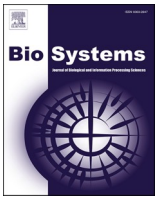
BioSystems 266 (2026) 105849

Contents lists available at ScienceDirect

BioSystems

journal homepage: www.elsevier.com/locate/biosystems



Protocode mats and living membranes: A new hypothesis for the origin of the standard genetic code

Alexander Nesterov-Mueller *, Dimitry Schmidt

Institute of Microstructure Technology, Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, 76344, Germany

ARTICLE INFO

Keywords:
Combinatorial fusion cascade
Origin of the genetic code
Codon assignments
Protocode
Symbiogenesis

ABSTRACT

The standard genetic code (SGC) maps 64 codons to 20 amino acids and three stop codons. The Combinatorial Fusion Cascade (CFC), introduced in 2020, identified three combinatorial rules that reproduce all 64 codon–amino acid assignments through a cascade of eight initial complementary pairs, four protocodes, and the SGC. Here we advance the CFC programme on several fronts not made explicit in the original publications. First, we draw a sharp distinction between the fusion rules — objective structural properties of the standard code table, readable independently of any historical assumption — and the CFC as a hypothesis about how these rules arose; the rules are facts, the cascade is one mechanism that could have generated them. Second, we formalize the rules as deterministic operators on the bases that regenerate all 64 assignments, and we introduce an entropy rank of codons under which the three cascade stages emerge as nested levels ($4 \subset 16 \subset 64$), making the increase of codon entropy along the cascade explicit. Third, we give an explicit set of falsifiable predictions, separating those testable now from those awaiting suitable methods: Mendeleev-like positions for amino acids that entered or left the code (canavanine as X2; the recently reported archaeal code in which UAG encodes pyrrolysine matching the predicted occupancy of X1); the presence of aromatic amino acids (Phe, Tyr) from the inception of the code; and a temporal ordering in which sulfur chemistry was integrated only after the initial complementary pairs were established (cysteine at the protocode stage, methionine after fusion). Fourth, and most centrally, we propose that the code arose not in lipid vesicles but in protocode mats — layered RNA–peptide communities on catalytic mineral surfaces, in which amphiphilic RNA–peptide complexes serve at once as code carriers and as a surface-associated boundary rather than a closed lipid container. This setting does not require a lipid membrane while the code itself was forming and yields a falsifiable geochemical roadmap, the temporal order of the cascade constraining the sequence of environments in which it unfolded.

1. Introduction: Koonin's admission

The standard genetic code (SGC), which maps 64 nucleotide triplets to 20 canonical amino acids and three stop codons, is arguably the single central informational invariant of all life on Earth. In a landmark review marking the fiftieth anniversary of Crick's frozen accident hypothesis, Koonin evaluated the three major theories of code origin and concluded that these approaches "have elucidated notable features of the standard code, such as high robustness to errors, but failed to develop a compelling explanation for codon assignments" (Koonin, 2017). Koonin and Novozhilov deepened this assessment, calling the problem "arguably the single central informational invariant of all life forms" and admitting: "The trouble is that the problem appears extremely and unusually hard" (Koonin and Novozhilov, 2017). They noted that phylogenetic analysis of aminoacyl-tRNA synthetases (aaRS) shows that at a stage when translation had already attained high fidelity, the correspondence between amino acids and codons was determined by RNA molecules, not by proteins. Importantly, they observed that attempts to decipher the primordial stereochemical code by comparative analysis of modern translation system components "are likely to be futile."

Even earlier, Koonin and Novozhilov had characterized the situation with rare candor: "Despite extensive and, in many cases, elaborate attempts to model code optimization, ingenious theorizing along the lines of the coevolution theory, and considerable experimentation, very little definitive progress has been made" (Koonin and Novozhilov, 2009). They identified a "dreary vicious circle": functional proteins cannot exist without a translation system, but the translation system itself requires functional proteins. Wolf and Koonin called the triplet code "the utmost

* Corresponding author.
E-mail address: alexander.nesterov-mueller@kit.edu (A. Nesterov-Mueller).

https://doi.org/10.1016/j.biosystems.2026.105849
Received 27 April 2026; Received in revised form 3 June 2026; Accepted 7 June 2026
Available online 8 June 2026

innovation of biological evolution" and emphasized that the coding principle "is not immediately dictated by any known physics or chemistry" (Wolf and Koonin, 2007). Szathmáry and Santos noted the irony that the RNA world hypothesis, within which translation becomes nearly intractable: "The path out of the RNA world seems to be very hard" (Szathmáry and Santos, 2018). Even a minimal translation system requires approximately 100 genes (Gil et al., 2004).

This article takes Koonin's assessment as its starting point. We analyze the conceptual framework that sustains the three major hypotheses and show that two principal assumptions have left the field with unresolved problems. We then present the Combinatorial Fusion Cascade (CFC), whose core consists of three combinatorial fusion rules that are objective structural properties of the code itself, and propose that the code arose in living membranes — layered RNA–peptide communities on catalytic mineral surfaces.

The CFC was introduced previously (Nesterov-Mueller and Popov, 2021; Nesterov-Mueller et al., 2021); the present paper sharpens and extends it on several fronts: the distinction between the rules as representation and the CFC as hypothesis; a deterministic formalization of the three rules as base-level operators that regenerate all 64 codon assignments; a quantitative formalization of the cascade through the entropy rank of codons, with the three stages emerging as nested rank levels ($4 \subset 16 \subset 64$); an explicit list of falsifiable predictions, separating those testable now from those awaiting methods; Mendeleev-type predictions for the open positions X1–X3; the prediction of aromatic amino acids from inception; the temporal ordering of sulfur chemistry; and the protocode-mat setting on catalytic mineral surfaces.

2. The dominant paradigm: the genetic code as a later evolutionary advantage

The prevailing view treats the genetic code as a product of molecular evolution – something that emerged gradually as a selective advantage for self-replicating RNA molecules in an RNA world. Within this paradigm, the code appeared after some form of primitive molecular life already existed. Replicating RNA molecules, through competition and natural selection, eventually developed a coding system that allowed them to produce useful peptides and then proteins, which improved their fitness.

This assumption dominates textbooks, grant applications, and the framing of nearly all experimental work on the code's origin. It contains two distinct sub-schools, which despite their different methodologies share the same foundational postulate.

2.1. The molecular evolution sub-school

The first sub-school seeks to model how populations of replicating molecules could evolve a code through selection pressure. Its roots lie in Eigen's theory of the hypercycle (Eigen, 1971). Higgs showed that a four-column theory with a primordial code of Gly, Ala, Asp, Glu, and Val fits well with coevolution predictions (Higgs, 2009). Massey (2016) proposed error minimization through gene duplication. Hartman and Smith (2014) proposed the gradual GC → GCA → GCAU expansion, supported by Kubyshkin and Budisa (2019). The coevolution theory, proposed by Wong in 1975 (Wong, 1975), is based on the idea that the code originally consisted of a few precursor amino acids subsequently replaced by their biosynthetic products. A controversial issue remains the principle of codon transfer from precursors to products (Di Giulio, 2008).

