bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Cryo-EM structure and functional landscape of an RNA polymerase ribozyme

Ewan K.S. McRae[1,3], Christopher J.K. Wan[2,3], Emil L. Kristoffersen[1,2], Kalinka Hansen[1], Edoardo Gianni[2], Isaac Gallego[2], Joseph F. Curran[2], James Attwater[2], Philipp Holliger[2,4], Ebbe S. Andersen[1,4]

[1]Interdisciplinary Nanoscience Center, Aarhus University, Aarhus, Denmark; [2]MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

*Communication to: ph1@mrc-lmb.cam.ac.uk & esa@inano.au.dk

[3] these authors contributed equally to this work

[4] jointly supervised the work

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Abstract

The emergence of an RNA molecule capable of replicating itself and other RNA sequences is a central pillar of hypotheses regarding the origin of life[1,2]. *In vitro* evolution has yielded polymerase ribozymes (PR) that can copy a range of RNA templates using nucleotide[3-10] or trinucleotide triphosphates (triplets)[11] as substrates and may give rise to a replicase activity. However, our understanding of PR function is encumbered by a lack of structural information beyond the progenitor class I ligase (cIL) ribozyme[12-14]. Here, we report the structure of the complete 5TU+t1 triplet polymerase ribozyme (TPR) apoenzyme and map its structure / function landscape. The TPR is an RNA heterodimer, comprising a catalytic (5TU) and a catalytically inactive (t1) subunit held together by two kissing loop interactions and its overall structure resembles a left hand with thumb and fingers at a 70° angle. While the 5TU subunit shows partial structural homology to the cIL, the t1 accessory subunit - despite sharing the same progenitor - exhibits a dramatically reorganized secondary and tertiary structure. Our combined structural and functional data suggest a model for templated RNA synthesis by the TPR holoenzyme and provide a foundation for a better understanding of RNA's potential for self-replication.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RNA catalysts (ribozymes) occupy central structural and catalytic roles in the function of modern cells including tRNA processing (RNaseP), mRNA splicing (spliceosome, group I / II self-splicing introns) and translation (ribosome peptidyl transferase center)[15]. In addition, a much wider variety of ribozyme activities not found in nature has been discovered by *in vitro* evolution, including polymerase ribozymes (PR) that are capable of synthesizing a complementary strand on an RNA template[3-7,11]. The capacity for RNA-catalyzed RNA-templated synthesis and replication is widely believed to have been a central pillar of the emergence of life's first genetic system and even life itself.

The earliest examples of nascent PR activity were found in self-splicing intron (SSI) ribozymes, in particular a variant of the *sunY* SSI ribozyme, which allowed single nucleotide triphosphate (NTP) extension[16] or the iterative ligation of RNA oligonucleotides on a complementary strand[17] including assembly of one of its subunits from RNA oligonucleotides[18,19]. The same *sunY* SSI ribozyme was also shown to incorporate short RNA trinucleotide substrates[20], but with relatively low fidelity.

A more fully developed RNA polymerase activity emerged from derivatives of the class I ligase (cIL) ribozyme[12,21], which after engineering and reselection could incorporate up to 14 NTPs in a template-dependent manner[3]. The polymerase activity of this first "true" PR was progressively improved by *in vitro* evolution to enable the synthesis of long RNAs (100-200 nts on some RNA templates)[5,8] as well as the synthesis of functional RNAs including a hammerhead ribozyme[4], tRNA[6], Broccoli fluorescent RNA aptamer[11] and the progenitor cIL ribozyme itself[9]. Recently, a variant utilizing trinucleotide triphosphates (triplets) as substrates (a triplet polymerase ribozyme (TPR)) emerged as a heterodimer from *in vitro* evolution[11]. This TPR displayed a remarkable ability to copy structured RNA templates including segments of its own sequence[11] as well as circular RNA templates by rolling circle synthesis[22].

However, despite the above examples of PRs, there is no structural information available beyond the crystal structure of the cIL ribozyme[13,14], the progenitor of the most advanced PRs including the TPR. While the cIL structure provided insights into the mechanism of phosphodiester bond formation and cIL interaction with the RNA substrate, it is unclear how and to what extend these features are retained in PRs, which diverge from the cIL not only by a number of mutations to the ribozyme core, but also by 5'- and 3'-extension sequences. A better understanding of how PRs perform accurate substrate selection, general RNA template interaction and templated RNA synthesis would therefore benefit from the structure of an active PR.

The structural challenge is defined by the highly dynamic nature and conformational malleability of large RNAs. Despite significant recent progress, in particular in cryogenic electron microscopy (cryo-EM) approaches to RNA structure

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

determination[23,24], RNA-only structures have remained challenging targets for structure determination. In the case of the TPR, conformational heterogeneity may be further exacerbated by its origins: this ribozyme was evolved (and functions best) in the eutectic phase of water ice[25], where subzero temperatures and high counterion concentrations present weak adaptive pressure for stable folding at ambient temperatures, potentially allowing for even more inherent structural dynamics and heterogeneity than RNAs derived from biology.

Leveraging RNA *in vitro* evolution (see below) as well as advances in RNA sample (see attached manuscript, McRae et al. 2022) and grid preparation[26] and image data processing[27], we report the cryo-EM structure of the complete, heterodimeric TPR apoenzyme determined at its optimal functional magnesium concentration ([Mg^{2+}] = 100 mM), together with a comprehensive fitness landscape of TPR function. Our structure shows the molecular anatomy of the catalytic polymerase subunit, and reveals the nature and potential origin of its mutualistic association with the catalytically inactive accessory subunit to form the fully active heterodimeric ribozyme. Our data reveals structural and functional details not previously described and provides the foundation for a better understanding of TPR function.

Cryo-EM structure of optimized TPR heterodimer

The heterodimeric TPR was evolved to use triplets as substrates (Fig. 1a). In order to improve activity and stability of the original t5+1 TPR[11], we executed further rounds of *in vitro* evolution using an adaptation of a previously described tethered template selection scheme[11] (SI Fig. 1a). Starting from random mutant libraries of the catalytic t5 subunit, we carried out 8 rounds of selection for triplet polymerase activity with increasing stringency in the presence of a conserved t1 accessory subunit (see Methods). We identified two mutations in t5 (ΔU38 and C110U) and combined these with 3 more t5 mutations (U117C, U132C, U148A) discovered in separate selection experiments (to be described elsewhere) (SI Fig. 1b). The resulting t5 variant, 5TU (t5: ΔU38, C110U, U117C, U132C, U148A) exhibited superior triplet polymerase activity compared to t5 (SI Fig. 1c) and remained receptive to activity enhancement by the t1 accessory subunit to copy longer templates (Fig. 1b).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

