



CHALMERS
UNIVERSITY OF TECHNOLOGY

From peptides to DNA: All required steps can be catalyzed

Downloaded from: <https://research.chalmers.se>, 2026-04-16 20:31 UTC

Citation for the original published paper (version of record):

Lei, S., Senissar, M., Leistner, K. et al (2026). From peptides to DNA: All required steps can be catalyzed. *Proceedings of the National Academy of Sciences of the United States of America*, 123(9). <http://dx.doi.org/10.1073/pnas.25343871231>

N.B. When citing this work, cite the original published paper.



From peptides to DNA: All required steps can be catalyzed

Lei Shi^{a,1} , Mériem Senissar^{b,1}, Kirsten Leistner^c, Carsten Jers^b, and Ivan Mijakovic^{a,b,2}

Edited by Jens Nielsen, BioInnovation Institute, Hellerup, Denmark; received December 3, 2025; accepted February 1, 2026

Ensuring information flow (heredity) and metabolic processes (catalysis) are two important prerequisites for early evolution. The widely accepted “RNA world” theory proposes that ancient RNAs ensured both heredity and catalysis during the transition from prebiotic to biotic evolution. However, alternative hypothetical molecules and processes have also been proposed, suggesting that catalytic peptides may have existed before polynucleotides, and that their sequences were later reverse translated into genes. Our objective was to experimentally address these alternative theories by asking whether the steps required for the hypothetical conversion of peptide sequences into DNA could be catalyzed by the existing molecular kit. The reactions we tested comprise i) step-wise degradation of peptides by a processive amino peptidase, sequentially releasing amino acids, ii) matching the identity of released amino acids to codons by aptazymes (RNA adapters that recognize amino acids and self-cleave and release specific codon triplets in response), and iii) ligating codon triplets into longer RNAs that can be reverse-transcribed into DNA. In a hypothetical processive system based on these reactions, the resulting DNA sequence would match the sequence of amino acids in the starting peptide. Our results suggest that all these steps can be catalyzed, and therefore the possibility of reverse translation occurring at some point in early evolution should not be disregarded.

reverse translation hypothesis | prebiotic evolution | catalytic RNA | processive aminopeptidase | amino acid-to-codon matching

As argued by Wolf & Koonin (1), the origin of translation of genetic information from nucleic acids to proteins in the RNA world is one of the most difficult problems in evolutionary biology. The “frozen accident” theory, postulated by Francis Crick (2), claimed that the assignment of codons to amino acids was random. This notion has been abandoned due to accumulating evidence of specific interactions between amino acids and RNA, which were proposed to form a basis for the assignment of codons (3). This line of investigation resulted in the discovery of RNA aptamers (4) that are capable of binding many (but not all) proteinogenic amino acids (5). Coevolution theory (6) postulated that not all proteinogenic amino acids were available in prebiotic conditions. Hence, many of them had to be produced by biosynthetic pathways, and this guided the gradual development of the genetic code during biotic evolution. Higgs’ and Massey’s theories proposed explanations for this code expansion, based on physicochemical properties of amino acids and error cost minimization (7, 8). Most of the current debate on the origin of translation consists of attempts to combine elements of the above-mentioned theories into a coherent narrative (9). However, to paraphrase a recent statement by Kun & Radványi (10) who urged empiricists to conduct more experiments: most papers on the origin of the genetic code are reviews and not original research.

Information flow (heredity) and metabolic processes (catalysis) are arguably the two central concepts featured in the transition from prebiotic to biotic evolution. The ancient RNAs in the RNA world theory unite both these concepts (11), having the capacity to ensure heredity and catalysis. In line with the RNA world theory, plausible scenarios have been proposed for transition from RNA-based to peptide-based catalysis (12, 13). One issue with the RNA world theory, as pointed out by Stanley E. Miller (14), is that no plausible prebiotic process for producing ribonucleosides or ribonucleotides has been demonstrated. An alternative molecule that can be synthesized in prebiotic conditions (14), the peptide nucleic acid (PNA), has been proposed as the first genetic molecule (15). PNA is essentially a DNA analogue with a peptide backbone that can hybridize to complementary oligonucleotides (16). In the PNA/RNA world hypothesis (15), the peptide-based PNAs serve at the molecules of heredity, and the RNAs enter the stage later, serving as catalysts. Further alternative theories exist, stipulating that autocatalytic systems existed first, and genetic heredity evolved later (17). Among others, this point has been argued by Markus Rasler in his concept of nonenzymatic metabolism (18), and William

Significance

Our understanding of the transition from prebiotic to biotic evolution and the formation of living cells is incomplete. Some theories propose that RNA molecules with catalytic properties were the key in this transition, ensuring that early cells can catalyze metabolic reactions and transmit genetic information. Alternative theories propose that peptides were the first catalysts, but peptides do not have the capacity to transmit genetic information. In this study, we demonstrate that genetic information can conceivably be transmitted from peptides to DNA, by using the existing molecular catalysts, both RNA and peptide based. Our results do not offer proof for any of the theories, nor do they challenge any of them. They merely contribute to assessing which scenarios are possible.