As noted by Nesterov-Mueller and Popov, "this view is only a hypothesis and cannot be used to negate other approaches" (Nesterov-Mueller and Popov, 2021). Meanwhile, Woese et al. concluded that aaRS evolution "certainly influenced the formation of the modern translation mechanisms, but did not shape the codon assignments" (Woese et al., 2000). Haig and Hurst believed that "the code could acquire its major features before the evolution of proteins" (Haig and Hurst, 1991). These ideas have not received much attention against the background of rapid genomic developments since 2000. It was first more important to understand how the genomes and proteomes of the most archaic microorganisms relate to the genetic code (Weiss et al., 2018).

2.2. The prebiotic chemistry sub-school

The second sub-school approaches from prebiotic chemistry. Miller's (1953) experiment demonstrated amino acid synthesis under simulated early Earth conditions, though the gas composition has been criticized (Bada, 2013). Despite the lack of proven connections between amino acid appearance on Earth and code entry, Miller's results are widely used as evidence for a gradually evolving code (Brooks et al., 2002; Higgs and Pudritz, 2009). Significant experimental progress includes Ferris's montmorillonite-catalyzed RNA oligomers (Ferris, 2006), Müller et al.'s peptide synthesis on complementary RNA fragments (Müller et al., 2022), and Szostak's work on non-enzymatic RNA copying (Radakovic et al., 2024). These are important contributions to chemistry, but as we argue in Chapter 3, they face fundamental contradictions when applied to the code origin problem.

2.3. The shared assumption and its consequences

Both sub-schools share the assumption that the code came after molecular life existed. This forces explanatory work into a framework where the code must be derived from pre-existing molecular processes, makes the code's origin contingent on solving prior unsolved problems, and excludes the possibility that the code's structure contains its own explanation.

The outstanding work of Vetsigian, Woese, and Goldenfeld gave a different perspective: "horizontal transfer of genes and perhaps other complex elements among the evolving entities, a dynamic far more rampant and pervasive than our current perception of horizontal gene transfer, is required to bring the evolving translation apparatus, its code, and by implication the cell itself to their current condition" (Vetsigian et al., 2006). This insight – that universality points to fusion among competing entities – is the direct intellectual precursor of the Combinatorial Fusion Cascade.

2.4. Recent developments within the old framework

In 2021, the CFC provided for the first time a complete mathematical description of all 64 codon assignments through three combinatorial rules (Nesterov-Mueller and Popov, 2021; Nesterov-Mueller et al., 2021). This raised a new standard: any subsequent theory must account for the CFC pattern as a structural property of the code. Two programs that appeared since 2021 illustrate how the field continues to operate within the old framework: Di Giulio's coevolution theory and Fontecilla-Camps's stereochemical model.

Di Giulio and the coevolution theory. Between 2023 and 2025, Di Giulio published nine papers in BioSystems extending the coevolution theory. His 2024 review argues that the code structure is an imprint of biosynthetic relationships between amino acids through pretran synthesis (Di Giulio, 2024a). Crucially, Di Giulio acknowledges that the coevolution theory cannot predict individual codon-amino acid assignments. It predicts only organizational patterns – clustering of biosynthetically related amino acids, precursor-product pairs differing by a single base – but cannot explain why UUC encodes phenylalanine, why arginine has six codons, or why there are exactly three stop codons. After fifty years, the coevolution theory remains a theory of correlations, not of assignments.

Di Giulio's 2024 paper on code timing (Di Giulio, 2024b) places the code's origin at the progenote-to-cell transition, claiming it has "nothing to do with the origin of life." This late timing deliberately disconnects the code from prebiotic chemistry entirely – a consistent application of

the "code as later advantage" assumption.

In 2025, Di Giulio reinterprets the discovery of a 62-sense-codon variant in methanogenic archaea (Di Giulio, 2025), where all UAG codons encode pyrrolysine (Kivenson et al., 2025), arguing this variant is ancestral. From the CFC perspective, this is entirely expected: the CFC predicts that X1 – the amino acid whose removal generated stop codons UAA and UAG – originally occupied these positions. The CFC explains the UAG/pyrrolysine connection as a structural property of the cascade; Di Giulio cannot explain why pyrrolysine occupies specifically UAG and not any other codon.

Fontecilla-Camps and the 8 A/U primordial code. Fontecilla-Camps proposed that the initial code consisted of 8 triplet codons composed entirely of A and U nucleotides, encoding 6 amino acids plus one stop signal (Fontecilla-Camps, 2023). His assignments are based on SELEX-measured stereochemical affinities. This makes specific predictions for 8 out of 64 codons but provides no rules for how the remaining 56 codons were assigned. The model reduces the problem from 64 unknowns to 56; it does not solve it. Moreover, as Koonin and Novozhilov noted (Koonin and Novozhilov, 2017), the strongest SELEX affinities are found for complex amino acids not available prebiotically, while Fontecilla-Camps's initial assignments include Phe and Tyr – among the least readily produced in prebiotic synthesis.

Both programs share three critical limitations. First, neither can predict the complete set of codon-amino acid assignments: Di Giulio predicts clustering patterns; Fontecilla-Camps predicts at most 8 out of 64. The CFC generates all 64 from three rules. Second, both rely on the code-as-later-advantage assumption. Third, neither addresses the CFC pattern – the invariance of the second codon position, the systematic wobble-position mutations in dominant entities, the first-position mutations in recessive entities – which any theory must ultimately explain.

Koonin and Novozhilov themselves, while providing the most penetrating critique of the three classical hypotheses, also proposed their own framework: partial optimization of a random code through codon swapping under selective pressure for error minimization (Koonin and Novozhilov, 2017; Novozhilov et al., 2007). Their model begins from a random initial assignment and improves it by swapping amino acids between codon blocks. This presupposes exactly the molecular chaos starting condition whose impossibility is analyzed in Section 3.6. Their critique of stereochemistry, coevolution, and error minimization is therefore correct – but their own proposal remains within the same paradigm and inherits its central limitation: it cannot explain why specific codons are assigned to specific amino acids.

A distinct conceptual boundary separates the genetic code expansion program from the question of the code's origin. Budisa and colleagues have demonstrated that codon assignments in modern organisms can be engineered – sense codons can be reassigned to non-canonical amino acids through orthogonal translation systems (Hoesl and Budisa, 2012; Luu et al., 2025). These are impressive achievements in synthetic biology, but they address a different question entirely: how to modify existing codon assignments in living cells, not how the original assignments arose. The genetic code expansion program presupposes the very system whose origin is in question. Modifying the code with engineered aaRS and recoded genomes tells us what the modern translation machinery can tolerate, not what produced it.

2.5. Summary: the paradigm and its consequences

The dominant paradigm – shared by molecular evolution models, prebiotic chemistry approaches, and recent theoretical developments alike – treats the genetic code as a late evolutionary product. This assumption forces the search for the code's origin into a framework where prior unsolved problems (origin of replication, origin of catalysis, origin of compartmentalization, origin of metabolism) must be solved first. It excludes by construction the possibility that the code's structure may be the primary datum from which the origin should be reconstructed. And it has proven unable, after more than fifty years of intensive work, to explain why specific amino acids are assigned to specific codons.

It is precisely this alternative – that the code's structure contains its own explanation – that we develop in Chapter 4. But first, we must examine in detail why the conventional approaches have left a specific question unresolved that appears not to be merely technical – one that has so far resisted more data, better experiments, or cleverer models within the existing framework.

