



# Article A Proposal of the Ur-RNAome

Miryam Palacios-Pérez <sup>1,2,3,\*</sup> and Marco V. José <sup>1,2,\*</sup>

- <sup>1</sup> Theoretical Biology Group, Instituto de Investigaciones Biomédicas, Universidad Nacional Autónoma de México, Ciudad de México 04510, Mexico
- <sup>2</sup> Network of Researchers on the Chemical Emergence of Life (NoRCEL), Leeds LS7 3RB, UK
- <sup>3</sup> NoRCEL's Latin America Hub, 113 Philosophy Hall, University of California, Berkeley, CA 94720, USA
- \* Correspondence: mir.pape@iibiomedicas.unam.mx (M.P.P.); marcojose@iibiomedicas.unam.mx (M.V.J.)

**Abstract:** It is widely accepted that the earliest RNA molecules were folded into hairpins or minihelixes. Herein, we depict the 2D and 3D conformations of those earliest RNA molecules with only RNY triplets, which Eigen proposed as the primeval genetic code. We selected 26 species (13 bacteria and 13 archaea). We found that the free energy of RNY hairpins was consistently lower than that of their corresponding shuffled controls. We found traces of the three ribosomal RNAs (16S, 23S, and 5S), tRNAs, 6S RNA, and the RNA moieties of RNase P and the signal recognition particle. Nevertheless, at this stage of evolution there was no genetic code (as seen in the absence of the peptidyl transferase centre and any vestiges of the anti-Shine–Dalgarno sequence). Interestingly, we detected the anticodons of both glycine (GCC) and threonine (GGU) in the hairpins of proto-tRNA.

Keywords: RNA; RNY code; evolution of genetic code; proto-tRNA; LUCA

# 1. Introduction

Before the emergence of catalytic proteins and DNA for heredity as separate molecules, RNA was the first biological molecule. Two of its characteristics, while prone to mistakes, allowed life to arise in a hypothetical "RNA world": it could store information and act as a catalyst for processes like self-excision [1–3]. Test tube experiments have showed the various catalytic properties of RNA, reinforcing the idea that the first biochemical systems could have been entirely centred on that molecule [4].

Since RNA is the most versatile of all the biological macromolecules, and based on physicochemical theoretical works, RNA is thought to have originated the genetic code  $\sim$ 4.36 + 0.1 billion years ago [5–9]. Eigen and Schuster [10] glimpsed that the primeval genetic code (PGC) consisted of ribonucleotide chains following the pattern RNY, in which R means purines (A/G) and Y means pyrimidines (C/U), while N symbolises any of the nucleobases (A/C/G/U) in accordance with the parity rule R:Y. It was also shown that RNY is the main pattern in ribosomal RNA (rRNA) subunit 5S (5S rRNA) for more than 200 varied species [11], and that only primitive transfer RNA (tRNA) molecules with the RNY pattern are susceptible to being efficiently replicated, translated, and therefore amplified [12].

Early RNA genes could have been very short, reaching a maximum length of 50 to 100 nt, with most probably configured into mini-helixes whose strands must are stable and therefore equivalent to each other in topology and chemical composition [5–7,10,13–19]. These probably functioned as proto-tRNAs [20–23]. Indeed, these proto-tRNAs have been found to be concatenated to form all the rRNAs [24], including the peptidyl transferase centre (PTC) of the ribosome [25–27]; the concatamers can be translated into functional, modern-like proteins [28]. It has been proposed that the Ur-gene was composed of RNY sequences, while the RNA was folded in hairpins and mini-helixes resembling proto-tRNAs, which were able to bind prebiotic amino acids (aa) [5–7,10,29] that were also encoded by



Citation: Palacios-Pérez, M.; José, M.V. A Proposal of the Ur-RNAome. *Genes* 2023, 14, 2158. https:// doi.org/10.3390/genes14122158

Academic Editor: Clifford J. Steer

Received: 14 October 2023 Revised: 16 November 2023 Accepted: 16 November 2023 Published: 29 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). RNY triplets i.e., aa encoded by the aforementioned PGC. Actually, the mini-helix is thought to be the most ancient historical domain of tRNAs (3.6–3.7 BYA) [30].

We previously obtained the phenotype of amino acids and proteins corresponding to the evolution of genetic code [29,31], but RNA evolution was simply ignored. In this work, we determine the 2D and 3D structures of early RNA molecules based on the PGC. We find that those RNAs can indeed fold into short hairpins, and we even capture the anticodon loop of some of the earliest tRNA isoacceptors able to carry prebiotic aa.

#### 2. Methods

We retrieved the RNAome from phylogenetically distant organisms, from which the triplets that did not belong to the early genetic code (RNY) were discarded. The sequences were then grouped according to type; then, each fragment was assembled into its original order in cases that the original gene had more than one fragment encoded by RNY triplets. To generate negative controls, the sequences were shuffled thrice. If more than one organism contained at least one RNA fragment encoded by RNY triplets, the fragments of each RNA type were arranged according to the original order in the gene; the RNAs were then aligned with each other to obtain a consensus sequence, and the corresponding logo sequences were generated. It is worth recalling that Ts were replaced by Us. Finally, we obtained the 2D and 3D structures of the RNAs encoded by RNY triplets. Each of the steps that we followed is detailed below, and a graphical flowchart can be found in Supplementary File S1.

