



Cite this: *Chem. Commun.*, 2021, 57, 7517

Received 29th April 2021,  
Accepted 2nd July 2021

DOI: 10.1039/d1cc02290c

rsc.li/chemcomm

# RNA diversification by a self-reproducing ribozyme revealed by deep sequencing and kinetic modelling†

Cyrille Jeancolas,<sup>‡ab</sup> Yoshiya J. Matsubara,<sup>‡c</sup> Mykhailo Vybornyi,<sup>a</sup> Camille N. Lambert,<sup>a</sup> Alex Blokhuis,<sup>‡ad</sup> Thomas Alline,<sup>a</sup> Andrew D. Griffiths,<sup>a</sup> Sandeep Ameta,<sup>‡c</sup> Sandeep Krishna<sup>c</sup> and Philippe Nghe<sup>\*a</sup>

**We demonstrate that a recombinase ribozyme achieves multiple functions in the same reaction network: self-reproduction, iterative elongation and circularization of other RNAs, leading to synthesis of diverse products predicted by a kinetic model. This shows that key mechanisms can be integrated and controlled toward Darwinian evolution in RNA reaction networks.**

The “RNA world” hypothesis surmises RNA as a primordial biopolymer capable of evolution due to its ability to store information and catalyze reactions.<sup>1</sup> An RNA polymerase ribozyme that catalyzes its own copying and the copying of other RNAs would support this hypothesis but has not yet been found.<sup>2</sup> Alternatively, ribozymes can make copies of themselves from fragments *via* energy-rich<sup>3</sup> or energy-neutral<sup>4</sup> chemistries. Autocatalytic reproduction has been engineered using the ~200 nt long *Azoarcus* group I intron ribozyme *via* recombination reactions between hybridized RNA fragments.<sup>5,6</sup> Such catalysis requires the binding of a 3 nt long internal guide sequence (IGS) of the ribozyme to a 3 nt long ‘tag’ sequence of the substrates.<sup>7</sup> Different IGS-tag combinations result in autocatalytic sets that differ in their dynamics and ability to reproduce.<sup>5</sup> However, open-ended evolution further requires self-reproducing RNAs to generate novel sequences. Group I intron ribozymes are promising candidates<sup>8–10</sup> for sequence diversification as they catalyze the recombination of other RNAs in addition to their own reproduction.

Here, we characterize sequence variation occurring through the combination of two mechanisms catalyzed by the self-reproducing *Azoarcus* ribozyme. A first mechanism, here called *repeated transfer* (RT), allows covalent additions of multiple RNA fragments to the 3′-end of an RNA substrate (**RNA1**, Fig. 1), leading to RNA elongation from 26 nt to at least 136 nt. A second mechanism, here called *terminal strand attack* (TSA), results in the ligation of hybridized RNAs (**RNA2**, Fig. 1) into hairpin and circular products. RT and TSA can be combined (**RNA3**, Fig. 1), leading to diversification of the RNA products together with ribozyme self-reproduction. For further mechanistic details, see ESI.† Furthermore, the reaction network dynamics parameterized from gel data can be integrated in a kinetic model that predicts species distributions revealed by sequencing (ESI†).

First, we show that **RNA1** is elongated by multiple additions of small fragments in the presence of the ribozyme. **RNA1** is composed of an 18 nt “stem” (S), a 3 nt tag, and a 5 nt “mobile unit” (Fig. 1). We observe a variety of products denoted S<sub>n</sub>, which length corresponds to the addition of *n* mobile units



**Fig. 1** RNA substrates are composed of a “stem” (blue) and a tag (red). **RNA1** stem differs from **RNA2** and **RNA3** stems by two nucleotides (bold). **RNA2** and **RNA3** have a self-complementary sequence (underlined), which can hybridize in experimental conditions, whereas **RNA1** does not. **RNA1** and **RNA3** have a “mobile unit” after the first tag which can be polymerized by the *Azoarcus* ribozyme.