Author affiliations: ^aSystems and Synthetic Biology Division, Department of Life Sciences, Chalmers University of Technology, Gothenburg 41296, Sweden; ^bThe Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kongens Lyngby 2800, Denmark; and ^cThe MathWorks Aktiebolag, Gothenburg 41755, Sweden

Author contributions: M.S., K.L., C.J., and I.M. designed research; L.S. and M.S. performed research; L.S., M.S., C.J., and I.M. analyzed data; and L.S., K.L., and I.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2026 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

¹L.S. and M.S. contributed equally to this work.

²To whom correspondence may be addressed. Email: ivan.mijakovic@chalmers.se.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2534387123/-/DCSupplemental>.

Published February 26, 2026.

F. Martin in the concept of autocatalytic chemical networks that preceded proteins and RNA in evolution (19, 20). Hartmann & Smith proposed the existence of early catalytic peptides in the evolutionary transition from a thioester world to a phosphate world, where sugar phosphates, nucleotides, and polynucleotides hypothetically appeared after the emergence of these catalytic peptides (21). In several of these theories, the problem of prebiotic synthesis of ribonucleosides or ribonucleotides is solved by catalytic peptides. But, if catalytic peptides existed before polynucleotides (14, 21) then a key question arises about information flow and heredity. Was the information contained in peptide sequences at some early point of evolution “written” into nucleic acid polymers?

Logically, the idea that reading (transcription of DNA sequences to mRNA and translation of mRNAs to proteins) is linked to writing, is not controversial. Yet, this hypothetical scenario of a process of inversed transcription/translation has not been extensively explored. The theoretical possibility of reverse translation of proteins into genes has been excluded from the standpoint of information theory by Yockey (22), based on the argument that the genetic code is degenerate. This argument was later refuted by Craig (23), who argued for the theoretical possibility of a reverse translation mechanism involving tRNAs, and essentially a reverse sequence of ribosomal reactions, coupled to elongation of mRNAs. To the best of our knowledge, the only experimental attempt to simulate such mRNA elongation was done by Masayuki Nashimoto (24), who demonstrated that engineered tRNA can be cleaved to transmit its anticodon to another RNA molecule. This concept had definite value but did not have a means of connecting codon transfer to the identity of a specific amino acid, nor was it able to repeatedly connect codons together based on the identity of amino acids. As a theoretical alternative to tRNA-based

reverse translation, Tozzi & Mazzeo proposed the existence of hypothetical two-dimensional peptides capable of amino acid–nucleotide matching (25).

In this paper, we propose a series of catalyzed reactions that represent all the steps required for a hypothetical process of reverse translation. Our proposed sequence of reactions (Fig. 1) starts with a processive amino peptidase, which sequentially releases amino acids from the N terminus of a peptide. The identity of released amino acids can be matched to codons by aptazymes, specific RNA adapters that contain an aptamer module that recognizes amino acids, and a ribozyme module, that can self-cleave and release specific codon triplets in response to amino acid binding. These codon triplets can be ligated into a longer RNA and reverse-transcribed into DNA. In a hypothetical processive system based on these reactions, the resulting DNA sequence would match the sequence of amino acids in the starting peptide.

Results

From Peptides to Amino Acids. To carry out the first step in the proposed sequence of reactions (Fig. 2A), we chose a rare processive bacterial aminopeptidase YsdC (26) (SI Appendix, Fig. S1). We chose to work with several synthetic peptides that were 12 amino acid-long. The peptides contained arginine (R) and/or glutamine (Q) in different ratios: R_{12} , Q_{12} , $(RRRQ)_3$ and $(QQQR)_3$ (Fig. 2B). The peptides were incubated with YsdC, and the reaction products were analyzed by mass spectrometry. The analysis of free amino acids revealed that YsdC successfully degraded the synthetic peptides into single amino acids, releasing only R from the peptide R_{12} , only Q from the peptide Q_{12} , and stoichiometrically correct mixtures of R and Q from the peptides $(RRRQ)_3$ and $(QQQR)_3$, respectively (Fig. 2B).

