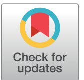
PNAS COMMENTARY



The central dogma in reverse

Natalia Mrnjavac[a] and William F. Martin[a,1]

Among all issues in early evolution, the origin of the genetic code is far and away the most poorly resolved. As such, there is no shortage of theories as to how it arose (1, 2). In PNAS, Shi et al. (3) report laboratory experiments that run counter to conventional views concerning the origin of the code, providing a new perspective on that long-debated issue. Their findings point to new ways to explore how information transfer might have evolved.

The genetic code was deciphered in the 1960s (4), underpinning the central dogma of molecular biology, which had it that information always flows from DNA to RNA to protein via transcription and translation. That dogma was broken in 1970 with the discoveries by Baltimore and Temin of reverse transcriptase, an enzyme that synthesizes DNA from RNA (5), making the route from DNA to RNA reversible (Fig. 1*A*). But in the decades following, the unidirectional flow of information from nucleic acids to protein remained unchallenged by counterexample, until now. Shi et al. (3) engineered the transfer of information from a protein into RNA and then, by using reverse transcriptase, from RNA into DNA, opening up a new chapter for the central dogma.

Their starting point for information transfer was protein. By using a simple peptidase to degrade dodecamer peptides, they released one amino acid at a time from the N terminus. The amino acid was then bound by an aptazyme (6), a small self-cleaving RNA molecule engineered from naturally occurring catalytic RNAs that releases a triplet upon amino acid binding to its aptamer region (3). There is no requirement for the aptazyme to replicate in their experiments, because their RNA molecule has a catalytic, not a replicating activity. The triplet is matched to the aptazyme by virtue of the latter's engineered amino acid binding specificity (Fig. 1*B*). All theories for origins allow RNA molecules to be naturally engineered via selection. The released triplet, or triplets after several rounds, are enzymatically ligated into RNA molecules. Again, all theories for origins allow RNA molecules to assemble via ligation from smaller RNA fragments. Then Shi et al. employed reverse transcriptase—an ancient enzyme activity—and PCR (polymerase chain reaction), the laboratory analog of primordial DNA replication (7), to synthesize double-stranded DNA that contains the aptazyme-derived triplet (codon) and its anticodon on complementary strands. Information from the peptide was thereby transferred to DNA (Fig. 1*B*)—the central dogma in reverse.

From the standpoint of early evolution, three aspects of the new findings need to be emphasized. First, Shi et al. (3) are very explicitly *not* saying that information processing or the genetic code actually arose this way. They are saying that the reaction sequence is *possible*, which their experiments bear out. Indeed, as Shi et al. (3) explain, there was a time when it was debated whether information flow from protein to DNA was possible at all. We now see that it is possible, and it seems unlikely that this will be the last report of its kind. Second, they do not obtain a long DNA sequence that mirrors the amino acid sequence of a random sequence peptide. They get the right nucleotide triplets, yes. But they do not obtain a long specific DNA sequence of triplets that would mirror a complex succession of amino acids in a dodecamer peptide consisting of 12 different amino acids. Their triplet ligation step does not preserve the order in which triplets were generated. Third, they are using modern enzymes in their experiments, whereby a critic might offer that there were no enzymes before the code existed. Again, Shi et al. are saying that the reactions they report are possible, not that they took place in that from on the early Earth, where many reactions could have been promoted by metals or metal ions alone (8) before the origin of enzymes.

The new findings provide reasons to reflect. What *would* we get if we ran the modern process of translation backward? After all, most biochemical reactions are, in principle, reversible and the central dogma is nothing more than a set of chemical reactions. If we were to run the modern process of translation in reverse, starting with a peptide as Shi et al. (3) did, we would start with a protein emerging from the exit tunnel at the ribosome, pulled back by one amino acid (Fig. 1*C*). This would cleave the C-terminal peptide bond, generating the aminoacyl-transfer RNA (aminoacyl-tRNA) of the last peptidyl transferase reaction, pushing the ribosome back one step on the messenger RNA (mRNA), and synthesizing guanosine triphosphate (GTP) from guanosine diphosphate and orthophosphate (P_i) at the steps catalyzed by the elongation factors EfTu and EfG in the process (9). Those GTP-dependent reactions are ancient (10) and, in principle, reversible but would require nonphysiological conditions to run in reverse because of the energy input required to generate GTP. The aminoacyl-tRNA would leave the ribosome and return to its cognate aminoacyl-tRNA synthetase (AARS) to generate the free tRNA plus aminoacyl-adenylate (aminoacyl-AMP), which in turn would release the free amino acid, synthesizing adenosine triphosphate (ATP) from adenosine monophosphate (AMP) and pyrophosphate (Fig. 1*C*). This is obviously not a thermodynamically favorable sequence of events, as it would involve synthesizing the

Author affiliations: [a]Institute for Molecular Evolution, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany

Author contributions: N.M. and W.F.M. wrote the paper.

