Molecular Dissection of Hairpin Ribozyme Catalyzed Self-Cleavage

Sélène Forget* and I. Ken Groupleader*

CPCV, Département de Chimie, École Normale Supérieure, PSL University, Sorbonne University, CNRS, 75005 Paris

E-mail: selene.forget@ens.psl.eu; guillaume.stirnemann@ens.psl.eu

Abstract

[...]

Introduction

The discovery of catalytic RNAs, or ribozymes, have drastically changed our view of biological catalysis, proving that cell activity can be undertaken by more than proteins. Yet the full comprehension of the chemistry at play in these systems remains uncertain, with notably remaining debates on how mechanistically does the catalyst perform the reaction. Ettoffer un peu ici. Peut être faire un petit inventaire de ce que l'on connait des autres ryb ?

Among the most-studied ribozymes outstands the hairpin ribozyme (HpR), which performs the cleavage of a phosphodiester bond upon its own backbone, with a 3'-5' polymeric linkage of its nucleic backbone replaced by a 2'-3' cyclic terminating linkage. The HpR is not the only ribozyme to be self-reactive, but it was found to be one of the few to be able to do so without requiring any ion. If we do have proof that all chemical components involved

in the self-splicing reaction are part of the RNA itself, we do not know exactly which RNA atoms take part, and neither do we know how exactly do they proceed the self-cleavage.

TESTET!! In this transphosphoesterification process, the central step is an addition-elimination step, which involved the formation of a pentavalent phosphorus transition state. For this to occur, two key questions arise: how is the nucleophilic group activated prior to the addition step? And how is the departing group activated prior to the elimination step?

Mutagenesis experiments,[?] reinforced by structural resolution studies,[?][?] have demonstrated that two surrounding nucleobases, G8 and A38, are (i) mandatory for the chemical activity and (ii) in the direct neighbourhood of the A-1 — G+1 junction. From there, a first mechanistic scenarios was proposed: the dianionic scenario proposes that both G8 and A38 are directly involved, acting respectively as the general base that activates the nucleophile O2' by capturing its proton, and as the general acid that reprotonates the leaving group O5' after the elimination step (see Figure 1 A). Despite supporting pH-dependency experiments, this cleavage scenario suffers from one major inconsistency, which is the necessary prior deprotonation but *in-vivo* impractical of the N1 atom of G8 nucleobase.

To address this insufficiency, a second scenario has been considered: the in-line monoanionic scenario (see Figure 1 B). Here, A38 and G8 nucleobases only contribute to the electrostatic stabilization of the transition state, being involved in several H-bonds with atoms of the scissible junction. En more, these non-covalent interactions force the catalytic site to arrange favorably for the initiation of the reaction. While being maintained in position, the deprotonation of the O2' atom of A-1 is performed by one of the non-bridging oxygens the phosphate group itself, leading to a monoprotic phosphate intermediate. However, this scenario also suffers from its apparent incompatibility with the experimental pH-dependency of the reaction, as argued by Kath-Schorr et al.[?]

All this debate remains unsettled so far because of the inherent self-reactiveness of the HpR. Indeed, because it is self-reactive, the hairpin ribozyme can not be structurally resolved in directly as such. Instead, mimics of the HpR in wich the chemical activity is inhibited,

etiher by removing the O2' oxygen nuclophilr, by methylating it or by replacing other chemical groups with non-reactive ones, are used to study the catalytic site. Thus, no one can directly observe the catalytic site in its reactive form. With arguments based on crystal structures of *mimics* of the catalytic states, and non-discriminative pH profiles, it has been hard to draw definitive conclusions from experiments only.

