nature chemistry

Perspective

https://doi.org/10.1038/s41557-024-01698-4

# Intellectual frameworks to understand complex biochemical systems at the origin of life

Received: 14 September 2022

Accepted: 14 November 2024

Published online: 6 January 2025

Check for updates

Burckhard Seelig [1,2] & Irene A. Chen [3]

Understanding the emergence of complex biochemical systems, such as protein translation, is a great challenge. Although synthetic approaches can provide insight into the potential early stages of life, they do not address the equally important question of why the complex systems of life would have evolved. In particular, the intricacies of the mechanisms governing the transfer of information from nucleic acid sequences to proteins make it difficult to imagine how coded protein synthesis could have emerged from a prebiotic soup. Here we discuss the use of intellectual frameworks in studying the emergence of life. We discuss how one such framework, namely the RNA world theory, has spurred research, and provide an overview of its limitations. We suggest that the emergence of coded protein synthesis could be broken into experimentally tractable problems by treating it as a molecular bricolage—a complex system integrating many different parts, each of which originally evolved for uses unrelated to its modern function—to promote a concrete understanding of its origin.

Understanding how life originated from a confluence of molecules is one of the most conceptually challenging problems in science, requiring an integration of efforts from multiple fields, including chemistry, biology and planetary science. Moreover, without a planet-sized experimental laboratory and half a billion years of experimental time, researchers must choose direction carefully. Intellectual frameworks are a critical ingredient for focusing efforts onto concrete questions. For example, one of the most successful frameworks in this field is the RNA world theory, proposed in the 1960s[1–4] (see ref. 5 for a historical review). This theory posits that an early living system used RNA (rather than DNA) to store genetic information and also used RNA (rather than proteins) to catalyse chemical reactions[6–10] (Box 1). The great elegance of this theory was a resolution of a famous chicken-and-egg paradox of the central dogma of molecular biology: if proteins are needed to synthesize DNA, but DNA is needed to encode proteins, how could the system get started in the first place? In the RNA world theory, this paradox is answered by the dual functionality of RNA as both the gene and catalyst, which would require only self-replicating, functional RNAs to create a sustainable system.

In this Perspective, we examine how the RNA world framework spurred research, and where it may reach its limitations when addressing the emergence of complex systems. We introduce Jacob's molecular bricolage (Box 1) as a possible intellectual framework through which to study the early complexification of life[11].

## Intellectual frameworks as a guide

Notably, the RNA world theory created a concrete intellectual framework with which to explore a possible origin of life. The RNA world theory itself was originally sparked by the newly determined structure of transfer RNA (tRNA)—the molecule that physically links specific RNA information (anticodons, which bind to messenger RNA (mRNA)) to particular amino acids. The structure of tRNA exhibited unexpected complexity compared with the double helix of DNA, including multiple stems and loops and stacking of secondary structure elements to

[1]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN, USA. [2]BioTechnology Institute, University of Minnesota, St. Paul, MN, USA. [3]Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, Los Angeles, CA, USA. e-mail: seelig@umn.edu; ireneachen@ucla.edu

give an overall L-shaped structure. The complexity in RNA folding was reminiscent of the protein folding seen in enzymes. Since sequence information determines structure, which determines function, this observation led to speculation that perhaps the complexity of the structure adopted by RNA indicates that RNA, too, could have complex functions such as catalysis. The proof that RNA could indeed catalyse reactions came decades later in the 1980s, in the laboratories of Sidney Altman and Thomas Cech, who shared the Nobel Prize in Chemistry in 1989 for the discovery of catalytic RNA, or ribozymes, in biology. Notably, both laureates devoted substantial time in their Nobel lectures to the topic of the origin of life. In retrospect, the discovery of catalytic RNA initiated sustained excitement about the RNA world theory as a conceptual lens through which experimentalists could probe the origin of life.

The RNA world theory prompted a variety of fascinating questions about RNA, which have stimulated a great amount of groundbreaking work. For example, if RNA were indeed a primordial molecule, how could RNA be synthesized prebiotically[12–16]? This fundamental question occupied the field for some time, as efforts from multiple laboratories failed until 2009, when an ingenious systems chemistry perspective led to the key demonstration of a prebiotically plausible synthesis of two of the RNA nucleotides[17], which has since been joined by additional synthetic routes[18,19] (see ref.16 for a recent review of this active research). Furthermore, even the decades of failed efforts led to innovations, as researchers pondered alternatives to RNA[20], such as xeno nucleic acids, whose unusual properties may enable new applications[21,22]. These are just some examples of research directions that sprung directly from the intellectual framework of the RNA world. Indeed, multiple questions surrounding the RNA world have spurred important research. Major directions include searching for RNA sequences to act as chemical catalysts for essential metabolic reactions in a primitive cell[23,24]; studying mechanisms by which RNA may be copied without enzymes (or ribozymes)[13,25,26]; developing self-replicating RNA systems[7,27]; and determining whether an RNA world would require simple peptides, cofactors or other important non-RNA components[28,29]. Although there may be energetic debates about the definition of an RNA world and how accurate the theory might be, it is undeniable that much progress in this field over the past few decades has been stimulated by the intellectual framework of the RNA world theory[1,2,4].

Now that exploration of the RNA world has matured as a field, the time is ripe to address areas of incompleteness. The RNA world is believed to have given rise to a major evolutionary transition in the creation of coded protein synthesis[30]—the one-way conversion of genetic information into coded protein sequences (Fig. 1). In this process, an mRNA sequence is read as a series of contiguous, non-overlapping, three-letter sequences (codons). One at a time, each codon is bound by a matching tRNA through complementary base pairing between the codon and an anticodon sequence on the tRNA. This matchmaking is facilitated by the ribosome—a large multisubunit protein–RNA complex with three distinct binding sites for consecutive codons and their tRNAs. Separately, each tRNA type is also charged with a specific amino acid by one of a family of aminoacyl-tRNA synthetase enzymes, such that each codon associates with a specific amino acid via the tRNA. Dozens of additional protein factors are also involved in this complex system. The invention of coded protein synthesis was so consequential that approximately half of cellular biomass is now protein. However, the system for decoding and translating mRNA into proteins is notoriously complex, and its origin has therefore been the subject of much speculation. Notably, while it serves as a starting point, the RNA world theory does not directly address this transition, in which the complexity of life would increase dramatically from a self-contained RNA world to a metabolism that could encode protein sequences. The reader is referred to recent reviews for a detailed survey of this area[31,32].

