Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Supporting information for: *Who stole the proton?* Suspect general base guanine found with a smoking gun in the pistol ribozyme

Şölen Ekesan* and Darrin M. York*

Laboratory for Biomolecular Simulation Research, Institute for Quantitative Biomedicine and Department of Chemistry and Chemical Biology, Rutgers University, Piscataway, NJ 08854, USA

E-mail: solen.ekesan@rutgers.edu; Darrin.York@rutgers.edu

*To whom correspondence should be addressed

Supporting Information

Herein we describe details of the knowledge-based approach through use of experimentally motivated distance and angle restraints to guide local rearrangements followed by sequential molecular dynamics refinement. In this way, we have used restraints as a way to guide the system to form the interactions of the L-platform/L-scaffold framework[S1] while at the same time ensuring certain catalytic fitness requirements[S2] and experimentally inferred functional roles (see Background of main text for discussion) are supported. Details for the series of restrained simulations are presented in the table below. As illustrated in the figure below, the resulting refolded structure resides in a minimum free energy basin where it stably fluctuates in *unconstrained* molecular dynamics simulations (5 independent simulations for 100 ns, and one extended to 200 ns).

Active site rearrangement simulation details

Model constructions based on L-Platform/L-Scaffold (L-P/S) motif were computationally carried out by introducing distance and angle restraints to guide the system to rearrange and proceed to maintain interactions experimentally established to be essential. Inline angle (nucleophile - P - leaving group), due its periodic nature was restrained with two bounds. The rest of the interactions for hydrogen bonding, base pairing, and base stacking were all maintained using distance restraints with half harmonic potentials. Usage of half harmonics (i.e. single bound restraint) provides flexibility inside the restraint limit where the penalty is zero. This enables the system to find new favorable basins, if such exist, and adapt to and stabilize the guided interaction when coupled with longer sampling.

Extra care needs to be given to application of stacking restraints as there are many placements between the two stacked nucleobases that would be favorable for the system, and while two or three distances orthogonal to base planes are used at a time to guide stacking, which distances will remain orthogonal once the system relaxes cannot be predicted. Therefore, stacking restraints need to be relaxed and redefined as the structure evolves.

If the system cannot support the introduced interactions, 1) it will continuously pull on the restraint limit throughout the trajectory, and 2) as soon as the restraint is removed it will snap back to a more favorable conformation. For this reason through the simulations the restrained distances are monitored, and upon completion restraints are relaxed/removed to test whether the system has adapted.

In all simulations the inline angle, general base - nucleophile hydrogen bonding, and Mg^{2+} direct coordination to G33:N7 are restrained to enable the system to find a conformation that supports what would become the active state in solution, and the "catalytic fitness" is not lost while local rearrangements are taking place. The list of simulations that yielded the final results reported in this study are summarized below with group headings explaining what structural changes have occurred. Each step uses the final structure of the previous step as the starting structure. First step uses the final structure of the relaxed and equilibrated system in solution. The six restrained simulations differ by the set of restraints used, and the list of their corresponding restraints are summarized in Table S1.

- **Rearrangement to form L-P/S**. Stacking and base pairing restraints are used to place G42 in the general base position: stacking between C41 and G-1, and base pairing with A32.
  Step 1: Restrained simulation 1, 50 ns.
  Step 2: Restrained simulation 2, 50 ns.
  Step 3: Free simulation, 100 ns.

- **Position G40 to stack between A39 and C41**. Stacking restraints are used. On top of other stacking restraints, here it was necessary to push G40 away from G42.
  Step 4: Restrained simulation 3, 100 ns.
  Step 5: Free simulation, 100 ns.

- **Pair G40 with C22**. Stacking and base pairing restraints are used to place C22 to pair with G40 and stack between U6 and G33. As this required a large shift for C22, the first 5 ps of the simulation used 1 kcal/mol$Å^2$ as force constant to prevent the system from blowing up. The rest of the simulations used 20 kcal/mol$Å^2$ same as all the other distance restraints.
  Step 6: Restrained simulation 4, 100 ns
  Step 7: Free simulation, 100 ns.