In parallel to the experimental efforts, the HpR has been studied using *in silico* tools. Complementary to X-ray crystal studies, molecular dynamics (MD) expands the knowledge gained through structural resolution by providing the dynamic behavior of the resolved object. Here, the challenge of studying self-reactive objects is no longer a problem, as one can simulate any reactive species without worrying about self-degradation. Computational studies of the HpR are of two sorts. First, extended simulation of the full HpR with classical force fields,[?][?][?] which alow for the exploration of the conformational landscape of the HpR.[?][?][?] Second, combined quantic and classic mechanics (QM/MM) simulations,[?][?][?][?][?] which investigate the chemical activity of the catalytic site by simulating the chemical steps of the reaction. These two investigations means are usually conducted sequentially, as QM/MM simulation require relevant pre-catalytic initial structures which can only be obtained from prior robust classical MD simulations.

The contributions of the classical MD studies can be classified into two opposing pitches. The first base iself on study of the catalytic site which has been explicitly constrained the catalytic site elements to ensure the formation of a specific H-bond net- work: G8:N1⋯A-1:O2', A38:N1⋯G+1:O5', G8:N2⋯G+1:Opro-Rp, and A38:N6⋯G+1:Opro- Rp. This aimed conformation was guided by chemical intuition on the s itself descriptions of the catalytic site arrange- ment have emerged. the general acid-base mechanism and its specific requirements for the arrangement of the catalytic site. Veenis et al.[?] and Mlynsky et al.[?] Our recent efforts to verify the convergence of the catalytic site conformational landscapefocus on the positions of G8:N1 relative to O2' and A38:N1 relative to O5'. or by analogy with artificial mimics of transition state structures [187] However, they do not address the deprotonation

process of G8:N1 and paradoxically describe a pre-catalytic arrangement that supports a general acid-base pathway while G8 remains protonated [211, 210].

Far from settling the debate, these *in silico* studies have rather highlighted the complexity of the catalytic site conformational landscape, Anyway, they suffer from the inherent limitations of MD: and on the other hand the limited conformational sampling of QM/MM simulations.

All we have is the knowledge that the ribozyme is self-reactive and can perform the cleavage without any cofactor. We are missing simulations of the cleaved states, which are necessary to understand the chemical activity of the ribozyme. fill the partial knwoledge led us to concieve ..

Althought The simulation will provide relevant structures of either reactive, product, or transition states, which are mandatory for subsequent quantum calculations.

General Base/ General Acid Dianionic Scenario
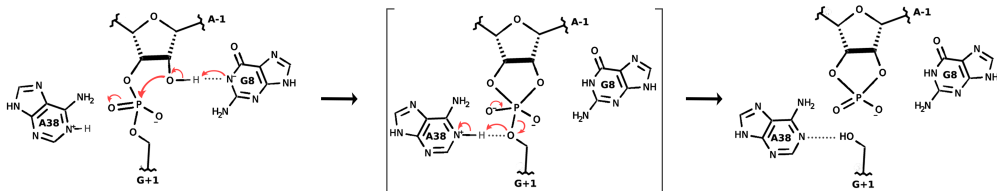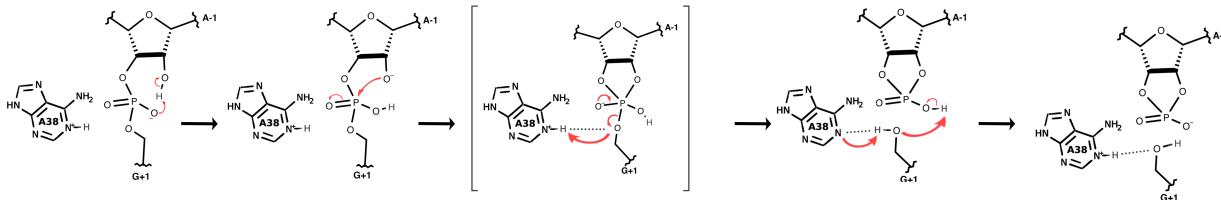

Inline Monoanionic Scenario

Figure 1: Detailed schemes of the three proposed mechanisms for the Hairpin Ribozyme self-cleavage reaction: the general acid-base mechanism, the inline mono-anionic mechanism, and the A38 mediated proton shuttle scenario. We chose here to represent the inline mono-anionic mechanism with the A38 mediated proton transfer in the elimination step. The direct reprotonation of O5' by the phosphate group is shown in Figure **??** in chapter **??**.