3. Unresolved problems of the "code as advantage" framework

Each of the central research directions within the current framework has produced important scientific results. Ribozyme research has revealed the catalytic potential of RNA. Protocell research has advanced our understanding of compartmentalization. Stereochemistry has mapped correlations between codons and amino acid properties. Aminoacyl-tRNA synthetase studies have illuminated the modern translation machinery. However, each of these programs answers a question that is distinct from the question of the genetic code: why are specific amino acids assigned to specific codons? When evaluated against this specific question, four recurring difficulties emerge, examined in turn below. To date, none of these approaches — individually or, on the evidence available, in combination — has reached the actual problem — the assignment of specific amino acids to specific codons.

3.1. Limits of the ribozyme approach

The RNA world hypothesis requires ribozymes capable of self-replication (Cech, 2000; Guerrier-Takada et al., 1983). Functional ribozymes require ~100-200 nucleotides. The most successful artificial RNA replicases (~200 nt) still cannot replicate sequences of their own length (Attwater et al., 2013; Horning and Joyce, 2016). The longest non-enzymatic oligomers on mineral surfaces reach ~40-50 nt (Ferris, 2006) – far below the minimum. RNA undergoes rapid hydrolysis under conditions required for catalysis: Mg^{2+} at 5-50 mM accelerates backbone cleavage (Li and Breaker, 1999). Catalytic rates of ribozymes (10^5-10^7-fold enhancement) are orders of magnitude below protein enzymes (10^{10}-10^{17}-fold) (Doudna and Cech, 2002). The longer the ribozyme, the less probable its formation and the faster its degradation – a double bind with no escape within the RNA world framework.

3.2. Limits of the lipid-membrane approach

Protocell research has made significant progress in understanding self-assembly and compartmentalization of simple vesicles. However, the question it answers – can lipid membranes form and encapsulate molecules? – is not the question of the genetic code. Even a perfect protocell says nothing about why UUC encodes phenylalanine.

Moreover, the protocell program faces its own internal contradictions. Metal ion incompatibility: Mg^{2+}/Mn^{2+}/Fe^{2+} essential for RNA chemistry destroy fatty acid membranes (Chen and Szostak, 2004; Monnard and Deamer, 2002). Citrate chelation reduces effective Mg^{2+} available for RNA chemistry (Adamala and Szostak, 2013). Membranes are impermeable to nucleotide monomers required for replication inside vesicle (Mansy et al., 2008). In addition, phospholipids are products of later biochemistry; archaeal and bacterial lipids differ fundamentally, suggesting membrane chemistry postdates LUCA (Łapińska et al., 2023). All of these difficulties are specific to a closed lipid membrane. We develop an alternative in chapter 4: a catalytic surface acting not as a container but as the low-entropy starting point for the code-generating cascade.

3.3. Limits of physicochemical derivation

Stereochemical and physicochemical studies have revealed real and

important correlations between codon properties and amino acid characteristics. The correlation between the second codon position and amino acid hydrophobicity (Copley et al., 2005) and the evidence for error minimization (Haig and Hurst, 1991; Freeland and Hurst, 1998) are genuine contributions to our understanding of the code's structure.

Wolf and Koonin stated that the triplet code "is not immediately dictated by any known physics or chemistry" (Wolf and Koonin, 2007). Studying nucleotide and amino acid properties does not reveal the code. Modern translation does not involve direct codon–amino-acid recognition (Koonin and Novozhilov, 2017). The SELEX paradox deepens the problem: the strongest RNA–amino-acid affinities are found for amino acids that were scarce or late on the early Earth (Arg, His, Phe, Trp), while prebiotically abundant amino acids show the weakest signals (Yarus et al., 2009). We stress that stereochemistry is the one programme that explicitly sets out to connect specific codons with specific amino acids, and the expectation that it should do so is entirely legitimate. Our claim is the narrower, empirical one: on the field's own data, it has not yet delivered those assignments, because the SELEX paradox places the strongest measured affinities on amino acids that were scarce on the early Earth. We therefore do not dismiss stereochemistry; we identify the specific obstacle that still separates its correlations from the assignment problem.

3.4. Aminoacyl-tRNA synthetases as products of the code

The discovery of two structurally unrelated aaRS classes (Eriani et al., 1990) and an ancient operational RNA code in the tRNA acceptor stem (Carter and Wills, 2018) illuminate how translation works today. We do not reject using the modern code. We reject only deriving its origin from today's translation machinery. The CFC instead explains the code from its own structure.

Woese et al. concluded that aaRS evolution "did not shape the codon assignments" (Woese et al., 2000). Lei and Burton stated: "There is no known mechanism to generate aaRS proteins until the code has evolved" (Lei and Burton, 2020). If the code preceded proteins (Haig and Hurst, 1991), then aaRS-based approaches are circular: using products to explain the origin of the system that produced them. Consistently, the CFC's dominant/recessive partition does not coincide with the aaRS class I/II division (Nesterov-Mueller et al., 2021); the two classifications are independent.

3.5. How the difficulties interlock

The four experimental difficulties are mutually reinforcing. The ribozyme impasse weighs against self-replicating RNA as the starting point; the membrane impasse against lipid compartments; the physicochemical derivation impasse against a direct physicochemical explanation; and the aaRS paradox against protein evolution as the mechanism. To our knowledge, no existing approach has yet yielded the specific codon–amino-acid assignments, singly or in combination. We do not claim that no combination of existing views could ever succeed — that would require a formal analysis we do not provide; we observe only that none has so far, which motivates exploring a different mechanism.

3.6. The problem of initial conditions

Molecular evolution theories assume a starting point: a random mixture of monomers from which the code gradually emerged. The space of possible coding tables is $\sim 10^{83}$ (Novozhilov et al., 2007). The probability of finding the SGC through random search is physically impossible. Eigen's threshold paradox (Eigen, 1971; Eigen and Schuster, 1979) deepens the problem: minimal complexity for natural selection requires prior natural selection. Wolf and Koonin noted: "In order to attain the minimal complexity required for a biological system to get on the Darwin-Eigen spiral, a system of a far greater complexity appears to be required" (Wolf and Koonin, 2007).

The code and translator must co-emerge: a code without a ribosome is a table with no reader. Koonin and Novozhilov (2009) called this the "dreary vicious circle". Special environments – hydrothermal vents (Martin and Russell, 2003), tectonic faults (Schreiber et al., 2012) – describe plausible settings for prebiotic chemistry but cannot assign UUC to phenylalanine. Crucially, Kudella et al. showed that templated ligation significantly reduces sequence space (Kudella et al., 2021), suggesting the prebiotic world may not have been chaotic – undermining the fundamental premise of molecular evolution from random initial conditions.

3.7. The prebiotic chemistry disconnect

Prebiotic chemistry has demonstrated that ~10 amino acids form abiotically (Miller, 1953; Pizzarello and Shock, 2010). The inference that prebiotically available amino acids entered the code first contains a categorical gap: demonstrating that a molecule could have existed says nothing about why it is assigned to a specific codon. The SELEX paradox (Yarus et al., 2009) shows strongest affinities for complex amino acids not available prebiotically. Prebiotic nucleotide synthesis pathways do not resemble biological pathways (Stairs et al., 2018). The Sutherland group's cyanosulfidic chemistry (Patel et al., 2015) achieves remarkable selectivity, but it requires hydrogen sulfide as a reductant from the outset — a condition that may correspond to a later geochemical stage than the one in which the initial complementary pairs formed. The CFC independently predicts that sulfur-containing amino acids entered the code only after the earliest stages (§4.4), suggesting that the chemical environment of the initial pairs was pre-sulfidic.

