# Supplementary Materials for

# Origin of life: protoribosome forms peptide bonds and links RNA and protein dominated worlds

**Authors:** Tanaya Bose<sup>1</sup>, Gil Fridkin<sup>1,2</sup> Chen Davidovich<sup>1</sup><sup>†</sup>, Miri Krupkin<sup>1</sup><sup>‡</sup>, Nikita Dinger<sup>1§</sup>, Alla Falkovich<sup>1</sup>, Yoav Peleg<sup>1</sup>, Ilana Agmon<sup>3, 4</sup>, Anat Bashan<sup>1</sup>, Ada Yonath<sup>1\*</sup>

Correspondence to: ada.yonath@weizmann.ac.il

#### This PDF file includes:

Supplementary Text Figs. S1 to S21 Tables S1 **Figure S1: A two-dimensional representation of the RNA constructs designed based on part of** *T. thermophilus* **23S rRNA around the PTC.** The nucleotides highlighted in yellow were used as the sequences of the RNA constructs tt\_A1, tt\_P1, tt\_A1P1, tt\_A2, tt\_P2 and tt\_A2P2, respectively. Green stretches indicate direct connection between nucleotides and green lines represents positions where the hyper stable CUUCGG hairpin was inserted. rRNA helices and nucleotide numbers (E. coli numbering) are designated in blue and black, respectively. A-loop and P-loop are pointed with red arrows.



**Figure S2: RNA constructs sequence alignment** (G- red, A- green, C- orange and U- blue). **A**. The alignment of A-reg short and long sequences. **B**. The alignment of P-reg minimal and extended sequences. GUGA addition to tt\_P1c and tt\_P2c is marked with a yellow box. **C**. The alignment of long sequences which represent the full proto-ribosome. **D**. The alignment of various P-reg minimal sequences. rRNA sequences were aligned using Clustal omega server (Sievers, Wilm et al. 2011) and imaged using Jalview (Waterhouse, Procter et al. 2009).