In this study, we provide insights to the mechanistic debate by investigating the conformational landscape of different "intermediates". We

Materials and Methods

Simulation of the HpR intermediates and Analysis Softwares

The different simulations of the hairpin ribozyme intermediates were performed using the GROMACS 2022.3[?] patched with PLUMED 2.7.[? ?] The visualization and analysis of the REST2 trajectories were performed using respectively

Simulation of Product States   The initial structures of all the cleaved states investigated in this study were constructed from the crystallographic structure of the 61-mer minimal hairpin ribozyme resolved at 2.2 Å, in which the G+1 phosphate group was removed (PDB ID: 2P7D, see Table **??**).[?] The missing phosphate (P, Opro-Rp, Opro-Sp atoms) was positioned according to the Vanadate Structure (PDB ID: 2P7E). After superposing the respective A-1 nucleotides, the Cartesian coordinates of the vanadate were extracted to be used for the phosphorus atom, and the same goes for the oxygen atoms.

Additionally, a conformational heterogeneity at nucleobase G8 in the cleaved product mimic crystal structure was observed, with 50% of the population with G8 in the syn conformation, 50% in the anti conformation.[?] Both conformations were reported in the crystal structure. We chose to keep the anti conformer, as this is the conformation of the ligated structure, the most populated conformer of the vanadate transition state mimic (PDB ID: 2P7E), and the one that was resolved in the pre-catalytic state (PDB ID: 2OUE) and in the product state of the 4WJ ribozyme (PDB ID: 1M5K).

The positions of the final catalytic site atoms (in the crystal structure) were compared to the 4WJ cleaved HpR equivalent (PDB ID: 1M5V) and demonstrated excellent consistency, with an RMSD of 0.93 Å calculated over all heavy atoms of residues A-1, G+1, G8 and A38 (see Figure 3).

Forcefields.

Preparation of Non-Standard Residues. The intermediates of the hairpin ribozyme with the most probable protonation states of the (isolated) chemical groups at *in vivo* pH. The existing force fields do not include the parameters for these non-standard states, which need to be generated specifically for this purpose. There are three possible situations:

Descriptors.

Similarly to the ligated states, the analysis of the cleaved catalytic site implies the monitoring of certain key variables, chosen for the relevancy regarding the mechanism. Here, we focus on to following observables:

- IAA and $d_{G+1:O5'-A-1:PC}$ which are the key observables for the nucleophilic attack,

- $d_{A-1:O2'-G8:N1}$, $d_{A-1:O2'-A38:N6}$, $d_{OproSp-G8:N2}$, $d_{OproRp-G8:N2}$, which informs on the positions of surrounding A38 and G8 regarding the 3' termini strand.

- $d_{G+1:O5'-A38:N1}$ which informs on the role of A38 in the proton transfer.

- $d_{OproSp-G+1:O5'}$, $d_{OproRp-G+1:O5'}$ which informs on the variant of the inline scenario proton transfer.

- The puckering of the G+1 ribose, symmetrically to A-1.

Convergence.

In the ligated state, convergence was established using both the eRMSD as global metric and conformational labeling described above as local metric. Figure **??** A-D details the time evolution of these metrics for the ligated states. In the cleaved state, we did not settle any satisfying local collective variable, and had to rely on the eRMSD global metric only (see Figure**??** E-F).

Results

Although unrelevant scientifically speaking, we choose here to consider accordingly every species that could appear along the different mechanistic pathways, starting with the mono-anionic scenario.

Mono-Anionic Scenario

The mono-anionic mechanism begins with a proton transfer from the O2' to one of the two oxygens of the scissile phosphate, leading to the formation of the "AP" intermediate (see Figure **??**). From there, the deprotonated O2' oxygen attacks the phosphorus atom, forming a pentavalent phosphorus transition state. The system subsequently proceed an elimination step, with the capture of the proton on the phosphorus oxygen and the departure of the substrate group.