#### 2.1. Data Sources

From https://ftp.ncbi.nlm.nih.gov/genomes/refseq/ accessed on 4 October 2023, we obtained all the RNAs (\*RNA\*.fna) of 13 bacteria (bac.) [Aquifex aeolicus VF5 (NC\_000918.1)  $\rightarrow$  "AqfxV", Bacillus subtilis 168 (NC\_000964.3)  $\rightarrow$  "Basub", Borreliella burgderfori B31  $(NC_{001318.1}) \rightarrow$  "Bobur", Deinococcus radiodurans R1  $(NC_{001263.1} \text{ y } NC_{001264.1}) \rightarrow$ "Derad", Escherichia coli K12 MG1655 (NZ\_CP025268.1)  $\rightarrow$  "EcoK12", Mycoplasma genital*ium* G37 (NC\_000908.2)  $\rightarrow$  "Mygen", Ca. *Pelagibacter ubique* HTCC1062 (NC\_007205.1)  $\rightarrow$ "Peubi", Shewanella piezotolerans WP3 (NC\_011566.1)  $\rightarrow$  "Shpz3", Streptococcus agalactiae A909 (NC\_007432.1)  $\rightarrow$  "SagA", Synechococcus CC9902 (NC\_007513.1)  $\rightarrow$  "SynCC", Thermotoga maritima MSB8 (NC\_000853.1)  $\rightarrow$  "Thmar", Thermus aquaticus Y51MC23 (NZ\_CP010822.1)  $\rightarrow$  "TaqY51", Thermus thermophilus HB8 (NC\_006461.1),  $\rightarrow$  "Ther2"] and 13 archaea (arc.) (NC\_018656.1) → "Nised", Haloarcula marismortui ATCC 43049 (NC\_006396.1 y NC\_006397.1)  $\rightarrow$  "Hamar", Haloferax volcanii DS2 (NC\_013967.1)  $\rightarrow$  "Hxvol", Haloquadratum walsbyi DSM 16790 (NC\_008212.1)  $\rightarrow$  "Haqwa", Korarchaeum cryptofilum OPF8 (NC\_010482.1)  $\rightarrow$  "Kocry", Methanocaldococcus jannaschii DSM\_2661 (NC\_000909.1)  $\rightarrow$  "Mejan", Methanosarcina acetivorans C2A (NC\_003552.1)  $\rightarrow$  "Macet", Pyrococcus furiosus DSM 3638 (NC\_003413.1)  $\rightarrow$ "Pyfur", Sulfolobus acidocaldarius DSM 639 (NC\_007181.1)  $\rightarrow$  "Sacid", Thermococcus gammatolerans EJ3 (NC\_012804.1)  $\rightarrow$  "Thgam", Thermococcus sibiricus MM739 (NC\_012883.1)  $\rightarrow$ "Thsib", Thermoplasma volcanium GSS1 (NC\_002689.2)  $\rightarrow$  "Tmvol"].

#### 2.2. Reconstruction of Sequences of Ancient RNAomes

To reconstruct the original arrangement of the RNAome, all RNAs were assembled one after the other, i.e., coding-wise (CW) with an ad hoc program, just as they are reported in the file \*RNA\*.fna, allowing a posterior alignment.

To generate an ad hoc filter for our arrangements, we also generated a random sequence from each assembled RNAome, shuffling the nucleotides thrice to eliminate the biological sense and information of the sequence. From each RNAome, and from its corresponding shuffled control, we discarded all triplets except those of the RNY type.

#### 2.3. Grouping and Assembly of RNAs

Using BLASTn [32] as a standalone version, we used the biological RNAomes that we constructed (previously mentioned) by concatenating all the RNA sequences of each

organism and their corresponding controls (the shuffled sequences) as queries. This allowed us to obtain the RNAs encoded by the PGC using the file \*RNA\*.fna\* of every organism as the databases. Since RNY possesses only a quarter of the number of triplets as the SGC, we adjusted the parameters to allow as many outcomes as possible from those commonly used in BLAST searches, while preserving the maximum E-value at 10. To determine a cut-off value for the RNAs retrieved, numerical comparisons of the E-values of each of the biological RNAomes in RNY with their corresponding controls (those previously shuffled) were performed, thus setting the cut-offs for each organism.

The length of the RNA molecules was not selected beforehand but was the result of using a BLAST alignment for two sequences so that the fragments were retrieved as they were encoded by RNY triplets in the RNA molecules of each organism.

We grouped all the retrieved fragments according to the RNA molecule to which each one belongs; for instance, all fragments belonging to 5S rRNA were grouped together, all RNA fragments of A-type RNase P were grouped together, etc., and this was performed for each organism. Additionally, the tRNAs were sorted according to their cognate aa and the anticodon of each, which we identified using the programs 'tRNA finder' [33] and 'tRNA scan' [34]. Note that we provide the anticodon, and not the codon, of each aa.

#### 2.4. MSAs of RNAs

Not all forms of RNA recovered have RNY triplet-encoded segments in more than one organism. In fact, only ribosomal RNAs (rRNAs) can be aligned with each other because several organisms contain more than one copy of the same gene, and the RNY-encoded portions are at similar positions. To align the small fragments encoded by RNY triplets of the ribosomal genes, we used the CHAOS-DIALIGN software (version 2.2.2) [35], as it works best with fragmentary sequences in local alignments. From the multiple sequence alignments (MSAs) generated, we obtained the consensus sequence using the UGENE suite [36].

#### 2.5. Sequence Logos

We generated a graphical representation, in the form of sequence logos [37], of the MSAs of the RNA molecules encoded by RNY triplets.

#### 2.6. Representation in 2D of the Recovered Fragments

We predicted the secondary structure of all our individual RNA sequences, or their consensuses, within the webserver RNAfold of the ViennaRNA suite [38], selecting the structure with the minimum free energy (MFE) under the Andronescu model, avoiding isolated base pairs, and leaving all other parameters the same. The 2D structures were visualised with the tool forna [39].

#### 2.7. Representation in 3D of the Reconstructed Fragments

For RNA molecules encoded by RNY triplets, we adopted the Vienna format (dotbracket notation) provided by the 2D-structure prediction program to construct de novo the corresponding 3D structure on the automated modelling server RNAcomposer (version 1.0) [40]. The structures were visualised using Chimera software (version 1.14) [41].