Muchowska et al. (2019) showed iron-promoted synthesis of metabolic precursors; Preiner et al. studied ancient metabolic pathways (Preiner and Asche, 2019). These address which molecules were available – not why specific molecules are assigned to specific codons. Wehbi et al. challenged the consensus on amino acid recruitment order, finding that small size predicts ancient enrichment better than prebiotic abundance (Wehbi et al., 2024). The correlation between Miller-experiment yields and code entry order appears weaker than assumed.

4. Molecular life originates around codon assignments: from rules to protocode mats

4.1. Combinatorial fusion cascade and the protocode–SGC transition

The CFC (Nesterov-Mueller and Popov, 2021; Nesterov-Mueller et al., 2021; Nesterov-Mueller, 2025) proceeds in three stages (Fig. 1): eight initial complementary pairs with monobasic codons (AAA/UUU or GGG/CCC) → four protocodes (16 codons each) → the SGC (64 codons), each stage multiplying the number of codons by four (4 → 16 → 64). The sixteen amino acids carried by the initial pairs match exactly the number proposed by Copley, Smith and Morowitz for a doublet-codon code (Copley et al., 2005).

The initial pairs and protocodes are split into dominant and recessive entities. The terms are borrowed from classical genetics and are purely operational, with no Mendelian meaning: a dominant entity retains its original codon/amino-acid assignments after fusion, whereas a recessive entity acquires new triplets.

The cascade is built entirely at the level of the bases, without invoking specific amino-acid properties; it is a striking feature that, once amino acids are placed on this base-level skeleton, the result coincides with the code table. The construction rests on three principles: competition of triplets; an increase of triplet entropy that preserves complementarity at each fusion event; and the wobble principle — the degeneracy of the code at the third position under substitutions within the purines (A, G) and within the pyrimidines (C, U).

Fig. 2 shows the fusion cascade without amino acids, so that the principles of competition and of entropy increase are visible directly. At

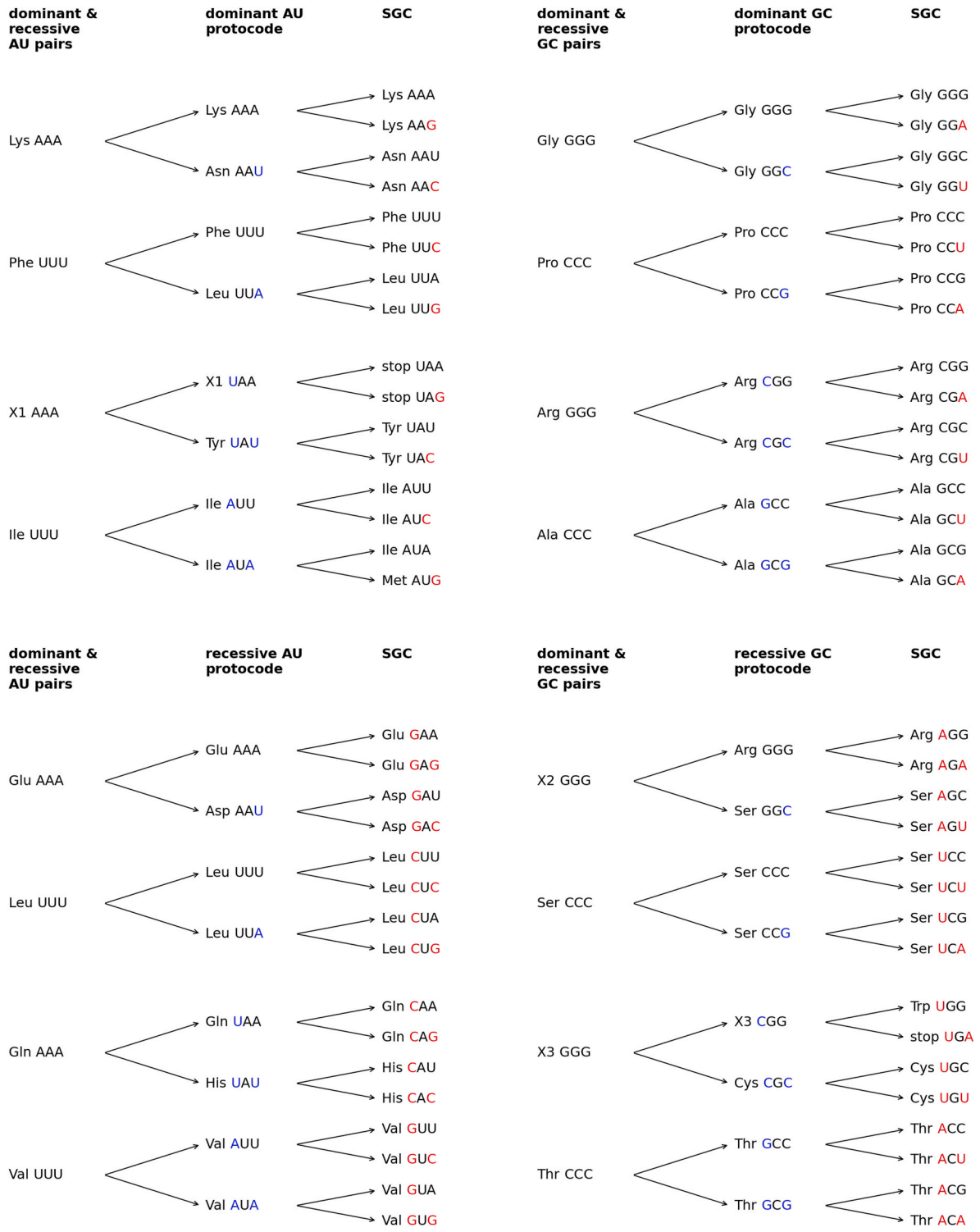


Fig. 1. Combinatorial fusion cascade of the canonical amino acids leading to the codon assignments in the SGC. The blue letters indicate the fusion rules for the dominant and recessive AAA/UUU- and GGG/CCC-pairs to the protocodes. The red letters indicate the fusion rules for dominant and recessive AU- and GC-protocodes to the SGC. The fusion pattern is identical for all amino acids: the third position changes in the codons of the dominant entities. The first or the first and the third positions change in the codons of the recessive entities (Nesterov-Mueller and Popov, 2021). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

the competing-pairs stage the triplets are homogeneous and complementary, and therefore of minimal entropy: minimality is reached both through complementary pairing between the entities (AAA/UUU, GGG/CCC) and through the monobasic composition of the codons. At the protocode stage each entity already contains eight codons, obtained by raising the entropy — diluting the homogeneous triplets with the complementary base of the same pair. The fusion of the protocodes into the SGC then introduces higher-entropy triplets, as codons built from a single complementary pair are joined by codons carrying bases from the other protocodes. This qualitative increase of entropy along the cascade can be made precise by assigning each triplet an entropy rank.