## Chemical mechanisms and the origin of translation

Several intriguing hypotheses have been put forth to explain the origin of different biochemical elements of translation[32]. One early influential idea for the origin of informational coding is the stereochemical hypothesis that direct, non-covalent binding of amino acids to RNA aptamers, acting as early anticodons, could template peptide formation in tandem (Fig. 2a). For example, in support of this idea, the human immunodeficiency virus *trans*-activator of transcription (HIV Tat) aptamer, which contains two arginine-binding sites, has been shown to promote the coupling of an *N*-carboxyanhydride arginine analogue to a peptide primer containing a terminal arginine[33]. It has been suggested that such aptamer-like templates might have evolved into adaptor RNAs, whereas their informational function was taken over by a coding strand (the precursor to mRNAs)[34]. Nevertheless, despite the attractiveness of this idea, efforts to validate the stereochemical hypothesis have met with mixed results[32].

An alternative idea is that the linkage between amino acids and RNAs might have been made covalently early on. A variety of in vitro evolution experiments have discovered RNAs that catalyse ester bond formation between an activated amino acid and an RNA[35–37] (Fig. 2b). Interestingly, one of these ribozymes is extraordinarily small, suggesting that a minimal amount of information is required to encode this activity[38]. Indeed, a systematic search of sequence space yielded self-aminoacylating RNAs with a frequency of approximately one in one billion sequences (or ~10 pg of RNA)[37]. Aminoacyl transfer can also be promoted simply by the steric environment of an RNA stem overhang[39]. Thus, whether by non-covalent or covalent interactions, it seems plausible that amino acids and RNA could form early informational associations in the RNA world.

## The question of why

Considering this case study of aminoacyl-RNA, it can be seen that experimental work has made inroads into the problem of establishing chemical mechanisms to create associations leading to the genetic code. This work focuses on demonstrating how associations would arise, but it generally does not address why they would evolve. In the absence of the full system of coded protein synthesis, why would activities leading to aminoacyl-RNA and other parts of the translation system survive and increase under natural selection? In other words, synthetic experimental approaches provide proximate explanations (for example, how ribozymes catalyse aminoacylation), but the question of ultimate explanations (that is, what evolutionary pressures select for aminoacylation) remains.

In general, one may contrast two scenarios to explain the evolutionary selection of a complex biological system. In one scenario, the elementary or partial stages of the system were selected for its final function (Fig. 3a). A classic example of this is the evolution of the eye, for which a gradual series of increasingly sophisticated stages can be identified, from a patch of light-sensitive cells to a focused lens eye[40–42]. In this case, evolution is believed to have been driven by selection for visual acuity throughout the stages leading to the modern eye. This type of explanation, having a single driving pressure, is intellectually parsimonious and is often an unspoken starting point for specific hypotheses about the origin of coded protein synthesis[43]. For example, the stereochemical hypothesis proposes a relatively simple system whose functional output, like that of the modern system, is coded peptides; therefore, selection for such peptides might have driven the evolution of both the early and modern systems. A disadvantage of this type of explanation is that a simple system must indeed have the activity of interest, but some complex biological systems exhibit a phenotype that would seem to require a number of interacting parts whose spontaneous emergence together appears infeasible. Although this appearance could be due to a failure of human understanding or imagination,

https://doi.org/10.1038/s41557-024-01698-4

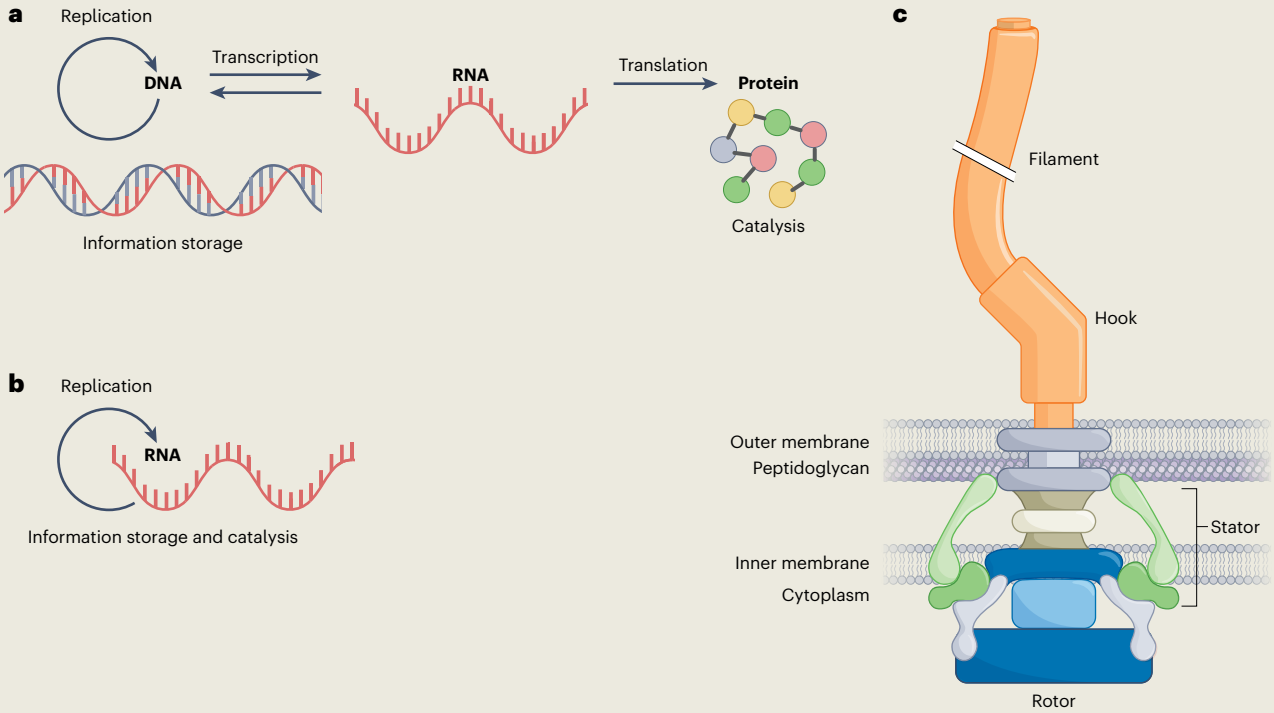
## BOX 1

# Intellectual frameworks for studying the origin and early evolution of life

**The RNA world theory and self-replication.** The central dogma of modern molecular biology states that genetic information flows in a single direction (panel **a** of the inset figure). The DNA (genotype) is transcribed to RNA, which is translated into protein (phenotype). Almost all reactions are catalysed by protein enzymes. DNA is replicated using protein enzymes, and RNA can be reverse transcribed back into DNA in some organisms.

In a hypothesized RNA world, the RNA functions both as the genotype and the phenotype (panel **b** of the inset figure). RNA enzymes (ribozymes) would replicate RNA and also catalyse all other reactions.