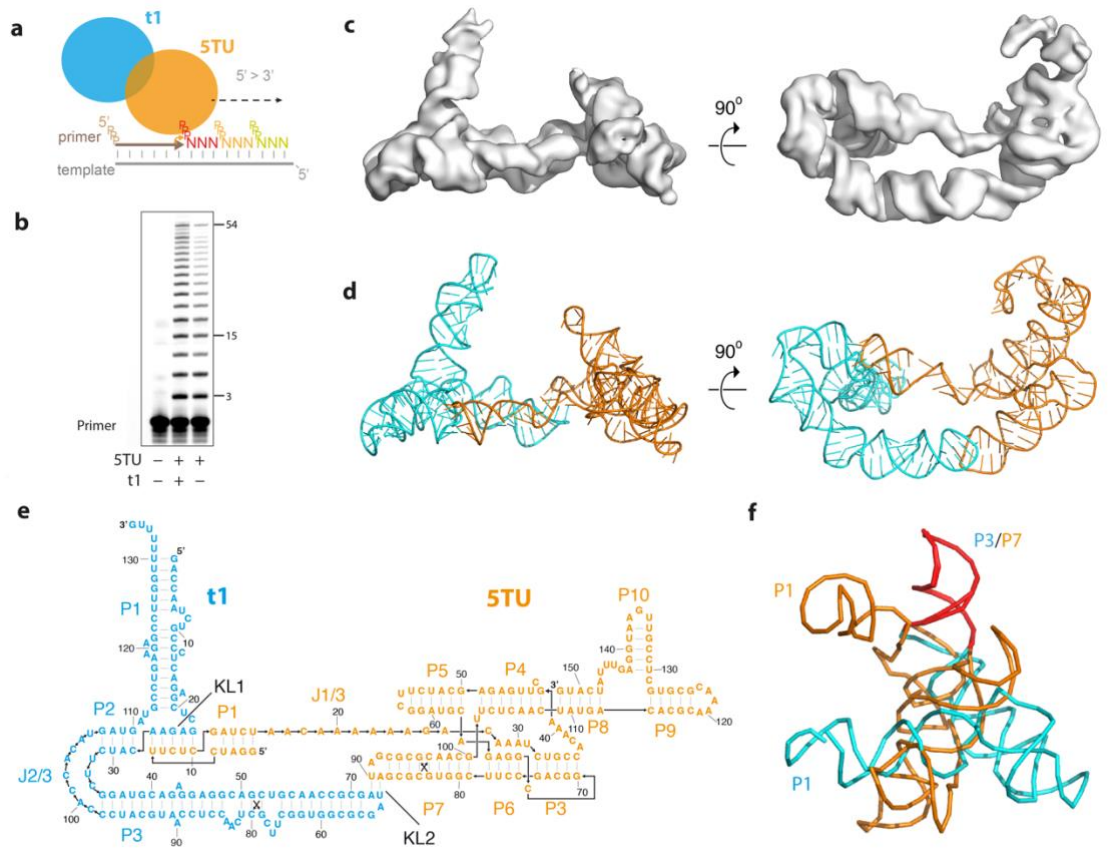

Figure 1. Structure of the Triplet Polymerase Ribozyme (TPR). (**a**) Schematics of the 5TU+t1 heterodimer, template and triplet substrates. (**b**) Activity of 5TU alone or in combination with t1 in copying a template encoding for $(GAA)_{18}$ after 15 hours. (**c**) Two views of the cryo-EM reconstruction shown in grey. (**d**) Two views of the atomic model for 5TU (orange) and for t1 (cyan). (**e**) The secondary structure diagram is shown for 5TU (orange) and for t1 (cyan). (**f**) Structural alignment of t1:P3 and 5TU:P7 stems shows major structural difference between the two subunits.

Next, we sought to determine the structure of the TPR heterodimer 5TU+t1 in its active form (at an optimal Mg^{2+} concentration of 100 mM). Using cryo-EM, we were able to reconstruct the full length 5TU+t1 RNA heterodimer complex to a global resolution of 5.9 Å (Fig. 1c and SI Fig. 2,3). The cryo-EM map allows for unambiguous placement of all double helices (P) based on the secondary structure predictions of 5TU and t1 (SI Fig. 4,5) and *de novo* assembly of the remaining joining (J) and loop (L) regions using DRRAFTER[24] (SI Fig. 6). The final model of the heterodimer was refined using molecular dynamics and energy minimizations (see Methods) and reached a map-to-model cross correlation of 8.3 Å at FSC=0.5 and 6.6 Å at FSC=0.143 (SI Fig. 7). At the global resolution of 5.9 Å, the data support the overall conformation of the RNA backbone, whereas exact base positions are the result of geometrical refinement and energy minimizations. 3D variability analysis revealed intra- and interdomain flexibility, which could explain why we were not able to obtain higher resolution (SI Fig. 8, SI Movie 1). Further support of the structural model was gained

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

by an independently determined low resolution (8 Å) map of the progenitor t5+1 ribozyme[11] at lower Mg^{2+} concentrations (SI Fig. 9).

The model revealed the overall structural anatomy of the TPR to resemble an upturned left hand, with the thumb formed by the t1 and fingers formed by the 5TU subunit at an approximate angle of 70°, with the palm formed by a bipartite interaction of the subunits through two distinct kissing loops (KL1, KL2). The model can be rationalized in a secondary structure diagram that shows helical stacking, pseudoknots and interaction sites (Fig. 1e). The 5TU subunit comprises the catalytic core domains P3-7, the template binding strand J1/3, and peripheral domains P1+P8-10. In contrast, the non-catalytic accessory subunit t1 adopts an extended secondary structure that contains only three main hairpin domains P1-3. Although both 5TU and t1 subunits are derived from the same starting sequence by *in vitro* evolution (with their core sequences diverging only by 7 mutations (SI Fig. 5a)), their secondary and tertiary structures have diverged radically with only a 22-bp segment of the t1:P3 hairpin domain retaining its original structure (5TU:P7) (Fig. 1f, SI Fig. 10).

Functional landscape of the TPR heterodimer

Next, in order to obtain information on the functional importance of the structural features observed in our model of the TPR heterodimer, we performed a comprehensive fitness landscape[28,29] analysis (Methods, Fig. 2 and SI Figs 11-16, SI Movie 2). Mutant libraries of 5TU and t1 with mutation rates of 3% per position (1% of each alternative base) were bottle-necked to about 10^6 members and subjected to one round of purifying activity selection (in triplicate), whereby only TPR mutants capable of successfully copying a given RNA template were recovered. Pre- and post-selection libraries were sequenced and changes in genotype abundance were quantified; we define ribozyme “fitness” as the log-transformed enrichment of a given genotype relative to the wild-type 5TU or t1 sequence. After filtering, we obtained the relative fitness of 128,708 ribozyme variants, comprising 79,702 5TU and 49,006 t1 genotypes, providing fitness estimates of all point mutants in t1, as well as 99.6% of all point mutants in 5TU. For both subunits, calculated fitness was strongly correlated across replicates (Pearson coefficient R= 0.89 (5TU) / 0.95 (t1), and R = 0.97 (5TU) / 0.95 (t1) if only single and double mutants are considered) (Fig. 2a, SI Fig. 11b). To our knowledge, this presents the first, comprehensive genotype / phenotype correlation dataset for a large, complex ribozyme.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

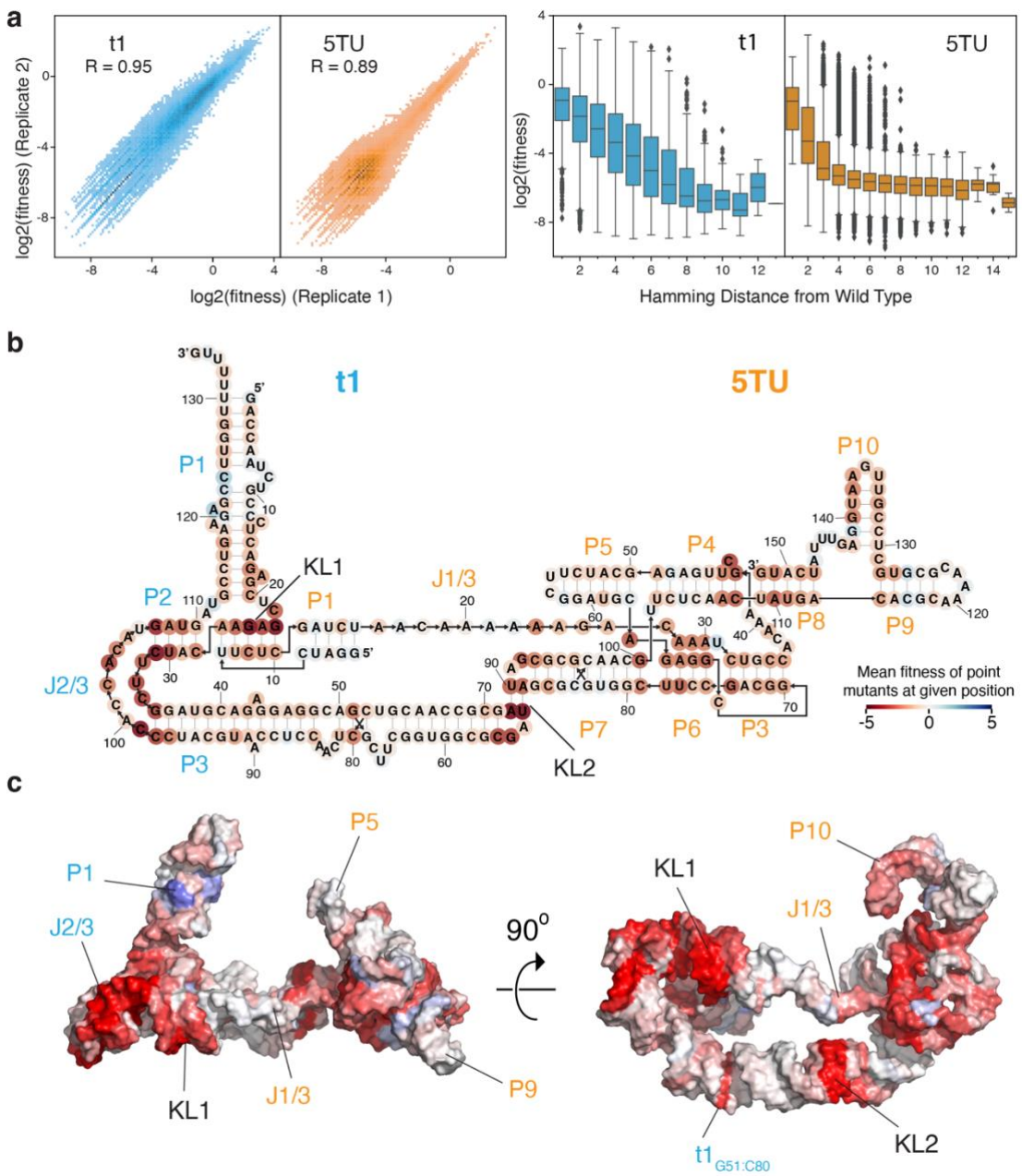

Figure 2. Fitness landscape of the TPR. (**a**) Reproducibility of fitness values between two replicates of t1 (cyan) and 5TU (orange) (for other replicates see SI Fig. 11), and fitness values as a function of Hamming mutational distance for t1 (cyan) and 5TU (orange) (n=3). (**b,c**) Average fitness values for a given nucleotide position in TPR secondary structure (b) and tertiary structure (c) (see SI Movie 2).