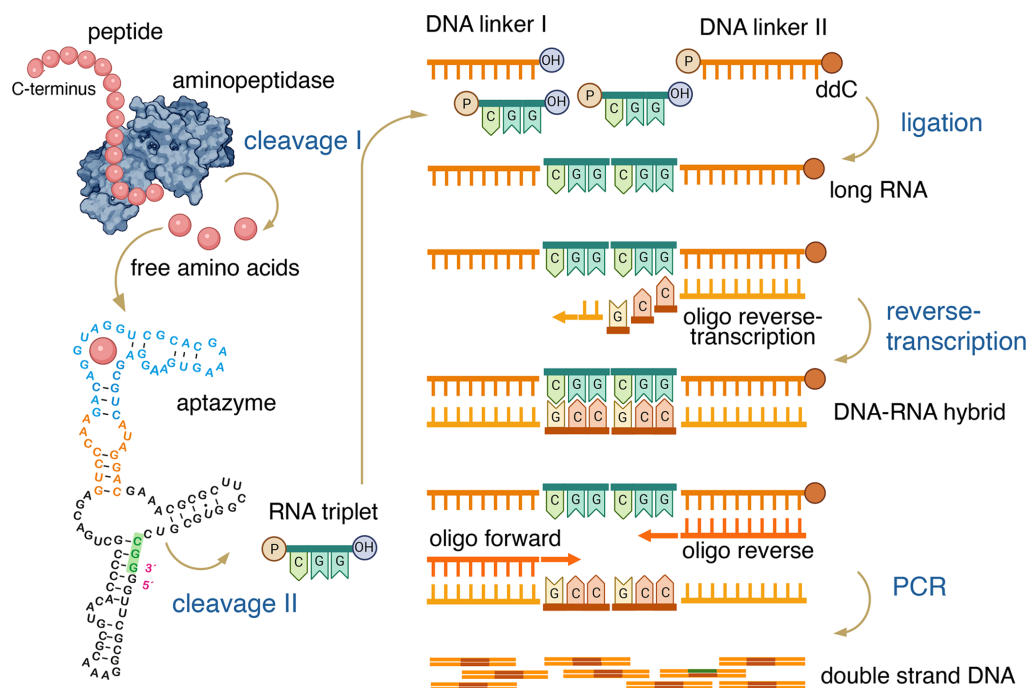
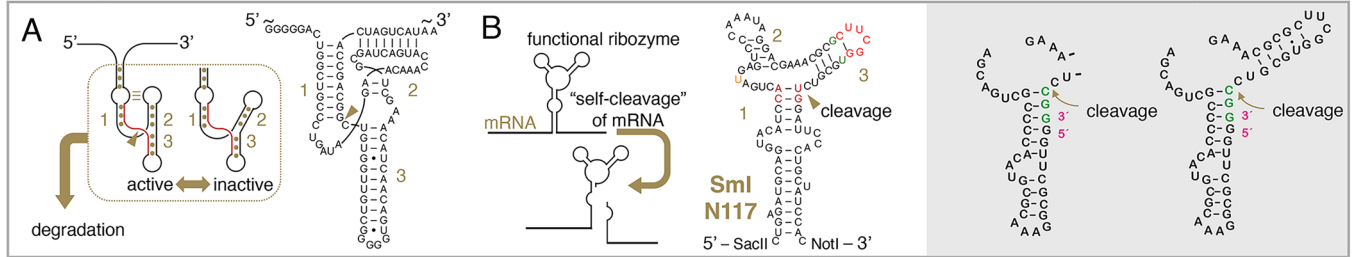
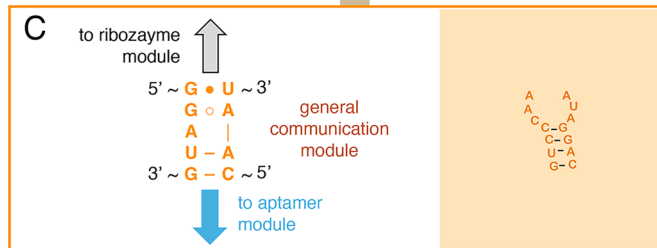


Fig. 1. Schematic representation of the proposed processes of reverse translation: from peptides to DNA. Cleavage I: the aminopeptidase (blue) processively cleaves the peptide bonds at the N terminus of a peptide (pink), releasing free amino acids. Cleavage II: the RNA aptazyme binds a free amino acid, which triggers its ribozyme activity, self-cleaving, and releasing the RNA triplet from its 3' end. Ligation: the RNA triplet is ligated into long RNA by T4 RNA Ligase 1 (ssRNA ligase). DNA linkers I and II are introduced in the ligation reaction to obtain the sites for pairing with the primer for reverse transcription, and thereafter for pairing with primers during PCR. DNA linker II is phosphorylated at 5' to enable the ligation with the codons and is modified with dideoxycytidine (orange solid circle) at 3' to block further ligation at that end. Reverse-transcription: long RNA is converted into RNA–DNA hybrid by reverse transcriptase. PCR: RNA–DNA hybrid is amplified by PCR. RNA sequences are shown in green and DNA in orange.