- **A-minor reorganization**. Stacking and hydrogen bonding restraints are used to reorganize the A-minor motif, that was strained upon C22 rearrangement
  Step 8: Restrained simulation 5, 100 ns.
  Step 9: Restrained simulation 6, 100 ns. This simulation removes the A-minor specific restraints while keeping others to let the A-minor motif adjust itself to the new environment.

- **Unrestrained independent simulations**
  Step 10: 5 independent free simulations (4x100 ns and 1x200 ns)

Table S1: List of restraints used in refolding simulations

Region/Purpose	Residues	Atoms	Criteria	Simulations					
				1	**2**	**3**	**4**	**5**	**6**
Core	C41 - G33	N3 - H1	< 1.9	x	x	x	x	x	x
		H41 - O6	< 1.9	x	x	x	x	x	x
	MG101 - G33	MG - N7	< 2.3	x	x	x	x	x	x
	G61 - U62	O2' - P - O5'	$140 < < 170$	x	x	x	x	x	x
L-P/S bp & H-bond	G42 - G61	N1 - HO2'	< 1.9	x	x	x	x	x	x
		N1 - O2'	< 3.2	x	x	x	x	x	x
	G42 - A32	O2' - H61	< 1.9	x		x	x		x
		N3 - H62	< 1.9	x	x	x	x	x	x
		H22 - N7	< 1.9	x	x	x	x	x	x
	G61 - A31	H22 - N1	< 1.9					x	x
		N3 - H61	< 1.9					x	x
L-P/S stack	G42 - C41	N1 - N4	< 4.2		x	x	x	x	x
		N7 - N1	< 4.2		x				
		O6 - C4	< 4.2			x	x	x	
		N9 - O2	< 4.2			x	x	x	
		N9 - N1	< 4.2						x
	G33 - A32	N9 - N9	< 4.2						x
		N1 - N6	< 4.2						
	A32 - A31	N9 - N9	< 4.2					x	
	G61 - G42	N9 - N2	< 4.2	x					
		N9 - N1	< 4.2		x				
		N7 - O6	< 4.5						x
G40 placement	C41 - G40	O2 - N3	< 4.2			x	x	x	
		N4 - O6	< 4.2			x	x	x	x
		N1 - N3	< 4.2						x
	G42 - G40	O6 - N2	$6 <$			x	x		
		O6 - N9	$6 <$			x	x		
		O6 - N3	$6 <$			x	x		
C23 bp & stack	G40 - C23	O6 - H41[a]	< 1.9				x	x	x
		H1 - N3[a]	< 1.9				x	x	x
		H21 - O2[a]	< 1.9				x	x	x
	G33 - C23	N9 - N1	< 4.2					x	
	A39 - U7	N1 - H3	< 1.9				x	x	x
		H61 - O4	< 1.9				x	x	x
A-minor	C4 - A20	O2' - N7	< 3.5					x	
		O2 - N6	< 3.5					x	
	C16 - A20	O2 - N6	< 3.5					x	
	C5 - A21	O2' - N7	< 3.5					x	
		O2 - N6	< 3.5					x	
	G15 - A21	N3 - N6	< 3.5					x	
	G6 - A22	O2' - O2'	< 3.2					x	
	G24 - A14	N6 - O2'	< 3.5					x	
	A22 - A24	O2' - O4'	< 3.5					x	
		N9 - N9	< 4.2					x	
		N6 - N6	< 4.2					x	
Other	G65 - C26	O6 - N4	< 3.5					x	
		N2 - O2	< 3.5					x	
	U62 - U29	O4 - H3	< 1.9					x	x
		H3 - O2	< 1.9					x	x
Total number of restraints used in simulation				10	11	17	22	37	24

Restraint lists are given in accord with the atom name and residue numbering in PDB 6r47, where numbers for residues upto 25 are +1 of their respective cannonical numbering and residues 61 and 62 correspond to G-1 and U+1 respectively. Distance (Å) restraints are half harmonics with 20 kcal/molÅ^2 force constants. Inline angle (degrees), is restrained with double bounds each with 50 kcal/molÅ^2 force constants. List of abbreviations: L-P/S (L-platform/L-Scaffold); bp (base pair); H-bond (hydrogen bonding). [a] Weights for these restraints were introduced as 1 kcal/molÅ^2 for the first 5 ps to ease the large conformational shift, then raised to the 20 kcal/molÅ^2 as other distance restraints.