**The concept of molecular bricolage for the origin of complex systems.** *Bricolage* is a French word for tinkering, implying a cobbling together of available parts. In art, bricolage refers to the technique of assembling diverse found objects to create a larger piece for a new purpose. In evolution, bricolage refers to a complex system that emerges through the co-option of parts whose initial functions were unrelated to their ultimate function in the system. A classic example of evolutionary bricolage is the bacterial flagellum—a large, multimeric organelle comprising dozens of parts[84] (panel **c** of the inset figure). The main function of the flagellum is motility. However, genetic analyses have indicated that the ancestral functions of its parts were a variety of cellular tasks unrelated to cell movement (for example, an export system developed around an ATPase (light blue); a motor with a rotor evolved from a protein secretion system (dark blue) and a stator evolved from active import proteins (green); proteins evolved to be filamentous and structural (orange); and a chemotactic response evolved from regulatory domains (not shown))[84]. The apparatus is drawn for a Gram-negative organism with outer (top) and inner (bottom) membranes[85,86]. Thus, one may infer that evolution of the flagellum actually involved multiple selection pressures, unrelated to motility, which first led to the evolution of certain proteins. Once these proteins evolved, they formed chance associations with each other that exhibited favourable—or at least neutral—phenotypes. At a critical point, a weak assemblage (that is, a bricolage) formed from these associations, exhibiting a new function related to motility. The selective advantages of motility then became the dominant force governing the evolution of proteins of the flagellum.




these difficulties also prompt consideration of an alternative explanation.

An alternative possibility—termed the molecular bricolage by François Jacob[11,44]—is that the parts of a complex system originally evolved for functions unrelated to their final functions (Fig. 3b and Box 1c). The term bricolage generally describes the process of creating a new work—particularly artwork—by combining miscellaneous existing materials. The molecular bricolage hypothesis for the evolution of complex biological systems therefore implies a scenario with at least two distinct stages. First, parts, such as ribozymes or proteins, evolve to perform initial functions unrelated to the modern function. These parts exist together in the organism and might exist in multiple copies, particularly if the replication process is uncontrolled. Any interactions among different parts at this stage would be incidental, but could be frequent, especially in the context of a crowded cell or protocell[45]. Then, in the second stage, a rudimentary assemblage of the parts would exhibit a new function that is favourable to the fitness of the organism, and that assemblage evolves as a system under selection for the ultimate function (Fig. 3b). For a highly complex system, a variation on the bricolage model is the hierarchical emergence of modules independently

https://doi.org/10.1038/s41557-024-01698-4

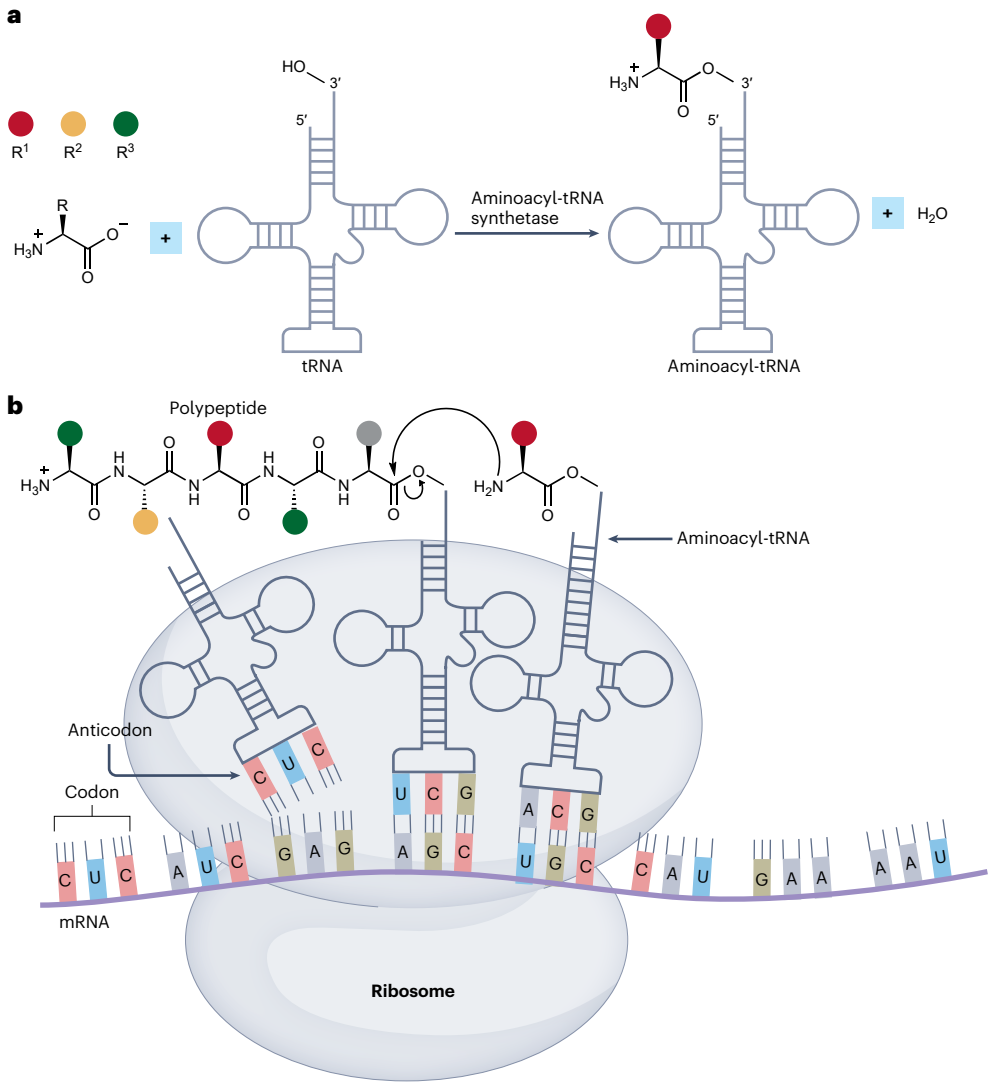



**Fig. 1 | Modern protein translation is conserved in all known life forms and requires complex molecular machinery with many interacting parts.** **a**, The charging of tRNA with its cognate amino acids requires at least 20 aminoacyl-tRNA synthetase enzymes to produce specifically aminoacylated tRNA. **b**, Protein translation of the genetic information encoded in mRNA is carried out by the ribosome. The ribosome is a >2.5 million Dalton assembly of >4,000 nucleotides of ribosomal RNA and 40–80 ribosomal proteins, comprising a large and small subunit. Formation of the peptide bond between the growing polypeptide chain and the incoming activated amino acid (aminoacyl-tRNA) takes place in the peptidyl transferase centre and is catalysed by RNA[82,83]. The sequence of the growing peptide chain is determined by base pairing of the codon in the mRNA with the complementary anticodon of the charged tRNA. For the proper functioning of the translation machinery, numerous additional proteins are required, including initiation factors, elongation factors and release factors (not shown here).

performing unrelated functions at first, before eventually combining to perform the ultimate function (a modular bricolage; Fig. 3c).