The detail of this last step is another cornerstone of the debate.

Pre-catalytic state

pro-Rp

Dianionic Scenario

Discussion

Comparison with Experimental Data

Our results contrast with some of the available experimental data. At the forefront of these experimental elements is the pKa of the phosphate oxygens, which, being around 1, could not easily accept a proton from a neutral A-1:O2'. We note that the pKa of the A-1:O2' atom was re-evaluated to be approximately 18.5,[?] making it highly unlikely for either G8 or the

phosphate oxygens to capture it. Whether through the pro-Rp or the pro-Sp pathway, the monoanionic theoretical pH-profile should exhibit high activity at low pH and a sharp drop around pH 5.5. Experimentally, however, we observe the opposite trend, and no alternative explanation has been proposed to support a monoanionic scenario.

The dianionic scenario on the other hand, has been argued as the compatible with the experimental pH rates.[?][?] However remains the question of the deprotonation of G8 at *in vivo* pHs. If the G8:H1 proton was to be deprotonated on the fly, then by whom would it be captured? And why would such a chemical specie not directly deprotonate the HO' itself? Our simulations did not show any evidence of unexpected chemical players, and left the question of the deprotonation of G8 unanswered.
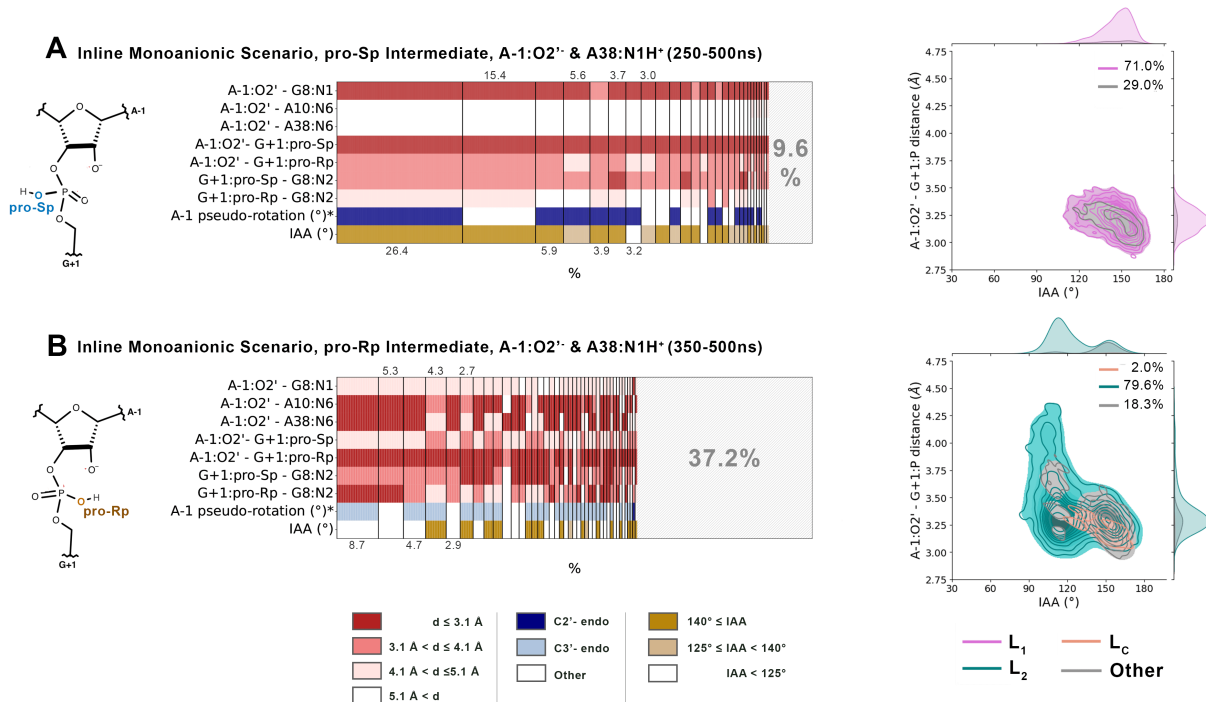