Let $B = \{A, U, G, C\}$ be the alphabet of bases. The bases split into two

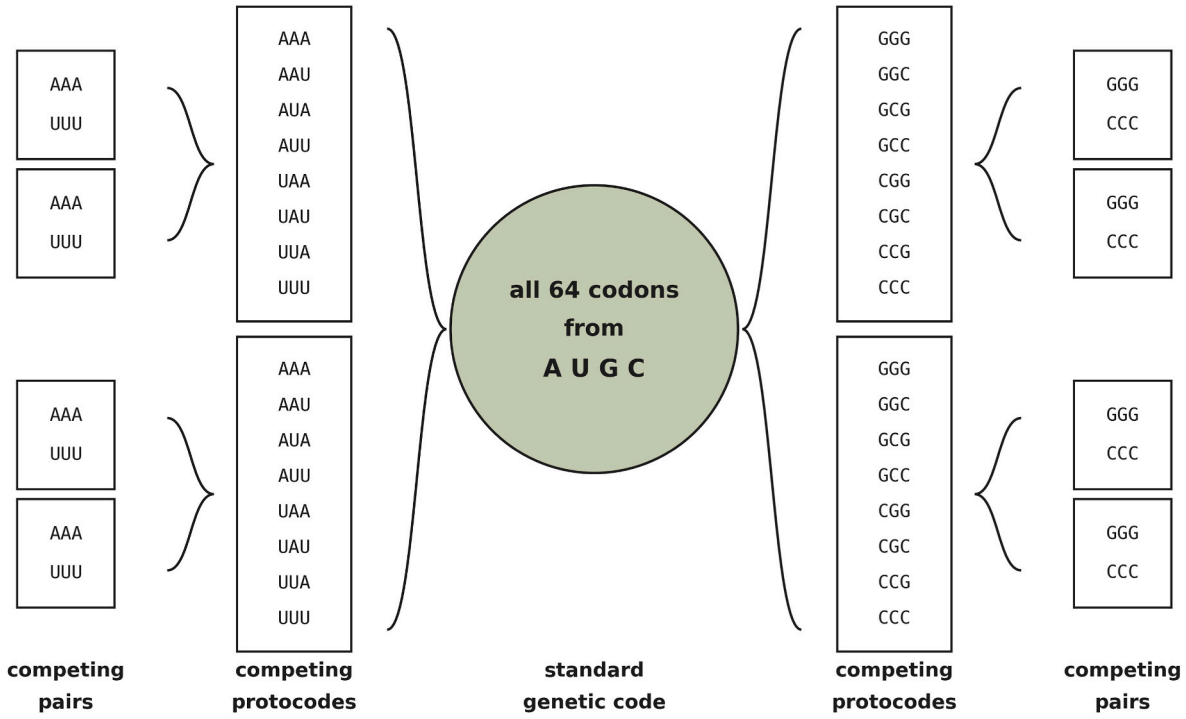


Fig. 2. A simplified, amino-acid-free version of Fig. 1, showing that the CFC operates entirely at the triplet level: competing complementary pairs fuse into competing protocodes and then into the standard genetic code, with the codon entropy increasing at each fusion step.

complementary pairs, {A, U} and {G, C}; the complementary class K maps each base to its pair: $K(\mathrm{A}) = K(\mathrm{U}) = \{\mathrm{A}, \mathrm{U}\}$, $K(\mathrm{G}) = K(\mathrm{C}) = \{\mathrm{G}, \mathrm{C}\}$. Let σ be the base exchange $\sigma(\mathrm{A}) = \mathrm{G}$, $\sigma(\mathrm{G}) = \mathrm{A}$, $\sigma(\mathrm{C}) = \mathrm{U}$, $\sigma(\mathrm{U}) = \mathrm{C}$, which sends each base into the opposite complementary class.

For two bases we define a relation cost $\kappa(x, y)$: 0 if $x = y$ (identical); 1 if $\{x, y\}$ is a complementary pair (A–U, G–C); 2 if the two bases are of the same ring class but not complementary (the transitions A–G and C–U); 3 otherwise (the transversions A–C and G–U). For a codon $c = (b_1, b_2, b_3)$ we take its three positional pairs and define the entropy rank

$$\rho(c) = \max\{\, \kappa(b_1, b_2), \kappa(b_1, b_3), \kappa(b_2, b_3) \,\}, \tag{1}$$

the costliest relation present in the codon, together with the cumulative entropy

$$S(c) = \kappa(b_1, b_2) + \kappa(b_1, b_3) + \kappa(b_2, b_3). \tag{2}$$

The three cascade stages are exactly the rank levels $\{\, c\colon \rho(c) \le r \,\}$: $\rho = 0$ gives the four homogeneous monobasic codons {AAA, UUU, GGG, CCC} (the initial pairs); $\rho \le 1$ gives the sixteen codons whose bases all lie in a single complementary pair, eight over {A, U} and eight over {G, C} (the protocodes); $\rho \le 3$ gives all 64 codons (the SGC). These sets are strictly nested, $\{\rho \le 0\} \subset \{\rho \le 1\} \subset \{\rho \le 3\} = B^3$, i.e. $4 \subset 16 \subset 64$.

The fusion raises the rank monotonically. Because σ sends a base into the opposite complementary class, $K(\sigma(b)) \neq K(b)$, a protocode codon (all bases of one class, $\rho \le 1$) acquires a pair of cost ≥ 2 as soon as σ acts on any position, so ρ jumps to ≥ 2. Fusion can therefore only increase the rank: $\rho_{\max} = 0 \to 1 \to 3$ along the three stages (equivalently max $S = 0 \to 2 \to 6$). The cascade thus runs from homogeneity (rank 0) through a single complementary pair (rank 1) to all relation types (rank 3), each stage strictly nested in the next.

For example, AAU has pairs A–A, A–U, A–U ($\kappa = 0, 1, 1$), so $\rho = 1$, $S = 2$ (protocode); GAA has G–A, G–A, A–A ($\kappa = 2, 2, 0$), so $\rho = 2$, $S = 4$ (SGC); GUA has G–U, G–A, U–A ($\kappa = 3, 2, 1$), so $\rho = 3$, $S = 6$ (SGC). The rank is raised by specific fusion operations, which we now state as three rules.

The fusion operators Σ_1, Σ_3, Σ_{13} apply σ in the first position, the third, or both, and never in the second:

$$\Sigma_1(b_1, b_2, b_3) = (\sigma(b_1), b_2, b_3),\ \Sigma_3(b_1, b_2, b_3) = (b_1, b_2, \sigma(b_3)),$$
$$\Sigma_{13}(b_1, b_2, b_3) = (\sigma(b_1), b_2, \sigma(b_3)). \tag{3}$$

Each fusion operator leaves position 2 unchanged. The system of a codon is fixed by its second position: an AU-system when $K(b_2) = \{\mathrm{A}, \mathrm{U}\}$, a GC-system when $K(b_2) = \{\mathrm{G}, \mathrm{C}\}$. The sixteen protocode codons are the eight over {A, U} and the eight over {G, C}. Each protocode codon h is assigned an amino acid; since the same codon appears in both a dominant and a recessive protocode, it has a dominant assignment $a_{\mathrm{D}}(h)$ and a recessive one $a_{\mathrm{R}}(h)$ — for AAA, Lys and Glu, respectively.

The protocode-to-SGC transition obeys three rules, identical for all amino acids.

Rule 1 (invariance of the second position). No fusion operator changes b_2. This is the most conserved feature of the code.

Rule 2 (dominant fusion). In a dominant entity the exchange σ is applied in the third (wobble) position only (Lei and Burton, 2020; Koonin and Novozhilov, 2017):

$$\Phi_{\mathrm{D}}(h) = \{(h, a_{\mathrm{D}}(h)), (\Sigma_3(h), a_{\mathrm{D}}(h))\}. \tag{4}$$

For example, in the dominant AU protocode the pair Lys AAA/Phe UUU expands to Lys AA (A/G) and Phe UU(U/C); the first and second positions stay fixed.