#### 2.8. Negative Controls

To generate control sequences, we shuffled each of the RNY-encoded fragments (or their consensus sequences) thrice and obtained their thermodynamic descriptions and 2D and 3D structures, as was performed for the biological sequences.

#### 3. Results and Analyses

We used the genomes of 26 organisms with different lifestyles (13 bacteria and 13 archaea) based upon the latest update of the tree of life (ToL), which places eukaryotes among the latter [42–47]. There are several challenges in modelling an RNA molecule de novo, and the difficulty increases as the length does by virtue of the fact that RNA folding depends on numerous parameters [48]. Accordingly, the folding of RNY-encoded fragments did not entail additional difficulties, as they are mostly short and self-complementary; however, this detail is particularly interesting because several authors have considered early RNAs to be folded like hairpins or mini-helixes.

Table S1T in Supplementary File S2 lists the MFE of the RNA fragments encoded by RNY triplets, as well as the corresponding controls (shuffled sequences). Notice that the MFEs of the majority of the negative controls are higher than the biological sequences from which they come, i.e., the biological structures are more stable than their controls. In some cases, the MFEs of the biological sequences and the negative controls are zero (or just slightly lower than the biological one), which indicates that the results are not artefactual.

In some organisms, the three rRNAs and nine tRNAs have well-defined portions encoded by RNY triplets. Moreover, the RNA moieties of some signal recognition particles (SRP-RNA), RNases P (RNA-P), and RNAs 6S, retain small antique portions. All figures not shown in the main text can be found in Supplementary File S3.

The 5' end is always the first nucleotide in 2D structures; however, we placed each 3D RNA structure with the 5' position towards the viewer and labelled both ends (5' and 3') in Supplementary File S3 (so as not to clutter the main text).

#### 3.1. Ribosomal RNAs

The ribosome is a ribonucleoprotein lair formed by two subunits—a large ribosomal subunit (LSU) and a small ribosomal subunit (SSU)—in which, in turn, peptide growth is enabled entirely by RNA and the structural scaffold is provided by ribosomal proteins. The 16S rRNA couples with the messenger RNA (mRNA) to be translated into proteins according to the codons in it. The most critical portion of the translation is the peptidyl transferase centre (PTC), embedded in 23S rRNA; the PTC is formed by the A site and P site, while the important E site does not belong to the PTC. Finally, 5S rRNA keeps the tRNAs positioned at the A and P sites until translation finishes. Both 23S and 5S rRNAs belong to the LSU, while 16S rRNA belongs to the SSU [49–53].

The three types of rRNA have portions encoded by RNY triplets. In some cases, only one organism has a recognisable sequence of this type, but for most of them the ribosomal RNAs from several organisms have PGC-encoded portions, so it makes sense to generate sequence alignments to obtain consensuses that can be modelled in 2D and 3D. In all cases, we can see that the biological sequence is more stable than its control (Figures 1–3 here below and Supplementary File S2), and they tend to conform into short or complex helixes.

Only the 5S rRNA of the archaeon "Mejan" has a portion encoded by RNY triplets at the end of the molecule, and this fragment is folded like a hairpin (Figure 1A). On the other hand, the 5S rRNAs of many of our bacteria have RNY-encoded regions in the first quarter and also mostly in the third quarter of the molecule (Figure 1B).

The portions encoded by RNY triplets in both 16S and 23S rRNAs are punctually dispersed throughout both types of sequences (Figures S2 and S3 in Supplementary File S3, all the corresponding logo sequences and their corresponding controls can be found).

As we can observe, the RNY-encoding of 16S rRNA conforms to a clearly defined region, with weak nucleotides (A and T/U) flanked by strong nucleotides (G and C) in the middle of the sequence in both archaea (Figure S2A in Supplementary File S3) and bacteria (Figure S2B in Supplementary File S3); additionally, the 3' end in bacteria is slightly less defined than in archaea.

On the other hand, RNY triplets recovered only a tiny fragment (60 nt) of 23S rRNA, with less conservation in the archaeal (Figure 3A) than in the bacterial (Figure 3B) reconstructions.



**Figure 1.** The 2D and 3D structures of RNY–encoded portions of 5S rRNA, as well as the logo sequences of the bacterial alignment; the MFE is also indicated in each case. On each panel, the structures and MFE values of the biological sequences are on the left side of the dotted line and the controls are on the right side. Archaea are shown in (**A**) and bacteria in (**B**).



**Figure 2.** In **(A)**, 2D and 3D structures of RNY–encoded portions of 16S rRNA from archaea; in **(B)**, 2D and 3D structures of RNY-encoded portions of 16S rRNA from bacteria. The MFE is indicated in each case.



**Figure 3.** In (**A**), 2D and 3D structures of RNY–encoded portions of 23S rRNA from archaea; in (**B**), 2D and 3D structures of RNY–encoded portions of 23S rRNA from bacteria. The MFE is indicated in each case.

#### 3.2. RNase P, SRP, and 6S

The RNA moieties of some types of RNase P (Figure 4), some types of SRPs (Figure 5), and one 6S RNA (Figure 6), have at least one portion encoded by RNY triplets. Each of these fragments is present in only one, although not the same, of the 26 organisms selected here.

RNase P is a ubiquitous ribonucleoprotein that catalyses the maturation of tRNAs by removing their extraneous 5' sequences. All species require the RNA moiety of the RNase P (RNA-P), but whereas in bacteria and archaea the protein portion is totally or just marginally dispensable, respectively, eukaryotes cannot survive without the proteins of their RNase P [54–56]. The RNA-Ps of the archaeon "Hxvol" (Figure 4A) and of the bacterium "Mygen" (Figure 4B) each have one portion encoded by RNY triplets. The archaeal RNA-P fragment is more stable than its control, whereas the MFE of the second bacterial RNA-P fragment is zero, as is that of its shuffled control (Supplementary File S2). On other hand, the RNA-P of the bacterium "SynCC" has two fragments encoded by the PGC. In Figure 4C, we observe the concatenation of both fragments in the same order as they appear in the original RNA molecule, and this construct is more stable than its corresponding control. When each fragment is modelled individually (Supplementary File S2), we see that the first one is more stable than its control, and that the stability of the control of the second fragment is just slightly higher than that of the biological sequence. All those RNA-P fragments encoded by RNY triplets, whose MFE is different from zero, tend to form mini-helixes or hairpin-like structures.