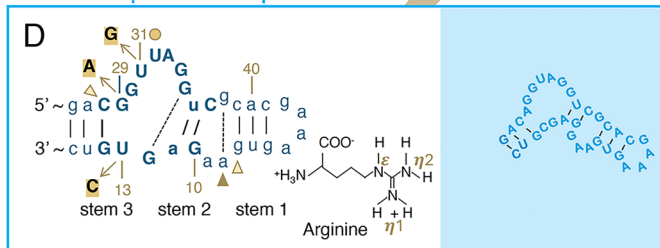
I: ribozyme



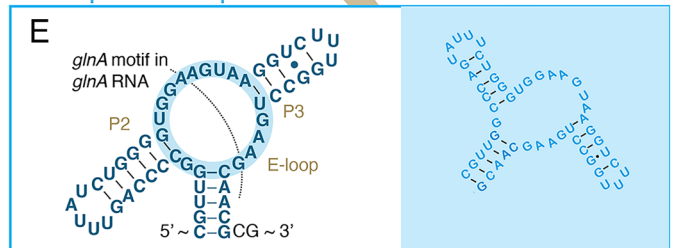
II: communication module



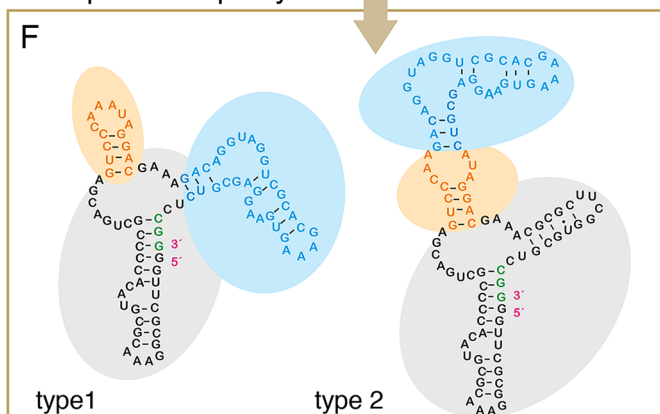
III: R-responsive aptamer



Q-responsive aptamer



R-responsive aptazyme



Q-responsive aptazyme

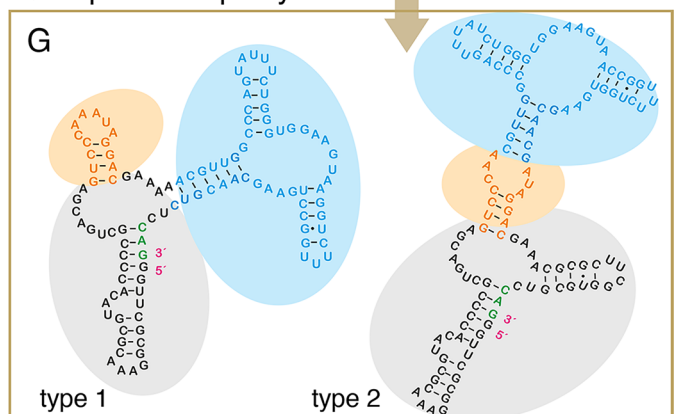


Fig. 3. Schematic representation of the design of R- and Q-responsive aptazymes. The three key modules of amino acid-responsive aptazymes are indicated with different colors: I, ribozyme (gray); II, communication module (orange); and III, amino acid-responsive aptamer (blue). The two R/Q-responsive aptazymes, type 1 and type 2, were obtained by combining each of the three elements, based on the available literature. (A) Schematic representation of the hammerhead ribozyme and a motif from *Y. lipolytica* (29). Cleavage site is indicated by an arrow. (B) Self-cleaving *Schistosoma* Sm1 ribozyme from the digenetic trematode *Schistosoma mansoni* (30). Cleavage site is indicated by an arrow. Our ribozyme design combined elements of both ribozymes (gray box). (C) The general aptazyme communication module as proposed by Kertsburg and Soukup (33). (D) Arginine-specific RNA aptamer that was obtained from an in vitro selection study for specific binding to L-citrulline, and then subjected to further selection for recognition and binding for L-arginine (34). (E) A *glnA* motif of the 67 *glnA* RNA from *Synechococcus elongatus*, binding to L-glutamine (35). (F) The resulting R-responsive aptazymes of type 1 and 2. (G) The resulting Q-responsive aptazymes of type 1 and 2.

to aptamers were tested (Fig. 3 F and G). In type 1 aptazyme, the communication module was placed within the ribozyme as an independent stem-and-loop structure separated from the aptamer. In type 2 aptazyme, the communication module was placed to directly bridge the ribozyme to the aptamer structure. The first attempt was made with type 1 aptazymes (SI Appendix, Fig. S2). This failed to provide the desired result because the Q-responsive aptazyme self-cleaved even in the absence of Q,

while the R-responsive aptazyme did not self-cleave even after addition of R. Our next attempt involved aptazymes of type 2 (Fig. 2 C and E). The Q-responsive aptazyme (Fig. 2 C, Right) did not self-cleave in the absence of either Q or Mg^{2+} (required for ribozyme reaction), but it did self-cleave in the presence of both, reaching >50% self-cleavage in 45 min. The R-responsive aptazyme (Fig. 2 E, Right) also strictly required the presence of R and Mg^{2+} for self-cleavage, reaching >50% self-cleavage in 8 h.

From RNA Triplets to DNA. Next, the RNA triplets, CAG for the Q-responsive aptazyme and CGG for the R-responsive aptazyme, were ligated with the help of DNA linkers (as outlined in Fig. 1) using a T4 RNA Ligase 1 (ssRNA Ligase). It is important to note that triplets released by the self-cleavage of aptazymes do not have a phosphorylated 5' end, which is necessary for subsequent ligation. This issue can be solved by polynucleotide kinases (36), which can catalyze 5' phosphorylation of short RNA oligonucleotides, and are known to be ancient enzymes (37). Alternatively, there are mechanisms of RNA cleavage that produce a phosphorylated 5' end, such as ribonuclease P (38), which is an ancient ribonucleoprotein. It contains an RNA component that is sufficient for catalyzing the cleavage in vivo (39). In our experiments, we used a 5' phosphorylated version of RNA triplets, thus established long RNA templates for the reverse transcriptase (Fig. 2 D and F). The DNA linkers were used to provide an anchor point for the following steps, reverse transcription and PCR. Following the ligation step, we observed the growth of long RNAs, although the RNA triplets themselves could not be visualized in TBE-urea gels (SI Appendix, Fig. S3). From this point onward, routine reactions of reverse transcription and PCR were carried out, producing an RNA–DNA hybrid and a double-stranded DNA, respectively (Fig. 1 and SI Appendix, Fig. S4 A–D). The synthesized double-stranded DNA was sequenced (SI Appendix, Fig. S4E), revealing the correct subsequent placement of poly(CAG) sequences issued from the “Q” RNA triplets, and poly(CG) sequences issued from the “R” RNA triplets.

