangular spectral windows (spectral extracts) were chosen individually to minimize contributions from unwanted signals while capturing all contributions from the peak transition (Extended Data Fig. 10g, shows one example for T8 H6-H7 cross correlations). Due to asymmetry of the transition, cross peaks involving identical spins but occurring on different sites of the diagonal were treated individually. In total 15 cross peaks belonging to 10 different nucleotides were selected for initial analysis. Linewidths in direct (λ_2) and indirect dimension (λ_1) of the initial (B_2) and final state(s) (C) were measured using reference spectra of the respective stable states recorded under identical conditions. The experimentally determined linewidths were used to simulate the respective time-resolved 2D spectral extract individually for each analyzed peak. Identical window functions as used in the processing of experimental data were also included in the simulation of the data (note that comparable results were also obtained without usage of window functions, however spectral crowding and processing artifacts were considerably more severe).
To simulate the time-dependent changes of the signal, population profiles (P(t)) representing either a 2-state or 3-state transition acting on the NMR signal (S(t)) were included in time-space before Fourier transformation into frequency space. Due to the time course of the transition and the experimental setup, time-dependent changes of the transition were only included in the indirect dimension and the direct dimension was simulated in frequency space. Initially, a Lorentzian peak shape was compared to a Gaussian peak shape to represent the direct dimension. Usage of Gaussian peak shapes ($G(\omega,\lambda)$) led to a slightly increased agreement to experimental data and was therefore used for all subsequent simulations. Overall, the simulated spectrum (representing the real part of the experimental spectrum) had the form:

$$S(t_1,\omega_2) = \sum_n P_n(k_i,t) \cdot \mathrm{e}^{\left(2\pi i\omega_{1,n}-\lambda_{1,n}\right)t} \cdot G\left(\omega_{n,2},\lambda_{n,2}\right) \quad \text{(I)}$$

For each considered cross correlation three different conformations (n) with different peak positions ($\omega_{1,2}$) and linewidth ($\lambda_{1,2}$) representing the states B_2, C_1 and C_2 in Fig. 3a were assumed. The population profile of the simulated transition (2-state model) had the form:

$$P_{B_2} = B_2 \cdot \mathrm{e}^{-k_1 t} \quad \text{(IIa)}$$

$$P_{C_1} = (1-C_2)(1-\mathrm{e}^{-k_1 t}) \quad \text{(IIb)}$$

$$P_{C_2} = C_2(1-\mathrm{e}^{-k_1 t}) \quad \text{(IIc)}$$

When considering 3-state transitions the population profiles need to be modified according to:

$$P_{B_2} = B_2 \cdot \mathrm{e}^{-k_1 t} \quad \text{(IIIa)}$$

$$P_{C_1} = (1-C_2)(1-\mathrm{e}^{-k_1 t}) \cdot (1-\mathrm{e}^{-k_2 t}) \quad \text{(IIIb)}$$

$$P_{C_2} = C_2(1-\mathrm{e}^{-k_1 t}) \cdot (1-\mathrm{e}^{-k_2 t}) \quad \text{(IIIc)}$$

The simulation itself was carried out using the Fast Fourier Transformation (FFT) algorithm implemented in Matlab (Matlab R2018b, MathWorks) starting with Eq. I and the respective population profiles. To account for the occurrence of multiple *C*-states with variable intensity ratios, C_2 (in Eqs. IIb,c and IIIb,c) was included as another variable parameter, in addition to k, via nested loops in the Matlab script. Note that this approach also (partly) compensates for possible changes in NMR intensities of the respective states due to different relaxation properties.

For each set of parameters (C_2, k_1 and k_2 where applicable), the simulated spectral extract was subtracted from the experimental data after normalizing both data sets to their maximal intensity. The parameter set for which this difference reaches its minimum was selected as the best fit condition. To assess the reliability of the best fit condition, the plot of the difference over the respective parameter range was evaluated. Only conditions resulting in a clear global minimum and displaying considerable changes when varying the respective parameter were classified as reliable (in total 13 peaks belonging to 9 different nucleotides were classified as reliable; see Extended Data Fig. 10j for example of reliable and not reliable parameters). Not reliable peaks were excluded from data interpretation and are not included in the data presented in the main text.