Next, we analysed the dataset for global properties and concordance with established TPR function (Fig. 2a). While mean fitness of both 5TU and t1 mutants was negatively correlated with Hamming distance from wild-type (wt) sequences (Fig. 2a), the fitness decline was noticeably steeper for 5TU than in t1. Furthermore, while the majority of 5TU genotypes showed a much-reduced fitness compared to wt, the t1 fitness distribution - while also negatively skewed - was considerably flatter (SI Fig. 11a). These results are consistent with the highly-evolved catalytic 5TU subunit

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

occupying a steeper fitness peak (in a more rugged adaptive landscape) compared to the more recently evolved t1 accessory subunit. Fitness landscape analysis furthermore revealed the functional relevance of both known structural features of functional importance in cIL[13] and novel structural features that are unique to the TPR (Fig. 2b,c, SI Fig. 12). Known structural features include: the template-binding nucleotides in J1/3 (positions 23-24), the active site cytidine in P4 (position 43), and the P6 triple helix-forming adenosines (positions 28-30). Novel features of importance to TPR function include: the P10 stem (positions 137-140), the kissing-loop interactions (KL1 and KL2) between the two subunits, as well as the internal loop region of t1:J2/3,J3/2 (positions 99-106, 32-34) and a G-C base pair (bp) in t1:P3 (position 51 and 80). These will be discussed below in relation to the structural analysis.

In addition to the near complete set of all possible TPR single mutations, our data also contained a large number of double and higher-order mutations, enabling analysis of their interactions (epistasis). Analyzing double mutants, we found that significant epistatic interactions in both 5TU and t1 were negatively biased (SI Fig. 13, 14) and rarer in t1 than in 5TU. Moreover, as the distance between residues increased (as calculated from our structural model of the apoenzyme), both the proportion of significant epistatic interactions and the magnitude of epistasis, decreased in both subunits (SI Fig. 15b). Finally, we found that the average epistatic value decreased as the fitness of the first point mutation increased in double mutants of both 5TU and t1 (SI Fig. 15a). All of these trends are consistent with previously determined fitness landscapes of a yeast tRNA[30], and snoRNA[31], suggesting that they may represent general features of RNA evolution.

Although our dataset does not comprehensively capture all double mutants in either ribozyme subunits, we nonetheless found a large number of double mutants at base-pairing positions which exhibit positive epistasis, particularly within t1 lending support to our structural model (SI Fig. 16a). Moreover, at base-pairing positions predicted by our secondary structure models of 5TU and t1, point mutations that result in a wobble base pair were consistently higher in fitness compared to base pair-disrupting point mutations (SI Fig. 16b). Finally, our data also revealed striking differences in the respective mutational tolerance of the two subunits (SI Fig. 11a). Even though 5TU and t1 share large stretches of identical sequence within their core domains (SI Fig. 5), they are evidently subject to very different selective pressures with 5TU much nearer to a fitness peak than t1, and epistasis governing 5TU evolution to a stronger extent than t1 (SI Fig. 13,14).

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dual dimerization pre-organizes template-binding site

Having validated our structure model and a functional sequence map, we sought a better understanding of the structural elements that support RNA synthesis by the TPR. The most striking feature is that the two divergent subunits are connected through two distinct kissing loop (KL1, KL2) interactions (Fig. 3a). KL1 tethers the long single-stranded segment of 5TU:J1/3 by forming a 5-bp interaction between the loop of the 5TU:P1 hairpin and the t1:J1/2 region, which stacks coaxially with t1:P2 (Fig. 3b,c) – reminiscent of a branched KL[32]. One of two nucleotides connecting P1 and the KL1 domain, t1:C22, has no obvious base pairing partner, though displays a strong effect on fitness of the TPR (SI Fig. 12a). Just before KL1, 5TU:G11 remains base stacked within the 5TU:P1 cap helix with a distinct lack of signal in our EM-map where its predicted base pairing partner, 5TU:C5, is expected to be (SI Fig. 17a). Instead, 5TU:C5 appears to be flipped out and stacking with t1:C22, stabilizing the base of the t1:P1 helix (SI Fig. 17b).

The other KL interaction (KL2) forms a second contact point between the two subunits, a 2-bp loop-loop interaction between the apical loops of 5TU:P7 and t1:P3 (Fig. 3d,e) and by its geometry enforces a rigid, extended conformation of the single stranded 5TU:J1/3 segment clearly visible in electron density in both our structures (Fig. 1c, SI Fig. 9). The stretching and spatial orientation of the 5TU:J1/3 could be important for template recognition and orientation in the active site. Similar KL interactions have previously been observed in the structures of dimeric ribozymes (Varkud), riboswitches (glycine riboswitch) and retroviral RNA genome dimerization (HIV-1, MoMuLV) (reviewed in [33]). Indeed, the MoMuLV structure was so strikingly similar to the TPR KL2 that it could be inserted directly into our EM map and used as a starting point for modelling. Importantly, heterodimer formation is essential for full triplet polymerase activity (Fig. 1b) and primer/template interaction enabling RNA synthesis activity without template tethering, which is obligatory for most other polymerase ribozymes[4].