Discussion

Our results suggest that each step of the hypothetical conversion of a peptide sequence into a DNA sequence (Fig. 1) can be catalyzed with existing types of macromolecules/biocatalysts. By consequence, it is not impossible to imagine that prebiotic peptides capable of synthesizing RNA precursors could have provided all the catalysts needed to carry out a similar process. Importantly, to offer an evolutionary advantage, such reverse translation systems would need to operate in the presence of forward translation. Current evolutionary theory stipulates that forward translation arose gradually from a set of primitive reactions (40, 41). Hence, it could be speculated that primitive versions of reactions leading to both forward and reverse translation may have coexisted at some point in early evolution.

The key component of the proposed hypothetical process of reverse translation would be an appropriate “converter” of amino acid to nucleotide triplet identities, which in our case was an RNA aptazyme. While we demonstrated catalysis of each step in the hypothetical process of reverse translation, our current system is not processive. Processivity would require a more complex molecular structure where all the above-mentioned steps are sterically confined and coupled. It is plausible to presume that such processes would start with low processivity in more primitive stages and evolve efficiency over time. Many authors discussed physical compartmentation from the environment and self-organization as an important step in protocells, forming early catalytical compartments, giving rise to gradually more processive and efficient metabolic pathways (42, 43). In the absence of physical confinement, other mechanisms, such as clay hydrogels, may have played the role of ensuring high local concentrations of biomolecules, allowing for processive reactions (44). A good definition of such microspaces in which protometabolites and protobiopolymers could coevolve and gradually increase catalytic processivity and efficiency has been provided by Jan Spitzer (45). Of particular interest in this respect is the metabolosome model proposed by

Gibson & Lamond (46), in which RNA structures function as coordination centers/scaffolds binding small molecules, cofactors, and potentially amino acids that participate in primitive metabolic pathways.

In conclusion, we propose that our results allow for the possibility of an evolutionary scenario in which sequences of early catalytic peptides could get written into DNA (23–25). In such a hypothetical scenario, at some early point after the emergence of the first genomes, new evolutionary forces would have kicked in, balancing genome stability versus plasticity. Once functional protogenomes have been formed, maintaining the function of “genome writers” would gradually be counterselected, since they would compromise genome stability. Inactivation of biological pathways commonly occurs by blocking the first step of the process. In the case of our putative genome writer system, this would mean blocking, or counterselecting, the activity of processive aminopeptidases. From this perspective, it is interesting to point out that extant cells degrade most of their polymers processively. For example, nucleic acids and polysaccharides are degraded one nucleotide (47) or sugar monomer (48) at a time, which makes perfect sense energetically. Not so for protein degradation. Cells break down proteins into increasingly shorter peptides, using complex and energetically costly multienzyme assemblies (49). N-terminal peptidases with any sort of processivity are very rare in nature. Could it be that processive aminopeptidases were counterselected early in the genomic era? We hope that our findings and the questions they raise will meaningfully contribute to the debate about early evolution. In addition, they could have potential uses in synthetic biology, as a basis for a method to synthesize genome components based on known protein sequences.

Materials and Methods

Aminopeptidase Cleavage. Aminopeptidase gene *ysdC* was amplified from *B. subtilis* 168 genomic DNA by primer *ysdC_Bam_F* (5'-ATATGGATCCGCAAAATTAG ATGAAACATTGACC-3') and *ysdC_Hind_R* (5'-ATATAAGCTTCTATTGGTACGTAATTCG TCAAC-3'), then inserted in pQE30 between *Bam*HI and *Hind*III. 6xHis-tagged YsdC was purified using Ni-NTA agarose (Qiagen) as described previously (26). Aminopeptidase reaction was performed by mixing 160 μM peptides with 40 μM aminopeptidase in a 50 μL reaction solution. For reactions with R₁₂(RRRQ)₃ and (QQQR)₃, the solution contained 100 mM Tris-HCl pH 7.5, 100 mM NaCl, 50 mM KCl, 1 mM magnesium acetate, 5% glycerol and 0.05% Triton X-100. For reaction with Q₁₂, 100 mM Tris-HCl pH 7.5 was replaced by 100 mM Tris-HCl pH 8.8, while the other components remained the same. The reactions were incubated overnight at room temperature. The control reactions were stopped immediately by being frozen at –20 °C directly after mixing.