Plausibility considerations of 2-state vs. 3-state transition and NMR-detectable states

In general, the 2-state model assumes a transition from state B_2 simultaneously to state C_1 and C_2, requiring a single rate constant k. Using the respective population profiles (Eqs. II) the simulations can explain the most dominant experimentally detected peak shape modulations. The observation that using this approach the nucleotides in close proximity to the cleavage site show faster rates of transition (Extended Data Fig. 10h) suggests that substrate cleavage is indeed the first detectable change in the system. While this is not unexpected, it is difficult to conceive a plausible mechanism in which the slower rates can be explained in a 2-state model, which would require the slower rates to be independent of the faster rates but be initiated simultaneously. Knowledge of the system points to a picture where the substrate is cleaved and, subsequently, other regions are affected. This, however, requires a 3-state model with two independent but subsequently acting rate constants for the respective transitions. In this picture, only the cleavage site is explicitly affected by the first transition (i.e., cleavage), and the rate constant for this transition (k_1) should match the fast rate determined in the 2-state model. All other nucleotides in the intermediate state B_0 should not experience NMR-detectable changes in peak positions. Consequently, for simulation of the spectrum of the initial state for these nucleotides, Eq. IIIa needs to be adapted to:

$$P_{B_2+C_0} = P_{B_2} + B_2(1 - \mathrm{e}^{-k_1 t}) \cdot (\mathrm{e}^{-k_2 t}) \qquad \text{(IV)}$$

Interestingly, usage of this NMR-adapted 3-state population profile (Eq. IV and Eqs. IIIb,c) leads to a small but noticeable improvement in simulating the experimental data for nucleotides that display slow rates in the 2-state model (Extended Data Fig. 10i).

While the obtained data cannot exclude more complex transitions, the outlined plausibility considerations and the obtained fitting improvements suggest that the discussed 3-state transition model is sufficient to reproduce the key features encoded in the experimental data adequately.

The population profile that reflects the best fit according to this model (using the arithmetic mean of all resulting k_1 and k_2 values for the respective resolved peaks) is shown in Fig. 3e.

Interpretation of result for rational-design strategies of next-generation DNAzymes

The origin of the Dz's limited in vivo activity is likely the limited availability of free intracellular divalent metal ions[2,10]. Our results identified three important roles of metal ions in the function of the Dz. These roles are: scaffolding of the core region, activation of the loop region, and mediation of catalysis (as explained in more detail in the manuscript, Fig. 2a,b. It can be anticipated that all three roles can be exploited for rational design strategies that could involve:

(i) reducing the electrostatic repulsion of the DNA backbones in metal-ion binding site I to minimize the need for metal ions in the scaffolding step
(ii) stabilizing the loop in an active conformation via modification of the nucleotides in metal-ion binding site II to either increase binding affinities for divalent metal ions and/or covalently attach suitable metal-ion mimetics and/or stabilize favorable arrangements of nucleotides, e.g., by promoting a flip out of T4
(iii) introducing functional groups at the catalytically relevant positions (i.e., C13, G14) to increase the occupancy of cleavage-competent conformations.

Besides natural nucleotides, chimeric DNA and Peptide Nucleic acids (PNA) and/or Phosphorodiamidate Morpholino Oligomer (PMO) binding arm regions may help to realize point (i). Non-natural nucleobases may be key to achieve (ii) and (iii).

Note that both reducing the need for divalent metal ions and/or increasing the affinity to outcompete other intracellular metal-ion binders may lead to the desired result. In addition, established methods to increase stability and hybridization kinetics, as for example recently presented[12], should be included in future DNAzyme designs. In our manuscript, we provide an initial example of targeting the above-mentioned point (iii), and our structural insights identify and considerably narrow down the hot-spot regions for points I–III. Nevertheless, a sizable number of promising modifications and their combinations need to be evaluated in suitable in vitro and in vivo systems. Overall, we aim to stimulate the iterative development of next-generation DNAzymes via combined rational and systematic approaches to obtain DNAzyme variants with optimal in vivo activity.