Rule 3 (recessive fusion). In a recessive entity the exchange σ is applied in the first position alone, or in the first and third positions together:

$$\Phi_{\mathrm{R}}(h) = \{(\Sigma_1(h), a_{\mathrm{R}}(h)), (\Sigma_{13}(h), a_{\mathrm{R}}(h))\}. \tag{5}$$

For example, in the recessive AU protocode Glu AAA/Leu UUU generates Glu GA (A/G) and Leu CU(U/C); the same exchange appears in positions 1 and 3, never in position 2. The generated code is

$$\mathrm{SGC} = \cup\,(\Phi_{\mathrm{D}}(h) \cup \Phi_{\mathrm{R}}(h)) \tag{6}$$

the union over the sixteen protocode codons h, where $\cup$ is the standard set-union operator.

The complementary classes partition all 64 codons into two disjoint

groups: $K(b_1) = K(b_2)$ selects the dominant codons, whose first position agrees with the second, and $K(b_1) \neq K(b_2)$ selects the recessive codons, whose first position is crossed with the second. This split follows from the definition of the classes alone — every codon belongs to exactly one group, and the dominant and recessive expansions never collide on the same codon.

The three rules thus constitute a deterministic procedure on an a-priori lattice: their application reproduces the SGC from the competing protocodes, and replacing any single assignment destroys the one-to-one coverage of all 64 codons (producing either collisions or uncovered codons).

The off-table post-cascade events are given by an explicit list of reassignments on the completed code: AUG ↦ Met, UGG ↦ Trp, and removal of X_1 (UAA, UAG ↦ stop) and X_3 (UGA ↦ stop).

The dominant/recessive split can also be formalized through the primacy of the first two positions: codons that stay within a single complementary pair (low entropy rank) fall into the dominant domain, whereas the more entropic codons, carrying cross-class transition pairs, fall into the recessive one. This observation naturally suggests an initial two-base code. It would then be necessary to explain the entropy asymmetry between the entities — dominant protocodes are built from the bases of one complementary pair ($\rho \leq 1$), whereas recessive codons carry cross-class transition pairs ($\rho = 2$) — as well as the absence of competition between the codes. Against this background, the CFC produces the same entropy increase for every entity, achieved through the competition of codons.

4.2. Explanatory power

The CFC explains several features of the standard genetic code that have resisted explanation within the conventional framework:

(a) *Even and odd codon numbers.* In the SGC, most amino acids have an even number of codons (2, 4, or 6). This is a direct consequence of the cascade structure: each amino acid in the initial pairs ideally receives four codons after fusion. The odd numbers (1 codon for Met, Trp; 3 codons for Ile) arise from specific events – late entry or removal of amino acids at particular cascade positions. No other theory explains why even codon numbers dominate.

(b) *Six-codon amino acids.* Arg, Leu, and Ser each have six codons – an anomaly that has puzzled researchers for decades. Wong's (1975) coevolution theory explained this only ad hoc. In the CFC, six codons arise naturally: these amino acids received four codons in their primary protocode and two additional codons from a second protocode through dual transfer during fusion. This is a necessary consequence of the cascade geometry, not an adjustment.

(c) *Stop codons.* The three stop codons (UAA, UAG, UGA) arise from the removal of amino acids X1 (generating UAA and UAG) and X3 (generating UGA) from the cascade. Pyrrolysine occupying UAG in archaea (Srinivasan et al., 2002) and selenocysteine occupying UGA (Donovan and Copeland, 2010) are commonly interpreted as late evolutionary additions. The CFC offers the opposite interpretation: X1 and X3 originally occupied these positions in the cascade, were subsequently lost – generating the stop codons UAA, UAG, and UGA – and were later reoccupied by the same or chemically similar amino acids in organisms whose environments resembled the original protocode conditions.

(d) *Temporal order.* Trifonov (2000) compiled a consensus temporal order of amino acid entry into the code from 40 independent criteria. The CFC's cascade stages produce an entry order congruent with Trifonov's consensus (Fig. 3). Early amino acids

(a)

Amino acid chronology↓	Dominant protocodes	Recessive protocodes
I. Stage of the initial pairs	Lys, Phe	Glu, Leu
	X1, Ile	Gln, Val
	Gly, Pro	X2, Ser
	Arg, Ala	X3, Thr
II. Stage of the coexisting protocodes	Asn	Asp
		His
	Leu (uua/g)	Cys
		Arg (aga/g)
	Tyr	Ser (agu/c)
III. After-fusion-stage	Met	Trp
	stopcodon (uaa/g)	stopcodon (uga/g)

(b)

Amino acid chronology↓
Gly
Ala
Val
Asp
Pro
Ser
Glu
Leu
Thr
Arg
Ser (agu/c)
Arg (aga/g)
Asn
Lys
Gln
Leu (uua/g)
Ile
Cys
His
Phe
Met
Tyr, stop
Trp, stop

Fig. 3. (a) Amino acid chronology of the combinatorial fusion cascade; **(b)** amino acid chronology according to the consensus temporal order after Trifonov (Trifonov, 2000). Both the combinatorial fusion cascade and the consensus temporal order indicate a later acquisition of the codons UUA/UUG by the amino acid Leu, AGA/AGG by Arg, and AGU/AGC by Ser. As a result, each of these amino acids acquired six codons in the SGC. The color denotes the belonging of the amino acid to the stage of the combinatorial fusional cascade: red – the stage of the initial pairs, blue—the stage of coexisting protocodes, brown – after-fusion-stage (Nesterov-Mueller and Popov, 2021). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

(Gly, Ala, Asp, Glu, Val) occupy dominant initial pairs; later amino acids (Met, Trp) entered post-fusion.

4.3. Unfolded positions: the mendeleev principle

The CFC leaves structural positions for amino acids that entered or left the code (X1, X2, X3). X2, proposed as canavanine (Emmert et al., 1998), a guanidinooxy analogue of arginine. Pyrrolysine (Pyl) and lysine (Lys) are in the same protocode; selenocysteine (Sec) and cysteine (Cys) are in the same protocode. Preiner et al. noted the same synthesis pathway for Lys, Pyl, Met, Ile, and Asn (Preiner and Asche, 2019) – all in the dominant AU protocode. This parallels Mendeleev's periodic table: empty places with predicted properties.

4.4. The geochemical timing of sulfur in the code

Cys is the first sulfur-containing amino acid in the cascade, entering only at the protocode stage – not among the initial pairs. This indicates that sulfur chemistry was unavailable during the earliest stages of the code and became accessible only after the pre-existing initial pairs were exposed to sulfur through geological events such as the impact of a differentiated planetary embryo (Grewal et al., 2019) or late accretion of chondritic material (Wang and Becker, 2013). Met, the second sulfur-containing amino acid, entered even later – post-fusion – confirming that sulfur integration into the code was a stepwise process following the geochemical enrichment of the surface environment. Jenne et al. experimentally excluded Met and Trp from primordial RNA-peptide screening based on this CFC reasoning (Jenne et al., 2023).