			10	20	30	40	50
A	tt_A1 tt_A2	GAUGUCGG G <mark>AU</mark> GUCGG	UCGUCGCUUC UCGUCGCAUC	c <mark>uucggguu</mark> c	GGCUGUUCG	CCAUUAAAGO	GGCACGCGAGC GGCACGCGAGC
			70	80	90		
	tt_A1 tt_A2	UGGGUUCAC UGGGUUCAC	SAACGUCGUGA SAACGUCGUGA	A G A C A G U U C G C A G A C A G U U C G C	UC UC		
			10	20	30	40	50
В	tt_P1 tt_P1c	GAAGACCCC GAAGACCCC	C <mark>UGGAGCU</mark> C <mark>UGGAGCU</mark>	<mark>CUUCGG</mark> - <mark>CUUCGG</mark> -			<mark>AGUUAC</mark> <mark>AGUUAC</mark>
	tt_P2 tt_P2c	GAAGACCCC GAAGACCCC	GUGGAGCUUU GUGGAGCUUU	IACUCUUC <mark>GG</mark> A IACUCUUC <mark>GG</mark> A	GUUUGACUGG GUUUGACUGG	GGCGG <mark>UCC</mark> GG GGCGG <mark>UCC</mark> GG	AUAAAA <mark>GUU</mark> AC AUAAAA <mark>GUU</mark> AC
			70	80	90	100	
	tt_P1 tt_P1c	CCCGGGGGAL	JAACAGGCUGA JAACAGGCUGA	UCUCUUC <mark>GG</mark> A UC <mark>G</mark> <mark>UG</mark> A	GGUUUGGCAC	cuc cuc	
	tt_P2 tt_P2c	CCCGGGGGAL CCCGGGGGAL	JAACAGGCUGA JAACAGGCUGA	UCUCUUC <mark>GG</mark> A UC <mark>G</mark> UGA	GGUUUGGCAC GGUUUGGCAC	cuc cu-	
-				20	30	40	50
C	tt_AIPI tt_A2P2	GAAGACCC GAAGACCC	CGUGGAGCU		A <mark>GUUU</mark> GACUGO	GGGCGG <mark>U</mark> CCGG	AUAAAAGUUAC
			70	80	00	100	110
	tt_A1P1	CCCGGGGA					GCOCCOCCCOC
	tt_A2P2	CCCGGGGA	UAACA <mark>GGCUG</mark>	AUCUCUUCGG	AGGUUUGGCAG	CCUCGAUGUCC	<u> GCUCGUCGCAU</u>
			130	140	150	160	170
	tt_A1P1 tt_A2P2	c cc <mark>uu</mark> c <mark>ggg</mark>	UU <mark>GGGC</mark> U <mark>G</mark> UU	C <mark>G</mark> CCCAUUAA	GGCGGCACGCO AGCGGCACGCO	GAGCUGGGUUQ GAGCUGGGUUQ	CAGAACGUCGUG CAGAACGUCGUG
	tt_A1P1	AGACAGUU					
	tt_A2P2	A <mark>G</mark> A C A <mark>G</mark> U U	<mark>cggu</mark> c				
D	tt_P1	GAAG <mark>ACCC</mark>		20 UUCGGAGUUA	<sup>30</sup> cccc <mark>ggggau</mark>	40 A A C A GG C U G A I	<sup>50</sup> U <mark>CUUCGGAGG</mark>
	ec_P1c tt_P1m	AAAGACCC GAAGACCC	CGUGAACCUC CGUGG <mark>AG</mark> CUU	UUC <mark>GG</mark> AGGUA U AA - <mark>G</mark> UUA	CUCCGGGGGAU CCCCGGGGGAU	A A C A <mark>G G C U G</mark> A I A A C A <mark>G G C U G</mark> A I	U A G U G A U G U C G U G A G G
	tt_P1n tt_P1c	GAAGACCC GAAGACCC		U -  -  A A A G U U A U U C <mark>G G</mark> A <mark>G</mark> U U A	CCCCGGGGGAU CCCCGGGGGAU	A A C A G G C U G A A A C A G G C U G A	UCGUG <mark>A</mark> GG UCGUG <mark>A</mark> GG
	sa_P1c ef_P1c	GAAGACCC GAAGACCC	CGUGGAGCUC CAUGGAGCUC	UUCGGAGCUA UUCGG <mark>AGCU</mark> A	CCCCGGGGGAU CCC <mark>U</mark> GGGG <mark>AU</mark>	A A C A <mark>G G C U U</mark> A A A C A <mark>G G C U U</mark> A	U C G U G A G G U C G U G A G G

**Figure S3: Design schemes of the constructs tt\_P1c, tt\_P1n and tt\_P1m**. **A.** Incorporation of a GNRA tetraloop in construct tt\_P1 afforded tt\_P1c. **B.** Constructs tt\_P1n and tt\_P1m design. Natural shorter UUAA and UUA loops were added to H74 as shown in the upper image.





**Figure S4: 2D structure predictions of the protoribosome constructs.** All predictions were calculated using mfold (*66*) where  $\Delta G$  values are refined at 37<sup>o</sup>C.



Figure S5: 2D structure predictions of the protoribosome constructs. All predictions were calculated using RNA structure (67) where  $\Delta G$  values are refined at 37<sup>o</sup>C.



**Figure S6. Denaturing gels of the RNA constructs: A.** Purified tt\_P1, tt\_P1c, tt\_P2, tt\_P2c, tt\_A1P1and tt\_A2P2 RNA constructs (lanes 4-9). **B.** Purchased ec\_p1c, hm\_p1c, tt\_p1m, tt\_p1n, tt\_A1 and tt\_A2 RNA constructs (lanes 4-9). **C.** Purified sa\_p1c and ef\_p1c RNA constructs (lanes 4-5). **D.** Denaturing gel of the active constructs after fragment reaction of 24h at 37<sup>o</sup>C (lanes 4-8). A single strand RNA (ssRNA) ladder and two DNA ladders (ultralow range (ULR) and 50 base pair (50BP) were used as reference (lanes 1-3). 10% polyacrylamide was used for all gels presented.