## Protein translation as a bricolage

In coded protein synthesis, certain elements have been hypothesized to have evolved for unrelated functions[46]. For example, the coding coenzyme handle hypothesis suggests that amino acids were originally cofactors that improved the activity of ribozymes, leading to the evolution of RNAs that attach to amino acids at one site and bind the ribozyme at another site (a primitive codon–anticodon interaction; Fig. 4a)[32,47]. In this scenario, tRNAs first evolved as cofactor handles rather than informational adaptors. Although ribozymes using amino acid or peptide prosthetic groups are not known, there are examples of non-covalent cofactors, including a modified peptide for a self-triphosphorylating ribozyme[48] and histidine for an RNA-cleaving deoxyribozyme[49], where the amino acid appears to act as a general base. In addition, the original function of the ribosome may be different from its role today. One speculation to explain the emergence of the ribosome is the triplicase hypothesis, which suggests that the ribosome may have been originally selected as an RNA replicase that ligated trinucleotides together, directed by an RNA template (Fig. 4b)[50], with RNA trimers appended to the 3′ end being transferred from one proto-tRNA to the next. Another version of this hypothesis suggests that the primitive ribosome might have catalysed the polymerization of anticodons delivered by tRNA, through successive cleavage and ligation reactions (Fig. 4c)[51]. According to these hypotheses, the triplicase ribozyme would later be co-opted to transfer aminoacyl groups from one tRNA to the next, developing coded protein synthesis. The ribosome is indeed surprisingly promiscuous with regard to substrates[52], consistent with its proposed mechanism as an entropy trap[53,54]. This understanding has led to a different proposal that the ribosome's original function might have involved non-specific (non-coded) peptide polymerization[55,56] performed by

https://doi.org/10.1038/s41557-024-01698-4

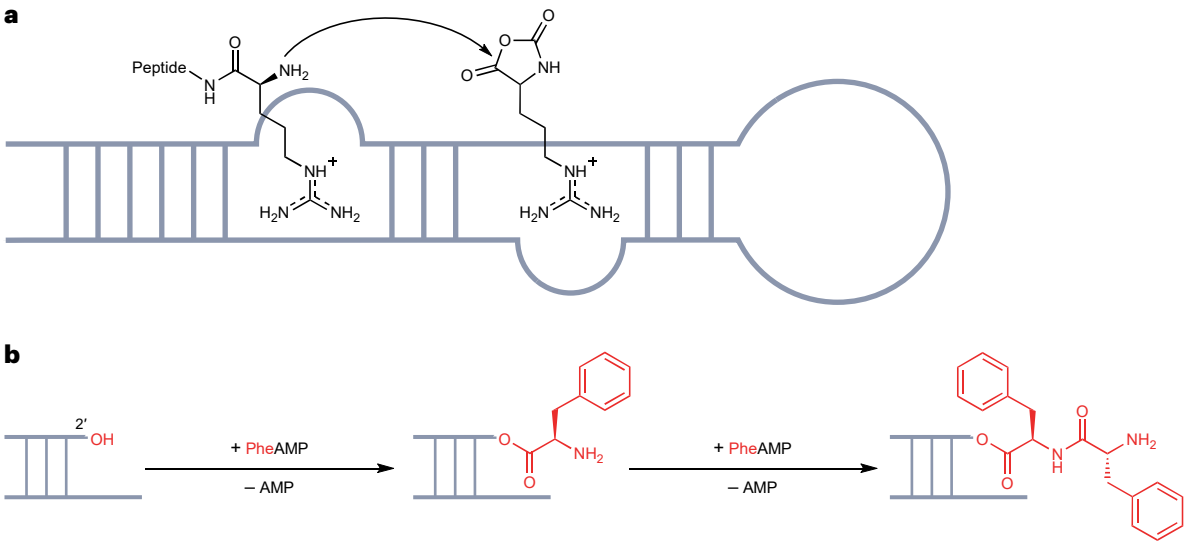



**Fig. 2 | Hypothesized routes for the origin of biochemical elements of translation. a**, Schematic of the stereochemical hypothesis of primitive coded translation through RNA-templated amino acid coupling[33]. An RNA aptamer (grey) that was selected in vitro to bind the arginine-rich human immunodeficiency virus *trans*-activator of transcription (HIV Tat) peptide was found to contain two arginine-binding sites. The two sites non-covalently bind the amino-terminal arginine of a short peptide and an *N*-carboxyanhydride-activated arginine and thereby catalyse the formation of a peptide bond. **b**, Synthesis of aminoacyl-RNA and peptidyl-RNA catalysed by a ribozyme[38]. A five-nucleotide ribozyme (5′-GUGGC; bottom strand; grey) catalyses the aminoacylation of a four-nucleotide RNA (top strand; grey) at its ribose 2′-hydroxyl group. The amino acid phenylalanine (Phe; red) is provided as a carboxyl-activated (L)-5′-phenylalanyl adenylate. The reaction yields both aminoacyl-RNA containing a single Phe residue (middle panel) and peptidyl-RNA containing two Phe residues (right panel). The grey vertical lines symbolize that only the first three nucleotides of the ribozyme and the four-nucleotide RNA form base pairs.