**Figure 4.** The 2D and 3D structures of RNY–encoded portions of the RNA moiety of RNase P (RNA-P). In (**A**), the archaeon "Hxvol"; in (**B**), the bacterium "Mygen"; in (**C**), the two sequences from bacterium "SynCC". The corresponding controls are in Supplementary File S2.



**Figure 5.** The 2D and 3D structures of RNY–encoded portions of the RNA moiety of SRP. In (**A**), the archaeon "Kocry"; in (**B**), the bacterium "Basub". The corresponding controls are in Supplementary File S2.





The signal recognition particle (SRP) is a widely distributed GTP-dependent ribonucleoprotein that helps direct the protein synthesis towards the membrane when needed. This SRP-RNA serves as the scaffold on which all its proteins will be assembled. SRP has so far been described in two variants in bacteria and one in archaea, as well as many more versions in eukaryotes. The SRP-RNA has several self-complementary regions that can fold into a few or many helixes [57,58]. The RNY triplets partially encode the RNA moieties of the SRP of the archaeon "Kocry" (Figure 5A) and of the bacterium "Basub" (Figure 5B). The archaeal SRP-RNA is much more stable than its control (Figure S5A in Supplementary File S3); the bacterial SRP-RNA and its control (Figure S5B in Supplementary File S3) have an MFE of zero, although the entropy is slightly higher than in the shuffled sequence.

The 6S RNA molecule is a widespread small global regulator of bacterial transcription that mimics B-form DNA and then binds to the active site of RNA polymerase (RNApol), thus blocking the transcription and enabling the release of the enzyme RNApol. It folds into a single, long self-complementary structure with some internal loops along the length of the molecule [59–64]. The molecule 6S RNA of the bacterium "Basub" is the only one of its kind with a portion encoded by RNY triplets (Figure 6), and it has a slight hairpin-like folding; although the paired bases are too few to achieve this, their entropy is certainly lower than that of its corresponding control.

## 3.3. tRNAs

The tRNAs are molecules ranging from 76 to about 90 nt in length that fold into a 2D cloverleaf or a 3D L-shape. The tRNAs serve as the physical adaptors between the genetic code "read" by the anticodon in the middle of the molecule and the phenotype in the form of the corresponding as charged in the distal 3' portion [65–67]. We found 12 tRNAs with one portion each encoded by RNY triplets (Figure 7, with only a few examples; the complete catalogue 7A to 7L is in Supplementary File S2); five of them can fold into small hairpins (sort of helix-like), while the other tRNA fragments remain unfolded.



**Figure 7.** The 2D and 3D structures of RNY–encoded portions of some tRNAs. In (**B**), Asn-5'GUU from "Thgam"; in (**E**), Cys-5'GCA from "Derad"; in (**H**), Gly-5'UCC from "Derad"; in (**I**), Gly-5'<u>GCC</u> from "Peubi" with the anticodon circled in red; in (**K**), Thr-5'<u>GGU</u> from "SagA" with the anticodon circled in red; in Supplementary File S2.

In most cases, the biological sequences are more stable than their corresponding controls, such as the tRNA of the archaea "Haqwa" for Gln-UUG (Figure S7A in Supplementary File S3 or of "Thgam" for Asn-GUU (Figure S7B below). Some other cases are only slightly more stable than the controls, such as the tRNA of the bacterium "Derad" for Cys-GCA (Figure S7E in Supplementary File S3). On a few occasions, the MFE of the biological sequence is zero, as is that of its control sequence, as in the case of bacteria "Bobur" for Gln-UUG (Figure S7D in Supplementary File S3) or "Derad" for Gly-UCC (Figure S7H below).

Remarkably, three of the RNY-encoded fragments capture the anticodons of their corresponding tRNAs. To wit, Gly-tRNA\_GCC of the bacterium "Peubi" (Figure S7I below) and Thr-tRNA\_GGU of the bacterium "SagA" (Figure S7K below) totally capture the anticodons (letters underlined in the text and circled in red in the corresponding figures) that are located just in the middle of the fragments and therefore at the loop of hairpin. Moreover, the bases of the anticodons point outwards as in the full tRNA molecules; each of the activating amino acids of these anticodons is also encoded by the PGC (Gly and Thr). This contrasts with Gly-tRNA\_UCC of the bacterium "Derad" (Figure S7H below) because, even if Gly were encoded very early, the PGC would not include the anticodon UCC. On the other hand, the fragment of tRNA-Gln\_CUG of the bacterium "Derad" (Figure S7F in Supplementary File S3) encoded by RNY triplets remains unfolded, as we mentioned earlier; the contrast relies on the fact that the anticodon is only partially included (only the letters underlined) in the fragment encoded by the PGC and that glutamine is not encoded by RNY triplets.

Finally, in several cases the control sequence is even more stable than the biological one but has a minor difference, as with the tRNA of the bacterium "SagA" for Asn-GUU (Figure S7H below) or "Thmar" for Phe-GAA (Figure S7L in Supplementary File S3).

## 4. Discussion

Though not all organisms have RNA with RNY-encoded portions, and although such fragments are very small and can barely be aligned, it is noteworthy that all the RNA molecules directly involved in the modern translation process withhold a snippet encoded by the PGC. We found that such RNY triplet-encoded RNA snippets can fold into small hairpins shorter than the size proposed for early functional genes (or even for proto-tRNAs) able to shape all the other biomolecules [24], suggesting an earlier stage in the evolution [68] that probably constituted the beginnings of these RNA molecules. Moreover, when we compare the predicted structure of the sequence encoded by RNY triplets of Gly-tRNA\_GCC with its modern structure (PDB ID 4mgn), we observe that not only the whole structures, but even the anticodon bases of both, are in almost the same outward positions (Figure 8). In contrast, in the case of the Gly-tRNA\_UCC fragment, the anticodon is not retrieved, a fact already reckoned with in 1981 [20,21] regarding a differential emergence of tRNA isoacceptors.