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

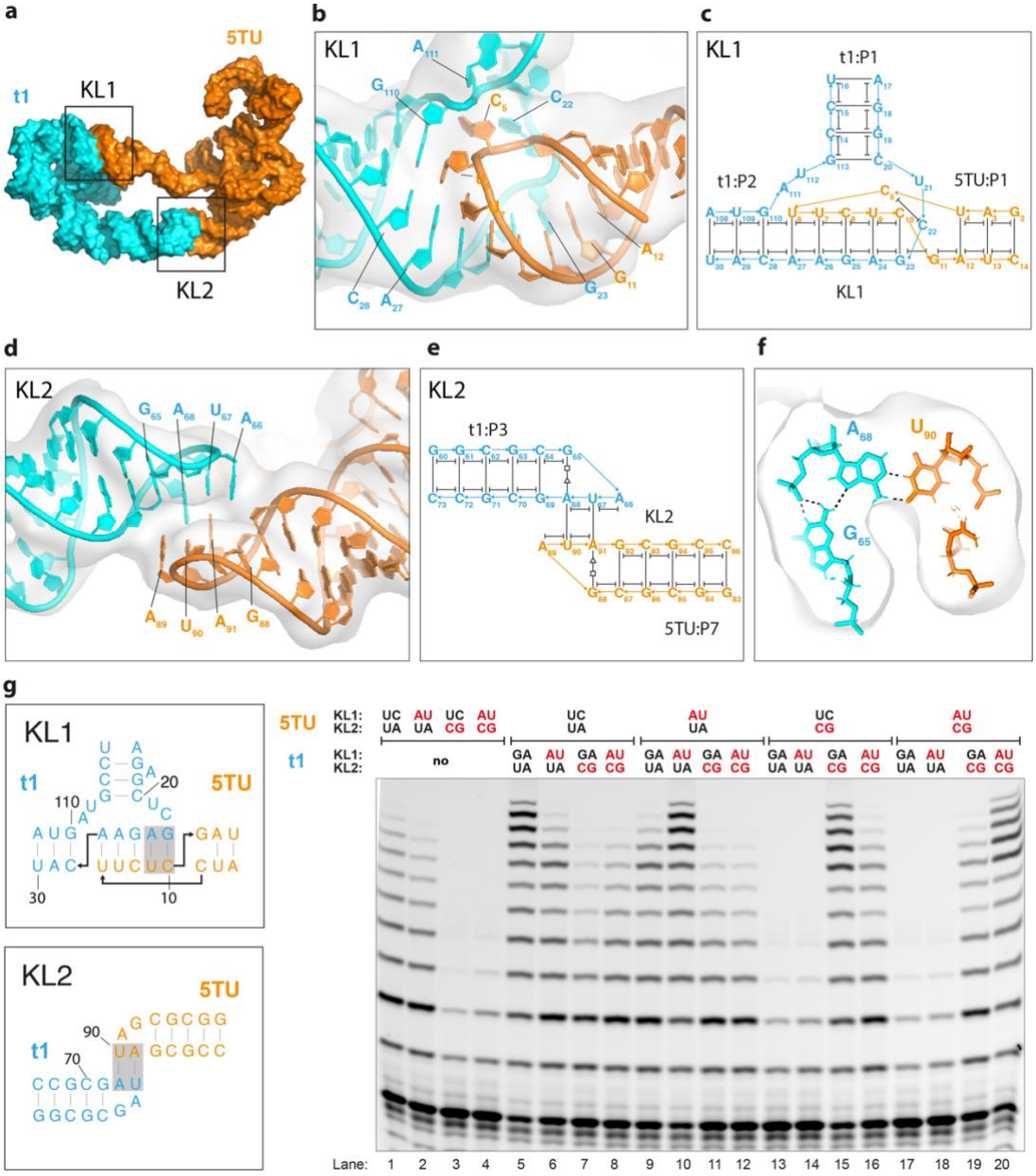

Figure 3. Kissing loop structure and interaction. (**a**) Model of TPR with 5TU in orange and t1 in cyan. KL1 and KL2 indicated by boxes. (**b**) Tertiary structure of KL1 with 5TU coloured orange and t1 coloured cyan and the EM volume as a transparent grey surface. (**c**) Structure diagram of KL1 showing base pairs (lines) and stacks (capped lines). (**d**) Tertiary structure of KL2 with 5TU coloured orange and t1 coloured cyan and the EM volume as a transparent grey surface. (**e**) Structure diagram of KL2 showing base pairs (lines) and stacks (capped lines). (**f**) Structural detail of KL2 showing triple base pair interaction of G65. (**g**) TPR primer extension activity of wild-type as well as mutant KL sequences. Left shows KL1 and KL2 mutations. Right shows primer extension gel electrophoresis, where mutation of 5TU KL1 and in particular KL2 reduce TPR activity, but activity can be restored by compensating mutations in cognate loops in t1.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Fitness landscape analysis provides clear evidence for the functional importance of the KL interaction (Fig. 2, SI Fig. 12). For example, KL1 base pairing between 5TU:U7-C10 and t1:G23-A26 shows a clear functional signal, since mutation in these regions are detrimental, and provides evidence of base pairing, since mutations to the wobble GU is less severe (SI Fig. 18a). As a direct consequence of the shared evolutionary ancestry between 5TU and t1, KL2 is composed of identical sequence GAUA between the terminal loops of 5TU:P7 and t1:P3 (SI Fig. 10). Correspondingly, the fitness of 5TU and t1 point mutants in the two loops are virtually identical (SI Fig. 18b). Besides the two inter-strand base-pairing residues, 5TU:C87-G88 and t1:C64-G65 also appear to be very sensitive to mutation. This pattern is consistent with the GAUA KL of the TPR structure and the GACG KL of the MoMuLV NMR structure[34]. In both cases the first G of the tetra loop forms a non-Watson-Crick bp with the purine involved in the inter-strand bp (Fig. 3f), and stacks on the previous pyrimidine base, which according to the fitness data can accommodate a wobble bp. 5TU:G92-C93 and t1:G69-C70 are sensitive to purine mutations, which can also be explained by the structural models that show the stacked 5TU:A89 and t1:A66 has a cross strand interaction with the phosphate backbone in this region.

Next, we sought to probe KL contribution to TPR activity by targeted mutation (Fig. 3g). While the activity of 5TU both in the absence and presence of t1 was severely affected by a double mutation in the KL2 core (5TU:U90C,A91G) (lane 3 and 13), a KL1 double mutant (5TU:U9A,C10U) only modestly impacted activity (lane 2 and 9), maybe due to the remaining base-pairing interactions in KL1. Nevertheless, in both cases (including KL1,2 double mutants), compensatory mutations in t1 KL1 (t1:G23A,A24U) and t1 KL2 (t1:U67C,A68G) restored TPR activity to “wild-type” levels (Fig. 3g, lane 10, 15, 20), showing 1) the importance of the KL1 / 2 interactions irrespective of the precise sequence and 2) confirming the importance of the central two base-pairs, in particular in KL2, in stabilizing the fully active TPR heterodimer configuration.

Conserved and extended features of 5TU

Despite of several mutations and unique 5’ and 3’ extensions, the catalytic core of the 5TU subunit retains close resemblance to the original cIL structure (Fig. 4a,b, SI Fig. 19). Indeed, the fitness landscape analysis revealed a low tolerance for mutation in congruence with hallmark features of the cIL function (Fig. 2, SI Fig. 12) including key tertiary interactions that form the putative active site (Fig. 4b,c)[13,14]. The overall arrangement of the two coaxially stacked helix segments, P5-P4 and P7-P6-P3, that serve as a scaffold for the active site, is preserved in the absence of the cIL:P1-P2

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

substrate helix. However, the 5TU EM map suggests that the two main connections (J3/4 and J5/6,J7/4) between the two helix segments are different from those found in cIL.

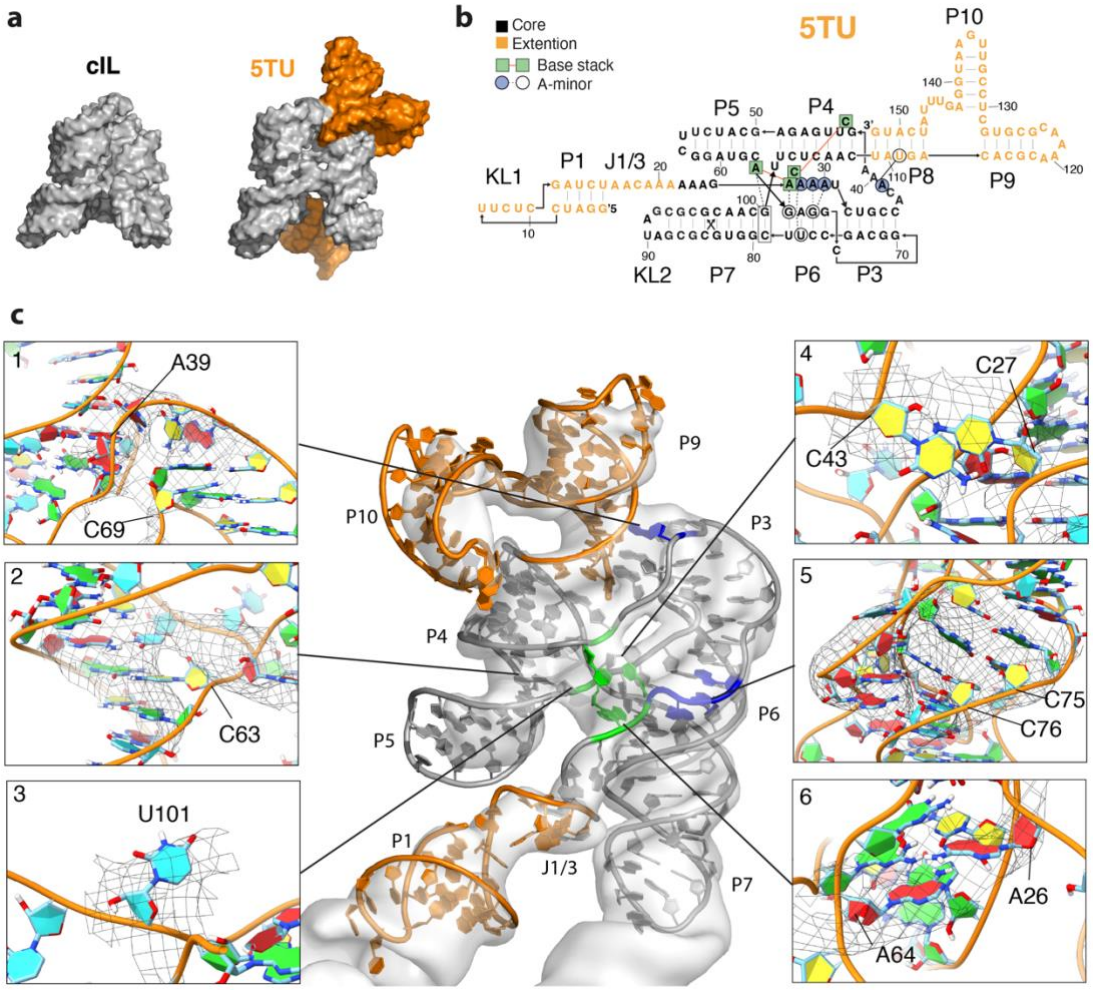

Figure 4. Structural features of 5TU. (**a**) Comparison of class I ligase (cIL) and 5TU, where the core domain is grey and extension domains of 5TU are orange. (**b**) Secondary structure model of 5TU showing core domain (black), extension domains (orange), active site base stacks (green boxes) and A-minor motifs (blue/white circles). (**c**) 5TU subunit model shown in cryo-EM map coloured similar to b. Detailed views of core nucleotides with EM map shown as a mesh selectively 3 Angstrom around the residues of interest: (1) J3/4 with C69, (2) unpaired C63, (3) flipped out U101, (4) C27-C43 base stack, (5) the P6 triple helix, (6) A26-A64 base stack. Refined model is shown with a ribbon cartoon backbone and bases coloured by identity (Yellow - Cytosine, Green - Guanine, Red - Adenine, Cyan - Uracil).