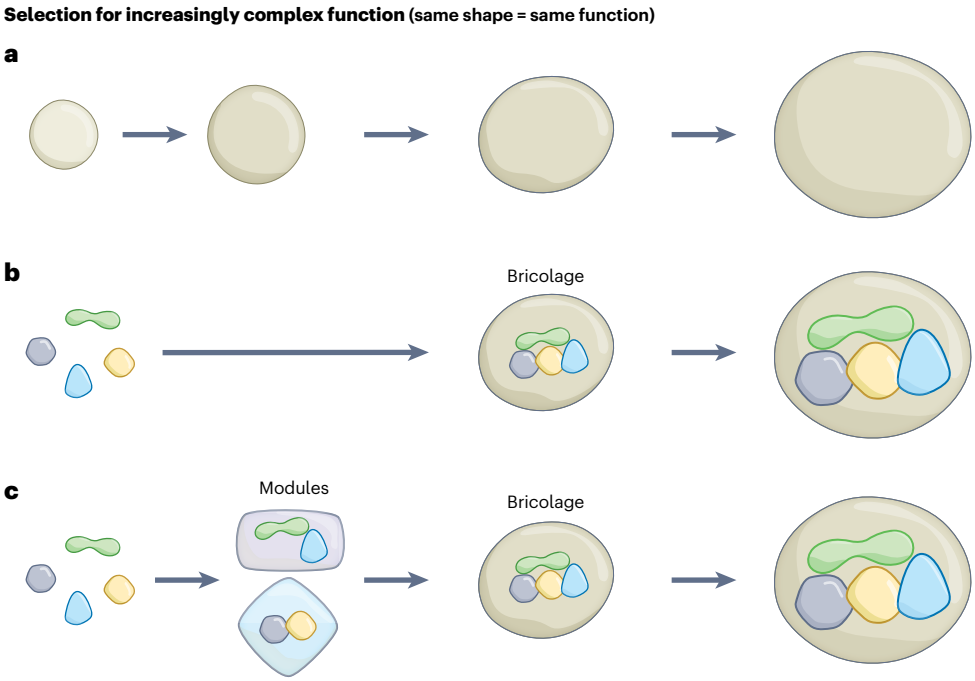



**Fig. 3 | Illustration of single selection pressure versus the molecular bricolage. a**, Direct selection. An increasingly sophisticated system is consistently selected for the same function (symbolized by the same oval shape; for example, evolution of the eye). **b**, Molecular bricolage. Evolution of parts selected for various different functions (symbolized by different shapes), which—when combined in a bricolage—yield a new complex function (oval shape; for example, evolution of the flagellum). **c**, Bricolage with intermediate modules. Initially selected functional parts could combine in limited systems selected for a new function that is still different from the ultimate function (modules). Two or more of these new functions in combination could yield the ultimate function (bricolage).

the large ribosomal subunit, or, alternatively, synthesis of non-peptide polymers such as polyesters[57–59]. Although evidence for these particular hypotheses is limited at present[32,46], it is nevertheless important that their concrete nature is amenable to experimental testing and computational analysis[46]. Indeed, a recent study[60] demonstrated experimentally that aminoacyl-RNA could have had a function in promoting

https://doi.org/10.1038/s41557-024-01698-4

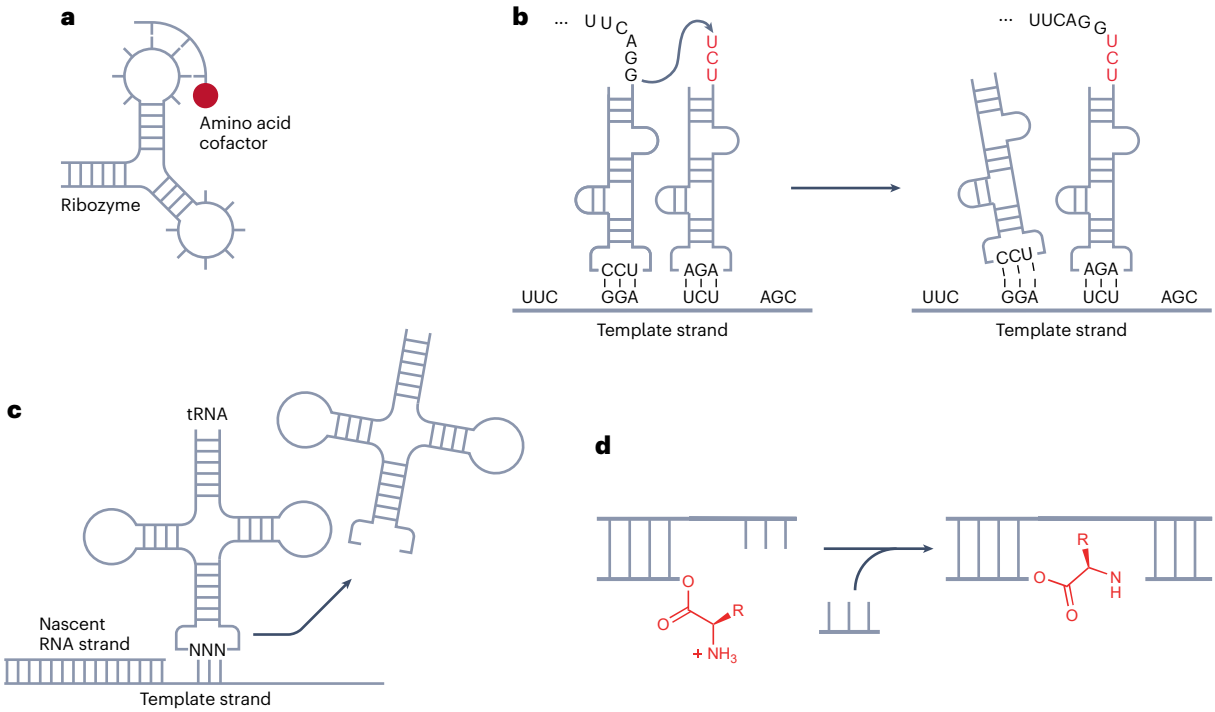



**Fig. 4 | Elements needed for coded protein synthesis that have previously been hypothesized to have evolved for unrelated functions. a**, Amino acids as cofactors of ribozymes, attached via their trinucleotide handle[47]. **b**, RNA replicase ribozyme that synthesizes RNA by ligating trinucleotides (triplicase)[50]. **c**, A variation of the triplicase hypothesis. The tRNA precursor delivers the nucleotide triplet (NNN), which is then added to the growing RNA chain through cleavage and ligation similar to catalysis by the modern spliceosome[51]. **d**, Aminoacyl-RNA promotes non-enzymatic RNA copying by ligation and primer extension[60].

non-enzymatic RNA copying, as aminoacyl-RNAs react quickly to join a downstream RNA oligonucleotide when both are annealed to a complementary template (Fig. 4d). In this example, the emergence of aminoacyl-RNA might therefore be driven by natural selection for faster RNA copying—a molecular function with little relation to translation.