4.5. Living membranes: a prediction of the CFC

By a living membrane we mean a surface-associated organization whose principal components are the amphiphilic RNA–peptide complexes that also carry proto-informational function — the complementary pairs and protocodes. The term denotes that the same molecules provide both information and partial spatial demarcation; it does not denote a closed lipid envelope or a biological organism. “Living” is used by analogy with “living polymerization” in chemistry — a self-maintaining, persistent, organized process — not in a cellular sense. Crucially, a living membrane is not an impermeable compartment: it demarcates a community on a catalytic surface without sealing it off, and therefore does not inherit the difficulties of lipid vesicles discussed in §3.2.

In independent simulations, Yarus reported that a code can assemble through the fusion of competing coding entities (Yarus, 2022, 2023, 2024) — the same symbiogenetic mechanism that defines the CFC. His model also requires a condition we build on here: the competing entities must persist long enough to fuse. His work thus provides independent support for the mechanism and for this persistence requirement, while the mineral-surface setting is inferred from the CFC itself — from the homogeneity of the seed triplets — and is not part of his model.

Yarus's model makes the persistence requirement quantitatively precise. In his “crescendo of competent coding” (c^3), approximately 1 in 22 fusion environments reaches codes with $\geq$20 assignments and $\leq$3 differences from the SGC, while approximately 90% of all codes are annihilated through incompatible fusions — only 9.4% survive in late environments (Yarus, 2022). The fossil record independently shows that the most ancient communities were densely layered — “close laminations, with different organisms densely layered” in 3.43 Ga stromatolites (Allwood et al., 2006) and “compact bundles of filaments” 3.75–4.28 Ga in seafloor jasper (Papineau et al., 2022) — a colonial organization that would favour the association of entities of different origin and competency, and hence the fusion of their codes that Yarus's model requires.

This persistence requirement has a consequence for the CFC. If the competing entities — complementary pairs and protocodes — are to persist as coherent units rather than disperse, they require protection from degradation and dilution; surface anchoring with an associated boundary is a natural way to provide it. We stress that this requirement is general: it applies to any persistent-RNA scenario, so we treat it as a necessary condition, not as a derivation specific to the CFC.

The setting itself, likewise, is not new, and it combines two aspects that have been studied largely apart and not from the standpoint of the genetic code: (i) RNA on mineral surfaces as catalytic support for RNA oligomerization (Ferris, 2006; Jerome et al., 2022; Closas et al., 2025) — a line that is itself contested, since adsorption can also constrain RNA folding — and (ii) peptide chemistry on complementary RNA strands (Müller et al., 2022; cf. Wong, 1991). Neither aspect, on its own, addresses codon assignments. We adopt both as established results rather than claim them.

What is specific to the CFC is the link between these two strands, which we make explicit in three steps, in order of strength.

First, the RNA–peptide communities are the code elements. In the CFC the initial pairs and protocodes are not abstract tables but the amphiphilic RNA–peptide complexes themselves — an amino acid associated with its cognate homo-triplet. The community is therefore not a container for the code but its physical realization. This is a direct, defining link, not an analogy.

Second, the homogeneity of the seeds requires a selective surface. The cascade is seeded by homogeneous monobasic triplets (AAA, UUU, GGG, CCC). Such sequence — and plausibly chiral — homogeneity is reached far more readily on a selective catalytic mineral surface than in bulk solution, which ties the start of the cascade to the mineral setting.

Third, the dependence on the surface decreases along the cascade. In the entropy-rank formalization (§4.1) the codon rank rises monotonically across the stages ($0 \rightarrow 1 \rightarrow 3$): the seeds (rank 0) demand maximal selectivity of the environment, whereas fusion raises the codons' own entropy, so the system becomes progressively self-sustaining and less dependent on the surface. The cascade is thus simultaneously an increase in codon entropy and a gradual detachment from the mineral substrate.

Living membranes are the predicted precursors of bacterial mats — the oldest known macroscopic biological structures on Earth (stromatolites, ~3.5 Ga). The transition from living membranes to bacterial mats is the transition from prebiotic chemistry to biology. Protocode mats — communities of competing living membranes on mineral surfaces — are the environment in which the CFC is proposed to have occurred.

4.6. Protocode mats: from bacterial mat analogy to falsifiable prediction

By a *protocode mat* we mean a surface-associated community of competing RNA–peptide complexes (the complementary pairs and protocodes) on a catalytic mineral surface. The comparison with bacterial mats is an analogy of organizing principle — chemical stratification and vertical exchange, with an emergent function no single layer achieves — not of morphology: we do not imply that protocode mats resembled stromatolites or formed mineral-layered fossil structures. The interaction between competing protocodes within layered structures can be understood by analogy with modern bacterial mats. In a bacterial mat, diverse microbial species form vertically stratified layers, each occupying a distinct chemical niche: phototrophs at the surface, sulfur oxidizers below, sulfate reducers deeper, methanogens at the base. These layers are not independent – they interact through vertical exchange of metabolites, forming a functionally integrated community that no single species could sustain alone. The CFC proposes that protocode mats operated analogously. A specific physical stratification is one illustrative possibility only — for example, the more hydrophobic AU-protocodes facing the aqueous medium and the GC-protocodes nearer the mineral surface — and is not part of the claim. What the analogy asserts is functional: distinct protocodes occupying distinct chemical niches with exchange between them.

In bacterial mats, the community achieves innovations that no single

species can: nitrogen fixation, sulfur cycling, and complex electron transport chains emerge from the interaction between layers, not from individual organisms. Similarly, the CFC predicts that the standard genetic code emerged from the interaction and fusion of protocode layers – a combinatorial innovation impossible for any single protocode acting alone. This hypothesis concerns only the pre-cellular stage — the assembly of the code in surface-associated RNA–peptide communities. The subsequent transition from such communities to membrane-bounded cells lies outside its scope and is treated as an open problem (§4.11).

Goldman, Fournier, and Kaçar found that *every* universal paralog predating LUCA relates to membrane function or protein synthesis – nothing else (Goldman et al., 2026). Membrane function is as ancient as the code itself. Moody et al. dated LUCA to ~4.2 Ga with membrane-bound ATP synthase already present (Moody and Álvarez-Carretero, 2024). Hegde and Keenan (2024) confirmed that two membrane protein insertion families trace to LUCA. These findings are exactly what the CFC predicts: membranes were not added to the code later as packaging – they were co-emergent with the code as part of the same generating process.

The following logical chain connects Yarus's computational results to a specific physicochemical prediction. First, fusion requires stability: competing codes must persist as coherent entities, and approximately 90% are annihilated (Yarus, 2022). Second, stability requires spatial organization: free-floating molecules in solution cannot maintain the persistence needed. Third, the organized layers must be functional code elements, not incidental mineral coatings – otherwise the zero-probability problem of Section 4.11 returns through the back door. Fourth, RNA layers on mineral surfaces require ion compensation: the phosphodiester backbone carries one negative charge per nucleotide, which must be neutralized by divalent cations (Mg^{2+}, Ca^{2+}, Fe^{2+}) or by positively charged amino acids. This creates natural selection pressure for the earliest codes to include positively charged amino acids – exactly as the CFC predicts: both protocodes contain one (Lys in AU, Arg in GC). Fifth, catalytic mineral surfaces solve the membrane transparency problem that plagues the conventional protocell program (Section 3.2). Protocode mats are open to nutrient exchange from the aqueous environment while maintaining spatial organization through surface binding and ionic interactions. The mineral surface itself provides catalytic activity for RNA polymerization (Ferris, 2006; Closas et al., 2025; Jerome et al., 2022). This chain of reasoning converts Yarus's computational observation into a falsifiable prediction: the genetic code arose in layered molecular communities on catalytic mineral surfaces where stability, ion compensation, catalytic activity, and nutrient access were simultaneously available. The lipid membrane came later, as a product of the code, not its precondition.