Figure 2: Overview of catalytic site behavior during the converged part of REST2 simulations of the HpR intermediate states in the inline monoanionic scenario: intermediate of the pro-Sp pathway (A), intermediate of the pro-Rp pathway (B). Left: Main combinations of the cleavage pathway key observables (9-key observables combination analysis). Right: 2D distribution of Inline Attack Angle (IAA) and A-1:O2'···G+1:P distance ($d_{O2'-P}$). See Methods for more details.

Encouragingly, other experimental findings do not directly contradict our monoanionic-favoring perspective. Notably, the crystallographic resolution of the HpR in its transition state analog[?] aligns closely with the H-bond network described by the highly stable pro-Sp intermediate, featuring the G8:N2⋯G+1:Opro-Sp and G8:N1⋯A-1:O2' H-bonds. Our results are consistent with G8 and A38H$^+$ acting as stabilizing agents. Their key roles in terms of electronic stabilization could explain the pH-dependence of the catalytic rate[?,?,?,?,?] and the sensitivity of the reaction to the mutations of these residues.[?,?] They need not be directly involved in catalysis.

It is interesting to note that the monoanionic scenario, with phosphate-assisted proton transfers, is actually reminiscent of the uncatalyzed mechanism in bulk water, that was recently studied by some of our group using deep-neural network inter-atomic potentials in particular,[?,?] and by others before.[?,?] The role played by the ribozyme active site residues here would be to provide an electrostatic environment minimizing the free-energy barriers of the reaction,[?] but this reaction would proceed though a very similar pathway as compared to the bulk.

Our approach has fulfilled its goal, which was to provide a dynamic view at the atomistic level of the ribozyme, when most of the experimental data are rather static and non-discriminative between the structure's misfolding and chemical inactivity. Nonetheless, we find that the available experimental data remain insufficient to discriminate between the two scenarios. Furthermore, our *in silico* insights have not bridged these gaps; instead, they have highlighted scenarios that seem improbable based on current experimental observations. Importantly, we note that no experiments have been specifically designed to test the monoanionic scenario.

Even if future QM/MM studies validate the enhanced-sampling MD-derived theory of the hairpin ribozyme mechanism, it is essential to acknowledge the experimental data that our results cannot explain, and the settlement of the debate will not be achieved without a solid experimental basis.

Are Equilibrium Molecular Dynamics Simulations Relevant?

The interpretation of our observations should be done with caution.

The approach itself invites scrutiny: intermediate species, by definition, are inherently unstable, and understanding their equilibrium behavior does not necessarily yield meaningful insights into a mechanistic pathway, which is inherently of high energy. This critique is especially pertinent to dianionic intermediates, whose very existence is debated, as the mechanism is often considered concerted.[?][?][?] In the end, distorted and unstable intermediates are both predictable and expected, offering little insight into the validity of this pathway.

Another critique, already addressed in the previous chapter, concerns the force field. Without entering into the debate of the force field accuracy, we will only highlight how force field biases, which may distort the depiction of the conformational landscape, make it impractical to rely solely on conformational sampling to resolve this debate. This sensitivity to parametrization is even more critical when studying intermediates, for which the charge redistribution was determined by association and could easily be questioned.

Building a QM/MM study on the Sampled Intermediates

Despite its limitations, our conformational sampling effort remains indispensable for the setup of a comparative QM/MM study. The mechanistic debate cannot advance without evaluating, within a unified framework, the energy barriers of the different mechanistic pathways to identify the most favorable chemical route.

As noted earlier, this endeavor requires a robust starting point. Yet defining a 'good' starting conformation for ribozyme reactivity studies presents significant challenges, with the critical need for a reliable strategy to characterize the conformational ensemble before undertaking any direct investigation of the chemical steps.