Amino Acid Quantification. The quantification of free amino acids released into reaction solution was determined by aTRAQ™ Kit Physiological 50 Assay kit (SCIEX) in an LC-tandem MS (LC-MS/MS) system QTRAP 6500+ (SCIEX) with an amino acid analyzer (AAA) C18 column (SCIEX). The concentration of free amino acids was calculated according to aTRAQ™ Internal Standard (SCIEX). In brief, the reactions were first mixed with sulfosalicylic acid solution to precipitate and remove proteins. The obtained soluble samples were then labeled with aTRAQ Reagent Δ8 and mixed with the internal standards, which were pre-labeled with aTRAQ Reagent Δ0. The mixtures were analyzed by LC-MS/MS using QTRAP System operating in multiple reaction monitoring (MRM) mode. The amino acids labeled with Δ8 or Δ0 have identical structures but with different numbers of isotopes, and therefore they exhibit the same retention times but can be differentiated by MS/MS.

Aptazyme Cleavage. Four aptazymes, Gln aptazyme type 1 (5'-GGTTCGGG AAACCGGTACACCCTGCTGACGAGTCCCAAATAGACGAAAACGTTGGCCAGTATTCT GGGTGGAAGTAAGGTCTTTGGCCTGAAGCAACGCTCCAG-3') and type 2 (5'-GGTTCG CGGAAACCGGTACACCCTGCTGACGAGTCCCAACGTTGGCCAGTTATCTGGGTGGAAG TAACCGTTTCTGGTGAAGCAACGCTAGGACGAAACGCGCTTCGGTCCGCTCCAG-3'), Arg

aptazyme type 1 (5'-GGTTCGCGAAACGCGTACACCCCGCTGACGAGTCCCAATAGGACGAAAGACAGGTAGGTGCGCACGAAAGTGAAGGAGCGTCTCCGG-3') and type 2 (5'-GGTTCGCGGAAACGCGTACACCCCGCTGACGAGTCCCAAGACAGGTAGGTGCGCACGAAAGTGAAGGAGCGTCATAGGACGAAACGCGTTCGGTGCCTCCGG-3') were synthesized by Integrated DNA Technologies. Self-cleavage of Gln type 1 and 2 was performed at 50 °C in reactions containing Tris-HCl 25 mM, NaCl 50 mM, glycerol 5%, MgCl₂ 0.5 mM, and Gln 50 mM as indicated in figures. Self-cleavage of Arg type 1 and 2 was performed at 12 °C in reactions containing Tris-HCl 25 mM, NaCl 50 mM, glycerol 5%, MgCl₂ 0.5 mM, and Arg 50 mM as indicated in the figures. Incubation times for all cleavage reactions are indicated in respective figure legends. To stop all reactions, they were mixed with Novex™ TBE-Urea Sample Buffer (2×) (Invitrogen) and the reaction products were separated on 15% Novex™ TBE-Urea Gels (Invitrogen) for 6.5 h at 180 V in 4 °C. RNA was visualized by SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen).

RNA Triplet Ligation. RNA triplets CAG (5'-Phosphorylation CAG-3') and CGG (5'-Phosphorylation CGG-3') with phosphorylated 5' end were synthesized by GenScript. RNA ligation was performed using T4 RNA Ligase 1 (ssRNA Ligase) (New England Biolabs), according to the manufacturer's instructions. DNA linker I (5'-CTCGAGTTCGACGATTCGATAAGCTTGATATC-3') and II (5'-phosphorylation AATCTCTGACGCCGGGGATCCACTAGTCTAGA dideoxycytidine-3') were included in the ligation reaction to introduce the sites for following reverse-transcription and PCR. Their sequences were selected to match the MCS region of the plasmid pBluescript to enable integration of the reaction products into the plasmid. Briefly, a 10 μL reaction contained 1× T4 RNA Ligase Reaction Buffer, 20% (wt/vol) PEG 8000, 0.5 μL RNase Inhibitor (Muriine), 1 μL T4 RNA Ligase 1, 1 mM ATP, 20 μM RNA triplets, 0.2 μM DNA linkers I and II. The reaction was incubated at 16 °C overnight, then mixed with Novex™ TBE-Urea Sample Buffer (2×) (Invitrogen) and incubated at 70 °C for 2 min to remove secondary structures. Samples were separated by 15% Novex™ TBE-Urea Gels (Invitrogen) for 1 h, unless otherwise

specified in figure legends, at 180 V at 4 °C. Reaction products were visualized by SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen). For DNA digestion, the overnight RNA ligation reaction was subjected to RNase-free DNase Set (Qiagen). For RNA digestion, the overnight RNA ligation reaction was subjected to RNase Cocktail™ Enzyme Mix (Invitrogen).

Reverse-Transcription. Before reverse transcription, the RNA ligation reaction was subjected to column cleanup with Monarch® RNA Cleanup Kit (Qiagen). Reverse-transcription was performed by the RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) with the oligo for reverse-transcription (5'-TCTAGAACTA GTGGATCCCCGGGCTG CAGGAAT-3'). Reverse-transcription reaction was mixed with Novex™ TBE-Urea Sample Buffer (2×) (Invitrogen) and subjected to 70 °C for 2 min, then separated on 15% Novex™ TBE-Urea Gels (Invitrogen) for 0.5 h at 180 V and 4 °C. RNA/DNA was visualized by SYBR™ Gold Nucleic Acid Gel Stain (Invitrogen).