In the cIL, J3/4 formed a hairpin cap at the end of P3 and, with one bridging G, entered P4 (SI Fig. 19a). The sequence of J3/4 in 5TU is different, and DRRAFTER modelling of this junction shows that J3/4 appears to insert into the minor/shallow groove of the P8 helix of the new 5TU accessory domain (Fig. 4c, panel 1). While the

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cIL has a GU base stack (cIL:G45,U76) between J3/4 and J6/3 important for connecting the top of the tripod, this was not predicted by our DRRAFTER model, where instead the corresponding nucleotides in TPR (A41 and C69) are relatively distant, and were further not severely affected by point mutations (Fig. 2). Out of the entire 5TU:J3/4 only A39 is functionally conserved and our strand path places it near the minor groove of 5TU:P8, hinting that it could form an A-minor type interaction with the proximal U111:A150 bp – thus substituting the function of the GU base stack.

In 5TU the antiparallel crossover of J5/6 and J7/4 differs from cIL by a A50G mutation in P4, which provides a base paring partner for C63. However, mutation to C63 is well tolerated (Fig. 2, SI Fig. 12), and although C63 is shown as base paired in our energy minimized model, the EM map has a conspicuous absence of signal at this residue (Fig. 4c, panel 2), which suggests that C63 is unpaired as in cIL. 5TU also has an unpaired nucleotide, U101, between P7 and P4 at the crossover junction that is base-paired in the cIL ribozyme[13]. We found U101 to be tolerant of mutation to any other residue (SI Fig. 12), and the EM map shows weak signal that can fit a flipped-out nucleotide in this location (Fig. 4c, panel 3), indicating that the base does not form any key interactions. Despite of the differences of J3/4 and J5/6,J7/4 in 5TU and cIL, the spacing between the connection sites is similar, which indicates that the overall conformation of the core domain is preserved to serve as a scaffold for the active site.

The key tertiary interactions that form the putative active site (Fig. 4b,c) are supported by the EM map, DRRAFTER modelling and fitness data. The core triple helix region composed of P6 and the A-minor triad (A28-A30) is observed in our EM map (Fig. 4c, panel 5) and was modelled by DRRAFTER. The fitness data show that A67 and U77 of P6 are highly sensitive to mutation, as they directly contact A28-A30. Additionally, A-minor interactions of A26 and A64 with P6 and P7 are similarly sensitive to mutation. In contrast, C75, C76 and U78 of P6 that are not involved in A-minor interactions are less sensitive; in fact, both C75U, C76U and U78C are as functional as wild-type, presumably because they preserve the helical structure by allowing wobble pairing. The A26:A64 and C27:C43 stacks from the putative active site are also apparent in our EM map (Fig. 4c, panel 4,6), where A64 and C43 are more sensitive to mutation than the A26 and C27 nucleotides, but mutation to any of these is detrimental to fitness.

Of the domains unique to 5TU, P8 and P9 extends the coaxial stack of P5-P4, while P10 projects from the side of the helix and bends towards the active site (Fig. 4c). P10 is considerably less tolerant to mutations compared to P9 (or P5 and P7) (Fig. 2, SI Fig. 12), consistent with the hypothesis that P10 (formerly termed “epsilon” domain[11]) may mediate interactions with the incoming trinucleotide triphosphate

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(triplet), while P9 and P5 point away from the active site and, likely, have little involvement in interacting with the primer/template duplex or triplet substrate.

Structural features of the scaffolding t1 subunit

To understand how 5TU interactions with t1 enabled full TPR activity, we investigated the structure of the t1 accessory subunit. The t1 subunit adopts a unique tripod-like structure (Fig. 5a): The t1:P2 and t1:P3 form two parallel legs connected by a U-turn motif formed by the J2/3 and J3/2 joining regions. The t1:P2 leg is extended by coaxial stacking by KL1 and 5TU:P1 leading into the single stranded 5TU:J1/3. The third leg is formed by t1:P1, which is connected to the other two helices by the branched KL1 to t1:P2 and by an A-bulge tertiary contact to P3. The P3 helix forms a long stem with a characteristic 120° bend and connects rigidly to 5TU by KL2. Our t1 structure model is furthermore supported by fitness landscape and epistasis analysis of both standard and non-canonical base pairs and stacks (Fig. 5b).

The U-turn of J2/3 and J3/2 is a tertiary motif with several non-Watson-Crick base pairs and stacking interactions (Fig. 5c). The J2/3 and the terminal regions of P2 and P3 that lead into J2/3, are highly sensitive to mutation, supporting the idea that these interactions may be important for t1 folding and function (Fig. 2, SI Fig. 12). Indeed, a clear trend observed in the fitness landscape is the strong conservation of the pyrimidine tracts on the outside edge of the helices leading into the junction (t1: U30-C34, U96-C99) (marked in yellow in Fig. 5d). These result in a narrowing of the shallow groove and a bending of the outside (5'-3') strands towards each other. Further narrowing the helix at the junction appears to be caused by the presence of two non-canonical pyrimidine-pyrimidine base pairs t1:C34,C99 and t1:U32,U106. Interestingly, t1:C34,C99, which sit at the apex of the turn and the last bp in P3 (Fig. 5d, panel 1), are two of the most invariant nucleotides in the structure with mutation of either to a G being highly deleterious (SI Fig. 12). In contrast, t1:U32,U106 in P2 are both more tolerant to mutation as long as they maintain the outer pyrimidine tract is maintained (i.e. mutation of U106A and U32C are neutral, while U32A is deleterious) (SI Fig. 12). Both P2 and P3 helices have a moderate preference for adenines (A103 and A100, respectively) that stack on the end of the helices and the unpaired nucleotides C101 and C104 of J3/2 also have a strong pyrimidine bias (SI Fig. 12).

The t1:P3 helix is interrupted by a noncanonical G42,A91 base pair and a A43,A90 double bulge (Fig. 5d). The connectivity in our EM-map between t1:P1 and this region in t1:P3 remains until high contour levels, suggestive of a tertiary interaction (Fig. 5d, panel 2). In our model, the A43,A90 double bulge is positioned appropriately to allow the unpaired t1:A18 from the t1:P1 helix to flip out and insert into the bulge, stacking

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

between A43 and A91 (Fig. 5d). This presumed A18,A43 stacking interaction is furthermore supported by fitness landscape analysis as both A18C and A43U mutations are individually deleterious, but jointly restore fitness. The A-bulge is placed approximately one helical turn from the J2/3 crossover of the U-turn and are likely cooperative interactions that reinforce this unique conformation.