**Figure S7: Labeling of RNA oligonucleotides with Fluorecein-5-thiosemicarbazide (FTSC).** Sodium periodate was used to cleave the cis diol (3' termini) of RNA into 2',3'-dialdehyde and then, fluorescein-5-thiosemicarbazide, was added to the quenched reaction mixture to label the 3' terminus of RNA through a condensation reaction between the carbazide and the aldehyde.



**Figure S8: MALDI spectra of the fragment reaction between CCA- pcb and C-Pmn in the presence of the RNA constructs at 37°C. A.** Positive (50S) and negative controls (no RNA) and CHCA matrix. **B.** Constructs tt\_A1, tt\_P1. **C.** Constructs sa\_P1c, tt\_P1c. **D.** Constructs tt\_A1P1, tt\_A2P2. **E.** Constructs tt\_A2, ec\_P1c. **F.** Constructs tt\_P1n and tt\_P1m. **G.** Constructs tt\_P2 and tt\_P2c. **H.** Construct cm\_tt\_P1c. The starting material CCA- pcb and the product C-Pmn-pcb are designated by S and P throughout.

#### Α





В













F



G





606.2431 100-1261,4913 tt\_A1P1 417.1096 1262.4962 2 568.3958 1003,3989 439.0918 607.2462 509.7330 1263.4962 1004.3997 1020.4829 739.5688 1291.5134 645.7545 739.7173 835.6095 919.6215 750 800 850 900 950 478,6114 450 500 <u>1119.9836</u> 1050 1100 1150 1200 0 <del>| ..</del> 400 650 1300 1350 m/z 550 700 т 600 1000 1250 606.2445 100tt\_P1 1003.3989 % 1004.4014 1291,5032 607.2458 751.3771 1261.4913 1,292.4994 476.9836 628.2240 745.4234 1005.4073 1,293.5166 934.3232 900 950 573.2935 1300 1350 m/z 550 600 650 700 500 800 850 1050 1100 1150 1200 1250 750 1000

Figure S9: LC-MS spectra of A. tt\_A1P1 and tt\_P1 B. sa\_P1c and tt\_P1c

A



В



Figure S10: UPLC chromatograms of 50S, ef\_P1c, tt\_P1, sa\_P1c, tt\_A1P1, tt\_P1c





Figure S11: Verification of the product identity; presence of a peptide bond. A. Schematic representation of the product basic hydrolysis. B. MALDI spectra of the reaction mixture of construct tt\_P1c before and after base treatment.



Α







Figure S12: MALDI spectra of the fragment reaction for various time durations A. tt\_P1c B. 50S.

### Α









Figure S13: MALDI spectra of fragment reactions using 10, 20, 30, 40, 50 and 100  $\mu$ M concentration of construct tt\_P1c.





Figure S14: MALDI spectra of the reactions at 2.8  $\mu M$  concentration of tt\_P1c and different substrate concentrations.







В







**Figure S16: MALDI spectra of the reaction mixtures at 25°C. A.** Positive control (50S) and the construct tt\_A1P1. **B.** The constructs tt\_P1c and tt\_P1. **C**. The constructs ef\_P1c and sa\_P1c.







## В



**Figure S17: MALDI spectra of the reaction mixtures at 50°C. A.** positive control (50S) and the construct tt\_A1P1 **B.** The constructs tt\_P1c and tt\_P1. **C.** The constructs sa\_P1c and ef\_P1c.