**Figure 8.** Structural comparison among the predicted structure of the sequence encoded by RNY triplets of Gly-tRNA\_GCC and its modern structure (PDB ID 4mgn). The structure encoded by the PGC is in light green, and its anticodon is in forest green; the crystallographic structure is in mid-blue, and its anticodon is in deep blue.

It is strikingly important to have discovered that the anticodon stem of tRNA-Gly\_GCC is purely encoded by RNY triplets because the whole tRNA cloverleaf may possibly have been formed via the ligation of proto-tRNA mini-helixes, 3–31 nt in length (one of which encoded for glycine\_GCC [69]), which resemble some of the other small RNA molecules found to be encoded using RNY triplets. Without going any further, any modern tRNA could have its origin in mini-helixes [29,69–71] that could be replicated themselves [72], combining the operational code in an ancient anticodon helix with the informational code in an early acceptor helix [19], prior to the appearance of contemporary tRNA specificities, and quite before the three domains of life diverge [73].

Life most probably originated when proteins began to be translated, for which a wellestablished PTC (as well as the respective anti-Shine–Dalgarno sequence) is the sine qua non [27,74,75]; however, we did not find any of them encoded by RNY triplets, which places our work in the realm of the protobiotic stage, and the small RNA hairpins encoded by RNY as the primordial seeds that eventually grew and ligated to each other to form more recognisable modern RNA molecules.

RNA could then have polymerised and randomly generated short ribonucleic chains in which the RNY pattern gradually began to prevail, as if it was a quasi-species [6,58] that evolved through cooperative interaction via cyclic coupling, i.e., hypercycles. Those RNA short sequences and their limited diversity supported prebiotic, autocatalytic reproduction by means of hypercycles [5,10,15–18,21]. Lastly, it is safe to assume that the Ur-RNA proposed here, encoded by the PGC, emerged before the so-called "First Universal Common Ancestor" (FUCA), because the PTC cannot be found encoded by RNY triplets [69,70].

RNA evolved as one of the first phenotypic biomolecules and the primordial genotypic biomolecule, and the PGC of such RNAs followed the pattern RNY. The modern translation molecules have their origins from short RNA hairpins formed by triplets pertaining to the PGC. All the small hairpins here described possibly constituted the beginnings of the corresponding modern RNA molecules and were probably part of a larger pool of RNA molecules that served as the seeds of more complex molecules. The lengths of our RNA sequences (20–30 nt) are far from the error catastrophe limit [5], and the Ur-gene is proposed [20,21] to have a length between 50 and 100 nt.

Speaking of contemporary issues, synthetic genetic codes can be designed to generate new proteins, and it is known that mutations of tRNA are associated with several diseases. For instance, cellular and mitochondrial tRNA overexpression and mutation relate to a wide range of human diseases [76–78], such as breast cancer [79] and neuro-gastrointestinal encephalopathy [80].

The results presented here provide astounding evidence that our approach can detect molecular structures from the protobiotic stage >3.7 billion years ago with surprising confidence.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/genes14122158/s1. File S1 contains the graphical flowchart of the Methodology. File S2 contains a table (S1T) of the MFE of each of the structures in this work and their controls. File S3 contains all the RNA sequence logos, and the 2D and 3D structures of each of the RNAs declared in this work.

**Author Contributions:** Conceptualization, M.V.J.; study design and analysis, M.P.-P. and M.V.J.; manuscript writing, M.P.-P. and M.V.J.; writing—review and editing, M.V.J.; funding acquisition: M.V.J. All authors have read and agreed to the published version of the manuscript.

**Funding:** M.P.-P. is a postdoctoral researcher at the Universidad Nacional Autónoma de México (UNAM) supported by fellowship 694877 of the Consejo Nacional de Humanidades, Ciencias y Tecnologías (CONAHCYT). M.V.J. was funded by the Dirección General de Asuntos del Personal Académico (DGAPA), Universidad Nacional Autónoma de México, UNAM (PAPIIT-IN200823).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are contained within the article and Supplementary Materials.

Acknowledgments: The authors thank Juan R. Bobadilla for his invaluable computer and technical assistance, and the members of the Theoretical Biology Group for their generous feedback. Authors thank also to anonymous reviewers for helping to improve this work. Molecular graphics and analyses were performed with UCSF Chimera, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311.

Conflicts of Interest: The authors declare no conflict of interest.

#### References

- 1. Kruger, K.; Grabowski, P.J.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. Self-Splicing RNA: Autoexcision and Autocyclization of the Ribosomal RNA Intervening Sequence of Tetrahymena. *Cell* **1982**, *31*, 147–157. [CrossRef]
- Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. The RNA Moiety of Ribonuclease P Is the Catalytic Subunit of the Enzyme. *Cell* 1983, 35, 849–857. [CrossRef]
- 3. Gilbert, W. Origin of Life: The RNA World. Nature 1986, 319, 618. [CrossRef]
- 4. Brown, T.A. Genomes, 2nd ed.; Wiley-Liss: Oxford, UK, 2002; ISBN 978-0-471-25046-3.
- 5. Eigen, M.; Gardiner, W.; Schuster, P.; Winkler-Oswatitsch, R. The Origin of Genetic Information. *Sci. Am.* **1981**, 244, 88–119. [CrossRef] [PubMed]
- 6. Wächtershäuser, G. The Place of RNA in the Origin and Early Evolution of the Genetic Machinery. *Life* **2014**, *4*, 1050–1091. [CrossRef]
- Chatterjee, S.; Yadav, S. The Origin of Prebiotic Information System in the Peptide/RNA World: A Simulation Model of the Evolution of Translation and the Genetic Code. *Life* 2019, *9*, 25. [CrossRef] [PubMed]