<sup>a</sup> Laboratoire de Biochimie, UMR CNRS-ESPCI 8231, Chimie Biologie Innovation, PSL University, ESPCI Paris, 10 rue Vauquelin, Paris 75005, France.  
E-mail: philippe.nghe@espci.psl.eu

<sup>b</sup> Laboratoire d'Anthropologie Sociale, Collège de France, 52 rue du Cardinal Lemoine, Paris 75005, France

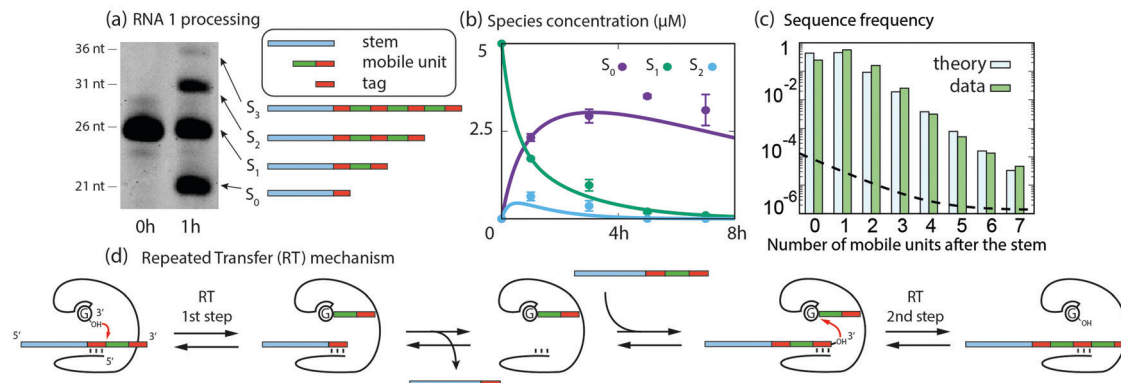
<sup>c</sup> Simons Centre for the Study of Living Machines, National Centre for Biological Sciences, Bellary Road, Bangalore 560 065, Karnataka, India

<sup>d</sup> Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen 9747 AG, The Netherlands

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1cc02290c

‡ These authors contributed equally.

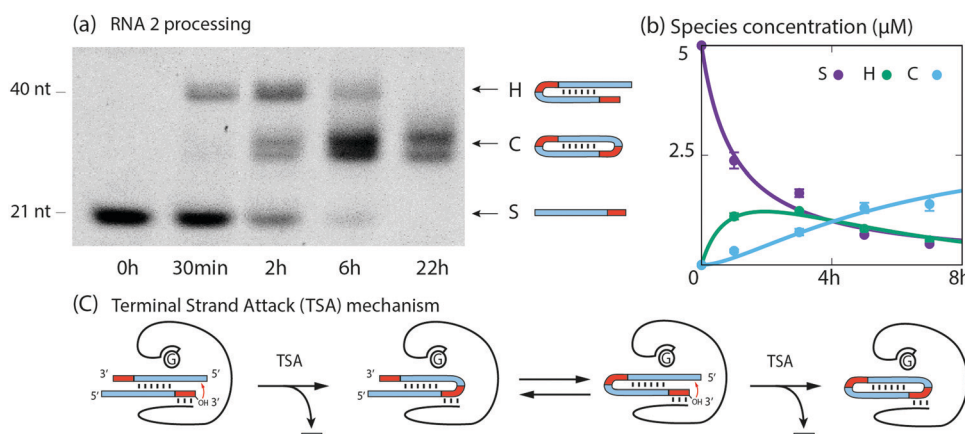




**Fig. 2** Ribozyme-based RNA elongation by RT. (a) Gel image showing the length diversity of products after 1 h of reaction (**RNA1** with *Azoarcus* ribozyme). (b) Evolution of the concentrations of **RNA1** ( $S_1$ ), products without a mobile unit ( $S_0$ ), and products with two mobile units ( $S_2$ ): experimental values from gel band intensities (dots) and kinetic model predictions (solid lines). Error bars are standard deviations from triplicates. (c) Fraction of products versus number of mobile units, after 45 min of reaction: RNA sequencing data (green bars) and kinetic model (blue bars, ESI<sup>†</sup>) fitted to gel electrophoresis data. The sequencing data is calibrated for length biases due to sequencing library preparation (see Fig. S1, ESI<sup>†</sup>). The dashed horizontal line indicates the calibrated limit of detection (materials and methods, ESI<sup>†</sup>). (d) Schematic of the RT mechanism accounting for the elongation of **RNA1**.