The authors declare no competing interest.

Copyright © 2026 the Author(s). Published by PNAS. This article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

See companion article, "From peptides to DNA: All required steps can be catalyzed," 10.1073/pnas.2534387123.

[1]To whom correspondence may be addressed. Email: bill@hhu.de.

Published April 6, 2026.

Downloaded from https://www.pnas.org by UNIVERSITAETS DUESSELDORF on April 6, 2026 from IP address 91.8.106.114.

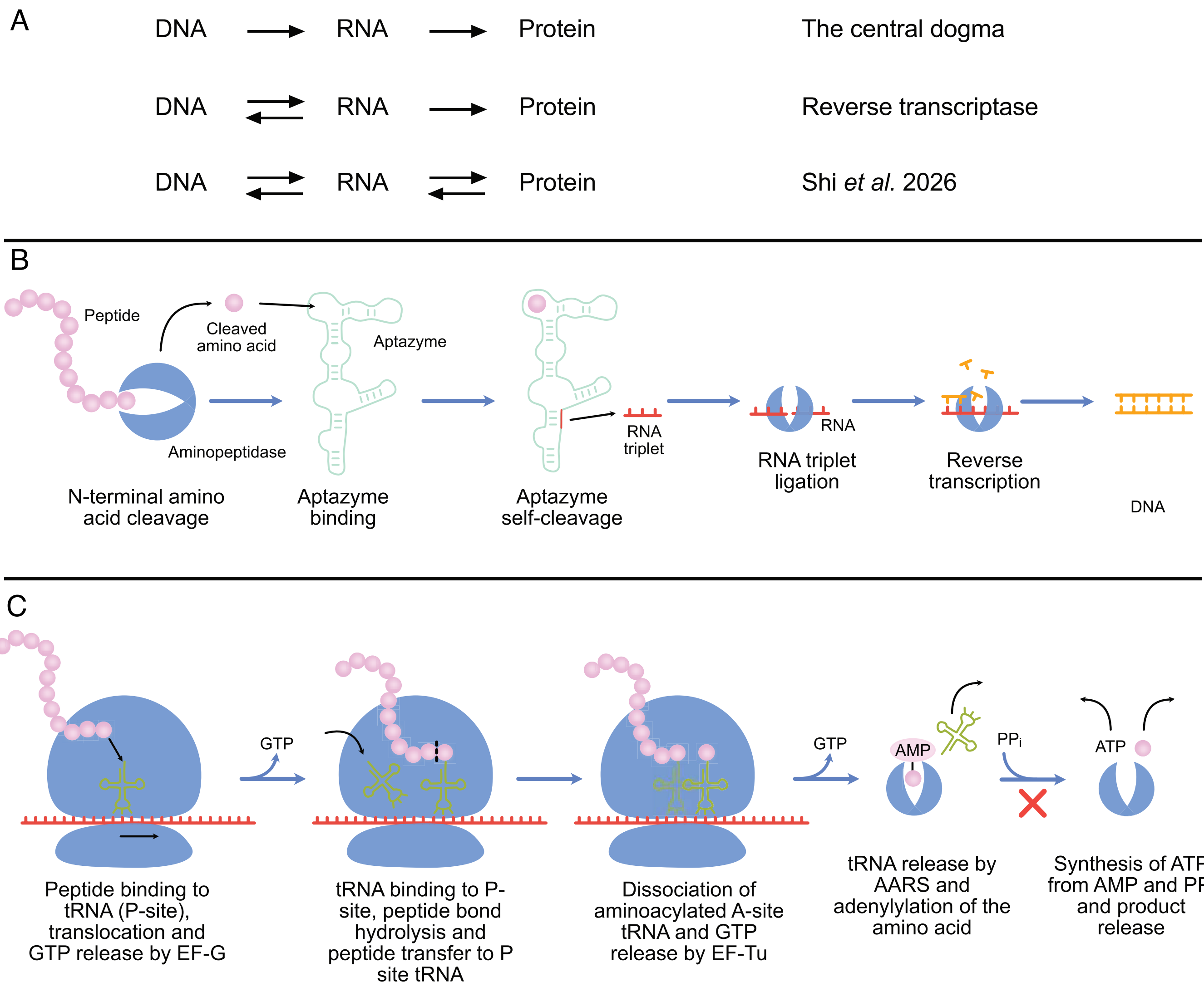


Fig. 1. Back and forth in the central dogma. (*A*) Reverse transcriptase uncovered a path from RNA to DNA, Shi et al. (3) report a path from protein to DNA. (*B*) Schematic depiction of the experiments in Shi et al. (3) (see text). (*C*) A theoretical process of reverse translation does not lead to nucleic acids. In the forward direction, the PP_i forming step of tRNA aminoacylation is irreversible under physiological conditions (see text).