- 8. Benner, S.A.; Bell, E.A.; Biondi, E.; Brasser, R.; Carell, T.; Kim, H.-J.; Mojzsis, S.J.; Omran, A.; Pasek, M.A.; Trail, D. When Did Life Likely Emerge on Earth in an RNA-First Process? *ChemSystemsChem* **2020**, *2*, e1900035. [CrossRef]
- 9. Lehman, N. The RNA World: 4,000,000,050 Years Old. Life 2015, 5, 1583–1586. [CrossRef]
- Eigen, M.; Schuster, P. The Hypercycle—A Principle of Natural Self-Organization Part C: The Realistic Hypercycle. *Naturwissenschaften* 1978, 65, 341–369. [CrossRef]
- Eigen, M.; Lindemann, B.; Winkler-Oswatitsch, R.; Clarke, C.H. Pattern Analysis of 5S rRNA. Proc. Natl. Acad. Sci. USA 1985, 82, 2437–2441. [CrossRef] [PubMed]
- 12. Lehmann, J. Amplification of the Sequences Displaying the Pattern RNY in the RNA World: The Translation → Translation/Replication Hypothesis. *J. Theor. Biol.* **2002**, *219*, 521–537. [CrossRef]
- 13. Eigen, M.; Winkler-Oswatitsch, R. Transfer-RNA: The Early Adaptor. Naturwissenschaften 1981, 68, 217–228. [CrossRef]
- 14. Eigen, M.; Winkler-Oswatitsch, R. Transfer-RNA, an Early Gene? *Naturwissenschaften* **1981**, *68*, 282–292. [CrossRef] [PubMed]
- 15. Delaye, L.; Lazcano, A. Prebiological Evolution and the Physics of the Origin of Life. *Phys. Life Rev.* 2005, 2, 47–64. [CrossRef] [PubMed]
- Kun, A.; Szilágyi, A.; Könnyű, B.; Boza, G.; Zachar, I.; Szathmáry, E. The Dynamics of the RNA World: Insights and Challenges. Ann. N. Y. Acad. Sci. 2015, 1341, 75–95. [CrossRef] [PubMed]
- Szilágyi, A.; Zachar, I.; Scheuring, I.; Kun, Á.; Könnyű, B.; Czárán, T. Ecology and Evolution in the RNA World Dynamics and Stability of Prebiotic Replicator Systems. *Life* 2017, 7, 48. [CrossRef] [PubMed]
- Szilágyi, A.; Könnyű, B.; Czárán, T. Dynamics and Stability in Prebiotic Information Integration: An RNA World Model from First Principles. Sci. Rep. 2020, 10, 51. [CrossRef]
- 19. Schimmel, P.; de Pouplana, L.R. Transfer RNA: From Minihelix to Genetic Code. Cell 1995, 81, 983–986. [CrossRef]
- Ribas de Pouplana, L.; Buechter, D.; Sardesai, N.Y.; Schimmel, P. Functional Analysis of Peptide Motif for RNA Microhelix Binding Suggests New Family of RNA-Binding Domains. *EMBO J.* 1998, 17, 5449–5457. [CrossRef] [PubMed]
- Mizuuchi, R.; Lehman, N. Limited Sequence Diversity Within a Population Supports Prebiotic RNA Reproduction. *Life* 2019, 9, 20. [CrossRef]
- 22. Giulio, M.D. On the Origin of the Transfer RNA Molecule. J. Theor. Biol. 1992, 159, 199–214. [CrossRef] [PubMed]
- Dick, T.P.; Schamel, W.W.A. Molecular Evolution of Transfer RNA from Two Precursor Hairpins: Implications for the Origin of Protein Synthesis. J. Mol. Evol. 1995, 41, 1–9. [CrossRef]
- 24. de Farias, S.T.; José, M.V. Transfer RNA: The Molecular Demiurge in the Origin of Biological Systems. *Prog. Biophys. Mol. Biol.* **2020**, 153, 28–34. [CrossRef]
- Farias, S.T.; Rêgo, T.G.; José, M.V. Origin and Evolution of the Peptidyl Transferase Center from Proto-tRNAs. FEBS Open Bio 2014, 4, 175–178. [CrossRef] [PubMed]
- 26. Torres de Farias, S.; Gaudêncio Rêgo, T.; José, M.V. Peptidyl Transferase Center and the Emergence of the Translation System. *Life* **2017**, *7*, 21. [CrossRef]
- Prosdocimi, F.; Zamudio, G.S.; Palacios-Pérez, M.; Torres de Farias, S.; José, M.V. The Ancient History of Peptidyl Transferase Center Formation as Told by Conservation and Information Analyses. *Life* 2020, *10*, 134. [CrossRef]
- de Farias, S.T.; do Rêgo, T.G.; José, M.V. Evolution of Transfer RNA and the Origin of the Translation System. *Front. Genet.* 2014, 5, 303. [CrossRef]
- Palacios-Pérez, M.; Andrade-Díaz, F.; José, M.V. A Proposal of the Ur-Proteome. Orig. Life Evol. Biosph. 2018, 48, 245–258. [CrossRef] [PubMed]
- Maizels, N.; Weiner, A.M. Phylogeny from Function: Evidence from the Molecular Fossil Record That tRNA Originated in Replication, Not Translation. Proc. Natl. Acad. Sci. USA 1994, 91, 6729–6734. [CrossRef]
- José, M.V.; Zamudio, G.S.; Palacios-Pérez, M.; Bobadilla, J.R.; de Farías, S.T. Symmetrical and Thermodynamic Properties of Phenotypic Graphs of Amino Acids Encoded by the Primeval RNY Code. Orig. Life Evol. Biosph. 2015, 45, 77–83. [CrossRef]
- Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic Local Alignment Search Tool. J. Mol. Biol. 1990, 215, 403–410. [CrossRef] [PubMed]
- Kinouchi, M.; Kurokawa, K. tRNAfinder: A Software System to Find All tRNA Genes in the DNA Sequence Based on the Cloverleaf Secondary Structure. J. Comput. Aided Chem. 2006, 9, 116–124. [CrossRef]
- Lowe, T.M.; Chan, P.P. tRNAscan-SE On-Line: Integrating Search and Context for Analysis of Transfer RNA Genes. Nucleic Acids Res. 2016, 44, W54–W57. [CrossRef]
- Brudno, M.; Steinkamp, R.; Morgenstern, B. The CHAOS/DIALIGN WWW Server for Multiple Alignment of Genomic Sequences. Nucleic Acids Res. 2004, 32, W41–W44. [CrossRef]
- Okonechnikov, K.; Golosova, O.; Fursov, M. Unipro UGENE: A Unified Bioinformatics Toolkit. *Bioinformatics* 2012, 28, 1166–1167. [CrossRef] [PubMed]
- Crooks, G.E.; Hon, G.; Chandonia, J.-M.; Brenner, S.E. WebLogo: A Sequence Logo Generator. *Genome Res.* 2004, 14, 1188–1190. [CrossRef]
- Gruber, A.R.; Lorenz, R.; Bernhart, S.H.; Neuböck, R.; Hofacker, I.L. The Vienna RNA Websuite. Nucleic Acids Res. 2008, 36, W70–W74. [CrossRef]
- Kerpedjiev, P.; Hammer, S.; Hofacker, I.L. Forna (Force-Directed RNA): Simple and Effective Online RNA Secondary Structure Diagrams. *Bioinformatics* 2015, 31, 3377–3379. [CrossRef]