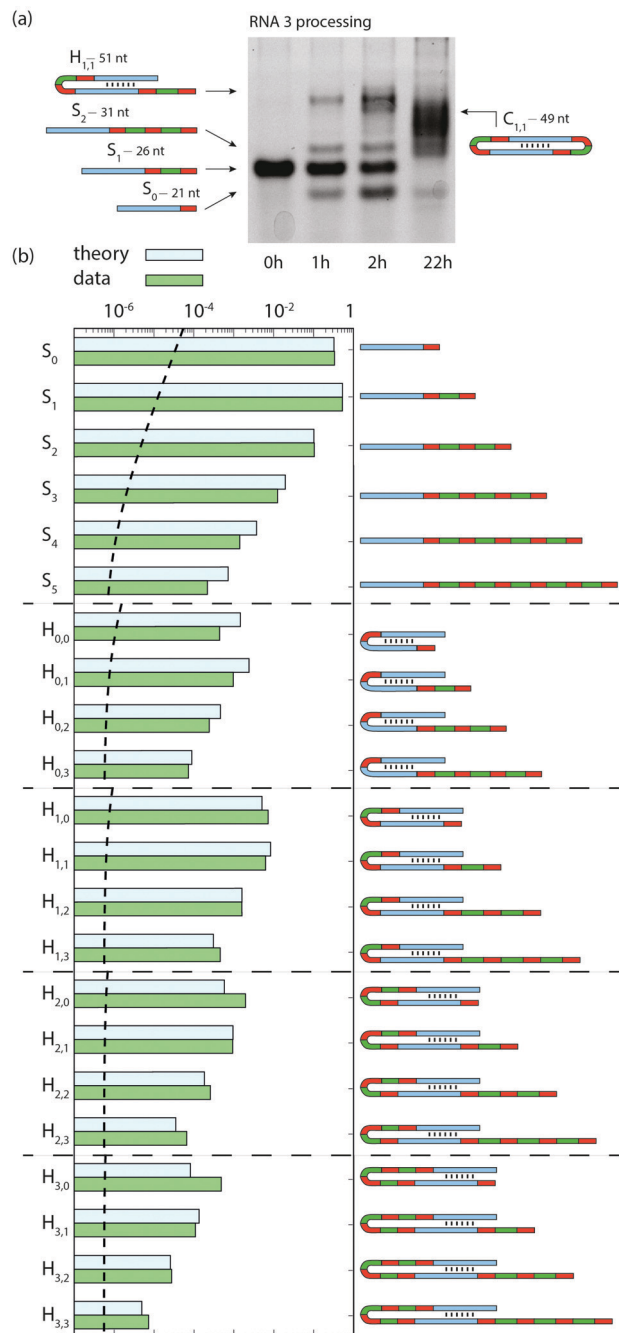
appended to the stem. Polyacrylamide gel images reveal four bands (Fig. 2a) after 1 hour of reaction, corresponding to  $S_0$  (21 nt),  $S_1$  (26 nt),  $S_2$  (31 nt) and  $S_3$  (36 nt). The concentrations of putative  $S_0$ ,  $S_1$  and  $S_2$  species were measured over 7 h by gel electrophoresis, showing the transient formation of  $S_2$  during the first 4 h (Fig. 2b). Sequencing after 45 min of incubation confirmed that RNA sequences result from iterative additions of mobile units (Fig. 2c and see Fig. S2 for all  $S_n$  species detected, ESI<sup>†</sup>). Such additions can be explained by formerly identified reactions catalyzed by the *Azoarcus* ribozyme,<sup>7</sup> here performed in 2 steps: first forward, then reverse after substrate exchange (Fig. 2d and see ESI<sup>†</sup> for details). Concentrations could be quantified by sequencing up to  $S_7$ , and up to  $S_{23}$  sequences were detected (Fig. S2, ESI<sup>†</sup>).  $S_n$  concentration decreases exponentially with  $n$  (Fig. 2c) consistent with earlier theoretical predictions.<sup>11</sup> We devised a kinetic model comprising the transfer reactions  $S_n + S_m \rightarrow S_{n+p} + S_{m-p}$ , using the same