Applying the idea of molecular bricolage naturally deconstructs a complex system into experimentally manageable parts. To understand the first bricolage stage, one may ask what evolutionary advantage is conferred by the parts of coded protein synthesis. For example, compartmentalization of RNAs appears to be essential for assembling primitive life-like systems, maintaining local macromolecular concentrations and preventing emerging RNA functions from being diluted into the bulk solution[7,61,87]. Therefore, peptidylation of RNA might help associate RNA with lipid membranes, as attachment of a lipophilic peptide can anchor the RNA in the membrane[62], such that selection for peptidyl-RNA might be driven by the advantages of compartmentalization. Such phenomena raise new questions as well. What are the minimal requirements for a molecular part to confer this fitness benefit? Are the reactions producing or assembling the parts prebiotically plausible? How do molecules and sequences optimized for the putative original function relate to the versions optimized for their later role in coded protein synthesis? In the second stage, the bricolage assemblage exhibits its final function for the first time. This stage raises the equally fascinating issue of what selective benefit might have been provided by the earliest coded peptides. A bricolage that formed the first coded protein synthesis machine from unrelated functional parts would only persist during natural selection if its product, the first coded peptide (that is, a short protein), provided an immediate selective advantage to the survival of the cell. A particularly strong selective advantage would have resulted if the cell, at that stage of evolution, had already become addicted to the functions provided by those initial coded peptides. Those pre-coding functional peptides might have emerged from random sequences[63,64] of a limited set of amino acids[65,66]. For example, short peptides might have served as important cofactors, but were produced in limiting quantities by non-coded synthesis. The coded synthesis would then be able to provide more of these peptides, on-demand and with a well-defined sequence, allowing a protocell containing such an innovation to edge out its competition. Hypotheses for such transitions that are currently not understood could be developed and probed in experimental systems.

Several steps in the bricolage scenario as an explanation for the origin of coded protein synthesis have already been shown to be feasible in terms of chemical synthesis. The abiotic synthesis of amino acids and the subsequent synthesis of peptide from those amino acids[27,67,68], as well as the formation of peptidyl-RNA[69], have been demonstrated under prebiotically plausible conditions, including conditions compatible with RNA synthesis[56,69]. One could envision many functions that non-coded peptides could have provided to the survival of a cell in an otherwise RNA-dominated world. For example, certain proto-peptides have been shown to increase the stability of RNA and vice versa (Fig. 5a)[70,71]. Furthermore, simple peptides could have served as cofactors for ribozymes (Fig. 5b) or interacted with lipid membranes to modify the stability or permeability of the cell (Fig. 5c). Overall, a molecular bricolage simplifies the problem of why the parts of the complex system would evolve by allowing a variety of initial functions.

## Bricolages elsewhere in the origin of life

The framework of a molecular bricolage could also be applied to the study of other complex systems that emerged during the origin of life. For example, a self-replicating RNA (an RNA replicase) might be a bricolage of multiple modules. A current experimental model for an RNA replicase is an RNA polymerase ribozyme[27,72–74]. In one evolutionary series, some ribozymes evolved within the pool to adopt a new function as a cofactor, enhancing the activity of partnered

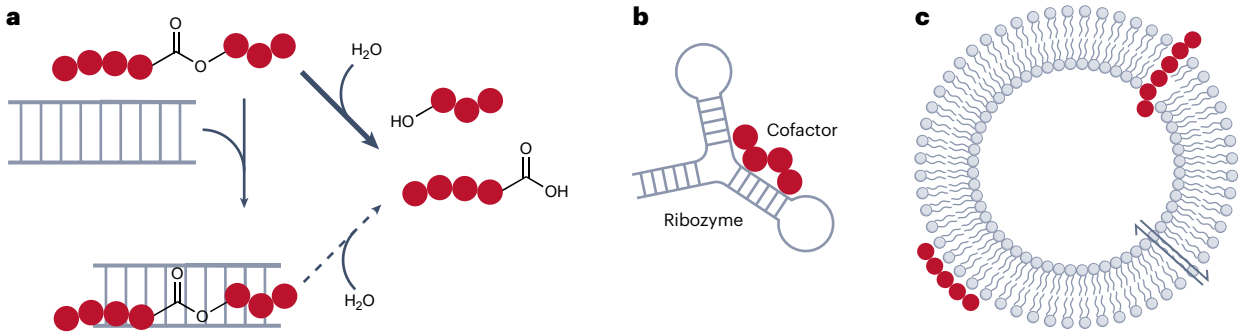



**Fig. 5 | Examples of possible functions of non-coded peptides. a**, Interactions between RNA and cationic proto-peptides (depsipeptides containing a mix of amide and ester bonds) have been shown to mutually stabilize both RNA and peptides[70]. **b**, Short peptides as cofactors for ribozymes. **c**, Peptides interacting with lipid membranes could modify the stability or permeability of the cell.

ribozymes. This evolutionary innovation spontaneously created a heterodimeric system with improved activity[75]. Promiscuous chemical activities in ribozymes, such as weak activities on a variety of substrates, could initiate developments of this type[52,76,77]. Such co-option events might drive the evolution of increasingly large macromolecular complexes. In another example, computational analysis predicted that metabolic phenotypes would have a high capacity for co-option. Using flux balance analysis on randomly generated networks based on *Escherichia coli* metabolism, the ability to utilize one carbon source (for example, glucose) was predicted to be highly correlated with the ability to utilize other carbon sources[78]. Although these correlations were highest for related functions (that is, biochemically similar molecules), this study illustrates that a metabolic network might be grown as a bricolage through the co-option of existing pathways for new substrates.

## Conclusions and outlook

Approaches to studying how life originated are often classified as being either observational or synthetic[79,80]. In the observational approach, the focus is on working backward from what we can observe and analysing the environmental settings, constraints and even sequences of early life based on natural history as a kind of retrosynthetic analysis applied to life. Alternatively, the forward synthetic approach attempts to experimentally simulate prebiotic events in order to discover pathways for creating biomolecules and stimulate emergent properties, rewinding and replaying the Gouldian tape of life from the beginning[81]. Frameworks such as the RNA world or molecular bricolage could be studied through either approach.

The molecular bricolage idea can break down the complex problem of the emergence of translation into more experimentally manageable pieces, from understanding the selection pressures for aminoacyl-RNA, peptidyl-RNA and other parts of the apparatus, to the selection for sub-modules of the system and up to the selection for simple functional peptides. We bring forward the idea of the molecular bricolage, in contrast with direct selection, as a useful hypothesis that need not be correct in order to fulfil the goal of stimulating hypothesis-driven experiments. Much successful effort in prebiotic chemistry has been devoted to figuring out how (for example, how to synthesize nucleic acids, amino acids and peptides prebiotically, or how to synthesize primordial lipids that can form protocells), but this approach has limits when trying to assemble complex systems with interacting parts. The time is ripe for chemists and biochemists to inquire into the evolutionary question of why.