- 40. Popenda, M.; Szachniuk, M.; Antczak, M.; Purzycka, K.J.; Lukasiak, P.; Bartol, N.; Blazewicz, J.; Adamiak, R.W. Automated 3D Structure Composition for Large RNAs. *Nucleic Acids Res.* **2012**, *40*, e112. [CrossRef]
- Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. J. Comput. Chem. 2004, 25, 1605–1612. [CrossRef] [PubMed]
- 42. Williams, T.A.; Foster, P.G.; Nye, T.M.W.; Cox, C.J.; Embley, T.M. A Congruent Phylogenomic Signal Places Eukaryotes within the Archaea. *Proc. R. Soc. B Biol. Sci.* 2012, 279, 4870–4879. [CrossRef]
- 43. Raymann, K.; Brochier-Armanet, C.; Gribaldo, S. The Two-Domain Tree of Life Is Linked to a New Root for the Archaea. *Proc. Natl. Acad. Sci. USA* 2015, 112, 6670–6675. [CrossRef]
- Spang, A.; Saw, J.H.; Jørgensen, S.L.; Zaremba-Niedzwiedzka, K.; Martijn, J.; Lind, A.E.; van Eijk, R.; Schleper, C.; Guy, L.; Ettema, T.J.G. Complex Archaea That Bridge the Gap between Prokaryotes and Eukaryotes. *Nature* 2015, 521, 173–179. [CrossRef] [PubMed]
- 45. Spang, A.; Ettema, T.J.G. Microbial Diversity: The Tree of Life Comes of Age. Nat. Microbiol. 2016, 1, 16056. [CrossRef] [PubMed]
- 46. Cornish-Bowden, A.; Cárdenas, M.L. Life before LUCA. J. Theor. Biol. 2017, 434, 68–74. [CrossRef]
- Imachi, H.; Nobu, M.K.; Nakahara, N.; Morono, Y.; Ogawara, M.; Takaki, Y.; Takano, Y.; Uematsu, K.; Ikuta, T.; Ito, M.; et al. Isolation of an Archaeon at the Prokaryote–Eukaryote Interface. *Nature* 2020, 577, 519–525. [CrossRef] [PubMed]
- 48. Mustoe, A.M.; Brooks, C.L.; Al-Hashimi, H.M. Hierarchy of RNA Functional Dynamics. *Annu. Rev. Biochem.* **2014**, *83*, 441–466. [CrossRef]
- 49. Moore, P.B.; Steitz, T.A. The Involvement of RNA in Ribosome Function. Nature 2002, 418, 229. [CrossRef]
- 50. Sievers, A.; Beringer, M.; Rodnina, M.V.; Wolfenden, R. The Ribosome as an Entropy Trap. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7897–7901. [CrossRef]
- Wallin, G.; Åqvist, J. The Transition State for Peptide Bond Formation Reveals the Ribosome as a Water Trap. Proc. Natl. Acad. Sci. USA 2010, 107, 1888–1893. [CrossRef]
- Jüttner, M.; Ferreira-Cerca, S. A Comparative Perspective on Ribosome Biogenesis: Unity and Diversity Across the Tree of Life. In *Ribosome Biogenesis: Methods and Protocols*; Entian, K.-D., Ed.; Methods in Molecular Biology; Springer: New York, NY, USA, 2022; pp. 3–22. ISBN 978-1-07-162501-9.
- 53. Wang, Q.; Su, H. A Tale of Water Molecules in the Ribosomal Peptidyl Transferase Reaction. *Biochemistry* **2022**, *61*, 2241–2247. [CrossRef] [PubMed]
- 54. Jarrous, N. Roles of RNase P and Its Subunits. *Trends Genet.* **2017**, *33*, 594–603. [CrossRef] [PubMed]
- 55. Daniels, C.J.; Lai, L.B.; Chen, T.-H.; Gopalan, V. Both Kinds of RNase P in All Domains of Life: Surprises Galore. *RNA* 2019, 25, 286–291. [CrossRef] [PubMed]
- 56. Di Giulio, M. The RNase P, LUCA, the Ancestors of the Life Domains, the Progenote, and the Tree of Life. *Biosystems* **2022**, 212, 104604. [CrossRef]
- 57. Larsen, N.; Zwieb, C. SRP-RNA Sequence Alignment and Secondary Structure. Nucleic Acids Res. 1991, 19, 209–215. [CrossRef]
- Rosenblad, M.A.; Larsen, N.; Samuelsson, T.; Zwieb, C. Kinship in the SRP RNA Family. RNA Biol. 2009, 6, 508–516. [CrossRef]
  [PubMed]
- 59. Barrick, J.E.; Sudarsan, N.; Weinberg, Z.; Ruzzo, W.L.; Breaker, R.R. 6S RNA Is a Widespread Regulator of Eubacterial RNA Polymerase That Resembles an Open Promoter. *RNA* 2005, *11*, 774–784. [CrossRef]
- 60. Wassarman, K.M. 6S RNA: A Small RNA Regulator of Transcription. Curr. Opin. Microbiol. 2007, 10, 164–168. [CrossRef]
- 61. Cavanagh, A.T.; Wassarman, K.M. 6S RNA, a Global Regulator of Transcription in Escherichia Coli, Bacillus Subtilis, and Beyond. *Annu. Rev. Microbiol.* **2014**, *68*, 45–60. [CrossRef] [PubMed]
- 62. Burenina, O.Y.; Elkina, D.A.; Hartmann, R.K.; Oretskaya, T.S.; Kubareva, E.A. Small Noncoding 6S RNAs of Bacteria. *Biochemistry* 2015, *80*, 1429–1446. [CrossRef]
- Chen, J.; Wassarman, K.M.; Feng, S.; Leon, K.; Feklistov, A.; Winkelman, J.T.; Li, Z.; Walz, T.; Campbell, E.A.; Darst, S.A. 6S RNA Mimics B-Form DNA to Regulate Escherichia Coli RNA Polymerase. *Mol. Cell* 2017, 68, 388–397.e6. [CrossRef] [PubMed]
- 64. Wassarman, K.M. 6S RNA, A Global Regulator of Transcription. Microbiol. Spectr. 2018, 6, 1–13. [CrossRef]
- 65. Berg, J.M.; Tymoczko, J.L.; Stryer, L. Biochemistry, 5th ed.; W H Freeman: New York, NY, USA, 2002; ISBN 978-0-7167-3051-4.
- 66. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. *Molecular Biology of the Cell*, 4th ed.; Garland Science: New York, NY, USA, 2002; ISBN 978-0-8153-3218-3.
- 67. Berg, M.D.; Brandl, C.J. Transfer RNAs: Diversity in Form and Function. RNA Biol. 2021, 18, 316–339. [CrossRef] [PubMed]
- 68. Hyeon, C.; Thirumalai, D. Chain Length Determines the Folding Rates of RNA. *Biophys. J.* 2012, 102, L11–L13. [CrossRef]
- 69. Root-Bernstein, R.; Kim, Y.; Sanjay, A.; Burton, Z.F. tRNA Evolution from the Proto-tRNA Minihelix World. *Transcription* **2016**, 7, 153–163. [CrossRef]
- Rodin, S.; Rodin, A.; Ohno, S. The Presence of Codon-Anticodon Pairs in the Acceptor Stem of tRNAs. *Proc. Natl. Acad. Sci. USA* 1996, 93, 4537–4542. [CrossRef] [PubMed]
- 71. Tamura, K.; Schimmel, P. Chiral-Selective Aminoacylation of an RNA Minihelix. Science 2004, 305, 1253. [CrossRef] [PubMed]
- 72. Guo, X.; Su, M. The Origin of Translation: Bridging the Nucleotides and Peptides. Int. J. Mol. Sci. 2023, 24, 197. [CrossRef]
- Widmann, J.; Di Giulio, M.; Yarus, M.; Knight, R. tRNA Creation by Hairpin Duplication. J. Mol. Evol. 2005, 61, 524–530. [CrossRef] [PubMed]