rate  $k$  for all  $n, m$  and  $p, p \leq m$ , and the degradation rates of stems and of mobile units (see “Model for RT” in ESI<sup>†</sup>). These 3 parameters were fitted using  $S_0$ ,  $S_1$  and  $S_2$  concentration at 0, 1, 3, 5 and 7 h from PAGE analysis (ESI<sup>†</sup>, Fig. 2b). The model reproduced the concentration time courses:  $S_1$  concentration decreases rapidly,  $S_2$  concentration reaches a maximum at around 45 min and decreases after 3 h of reaction, and  $S_0$  concentration increases until reaching a plateau after 3 h and subsequently slowly decreases. Without further fitting, the model predicted an exponential decrease of product frequencies as a function of their length at 45 min of reaction, as observed experimentally (Fig. 2c). From  $S_1$  to  $S_7$ , the measured and predicted decay factor  $\ln(S_n/S_{n+1})$ , with  $S_n$  being the fraction of  $S_n$ , agree quantitatively ( $-1.569 \pm 0.057$  and  $-1.589$ , respectively), with a correlation  $R^2 = 0.988$  ( $p < 10^{-5}$ ) between measured and predicted frequencies (Fig. S3, ESI<sup>†</sup>).



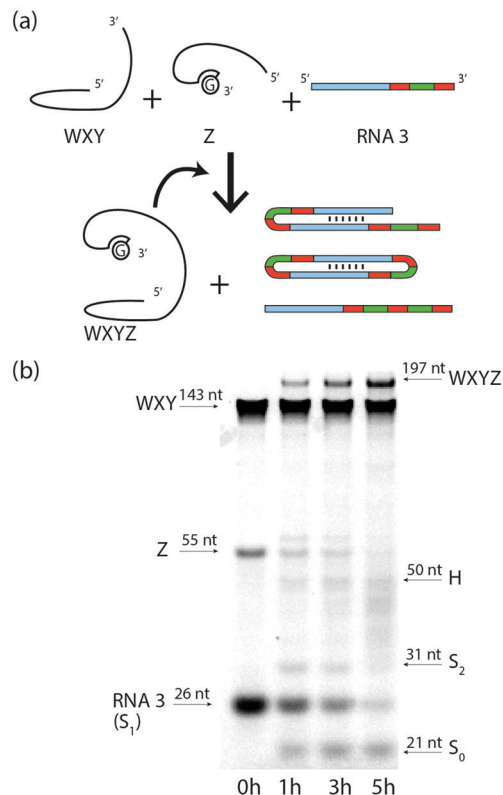
**Fig. 3** Dimerization and circularization by TSA. (a) Gel image showing the successive production of hairpins and circular RNAs starting from **RNA2**. Double bands are due to two sites for transesterification as confirmed in sequencing data. (b) Concentration evolution of **RNA2** ( $S$ ), hairpin ( $H$ ), and circular products ( $C$ ): experimental values from the band intensities (dots) and concentrations predicted by the model (solid lines). (c) Schematic of the TSA mechanism applied to **RNA2**.





**Fig. 4** Combination of RT and TSA reactions. (a) Gel image showing the bands corresponding to the addition of one mobile unit and hairpin formation (from **RNA3** substrate). At 22 h, diverse products longer than the substrate appear, among which hairpins and cyclic RNAs. (b) Relative proportion of expected products at 45 min measured by RNA sequencing compared to the predictions from the model, independently parameterized with gel data. Measured frequencies are calibrated for length biases (Fig. S1, ESI†). The dashed horizontal line indicates the calibrated limit of detection (materials and methods, ESI†).