Experimental findings, particularly the pronounced loss of ligation activity resulting from perturbations to the active site architecture, emphasize the importance of precise positioning and orientation in catalytic rate acceleration. Classical MD simulations, despite their

inherent limitations, remain central to accessing local arrangements and advancing our understanding of the mechanism. Currently, they provide the only practical means to explore large conformational changes that result in the complete reorientation of key nucleobases. As previously noted, these changes remain beyond the reach of structural resolution techniques, both due to their static nature and the inhibitory strategies employed to observe reactive species like the hairpin ribozyme.

Conclusion

The conformational landscapes of the different mecanstic intermediates examined in this paper have provided a comprehensive understanding of how changes in protonation states lead to complete rearrangements of the catalytic site of the hairpin ribozyme.

Acknowledgement

Supporting Information Available

The overall structure was similar to the equivalent pre-catalytic state used for ligated state simulations (PDB ID: 2OUE), with an RMSD calculated over the nucleic backbone of the 60 common residues evaluated at 0.58 Å. The product crystal structure differed only from the minimal methylated pre-catalytic state by the presence of a 9-triethylene glycol linker (S9L) in place of residue 14, which ensured the junction between H2 and H3. The linker was
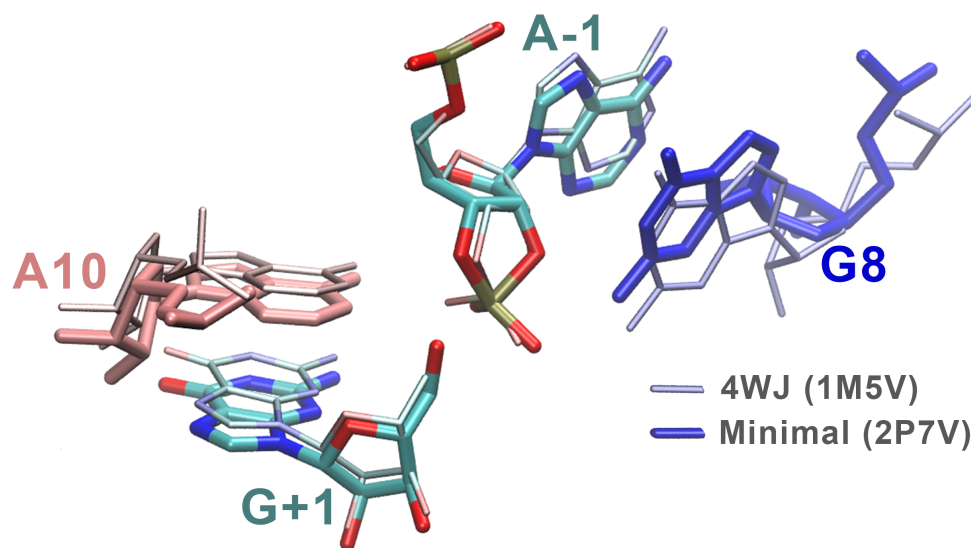
kept in the simulation.

Figure 3: Comparison of the Cleaved States Crystal Structures: the 4WJ hairpin ribozyme (PDB ID: 1M5V) in light colors and thin representation, the minimal hairpin ribozyme (PDB ID: 2P7D) used as initial structure for the cleaved states in this study is strong colors and thick representation. The two catalytic sites differ by only 0.93 Å, calculated over all heavy atoms appearing here, that is to say residues A-1, G+1, G8 and A38.

TOC Graphic

Some journals require a graphical entry for the Table of Contents. This should be laid out "print ready" so that the sizing of the text is correct.

Inside the `tocentry` environment, the font used is Helvetica 8 pt, as required by *Journal of the American Chemical Society*.

The surrounding frame is 9 cm by 3.5 cm, which is the maximum permitted for *Journal of the American Chemical Society* graphical table of content entries. The box will not resize if the content is too big: instead it will overflow the edge of the box.

This box and the associated title will always be printed on a separate page at the end of the document.