The P3 helix also contains two asymmetric 4-nt bulges, t1:C49,C82-A84 and t1:C53-U55,U78 (Fig. 5e), that are clearly visible in the EM map as distinct holes in the helix region (Fig. 5f, panel 1 and 2). The first bulge is modelled as a bifurcated bp where C49 H-bonds with both C82 and A84 resulting in a 60-degree bend. The second bulge is modelled as a bifurcated bp where U78 H-bonds with both C53 and U54 resulting in a 60-degree bend. Being placed with a spacing of a half turn the two bulges together result in a 120-degree bend of the helix. Mutations in these bulges does not appear to affect fitness to a large degree indicating that the asymmetry of the bulge with 1 nt across from 3 others are most important for maintaining the shape. However, the G51,C80 bp between the two bulges is highly sensitive to all mutations, except C80U, which would enable a G51,U80 wobble pair. Moreover, the G51C / C80G genotype is one of the strongest positive epistatic interactions within t1 double mutants. These data suggest that this bp may act as a crucial “clamp” between the t1:C49,C82-A84 and t1:U78,C53-U55 bulges to stabilize the structure and geometry of t1:P3 with respect to KL2.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

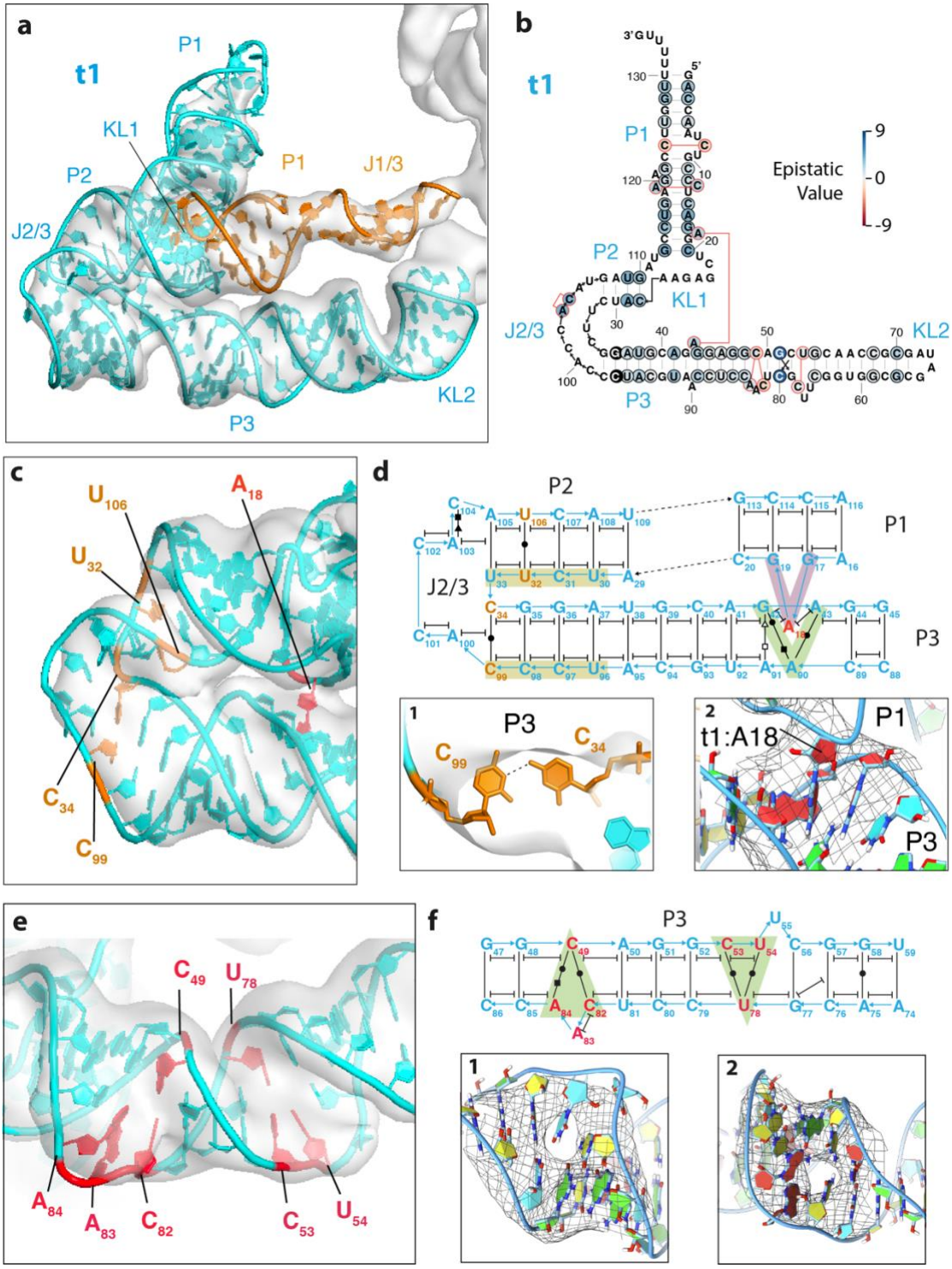

Figure 5. Structural features of t1. (**a**) EM map with t1 model (cyan) and 5TU:P1 and J1/3 (orange). (**b**) Epistasis of standard and non-canonical base pairs shown on secondary structure of t1. (**c**) U-turn motif of J3/2 highlighting the noncanonical C-C and U-U base pairs and A minor insertion motif between the t1:P1 and t1:P3 adenines. (**d**) Secondary structure diagram showing base pairs and stacks of U-turn and A-insertion motifs. Panel 1 shows UU bp. Panel 2 shows A-insertion motif. (**e**) Bulged regions of t1:P3. (**f**) Secondary structure diagram showing base pairs and stacks of 120-degree turn. Panel 1 and 2 shows observed gaps in the helical density.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The EM map of the t1:P1 stem is the least resolved region of the TPR structure, but its general base paired structure and bulges are supported by epistatic base changes (Fig. 5b). The low resolution of this region could be due to dynamics of t1:P1. 3D variability analysis suggests a continuous distribution of particles amongst our data that appear to represent dynamic movements within the TPR. We have attempted to reconstructed volumes using the particles from the tailing and leading edges of this distribution, which suggest the major movement is in the t1:P1 helix and the 5TU:P4 helix, with minor distortions observed elsewhere in the structure (SI Fig. 8). The t1:P1 helix appears to be supported at its base by two key tertiary interactions (the KL1 and A-bulge) that form a hinge allowing the large dynamic movement of t1:P1 (SI Movie 1). Because of the orientation of the hinge, the movement of the t1:P1 is in the direction of the 5TU active site. Based on the structural analysis we provide a full map of secondary and tertiary contacts within TPR (SI Fig. 20).

Model of the TPR holoenzyme

To further investigate the functional properties and build a model of templated RNA synthesis by the TPR holoenzyme, the catalytic 5TU subunit was first aligned to the cIL ribozyme crystal structure[13]. Comparison of 5TU to the cIL structure reveals that the active site and helices P3-P7 retain similar positioning relative to one another and even the placement of the long single stranded J1/3 appears conserved in the 5TU+t1 apoenzyme despite the loss of the intramolecular loop to the template helices cIL:P1-2 (see SI Fig. 19). Indeed, when only the P1 & P2 helices are removed from cIL, the structure fits in our 5TU cryo-EM map with a correlation coefficient of 0.8. We therefore built a model of the holoenzyme by aligning an elongated template-product helix to the P1 substrate helix in the clL structure (Fig. 6a, SI Fig. 21,22). Remarkably, this simple model allows placement of the primer-template duplex and triplet substrate 5' end in close proximity to features of the 5TU subunit known to interact with them, such as J1-3 segment, the active site, and the P10 domain (see below).

A notable feature of the TPR observed previously is its fidelity of 97% (per nucleotide position)[11], which is significantly higher than would be expected based on simple triplet binding thermodynamics. A significant contribution was ascribed to the P10 (formerly epsilon) domain that appears to enhance fidelity from the baseline 92% of a ΔP10 TPR[11]. Single atom replacement studies in the substrate triplet indicated that this fidelity boost likely relies on H-bonding with the shallow (minor) groove of the 3' base of the incoming triplet. More recent, functional data suggests, that the P10 domain may make even more extensive interactions. When only a single triplet is bound to the template 3' of the ligation junction, the P10 fidelity boost is lost but

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

regained in the presence of a second downstream triplet. However, in the absence of a downstream triplet, but using substrates of increasing length, P10-dependent fidelity gains are almost entirely restored when using a quadruplet ($pppN_4$) substrate, mimicking a duplex at only the first position of the downstream triplet, with minimal further fidelity increases seen upon incorporation of longer ($pppN_5$, $pppN_6$) substrates (SI Fig. 23). This suggests that the P10 domain forms functionally important contacts with the primer-template duplex extending at least 4 nucleotides downstream from the primer 3' end and the ligation junction. Indeed, our model positions P10 and specifically U135, G136 & A137 (Fig. 12, SI Fig. 19) in close proximity to the triplet substrate bound to template poised for interaction with the shallow groove (Fig. 6).