Next, we show that **RNA2** undergoes catalytic dimerization and circularization in the presence of the ribozyme *via* the TSA mechanism. **RNA2** is composed of an 18 nt long stem (S) with a 12 nt self-complementary region and a 3 nt tag (Fig. 1). Over 22 h, polyacrylamide gels (PAGE) indicate the transient



**Fig. 5** *Azoarcus* ribozyme self-reproduction and processing of **RNA3** in one pot. (a) Formation of WXYZ from WXY and Z fragments catalyzed by non-covalent (WXY:Z) and covalent (WXYZ) ribozymes, together with the formation of diversified RNA species through **RNA3** processing. (b) Gel showing the concomitant production of WXYZ through self-reproduction and the processing of **RNA3** in one pot.

formation of almost twice longer RNAs and the appearance of lower bands assigned to circular RNAs (Fig. 3a). The formation of hairpins (H) was confirmed by sequencing and of circular RNA (C) by differential mobility assays with 12% and 18% polyacrylamide gels<sup>12</sup> (Fig. S4, ESI†). Hairpins and cyclic RNAs result from the hybridization of two **RNA2** molecules *via* their self-complementary region and by the ribozyme catalyzed TSA reaction, enabled by the presence of 3' tags (Fig. 3c, see ESI† for details). We devised a kinetic model of the two TSA reactions  $S + S \rightarrow H$  and  $H \rightarrow C$  with rate constants  $k_1$  and  $k_2$ . We fitted these 2 parameters to the concentration of S, H and C at 0, 1, 3, 5 and 7 h of reaction from PAGE gel data. Theoretical and experimental concentration time courses of the three species agree over 7 h of reaction (Fig. 3b): S is roughly halved in the first hour, then decreases more slowly, H increases to a maximum after 2 h and then decreases to reach the same concentration as S, while C continuously increases during the 7 h of reaction.

Thereafter, we show the combination of ribozyme catalyzed elongation, dimerization, and circularization using **RNA3** as a substrate. **RNA3** combines **RNA2** self-complementary stem with **RNA1** mobile unit (Fig. 1), thus can undergo both RT and TSA. Over 22 h, species  $S_0$  and  $S_2$  were produced transiently along with hairpin RNAs, followed by the production of circular RNAs



(Fig. 4a) confirmed by RNA sequencing and gel mobility assay (Fig. S4, ESI†). Sequencing revealed a diversity of RNAs after 45 min:  $H_{m,n}$ , standing for hairpin (H) products with  $m$  and  $n$  mobile units after the first and the second stems, respectively (Fig. 4b and see Fig. S2 for all  $S_n$  species detected, ESI†). The 5 parameters of the combined models described above for **RNA1** and **RNA2** were fitted (Table S2, ESI†) to the concentration of  $S_0$ ,  $S_1$ ,  $S_2$  and H (the sum of all  $H_{m,n}$ ) measured on gel over 7 h of reaction starting with **RNA3** (Fig. S5, ESI†). The model was parameterized only with the gel data, which allowed quantification of circular RNAs absent from the sequencing data. The predicted proportions of  $S_n$  (with  $0 \leq n \leq 5$ ) and  $H_{m,n}$  (with  $0 \leq m, n \leq 3$ ) at 45 min reproduce the trends of each category of products (separated by horizontal dotted lines in Fig. 4b). Like for **RNA1**,  $S_n$  species decrease exponentially with their size, the measured and predicted decay factor being  $-1.74 \pm 0.09$  and  $-1.65$ , respectively. This trend is conserved with  $H_{m,n}$  species: at a given  $m$  (from 0 to 3),  $H_{m,n}$  species also decrease exponentially with  $n$  (from 1 to 3) with a similar exponential decay factor as  $S_n$  species (see “Model for RT + TSA” in ESI†). The correlation factor between measured and predicted frequencies is  $R^2 = 0.940$  ( $p < 10^{-5}$ , Fig. S3, ESI†).