Α







С

Figure S18: EtBr stained native gel of RNA constructs under various controlled conditions of dimerization. A-C. tt\_P1c, tt\_P1, and tt\_A1P1, respectively in the absence of salts. D. tt\_P1, tt\_P1c, and tt\_A1P1 with 30 mM KCl and no  $Mg^{2+}E$ . tt\_P1c, tt\_P1, and tt\_A1P1 with no salts but in the presence of the substrates CCA-pcb and C-Pmn. F. tt\_A1, tt\_A2 and ec\_P1c at 20  $\mu$ M RNA concentration. Gels were imaged at the wavelength of 254nm.



Figure S19: MALDI spectra of tt\_A1P1, tt\_P1 and tt\_P1c reactions performed in the absence of KCl and MgCl<sub>2</sub> (salts) or in the absence of MgCl<sub>2</sub>.

Intens. [a.u.] tt\_A1P1 in absence of MgCl<sub>2</sub> -1364.451 1309.160 S Intens. [a.u.] 0005 tt\_A1P1 in absence of salts S 1386.585 102.346 1182.300 1309.268 

m/z



С



#### Derivation of Equation used to calculate the dimerization constant (K<sub>d</sub>)

The reaction between a labelled and unlabelled oligomer can be represented as follows:

$$M + M^* \xrightarrow{Kd} MM^*$$
$$M + M \xrightarrow{Kd} MM$$

[M = Unlabelled oligomer, M\* = Labelled oligomer]

Since, the labelled and unlabelled oligomers are same RNA sequences, we assume that they have same  $K_{\rm d}$ 

$$K_d = \frac{[M][M^*]}{[MM^*]} = \frac{[M][M]}{[MM]}$$

 $\mathbf{M}_{\mathbf{0}}^* = [\mathbf{M}^*] + [\mathbf{M}\mathbf{M}^*]$   $\mathbf{M}_{\mathbf{0}}^* =$  Total Concentration of labelled oligomer

 $M_0 = [M] + 2[MM] + [MM^*]$   $M_0 = Total concentration of unlabelled oligomer$ 

$$\mathbf{Y} = \frac{[\mathbf{M}\mathbf{M}^*]}{[\mathbf{M}\mathbf{M}^*] + [\mathbf{M}^*]} * \mathbf{100} \text{ where } \mathbf{Y} = \text{Dimer Fraction \%}$$

Substituting the values in the above equation,

Dimer 
$$\% = \left\{ \frac{\sqrt{K_d(K_d + 8x)} - K_d - 2x}{2(K_d - X)} \right\} * 100$$
 (1)

Where  $X = M_0$  or concentration of the unlabelled oligomers Y = Dimer %

where, 
$$Dimer\% = \frac{[Dimer]}{[Dimer+Monomer]} * 100$$
 (2)

Figure S20: Determination of the dimerization constant (K<sub>d</sub>) of the RNA constructs using fluorescently imaged EMSA native gels. A-F. Dimerization was assayed at increasing RNA concentrations from 0.0 $\mu$ M (left) to 42 $\mu$ M (right) (lanes 1-12 RNA concentrations were 0, 0.08, 0.16, 0.24, 0.4, 0.72, 1.36, 2.46, 5.2, 10.4, 21 and 42 $\mu$ M respectively), with constant (0.15  $\mu$ M) Fluorecein-5-thiosemicarbazide labelled RNA concentration in all lanes. EMSA was performed for tt\_P1c, tt\_A1P1, tt\_P1m, sa\_P1c and ef\_P1c constructs, respectively.



Figure S21: EtBr staining of the native gels used for the determination of the K<sub>d</sub> of the RNA constructs. A-G. RNA construct dimerization is shown for tt\_P1c, tt\_A1P1, tt\_P1, tt\_P1m, tt\_P1n, sa\_P1c and ef\_P1c, respectively. Dimerization was assayed at increasing RNA concentrations from 0.0 $\mu$ M (left) to 42 $\mu$ M (right), with constant (0.15 $\mu$ M) labelled RNA concentration in all lanes (lanes 1-12 RNA concentrations are: 0, 0.08, 0.16, 0.24, 0.4, 0.72, 1.36, 2.46, 5.2, 10.4, 21 and 42  $\mu$ M, respectively). Gels were imaged at a wavelength of 254nm.