PCR, Plasmid Integration, and Sequencing. DNA fragments issued from the reverse transcription reaction were amplified with PCR oligo forward (5'-CTCGA GGTCGACGGTATCGATAAGCTTGATATC-3') and PCR oligo reverse (5'-TCTAGAACTAG TGGATCCCCGGGCTG CAGGAAT-3') using DreamTaq DNA Polymerase (Thermo Scientific). PCR products were digested by *Hind*III and *Bam*HI and inserted into pBluescript II SK. The inserts were analyzed by commercial DNA sequencing.

Data, Materials, and Software Availability. Study data are included in the article and/or *SI Appendix*.

ACKNOWLEDGMENTS. We are grateful to Prof. Jörg S. Hartig and his research group for precious advice on aptazyme design. We acknowledge funding from the Villum Fonden, grant number 00022883 (to I.M.), and the Novo Nordisk Foundation, grant number NNF20CC0035580 (to I.M.). Illustrations were created with the assistance of BioRender. Shi, L. (2026) <https://BioRender.com/3n9mce>.

1. Y. I. Wolf, E. V. Koonin, On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biol. Direct.* **4**, 14 (2007), 10.1186/1745-6150-2-14.
2. F. H. Crick, The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968), 10.1016/0022-2836(68)90392-6.
3. C. R. Woese, D. H. Dugre, W. C. Saxinger, S. A. Dugre, The molecular basis for the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **55**, 966–974 (1966), 10.1073/pnas.55.4.966.
4. M. Yarus, A specific amino acid binding site composed of RNA. *Science* **240**, 1751–1758 (1988), 10.1126/science.3381099.
5. M. Yarus, The genetic code and RNA-amino acid affinities. *Life* **7**, 13 (2017), 10.3390/life7020013.
6. J. T. Wong, A co-evolution theory of the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1909–1912 (1975), 10.1073/pnas.72.5.1909.
7. P. G. Higgs, A four-column theory for the origin of the genetic code: Tracing the evolutionary pathways that gave rise to an optimized code. *Biol. Direct.* **4**, 16 (2009), 10.1186/1745-6150-4-16.
8. S. E. Massey, A sequential 2–1–3 model of genetic code evolution that explains codon constraints. *J. Mol. Evol.* **67**, 510 (2006), 10.1007/s00239-005-0222-0.
9. M. Di Giulio, Theories of the origin of the genetic code: Strong corroboration for the coevolution theory. *Biosystems* **239**, 105217 (2024), 10.1016/j.biosystems.2024.105217.
10. A. Kun, A. Radványi, The evolution of the genetic code: Impasses and challenges. *Biosystems* **164**, 217–225 (2018), 10.1016/j.biosystems.2017.10.006.
11. T. R. Cech, The RNA worlds in context. *Cold Spring Harb. Perspect. Biol.* **4**, a006742 (2012), 10.1101/cshperspect.a006742.
12. S. D. Copley, E. Smith, H. J. Morowitz, A mechanism for the association of amino acids with their cocoon and the origin of the genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4442–4447 (2005), 10.1073/pnas.0501049102.
13. F. Müller *et al.*, A prebiotically plausible scenario of an RNA-peptide world. *Nature* **605**, 279–284 (2022), 10.1038/s41586-022-04676-3.
14. K. E. Nelson, M. Levy, S. L. Miller, Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 3868–3871 (2000), 10.1073/pnas.97.8.3868.
15. P. E. Nielsen, Peptide nucleic acid (PNA): A model structure for the primordial genetic material? *Orig. Life Evol. Biosph.* **23**, 323–327 (1993), 10.1007/BF01582083.
16. M. Egholm *et al.*, Pna hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566–568 (1993), 10.1038/365566a0.
17. M. Vitas, A. Dobovišek, In the beginning was a mutualism—On the origin of translation. *Orig. Life Evol. Biosph.* **48**, 223–243 (2018), 10.1007/s11084-018-9557-6.
18. M. Ralsler, An appeal to magic? The discovery of a non-enzymatic metabolism and its role in the origins of life. *Biochem. J.* **475**, 2577–2592 (2018), 10.1042/BCJ20160866.
19. M. C. Weiss, M. Premier, J. C. Xavier, V. Zimorski, W. F. Martin, The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *PLoS Genet.* **14**, e1007518 (2018), 10.1371/journal.pgen.1007518.
20. J. C. Xavier, W. Hordijk, S. Kauffman, M. Steel, W. F. Martin, Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. B.* **287**, 20192377 (2020), 10.1098/rspb.2019.2377.
21. H. Hartmann, T. F. Smith, Origin of the genetic code is found at the transition between a thioester world of peptides and the phosphoester world of polynucleotides. *Life* **9**, 69 (2019), 10.3390/life9030069.
22. H. P. Yockey, Can the central dogma be derived from information theory? *J. Theor. Biol.* **74**, 149–152 (1978), 10.1016/0022-5193(78)90295-3.
23. R. Craig, The theoretical possibility of reverse translation of proteins into genes. *J. Theor. Biol.* **88**, 757–760 (1981), 10.1016/0022-5193(81)90250-2.
24. M. Nashimoto, The RNA/protein symmetry hypothesis: Experimental support for reverse translation of primitive proteins. *J. Theor. Biol.* **209**, 181–187 (2001), 10.1006/jtbi.2000.2253.
25. A. Tozzi, M. Mazzeo, The first nucleic acid strands may have grown on peptides via a primeval reverse translation. *Acta Biotheor.* **71**, 23 (2023), 10.1007/s10441-023-09474-6.
26. N. Kishor, P. Guptasarma, Direct N-terminal sequencing of polypeptides using a thermostable bacterial aminopeptidase and MALDI-TOF mass spectrometry. *Anal. Biochem.* **488**, 6–8 (2015), 10.1016/j.ab.2015.07.006.
27. C. Rehm, B. Klausner, M. Finke, J. S. Hartig, Engineering aptazyme switches for conditional gene expression in mammalian cells utilizing an in vivo screening approach. *Methods Mol. Biol.* **2023**, 199–212 (2021), 10.1007/978-1-0716-1499-0_14.
28. S. K. Silverman, Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **9**, 377–383 (2003), 10.1261/ma.2200903.
29. L. A. Wurthaler, B. Klausner, J. S. Hartig, Highly motif- and organism-dependent effects of naturally occurring hammerhead ribozyme sequences on gene expression. *RNA Biol.* **15**, 231–241 (2018), 10.1080/15476286.2017.1397870.
30. G. Ferbeyre, J. M. Smith, R. Cedergren, Schistosoma satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18**, 3880–3888 (1998), 10.1128/MCB.18.7.3880.
31. M. Martick, W. G. Scott, Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**, 309–320 (2006), 10.1016/j.cell.2006.06.036.
32. L. Yen *et al.*, Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature* **431**, 471–476 (2004), 10.1038/nature02844.
33. A. Kertsburg, G. A. Soukup, A versatile communication module for controlling RNA folding and catalysis. *Nucleic Acids Res.* **30**, 4599–4606 (2002), 10.1093/nar/gkf596.
34. P. Burgstaller, M. Kochoyan, M. Famulok, Structural probing and damage selection of citrulline- and arginine-specific RNA aptamers identify base positions required for binding. *Nucleic Acids Res.* **23**, 4769–4776 (1995), 10.1093/nar/23.23.4769.
35. T. D. Ames, R. R. Breaker, Bacterial aptamers that selectively bind glutamine. *RNA Biol.* **8**, 82–89 (2011), 10.4161/rna.8.1.13864.
36. K. Randerath, E. Randerath, T. F. Danna, L. van Golen, K. L. Putman, A new sensitive ³²P-postlabeling assay based on the specific enzymatic conversion of bulky DNA lesions to radiolabeled dinucleotides and nucleoside 5'-monophosphates. *Carcinogenesis* **10**, 1231–1239 (1989), 10.1093/carcin/10.7.1231.
37. D. D. Leipe, E. V. Koonin, L. Aravind, Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* **333**, 781–815 (2003), 10.1016/j.jmb.2003.08.040.
38. S. C. Walker, D. R. Engelke, Ribonuclease P: The evolution of an ancient RNA enzyme. *Crit. Rev. Biochem. Mol. Biol.* **41**, 77–102 (2006), 10.1080/10409230600602634.
39. C. Guerrier-Takada, S. Altman, Catalytic activity of an RNA molecule prepared by transcription in vitro. *Science* **223**, 285–286 (1984), 10.1126/science.6199841.
40. K. Ruiz-Mirazo, C. Briones, A. de la Escosura, Prebiotic systems chemistry: New perspectives for the origins of life. *Chem. Rev.* **114**, 285–366 (2014), 10.1021/cr2004844.
41. S. A. Harrison *et al.*, Life as a guide to its own origins. *Annu. Rev. Ecol. Syst.* **54**, 327–350 (2023), 10.1146/annurev-ecolsys-110421-101509.