equivalent of four ATP—two GTP for ribosome movement on mRNA (9) plus two ATP for tRNA aminoacylation (11)—all driven by the removal of a single amino acid from a nascent peptide chain on the ribosome. Modern translation in reverse is not going to work. But the idea leads to more questions.

For example, what keeps the life process moving forward, rather than backward, in the first place? Energy is the obvious answer. Roughly 70% of a prokaryotic cell's energy budget is consumed by protein synthesis (12), such that the central dogma, the flow of information from DNA to protein, is also the prime energy consumer during growth. In cells, the energy supply comes from the harnessing of ion gradients via rotor-stator ATP synthases (13), whereby the energy to form those ion gradients stems from reactions of environmentally available chemicals or from light. But even if the environmental energy supply is low, cells still do not immediately unravel. One simple mechanism that keeps the vector of life pointing forward was identified by Kornberg (14). He emphasized that the synthesis of both nucleic acids and proteins involves steps that generate pyrophosphate (PP_i). The incorporation of (ribo)nucleoside triphosphates into nucleic acids and the ligation of amino acids onto tRNA by AARS enzymes both generate PP_i. Because all cells harbor ubiquitous pyrophosphatases, PP_i is rapidly hydrolyzed. This removes a substrate for the backreaction of PP_i-producing enzymatic steps (14), making the enzymatic reactions of both nucleic acid synthesis by RNA and DNA polymerases as well as aminoacylation of tRNA by AARS irreversible under physiological conditions.

Whether in the forward or reverse direction, the new findings (3) document a flow of genetic information. And where is the association between codons and amino acids—the genetic code—in all this? In cells, the covalent linkage of an amino acid to its cognate tRNA is the step that forges the genetic code out of structural information. The reaction is performed by Class I and Class II AARSs, complex and highly conserved enzymes that enable cells to synthesize proteins from information in genes (15–17). In the new study (3), the association of an amino acid to a triplet is performed by a small RNA molecule, an aptazyme engineered to release a triplet upon amino acid binding. This is reminiscent of Yarus's "escaped triplet" theory (1), which postulates steric and

Downloaded from https://www.pnas.org by UNIVERSITAETS DUESSELDORF on April 6, 2026 from IP address 91.8.106.114.

binding interactions between amino acids and triplets in small RNA molecules at the origin of the code. Yet Shi et al. (3) do not generate interactions between the amino acid and the triplet. Instead, they engineered the amino acid binding to one part of the aptazyme which, upon binding, self-cleaves at a separate region, literally generating an "escaped triplet," albeit one engineered in the laboratory.

"In PNAS, Shi et al. (3) report laboratory experiments that run counter to conventional views concerning the origin of the code, providing a new perspective on that long-debated issue."

In today's genetic code, the attachment of the correct amino acid to its cognate tRNA resides in the binding specificity of AARS enzymes for both substrates (15–17). That realization is hardly new (15), but it leads to the conclusion that the code was invented by AARS enzymes, which, paradoxically, require the genetic code for their own synthesis. That irreducible chicken-and-egg relationship casts a long shadow upon attempts to infer the origin of the code. Moreover, AARSs are not only needed for their own synthesis, they are needed for the synthesis of all proteins we know today. That in turn suggests that the AARS enzyme families were (among) the very first proteins ever synthesized by ribosomes, which some have been saying for quite some time (15–17). The aminoacylating function of AARS might have initially been performed by a small self-aminoacylating RNA (18).