## Table S1: Sequences of the RNA constructs

RNA	Sequence*					
Construct						
tt_A1	GAUGUCGGCUCGUCGCUUCGGCGGCACGCGAGCUGGGUUCAGAACG					
	UCGUGAGACAGUUCGGUC					
tt_P1	GAAGACCCCGUGGAGCU <mark>CUUCGG</mark> AGUUACCCCGGGGAUAACAGGCU					
	GAUCUCCGGAGGUUUGGCACCUC					
tt_A1P1	GAAGACCCCGUGGAGCUCUUCGGAGUUACCCCGGGGAUAACAGG					
	GAUCUCUUCGGAGGUUUGGCACCUCGAUGUCGGCUCGUCGCUUCGG					
	CGGCACGCGAGCUGGGUUCAGAACGUCGUGAGACAGUUCGGUC					
tt_A2	GAUGUCGGCUCGUCGCAUCCUUCGGGUUGGGCUGUUCGCCCAUUAA					
	AGCGGCACGCGAGCUGGGUUCAGAACGUCGUGAGACAGUUCGGUC					
tt_P2	GAAGACCCCGUGGAGCUUUACUCUUCGGAGUUUGACUGGGGCGGUC					
	CGGAUAAAAGUUACCCCGGGGAUAACAGGCUGAUCUCUUCGGAGGU					
	UUGGCACCUC					
tt_A2P2	GAAGACCCCGUGGAGCUUUACUCUUCGGAGUUUGACUGGGGCGGUC					
	CGGAUAAAAGUUACCCCGGGGAUAACAGGCUGAUCUCUUCGGAGGU					
	UUGGCACCUCGAUGUCGGCUCGUCGCAUCCUUCGGGUUGGGCUGUU					
	CGCCCAUUAAAGCGGCACGCGAGCUGGGUUCAGAACGUCGUGAGAC					
	AGUUCGGUC					
tt_P2c	GAAGACCCCGUGGAGCUUUACUCUUCGGAGUUUGACUGGGGCGGUC					
	CGGAUAAAAGUUACCCCGGGGGAUAACAGGCUGAUCGUGAGGUUUGG					
	CACCUC					
tt_P1c	GAAGACCCCGUGGAGCUCUUCGGAGUUACCCCGGGGAUAACAGGCU					
	GAUCGUGAGGUUUGGCACCUC					
ec-P1c	AAAGACCCCGUGAACCUCUUCGGAGGUACUCCGGGGAUAACAGGCU					
	GAUAGUGAUGUUUGGCACCUC					
sa P1c	GAAGACCCCGUGGAGCUCUUCGGAGCUACCCCGGGGAUAACAGGCU					
	UAUCGUGAGGUUUGGCACCUC					
ef P1c	GAAGACCCCAUGGAGCUCUUCGGAGCUACCCUGGGGAUAACAGGCU					
	UAUCGUGAGGUUUGGCACCUC					
tt_P1m	GAAGACCCCGUGGAGCUUUAAGUUACCCCGGGGAUAACAGGCUGAU					
	CGUGAGGUUUGGCACCUC					
tt_P1n	GAAGACCCCGUGGAGCUUUAAAGUUACCCCGGGGAUAACAGGCUGA					
	UCGUGAGGUUUGGCACCUC					
com_tt_P1c	GAGGUGCCAAACCUCACGAUCAGCCUGUUAUCCCCGGGGUAACUCC					
	GAAGAGCUCCACGGGGUCUUC					

\* Closing inserted loops are colored; hyper stable CUUCGG loop is colored red, GNRA (GUGA) tetraloop is colored cyan and H74 natural loop is colored green.