- 74. Prosdocimi, F.; de Farias, S.T. Entering the Labyrinth: A Hypothesis about the Emergence of Metabolism from Protobiotic Routes. *Biosystems* **2022**, 220, 104751. [CrossRef]
- 75. Prosdocimi, F.; de Farias, S.T. Origin of Life: Drawing the Big Picture. *Prog. Biophys. Mol. Biol.* 2023, 180–181, 28–36. [CrossRef] [PubMed]
- 76. Abbott, J.A.; Francklyn, C.S.; Robey-Bond, S.M. Transfer RNA and Human Disease. Front. Genet. 2014, 5, 158. [CrossRef]
- 77. Brandon, M.C.; Lott, M.T.; Nguyen, K.C.; Spolim, S.; Navathe, S.B.; Baldi, P.; Wallace, D.C. MITOMAP: A Human Mitochondrial Genome Database—2004 Update. *Nucleic Acids Res.* 2005, 33, D611–D613. [CrossRef] [PubMed]
- 78. Yarham, J.W.; Elson, J.L.; Blakely, E.L.; McFarland, R.; Taylor, R.W. Mitochondrial tRNA Mutations and Disease. *WIREs RNA* 2010, 1, 304–324. [CrossRef]
- 79. Pavon-Eternod, M.; Gomes, S.; Geslain, R.; Dai, Q.; Rosner, M.R.; Pan, T. tRNA Over-Expression in Breast Cancer and Functional Consequences. *Nucleic Acids Res.* 2009, *37*, 7268–7280. [CrossRef] [PubMed]
- Glatz, C.; D'Aco, K.; Smith, S.; Sondheimer, N. Mutation in the Mitochondrial tRNAVal Causes Mitochondrial Encephalopathy, Lactic Acidosis and Stroke-like Episodes. *Mitochondrion* 2011, 11, 615–619. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.