The ribosome, tRNAs, and AARS are all large macromolecules that are needed for the information in genes to be expressed. How could something so complex get started? The only tenable answer is as follows: from something simpler, and working models for all three components exist. For the ribosome, an ultraconserved core of ca. 70 nucleotides called the protoribosome (<3% of the 23S rRNA molecule) catalyzes peptidyl transferase reactions in vitro (19, 20). Similarly, tRNAs can be pruned down to functional "minihelices," RNA hairpins only 4 to 9 base pairs in length that comprise solely the stem of the cloverleaf, but are still functional and can be charged by AARS enzymes with surprising specificity (15–17). That property of minihelices highlights an often overlooked aspect of the code: The anticodon is only half the story. The tRNA acceptor stem contains vital information about the identity of the amino acid to be attached to the tRNA, with some information residing exclusively in the stem, and some residing in the anticodon (15–17). The information in the stem might well be the primordial form of the code (15). AARSs can, in turn, be pruned down to about 10% of their amino acid sequence (17) and still recognize—and charge—both tRNAs and minihelices. From such simpler, but functional, versions, more complex versions could have arisen. Simpler still is the system that Shi et al. (3) describe, which highlights molecular, information-transferring interactions from small peptides to small nucleic acids, interactions of a type that might, or might not, have existed in early molecular evolution.

1. M. Yarus, J. G. Caporaso, R. Knight, Origins of the genetic code: The escaped triplet theory. *Annu. Rev. Biochem.* **74**, 179–198 (2005).
2. M. Seki, On the origin of the genetic code. *Genes Genet. Syst.* **98**, 9–24 (2023).
3. L. Shi, M. Senissar, K. Leistner, C. Jers, I. Mijakovic, From peptides to DNA: All required steps can be catalyzed. *Proc. Natl. Acad. Sci. U.S.A.* **123**, e2534387123 (2026).
4. M. W. Nirenberg, J. H. Matthaei, The dependence of cell-free protein synthesis in *E. coli* upon naturally occurring or synthetic polyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1588–1602 (1961).
5. J. M. Coffin, H. Fan, The discovery of reverse transcriptase. *Annu. Rev. Virol.* **9**, 29–51 (2016).
6. J. M. Carothers, J. A. Goler, D. Juminaga, J. D. Keasling, JD, Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716–1719 (2011).
7. C. B. Mast, D. Braun, Thermal trap for DNA replication. *Phys. Rev. Lett.* **104**, 188102 (2010).
8. H. Kaur *et al.*, A prebiotic Krebs cycle analog generates amino acids with H_2 and NH_3 over nickel. *Chem* **10**, 1528–1540 (2024).
9. V. Ramakrishnan, Ribosome structure and the mechanism of translation. *Cell* **108**, 557–572 (2022).
10. N. Mrnjavac, W. F. Martin, GTP before ATP: The energy currency at the origin of genes. *BBA Bioenerg.* **1866**, 149514 (2024).
11. M. Yarus, Phenylalanyl-tRNA synthetase and isoleucyl-$tRNA^{Phe}$: A possible verification mechanism for aminoacyl-tRNA. *Proc. Natl. Acad. Sci. U.S.A.* **69**, 1915–1919 (1972).
12. A. H. Stouthamer, A theoretical study on the amount of ATP required for synthesis of microbial cell material. *Antonie Van Leeuwenhoek* **39**, 545–565 (1973).
13. A. Nakano, J.-I. Kishikawa, K. Mitsuoka, K. Yokoyama, Mechanism of ATP hydrolysis dependent rotation of bacterial ATP synthase. *Nat. Commun.* **14**, 4090 (2023).
14. A. Kornberg, "On the metabolic significance of phosphorolytic and pyrophosphorolytic reactions" in *Horizons in Biochemistry*, H. Kasha, B. Pullman, Eds. (Academic Press Inc., New York, 1962), pp. 251–264.
15. P. Schimmel, R. Giegé, D. Moras, S. Yokoyama, An operational RNA code for amino acids and possible relationship to genetic code. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8763–8768 (1993).
16. W. C. Carter Jr., R. Wolfenden, tRNA acceptor stem and anticodon bases form independent codes related to protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 7489–7494 (2015).
17. G. Q. Tang, H. Hu, J. Douglas, C. W. Carter Jr., Primordial aminoacyl-tRNA synthetases preferred minihelices to full-length tRNA. *Nucleic Acids Res.* **52**, 7096–7111 (2024).
18. M. Illangasekare, G. Sanchez, T. Nickles, M. Yarus, Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).
19. T. Bose *et al.*, Origin of life: Chiral short RNA chains capable of non-enzymatic peptide bond formation. *Isr. J. Chem.* **61**, 863–872 (2021).
20. M. Kawabata *et al.*, Peptide bond formation between aminoacyl-minihelices by a scaffold derived from the peptidyl transferase center. *Life* **12**, 573 (2022).

Downloaded from https://www.pnas.org by UNIVERSITAETS DUESSELDORF on April 6, 2026 from IP address 91.8.106.114.