Finally, we show that the *Azoarcus* ribozyme (noted WXYZ, where the letters refer to 4 contiguous parts, Fig. S6 ESI†)<sup>4</sup> self-reproduces from two fragments (WXY and Z) and simultaneously processes the **RNA3** substrate through RT and TSA (Fig. 5a). We observed by PAGE the appearance of the band corresponding to the covalent *Azoarcus* ribozyme as well as those corresponding to the products of TSA and RT mechanisms on **RNA3** (mainly  $S_1$ ,  $S_2$  and H) over 5 h of reaction (Fig. 5b). As bands characteristic of cyclic RNA could not be confidently detected in the product smear, we confirmed separately the ability of a WXY and Z mixture to generate a cyclic RNA from **RNA2** (Fig. S7, ESI†). The concentration dynamics of  $S_0$ ,  $S_1$ ,  $S_2$  and H are well described by the same model as for **RNA3**, and the same fitting procedure results in slightly reduced parameter values, which reflects the inhibitory effect of free Z strands and the lower catalytic efficiency of the WXY:Z non-covalent complex<sup>5</sup> (Fig. S8, ESI†). The concentrations of WXY, Z and WXYZ were used to fit a set of kinetic equations for *Azoarcus* self-reproduction, comprising formation of the WXY:Z complex and covalent bond formation  $WXY:Z \rightarrow WXYZ$  catalyzed by both WXY:Z and WXYZ.<sup>13</sup> The kinetics show that recombination reactions involving **RNA3** reduce the yield of self-reproduction of only  $\sim 10\%$  at 7 h, despite **RNA3** being initially 10 times more concentrated than ribozyme fragments (Fig. S9, ESI†).

This study shows mechanisms for ribozyme-mediated RNA sequence diversification, hairpin formation, circularization, extension, self-reproduction, and their integration in reaction networks. We have furthermore shown that the dynamics of these networks can be modeled quantitatively and that such a model predicts the diversity and concentration of products from the knowledge of substrate sequences. The ability of the *Azoarcus* ribozyme to catalyze its own formation from smaller fragments, diversify and elongate other fragments in an energy-neutral fashion is a notable advantage in the origin of life context. Furthermore, mechanisms for RNA elongation and structure formation are crucial to generate functional RNAs. Circularizing RNA could have been a way to store primordial genetic information,<sup>13</sup> thus favoring heredity as in viroids,<sup>14</sup> also of interest for the synthesis of cyclic siRNA for interfering therapies.<sup>15</sup> This study overall paves the way to engineer evolvable RNA systems by combining mechanisms enabling reproduction, diversification, and structure formation.

Funding from CEFIPRA, HFSP GRY0077/2019, ANR-10-IDEX-0001-02 PSL, IPGG ANR-10-LABX-31, ANR-10-EQPX-34, U. Paris, ED FIRE-Bettencourt, ERC grant agreement No. [101002075].

## Conflicts of interest

There are no conflicts to declare.

## References

- 1 M. P. Robertson and G. F. Joyce, *Cold Spring Harbor Perspect. Biol.*, 2012, **4**, a003608.
- 2 J. Attwater, A. Raguram, A. S. Morgunov, E. Gianni and P. Holliger, *eLife*, 2018, **7**, 1–25.
- 3 N. Paul and G. F. Joyce, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12733–12740.
- 4 E. J. Hayden and N. Lehman, *Chem. Biol.*, 2006, **13**, 909–918.
- 5 S. Ameta, S. Arsène, S. Foulon, B. Saudemont, B. E. Clifton, A. D. Griffiths and P. Nghe, *Nat. Commun.*, 2021, **12**, 842.
- 6 E. J. Hayden, G. Von Kiedrowski and N. Lehman, *Angew. Chem., Int. Ed.*, 2008, **47**, 8424–8428.
- 7 W. E. Draper, E. J. Hayden and N. Lehman, *Nucleic Acids Res.*, 2008, **36**, 520–531.
- 8 M. D. Been and T. R. Cech, *Science*, 1988, **239**, 1412–1416.
- 9 B. M. Chowrira, A. Berzal-Herranz and J. M. Burke, *EMBO J.*, 1993, **12**, 3599–3605.
- 10 A. J. Zaugg and T. R. Cech, *Science*, 1985, **229**, 1060–1064.
- 11 A. Blokhuis and D. Lacoste, *J. Chem. Phys.*, 2017, **147**, 094905.
- 12 T. R. Cech, A. J. Zaugg and P. J. Grabowski, *Cell*, 1981, **27**, 487–496.
- 13 J. Demongeot and A. Moreira, *J. Theor. Biol.*, 2007, **249**, 314–324.
- 14 M. C. Maurel, F. Leclerc, J. Vergne and G. Zaccari, *Viruses*, 2019, **11**, 1–17.
- 15 N. Abe, H. Abe and Y. Ito, *Curr. Protoc. Nucleic Acid Chem.*, 2012, **1**, 1–11.