Stability tests of the refolded structure with unrestrained molecular dynamics

The rearrangement path derived from these restraints described above is not meant to reflect the physical path the ribozyme goes through in rearranging itself, and as such, is in no way a commentary of the rearrangement mechanism. The purpose of this study is to show that a model active site configuration based on L-P/S is physically possible (i.e. achieved without breaking/making bonds) and stable (i.e. it resides in a local free energy basin with significant lifetime and does not snap back to a different configuration). The five independent unconstrained free simulations all maintained the interactions reported for the Model L-P/S construct after 100 ns (or 200 ns for sim 1), with the systems reaching equilibrium in 50ns as shown in Figure S1.

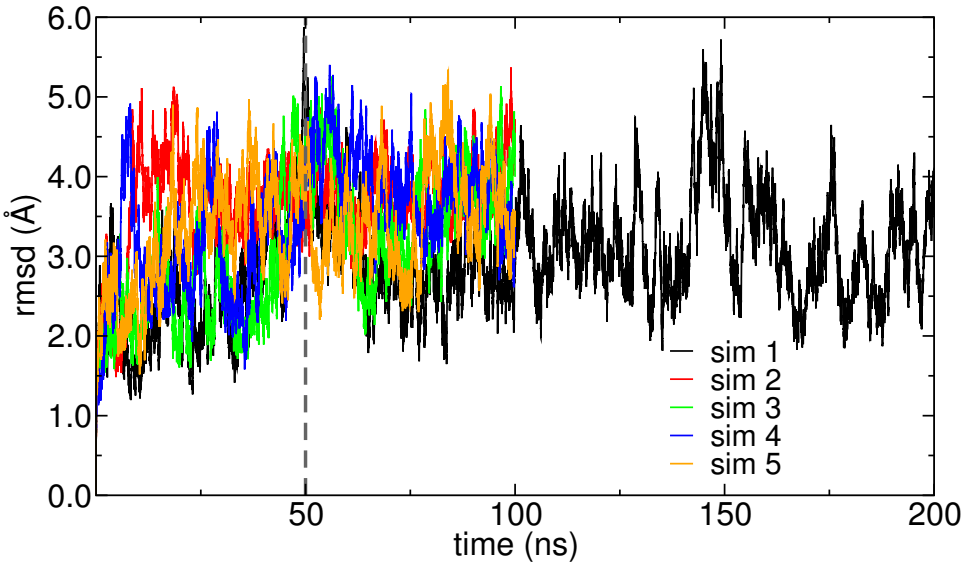

Figure S1: Root mean square deviations (rmsd) with respect to the starting structure from the unrestrained independent simulations of the pistol ribozyme Model L-P/S construct. Five independent simulations were carried out with the first one extended to 200 ns. All five reach equilibrium around 50 ns, and no structural drift is observed even after 200 ns.

References

(S1) Gaines, C. S.; Piccirilli, J. A.; York, D. M. The L-platform/L-scaffold framework: a blueprint for RNA-cleaving nucleic acid enzyme design. *RNA* **2020**, *26*, 111–125.

(S2) Bevilacqua, P. C.; Harris, M. E.; Piccirilli, J. A.; Gaines, C.; Ganguly, A.; Kostenbader, K.; Ekesan, Ş.; York, D. M. An Ontology for Facilitating Discussion of Catalytic Strategies of RNA-Cleaving Enzymes. *ACS Chem. Biol.* **2019**, *14*, 1068–1076.