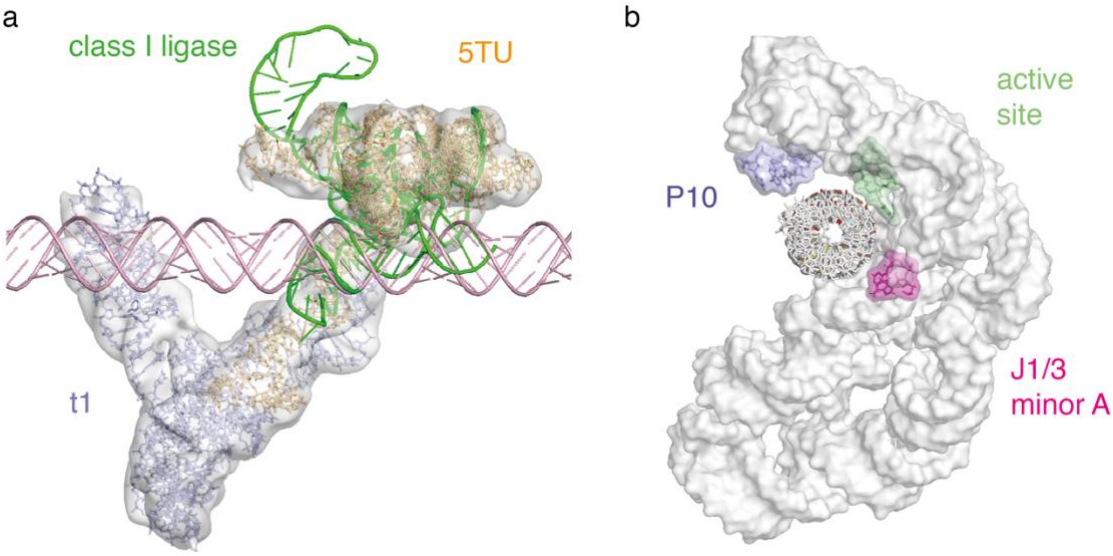

Figure 6. Structural model of the TPR holo-enzyme from alignments to the class I ligase. (**a**) TPR model (5TU (orange), t1 (blue) fitted into contour map with best fit alignment of class I ligase structure (and extruding U1A binding loop) (green ribbon) and an idealized double-stranded RNA template (pink) aligned to class I substrate helix. (**b**) Side of space-filling TPR model with idealized double-stranded RNA template (white) and putative contact sites active site (C43 (chartreuse)), J1/3 A-minor interaction (magenta) and P10 shallow groove interaction (pale blue).

Another remarkable feature of the TPR is its capacity to support non-canonical RNA synthesis modes such as triplet polymerization in the reverse 3'-5' direction[11]. Analysis of the 3'-5' mode of templated RNA synthesis by the TPR using deep sequencing (FidelitySeq, SI Fig. 24) suggests that - in contrast to the 5'-3' reaction - fidelity is reduced to 84%, even below the baseline fidelity of 5'-3' synthesis in the absence of the P10 domain (SI Fig. 25). Although the measured 3'-5' error rate may be both sequence-dependent and inflated by non-ribozyme derived errors from the sequencing workflow due to poor incorporation AU-rich triplets (SI Fig. 26), it is clear

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that TPR fidelity is significantly reduced for 3'-5' compared to 5'-3' synthesis. This loss of fidelity can now be rationalized in the light of our holoenzyme model as in the reverse 3'-5' mode (with the triplet triphosphate moiety positioned in the active site) P10 can neither interact with (nor stabilize) the substrate triplet, but instead is positioned to interact with the upstream (3') primer with no impact on triplet incorporation (SI Fig. 27).

Evolution of a mutualistic heterodimer

The structure of the 5TU+t1 TPR comprising a catalytic (5TU) and non-catalytic (t1) subunit (derived from the same progenitor) has interesting analogies with proteinaceous polymerases such as the HIV reverse transcriptase (RT) holoenzyme heterodimer. In HIV RT the non-catalytic p55 subunit appears to aid activity of the catalytic p65 subunit by positioning the primer/template duplex for optimal processive synthesis (SI Fig. 28). It is tempting to speculate that the non-catalytic t1 RNA subunit may serve a similar function. Indeed, our holoenzyme model (Fig. 6) indicates that RNA templates of 30 nucleotides (or longer) would be able to interact with t1:P1. From the 3D variability analysis (SI Movie 1) it is tempting to speculate that the hinge-like motion of t1:P1 could allow for docking of the template, followed by scanning for the correct positioning of the ligation junction near the active site.

We hypothesize that another role of the t1 accessory subunit may be to pre-configure 5TU:J1/3 in a conformation that allows for productive template docking. Indeed, the extended and rigid conformation of the single-stranded J1/3 linker segment is a noteworthy and unanticipated feature of the TPR structure. J1/3 is of particular interest because the equivalent positions to 5TU:A22-A24 are implicated in A-minor interaction with the substrate helix in the clL structure[13]. By analogy, one might expect there to be similar interactions in the TPR holoenzyme with the primer-template duplex. Indeed, our holoenzyme model positions the PT helix in proximity to J1/3 (Fig. 6, SI Fig. 27). Furthermore, functional data strongly suggests that the extended A-minor triad conformation (rather than the precise sequence of J1/3) is essential for full TPR function via enhancement by the t1 subunit. Lengthening or shortening this single-stranded region by as few as two nucleotides reduces TPR activity to the Δt1 baseline (SI Fig. 29). Thus the t1 domain and its KL interactions may together serve to hold J1/3 in this out-stretched conformation, as a longer or untethered single-stranded template-binding strand would likely be more dynamic and adopt a variety of conformations, increasing the entropic cost of template interaction. Analysis of the evolution of the related 52-2 polymerase ribozyme (which used NTPs as substrates)[7] suggest the emergence of a pseudoknot structure involving P7 and the J1/3 equivalent, which

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

might enhance PR activity via a similar restriction of the conformational freedom of this crucial sequence segment.

The TPR structure also reveals a number of RNA motifs not previously encountered. For example, the U-turn motif at t1:J2/3 appears to be a unique motif that combines a symmetrical incorporation of pyrimidine tracts with two pyrimidine-pyrimidine base pairs to create a tight turn (Fig. 5c,d). The t1:C34-C99 bp at the apex of this motif is extremely uncommon, with only 6 previous occurrences in the RNA structural database[35]. Furthermore, in each of the previous cases the C-C pair is in the middle of a co-axially stacked helix. NMR studies suggest C-C pairs to be highly mobile, often switching between one C as the H-bond donor to the acceptor and have been shown to be critically important for the activity of the HCV IRES[36] as well as the paromomycin binding motif[37]. Another interesting structural feature is the 5-bp branched kissing-loop (bKL) that connects 5TU and t1 (KL1, Fig. 3). Much like the 6-bp designer bKL structures recently described[38], this bKL is stabilized at its base by a trans-base stacking interaction (C5-t1:C22). In this case the branched helix (t1:P1) is further stabilized by a second tertiary interaction with P3, an apparent A-minor insertion. These A-bulge stabilizations appear to be a common feature of evolved RNA structures that have yet to be utilized in designer RNA structures. Incorporation of a 5-bp bKL with t1:J3/2 and accompanying A-minor insertion could offer a new motif for expansion of the RNA origami architecture[39].

The characterization of this motif also offers a potential explanation for the emergence of the mutualistic interaction between the catalytic and accessory subunits during *in vitro* evolution[11]. In the t1 progenitor RNA, the 3' sequence extension triggered a wholescale reorganization of the tertiary fold, abolishing its catalytic activity. Serendipitously, this exposed an RNA sequence capable of forming a kissing loop interaction with all other members of the selection library, which positioned the t1 5' selection cassette near to the active site of a bound catalytic subunit (5TU / t5 progenitor), allowing for mutualistic exploitation of its activity by t1. Over the course of the selection experiment, t1 gained further mutations to better associate and co-evolve with catalytically active subunits, and, in turn, active subunits that could exploit t1 complex formation thrived[11]. Thus, mutualism and eventual molecular symbiosis between the two subunits may have emerged by co-optation of an RNA parasite.

In conclusion, our results describe the structure and comprehensive structure-function analysis of the 5TU+t1 triplet polymerase ribozyme, a class of ribozyme for which no previous structures had been described. Our data provide a framework for a better molecular understanding of polymerase ribozyme function and RNA-catalyzed RNA replication, an enzymatic activity widely considered to be fundamental for the

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

emergence of life's first genetic system. Finally, our structure reveals structural and functional motifs with potential for applications in the construction of RNA nanotechnology objects and devices.