42. W. Martin, M. J. Russell, On the origins of cells: A hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **358**, 59–83 (2003), 10.1098/rstb.2002.1183.
43. N. M. Maraldi, In search of a primitive signaling code. *Biosystems* **183**, 103984 (2019), 10.1016/j.biosystems.2019.103984.
44. D. Yang *et al.*, Enhanced transcription and translation in clay hydrogel and implications for early life evolution. *Sci. Rep.* **3**, 3165 (2013), 10.1038/srep03165.
45. J. Spitzer, Emergence of life on earth: A physicochemical jigsaw puzzle. *J. Mol. Evol.* **84**, 1–7 (2017), 10.1007/s00239-016-9775-3.
46. T. J. Gibson, A. I. Lamond, Metabolic complexity in the RNA world and implications for the origin of protein synthesis. *J. Mol. Evol.* **30**, 7–15 (1990), 10.1007/BF02102448.
47. S. K. C. Korada *et al.*, Crystal structures of *Escherichia coli* exonuclease I in complex with single-stranded DNA provide insights into the mechanism of processive digestion. *Nucleic Acids Res.* **41**, 5887–5897 (2013), 10.1093/nar/gkt278.
48. L. N. Johnson, S. H. Hu, D. Barford, Catalytic mechanism of glycogen phosphorylase. *Faraday Discuss.* **93**, 131–142 (1992), 10.1039/fd9929300131.
49. H. Brandstetter, J.-S. Kim, M. Groll, P. Göttig, R. Huber, Structural basis for the processive protein degradation by tricorn protease. *Biol. Chem.* **383**, 1157–1165 (2002), 10.1515/BC.2002.127.