## References

1. Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367–379 (1968).
2. Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **38**, 381–393 (1968).
3. Rich, A. in *Horizons in Biochemistry* (eds Kasha, M. & Pullman, B.) 103–126 (Academic, 1962).
4. Woese, C. R., Dugre, D. H., Dugre, S. A., Kondo, M. & Saxinger, W. C. On the fundamental nature and evolution of the genetic code. *Cold Spring Harb. Symp. Quant. Biol.* **31**, 723–736 (1966).
5. Lazcano, A. The biochemical roots of the RNA world: from zymonucleic acid to ribozymes. *Hist. Philos. Life Sci.* **34**, 407–423 (2012).
6. Higgs, P. G. & Lehman, N. The RNA world: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17 (2015).
7. Joyce, G. F. & Szostak, J. W. Protocells and RNA self-replication. *Cold Spring Harb. Perspect. Biol.* **10**, a034801 (2018).
8. Müller, U. F. Re-creating an RNA world. *Cell. Mol. Life Sci.* **63**, 1278–1293 (2006).
9. Pressman, A., Blanco, C. & Chen, I. A. The RNA world as a model system to study the origin of life. *Curr. Biol.* **25**, R953–R963 (2015).
10. Robertson, M. P. & Joyce, G. F. The origins of the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, a003608 (2012).
11. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
12. Cafferty, B. J. & Hud, N. V. Abiotic synthesis of RNA in water: a common goal of prebiotic chemistry and bottom-up synthetic biology. *Curr. Opin. Chem. Biol.* **22**, 146–157 (2014).
13. Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
14. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
15. Szostak, J. W. The narrow road to the deep past: in search of the chemistry of the origin of life. *Angew. Chem. Int. Ed. Engl.* **56**, 11037–11043 (2017).
16. Yadav, M., Kumar, R. & Krishnamurthy, R. Chemistry of abiotic nucleotide synthesis. *Chem. Rev.* **120**, 4766–4805 (2020).
17. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
18. Kim, H. J. & Kim, J. A prebiotic synthesis of canonical pyrimidine and purine ribonucleotides. *Astrobiology* **19**, 669–674 (2019).
19. Becker, S. et al. Unified prebiotically plausible synthesis of pyrimidine and purine RNA ribonucleotides. *Science* **366**, 76–82 (2019).
20. Eschenmoser, A. Chemical etiology of nucleic acid structure. *Science* **284**, 2118–2124 (1999).
21. Taylor, A. I., Houlihan, G. & Holliger, P. Beyond DNA and RNA: the expanding toolbox of synthetic genetics. *Cold Spring Harb. Perspect. Biol.* **11**, a032490 (2019).

22. Eremeeva, E. & Herdewijn, P. Non canonical genetic material. *Curr. Opin. Biotechnol.* **57**, 25–33 (2019).
23. Lorsch, J. R. & Szostak, J. W. Chance and necessity in the selection of nucleic acid catalysts. *Acc. Chem. Res.* **29**, 103–110 (1996).
24. Wachowius, F., Attwater, J. & Holliger, P. Nucleic acids: function and potential for abiogenesis. *Q. Rev. Biophys.* **50**, e4 (2017).
25. Deck, C., Jauker, M. & Richert, C. Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nat. Chem.* **3**, 603–608 (2011).
26. Prywes, N., Blain, J. C., Del Frate, F. & Szostak, J. W. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife* **5**, e17756 (2016).
27. Martin, L. L., Unrau, P. J. & Müller, U. F. RNA synthesis by in vitro selected ribozymes for recreating an RNA world. *Life (Basel)* **5**, 247–268 (2015).
28. Neveu, M., Kim, H. J. & Benner, S. A. The "strong" RNA world hypothesis: fifty years old. *Astrobiology* **13**, 391–403 (2013).
29. Powner, M. W. & Sutherland, J. D. Prebiotic chemistry: a new *modus operandi*. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2870–2877 (2011).
30. Smith, J. M. & Szathmary, E. *The Major Transitions in Evolution* (Oxford Univ. Press, 1997).
31. Koonin, E. V. & Novozhilov, A. S. Origin and evolution of the universal genetic code. *Annu. Rev. Genet.* **51**, 45–62 (2017).
32. Kun, Á. & Radványi, Á. The evolution of the genetic code: impasses and challenges. *Biosystems* **164**, 217–225 (2018).
33. Harada, K. et al. RNA-directed amino acid coupling as a model reaction for primitive coded translation. *ChemBioChem* **15**, 794–798 (2014).
34. Ma, W. The scenario on the origin of translation in the RNA world: in principle of replication parsimony. *Biol. Direct.* **5**, 65 (2010).
35. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).
36. Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.* **7**, 28–33 (2000).
37. Pressman, A. D. et al. Mapping a systematic ribozyme fitness landscape reveals a frustrated evolutionary network for self-aminoacylating RNA. *J. Am. Chem. Soc.* **141**, 6213–6223 (2019).
38. Turk, R. M., Chumachenko, N. V. & Yarus, M. Multiple translational products from a five-nucleotide ribozyme. *Proc. Natl Acad. Sci. USA* **107**, 4585–4589 (2010).
39. Su, M. et al. Triplet-encoded prebiotic RNA aminoacylation. *J. Am. Chem. Soc.* **145**, 15971–15980 (2023).
40. Darwin, C. *On the Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).
41. Nilsson, D. E. & Pelger, S. A pessimistic estimate of the time required for an eye to evolve. *Proc. Biol. Sci.* **256**, 53–58 (1994).
42. Gregory, T. R. The evolution of complex organs. *Evol. Educ. Outreach* **1**, 358–389 (2008).
43. Lei, L. & Burton, Z. F. Evolution of life on Earth: tRNA, aminoacyl-tRNA synthetases and the genetic code. *Life (Basel)* **10**, 21 (2020).
44. Laubichler, M. D. Tinkering: a conceptual and historical evaluation. *Novartis Found. Symp.* **284**, 20–29 (2007).
45. Peng, H., Lelievre, A., Landenfeld, K., Müller, S. & Chen, I. A. Vesicle encapsulation stabilizes intermolecular association and structure formation of functional RNA and DNA. *Curr. Biol.* **32**, 86–96.e6 (2022).
46. Wolf, Y. I. & Koonin, E. V. 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* **2**, 14 (2007).
47. Szathmáry, E. Coding coenzyme handles: a hypothesis for the origin of the genetic code. *Proc. Natl Acad. Sci. USA* **90**, 9916–9920 (1993).
48. Sweeney, K. J. et al. Peptide conjugates with polyaromatic hydrocarbons can benefit the activity of catalytic RNAs. *Chem. Sci.* **14**, 10318–10328 (2023).
49. Roth, A. & Breaker, R. R. An amino acid as a cofactor for a catalytic polynucleotide. *Proc. Natl Acad. Sci. USA* **95**, 6027–6031 (1998).
50. Campbell, J. H. An RNA replisome as the ancestor of the ribosome. *J. Mol. Evol.* **32**, 3–5 (1991).
51. Poole, A. M., Jeffares, D. C. & Penny, D. The path from the RNA world. *J. Mol. Evol.* **46**, 1–17 (1998).
52. Janzen, E., Blanco, C., Peng, H., Kenchel, J. & Chen, I. A. Promiscuous ribozymes and their proposed role in prebiotic evolution. *Chem. Rev.* **120**, 4879–4897 (2020).
53. Sievers, A., Beringer, M., Rodnina, M. V. & Wolfenden, R. The ribosome as an entropy trap. *Proc. Natl Acad. Sci. USA* **101**, 7897–7901 (2004).
54. Bowman, J. C., Hud, N. V. & Williams, L. D. The ribosome challenge to the RNA world. *J. Mol. Evol.* **80**, 143–161 (2015).
55. Petrov, A. S. et al. History of the ribosome and the origin of translation. *Proc. Natl Acad. Sci. USA* **112**, 15396–15401 (2015).
56. Frenkel-Pinter, M., Samanta, M., Ashkenasy, G. & Leman, L. J. Prebiotic peptides: molecular hubs in the origin of life. *Chem. Rev.* **120**, 4707–4765 (2020).
57. Ohta, A., Murakami, H. & Suga, H. Polymerization of α-hydroxy acids by ribosomes. *ChemBioChem* **9**, 2773–2778 (2008).
58. Ohta, A., Murakami, H., Higashimura, E. & Suga, H. Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14**, 1315–1322 (2007).
59. Stan, A. & Mayer, C. Tethered ribosomes: toward the synthesis of nonproteinogenic polymers in bacteria. *ChemBioChem* **24**, e202200578 (2023).
60. Radakovic, A., Wright, T. H., Lelyveld, V. S. & Szostak, J. W. A potential role for aminoacylation in primordial RNA copying chemistry. *Biochemistry* **60**, 477–488 (2021).
61. Lai, Y. C. & Chen, I. A. Protocells. *Curr. Biol.* **30**, R482–R485 (2020).
62. Isaad, A. L. C. et al. A hydrophobic disordered peptide spontaneously anchors a covalently bound RNA hairpin to giant lipidic vesicles. *Org. Biomol. Chem.* **12**, 6363–6373 (2014).
63. Keefe, A. D. & Szostak, J. W. Functional proteins from a random-sequence library. *Nature* **410**, 715–718 (2001).
64. Tong, C. L., Lee, K. H. & Seelig, B. De novo proteins from random sequences through in vitro evolution. *Curr. Opin. Struct. Biol.* **68**, 129–134 (2021).
65. Newton, M. S., Morrone, D. J., Lee, K. H. & Seelig, B. Genetic code evolution investigated through the synthesis and characterisation of proteins from reduced-alphabet libraries. *ChemBioChem* **20**, 846–856 (2019).
66. Trifonov, E. N. The triplet code from first principles. *J. Biomol. Struct. Dyn.* **22**, 1–11 (2004).
67. Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **41**, 5416–5429 (2012).
68. Foden, C. S. et al. Prebiotic synthesis of cysteine peptides that catalyze peptide ligation in neutral water. *Science* **370**, 865–869 (2020).
69. Jauker, M., Griesser, H. & Richert, C. Spontaneous formation of RNA strands, peptidyl RNA, and cofactors. *Angew. Chem. Int. Ed.* **54**, 14564–14569 (2015).
70. Frenkel-Pinter, M. et al. Mutually stabilizing interactions between proto-peptides and RNA. *Nat. Commun.* **11**, 3137 (2020).
71. Delaye, L. & Lazcano, A. in *Astrobiology* (eds Chela-Flores, J. et al.) 285–288 (Springer, 2000).