Acknowledgements:

We thank our colleague K. Nguyen (MRC LMB) for helpful comments on the manuscript. The research at iNANO AU was supported by the Independent Research Fund Denmark (9040-00425B) (EKSM, ESA), the Novo Nordisk Foundation (NNF21OC0070452) (EKSM, ELK, KH, ESA), a fellowship from the Canadian Natural Sciences and Engineering Research Council (532417) (EKSM), a Carlsberg Foundation Research Infrastructure grant (CF20-0635) (ESA), and a Lundbeck fellowship (R250-2017-1502) (ELK). The research at MRC LMB was supported by the Medical Research Council, as part of United Kingdom Research and Innovation (also known as UK Research and Innovation (UKRI)) [MC_U105178804] (CJKW, EG, IG, JFC, JA, PH), a grant from the Volkswagen Foundation (96 755) (EG), a Herchel Smith studentship (2017) (CJKW), a Marie Curie fellowship (H2020-MSCA-IF-2018-845303) (IG), a Carlsberg fellowship (CF17-0809) (ELK). For the purpose of open access, the MRC Laboratory of Molecular Biology has applied a CC BY public copyright license to any Author Accepted Manuscript version arising.

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Gesteland, R. F., Cech, T. R. & Atkins, J. F. *The RNA World* (Cold Spring Harbor Laboratory Press, 2005).

2. Gilbert, W. Origin of Life - the RNA World. *Nature* **319**, 618-618 (1986).

3. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001). https://doi.org/10.1126/science.1060786

4. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209-212 (2011). https://doi.org/10.1126/science.1200752

5. Attwater, J., Wochner, A. & Holliger, P. In-ice evolution of RNA polymerase ribozyme activity. *Nat Chem* **5**, 1011-1018 (2013). https://doi.org/10.1038/nchem.1781

6. Horning, D. P. & Joyce, G. F. Amplification of RNA by an RNA polymerase ribozyme. *Proc Natl Acad Sci U S A* **113**, 9786-9791 (2016). https://doi.org/10.1073/pnas.1610103113

7. Portillo, X., Huang, Y. T., Breaker, R. R., Horning, D. P. & Joyce, G. F. Witnessing the structural evolution of an RNA enzyme. *Elife* **10** (2021). https://doi.org/10.7554/eLife.71557

8. Cojocaru, R. & Unrau, P. J. Processive RNA polymerization and promoter recognition in an RNA World. *Science* **371**, 1225-+ (2021). https://doi.org/10.1126/science.abd9191

9. Tjhung, K. F., Shokhirev, M. N., Horning, D. P. & Joyce, G. F. An RNA polymerase ribozyme that synthesizes its own ancestor. *Proceedings of the National Academy of Sciences of the United States of America* **117**, 2906-2913 (2020). https://doi.org/10.1073/pnas.1914282117

10. Tagami, S., Attwater, J. & Holliger, P. Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nature Chemistry* **9**, 325-332 (2017). https://doi.org/10.1038/Nchem.2739

11. Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E. & Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *Elife* **7** (2018). https://doi.org/10.7554/eLife.35255

12. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411-1418 (1993). https://doi.org/10.1126/science.7690155

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

13. Shechner, D. M. *et al.* Crystal Structure of the Catalytic Core of an RNA-Polymerase Ribozyme. *Science* **326**, 1271-1275 (2009). https://doi.org/10.1126/science.1174676

14. Shechner, D. M. & Bartel, D. P. The structural basis of RNA-catalyzed RNA polymerization. *Nature Structural & Molecular Biology* **18**, 1036-U1096 (2011). https://doi.org/10.1038/nsmb.2107

15. Cech, T. R. The RNA worlds in context. *Cold Spring Harb Perspect Biol* **4**, a006742 (2012). https://doi.org/10.1101/cshperspect.a006742

16. Zaug, A. J. & Cech, T. R. The Intervening Sequence RNA of Tetrahymena Is an Enzyme. *Science* **231**, 470-475 (1986). https://doi.org/10.1126/science.3941911

17. Doudna, J. A. & Szostak, J. W. Rna-Catalyzed Synthesis of Complementary-Strand Rna. *Nature* **339**, 519-522 (1989). https://doi.org/10.1038/339519a0

18. Doudna, J. A., Couture, S. & Szostak, J. W. A Multisubunit Ribozyme That Is a Catalyst of and Template for Complementary Strand Rna-Synthesis. *Science* **251**, 1605-1608 (1991). https://doi.org/10.1126/science.1707185

19. Green, R. & Szostak, J. W. Selection of a Ribozyme That Functions as a Superior Template in a Self-Copying Reaction. *Science* **258**, 1910-1915 (1992). https://doi.org/10.1126/science.1470913

20. Doudna, J. A., Usman, N. & Szostak, J. W. Ribozyme-Catalyzed Primer Extension by Trinucleotides - a Model for the Rna-Catalyzed Replication of Rna. *Biochemistry* **32**, 2111-2115 (1993). https://doi.org/10.1021/bi00059a032

21. Ekland, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373-376 (1996). https://doi.org/10.1038/382373a0

22. Kristoffersen, E. L., Burman, M., Noy, A. & Holliger, P. Rolling circle RNA synthesis catalyzed by RNA. *Elife* **11** (2022). https://doi.org/10.7554/eLife.75186

23. Kappel, K. *et al.* De novo computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nature Methods* **15**, 947-+ (2018). https://doi.org/10.1038/s41592-018-0172-2

24. Kappel, K. *et al.* Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods* **17**, 698-+ (2020). https://doi.org/10.1038/s41592-020-0878-9

25. Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. & Holliger, P. Ice as a protocellular medium for RNA replication. *Nature Communications* **1**, 76 (2010). https://doi.org/10.1038/ncomms1076

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

26. Russo, C. J. & Passmore, L. A. Electron microscopy: Ultrastable gold substrates for electron cryomicroscopy. *Science* **346**, 1377-1380 (2014). https://doi.org/10.1126/science.1259530

27. Punjani, A., Zhang, H. & Fleet, D. J. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. *Nat Methods* **17**, 1214-1221 (2020). https://doi.org/10.1038/s41592-020-00990-8

28. Pitt, J. N. & Ferre-D'Amare, A. R. Rapid construction of empirical RNA fitness landscapes. *Science* **330**, 376-379 (2010). https://doi.org/10.1126/science.1192001

29. Jimenez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc Natl Acad Sci U S A* **110**, 14984-14989 (2013). https://doi.org/10.1073/pnas.1307604110

30. Li, C., Qian, W. F., Maclean, C. J. & Zhang, J. Z. The fitness landscape of a tRNA gene. *Science* **352**, 837-840 (2016). https://doi.org/10.1126/science.aae0568

31. Puchta, O. *et al.* Network of epistatic interactions within a yeast snoRNA. *Science* **352**, 840-844 (2016). https://doi.org/10.1126/science.aaf0965

32. Liu, D. *et al.* Branched kissing loops for the construction of diverse RNA homooligomeric nanostructures. *Nat Chem* **12**, 249-259 (2020). https://doi.org/10.1038/s41557-019-0406-7

33. Bou-Nader, C. & Zhang, J. Structural Insights into RNA Dimerization: Motifs, Interfaces and Functions. *Molecules* **25** (2020). https://doi.org/10.3390/molecules25122881

34. Kim, C. H. & Tinoco, I., Jr. A retroviral RNA kissing complex containing only two G.C base pairs. *Proc Natl Acad Sci U S A* **97**, 9396-9401 (2000). https://doi.org/10.1073/pnas.170283697

35. Narayanan, B. C. *et al.* The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Research* **42**, D114-D122 (2014). https://doi.org/10.1093/nar/gkt980

36. Collier, A. J. *et al.* A conserved RNA structure within the HCVIRES eIF3-binding site. *Nature Structural Biology* **9**, 375-380 (2002). https://doi.org/10.1038/nsb785

37. Tavares, T. J., Beribisky, A. V. & Johnson, P. E. Structure of the cytosine-cytosine mismatch in the thymidylate synthase mRNA binding site and analysis of its interaction with the aminoglycoside paromomycin. *Rna* **15**, 911-922 (2009). https://doi.org/10.1261/rna.1514909

bioRxiv preprint doi: https://doi.org/10.1101/2022.08.23.504927; this version posted August 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

38. Liu, D., Shao, Y. M., Piccirilli, J. A. & Weizmann, Y. Structures of artificially designed discrete RNA nanoarchitectures at near-atomic resolution. *Science Advances* **7** (2021). https://doi.org/10.1126/sciadv.abf4459

39. Geary, C., Grossi, G., McRae, E. K. S., Rothemund, P. W. K. & Andersen, E. S. RNA origami design tools enable cotranscriptional folding of kilobase-sized nanoscaffolds. *Nat Chem* **13**, 549-558 (2021). https://doi.org/10.1038/s41557-021-00679-1