72. Cojocaru, R. & Unrau, P. J. Processive RNA polymerization and promoter recognition in an RNA world. *Science* **371**, 1225–1232 (2021).
73. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325 (2001).
74. McGinness, K. E. & Joyce, G. F. In search of an RNA replicase ribozyme. *Chem. Biol.* **10**, 5–14 (2003).
75. Attwater, J., Raguram, A., Morgunov, A. S., Gianni, E. & Holliger, P. Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife* **7**, e35255 (2018).
76. Janzen, E. et al. Emergent properties as by-products of prebiotic evolution of aminoacylation ribozymes. *Nat. Commun.* **13**, 3631 (2022).
77. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425 (1976).
78. Barve, A. & Wagner, A. A latent capacity for evolutionary innovation through exaptation in metabolic systems. *Nature* **500**, 203–206 (2013).
79. Mariscal, C. et al. Hidden concepts in the history and philosophy of origins-of-life studies: a workshop report. *Orig. Life Evol. Biosph.* **49**, 111–145 (2019).
80. Sasselov, D. D., Grotzinger, J. P. & Sutherland, J. D. The origin of life as a planetary phenomenon. *Sci. Adv.* **6**, eaax3419 (2020).
81. Gould, S. J. *Wonderful Life—the Burgess Shale and the Nature of History* (W. W. Norton & Company, 1991).
82. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science* **289**, 905–920 (2000).
83. Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920–930 (2000).
84. Pallen, M. J. & Matzke, N. J. From the origin of species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790 (2006).
85. Morimoto, Y. V. & Minamino, T. Structure and function of the bi-directional bacterial flagellar motor. *Biomolecules* **4**, 217–234 (2014).
86. Mandadapu, K. K., Nirody, J. A., Berry, R. M. & Oster, G. Mechanics of torque generation in the bacterial flagellar motor. *Proc. Natl Acad. Sci. USA* **112**, E4381–E4389 (2015).
87. Saha, R., Choi, J. & Chen, I. A. Protocell effects on RNA folding, function, and evolution. *Acc. Chem. Res.* **57**, 2058–2066 (2024).


## Acknowledgements

We thank J. W. Szostak and J. D. Sutherland for insightful conversations. This work was supported by the National Aeronautics and Space Administration (80NSSC21K0595 to B.S. and I.A.C. and 80NSSC18K1277 to B.S.), Simons Collaboration on the Origins of Life (290356FY18 to I.A.C. and 340762 to B.S.), Human Frontier Science Program (RGP0041 to B.S.), National Institutes of Health (GM 5R01GM108703 to B.S.) and National Science Foundation (EF-1935372 and EF-1935087 to I.A.C.).

## Author contributions

Both authors collaborated equally to develop the concept and research and write the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** should be addressed to Burckhard Seelig or Irene A. Chen.

**Peer review information** *Nature Chemistry* thanks Antonio Lazcano and the other, anonymous, reviewer for their contribution to the peer review of this work.