Sulfur chemistry may have been required in the broader prebiotic environment for basic building blocks (Patel et al., 2015) and for activation chemistry. How then could the initial complementary pairs have formed without it? The protocode-mat hypothesis offers a possible resolution: sulfur may have been abundant in the surrounding environment, while the mats themselves — at the stage of the initial pairs — remained locally protected from it. Cysteine entering at the protocode stage and methionine after fusion (§4.4) may then mark not the appearance of sulfur, but the moment when the mats became permeable to it. The mat may act as any biological membrane — admitting the chemistry the code requires, excluding what would degrade it.

4.7. Symbiosis as the main mechanism of bioinnovation

The CFC is fundamentally symbiogenetic. This places the origin of the genetic code within a broader biological principle. Margulis (1970) demonstrated that mitochondria and chloroplasts originated as endosymbionts. Kozo-Polyansky first formulated symbiogenesis as an evolutionary principle as early as 1924 (Margulis, 2010). In 2024, Coale et al. discovered the nitroplast, the fourth confirmed primary endosymbiosis, demonstrating that symbiogenetic innovation continues to the present day (Coale et al., 2024).

Vetsigian, Woese, and Goldenfeld showed that horizontal transfer among competing entities is required for a universal code (Vetsigian et al., 2006). Woese (2002) described the universal ancestor not as a single organism but as a diverse community of progenotes. The CFC represents the extreme form of this principle: complete combinatorial fusion of competing codes — not gradual optimization of a single lineage, but merger of distinct coding systems into a unified whole.

4.8. RNA–peptide interactions within protocodes: experimental evidence

The CFC predicts that protocodes were not abstract coding tables but physically realized RNA–peptide communities. This prediction has received direct experimental support. Müller et al. demonstrated prebiotically plausible peptide synthesis on short complementary RNA fragments under mild aqueous conditions (Müller et al., 2022), establishing that the RNA/peptide world concept has a concrete chemical basis. Crucially, their system operates on exactly the kind of short complementary oligonucleotides that the CFC posits as the molecular substrate of the initial competing pairs.

Jenne et al. tested this prediction systematically using high-density peptide arrays containing all combinatorial combinations of amino acids from the AU- and GC-protocodes (Jenne et al., 2023). The peptide libraries were incubated with fluorescently labeled 12-mer homo-oligonucleotides of adenine, uracil, guanine, and cytosine. The results revealed a striking pattern. In the dominant AU-protocode, the amino acids phenylalanine and tyrosine – both aromatic – dominated the strongest RNA-binding signatures. As peptide length increased, the number of Phe and Tyr residues in the strongest binders increased correspondingly. In the dominant GC-protocode, proline played an exceptional role, providing the strongest binding to 12-mer cytosine RNA, though polyprolines exhausted their binding potential at the 5-mer level.

These results have three implications for the CFC. First, the strongest RNA–peptide interactions occur between amino acids and their cognate homo-oligonucleotides in the dominant protocodes – exactly as predicted by the protocode partition. Phe and Tyr, encoded in the dominant AU-protocode by UUU and UAU respectively, bind most strongly to adenine-rich RNA. Pro, encoded in the dominant GC-protocode by CCC, binds most strongly to cytosine-rich RNA. Second, the dominance of aromatic amino acids in RNA binding supports the CFC prediction that aromatic amino acids were present from the inception of the code, not late additions as some prebiotic chemistry models suggest.

Radakovic et al. provided additional support from a different direction, demonstrating that RNA aminoacylation – the charging of RNA with amino acids – may have preceded its role in peptide synthesis (Radakovic et al., 2024). This suggests that the earliest function of the codon–amino acid association was not translation but the formation of RNA–amino acid complexes with structural and catalytic properties. Within the CFC framework, this is precisely what the initial competing pairs represent: amino acids associated with their cognate homo-triplets, forming amphiphilic complexes before any translation machinery existed.

4.9. Vertical transitions and their prebiotic chemistry interpretation

A striking feature of the CFC is that the three six-codon amino acids – Leu, Arg, and Ser – each acquired their additional codons through vertical transitions within a single protocode family. Leu transitioned exclusively within the AU system (receiving UUA/UUG from the recessive AU protocode). Arg and Ser transitioned exclusively within the GC system (Arg receiving AGA/AGG, Ser receiving AGU/AGC from the recessive GC protocode). This pattern may have a prebiotic chemistry interpretation.

Leu is the most abundant hydrophobic amino acid in modern

proteomes and has the broadest hydrophobic surface among branched-chain amino acids. In AU-rich RNA environments (adenine and uracil layers on mineral surfaces), Leu's hydrophobic character may have facilitated multiple modes of amphiphilic RNA-peptide interaction, which could explain why it occupies both dominant and recessive niches within the AU family. Arg has the strongest RNA-binding affinity of all amino acids (Yarus et al., 2009) and interacts preferentially with guanine-rich sequences. Its positive charge compensates the negative phosphate backbone of GC-rich RNA. Arg's ability to bridge between different GC-rich structures is consistent with its expansion within the GC system. Ser is the smallest amino acid with a hydroxyl group, giving it hydrogen-bonding versatility. In GC-rich environments, Ser may have served as a flexible structural linker between RNA and peptide components. Ser is notably overrepresented in intrinsically disordered protein regions – suggesting an ancient role as a flexible connector.

The vertical-transition pattern is consistent with a prebiotic chemical principle: amino acids expand their codon territory within the nucleotide environment to which they are chemically best adapted. This connects to modern prebiotic concepts: Ferris's (2006) montmorillonite-catalyzed RNA polymerization, Müller et al.'s peptide synthesis on complementary RNA (Müller et al., 2022), and Closas et al. (2025) clay-induced RNA replication all demonstrate that surface chemistry is selective, not chaotic.

4.10. A prebiotic chemistry roadmap derived from the CFC

The temporal order of the cascade yields a set of falsifiable predictions about the chemical environments in which the code was assembled. We state each as a prediction together with the observation that would falsify it, and we separate those testable with current methods from those awaiting suitable techniques. These chemical predictions are distinct from, and additional to, the quantitative structural predictions of the CFC discussed above (the deterministic generation of all 64 codon assignments, the 32/32 dominant–recessive partition, the six-codon families, and the stop-codon positions), which are already testable against the code table itself.

Prediction 1 (testable now): aromatic amino acids from inception. The CFC predicts that the aromatic amino acids Phe and Tyr were present at the inception of the code, not late additions. This is already partially supported: on high-density peptide arrays, the dominant-AU aromatics show the strongest cognate RNA binding (Jenne et al., 2023). The prediction would be weighed against by the absence of early aromatic association in further array and in-vitro selection experiments. Prediction 2 (testable now): sulfur ordering. The CFC predicts that no sulfur-containing residue was present among the initial complementary pairs — Cys enters only at the protocode stage and Met only after fusion (Section 4.4). The prediction would be falsified by evidence that a sulfur-containing residue was in fact part of an initial pair.

Prediction 3 (testable now): occupancy of the open positions X1, X2, X3. The CFC predicts that the present stop codons originally carried amino acids and leaves three structural positions (X1 at UAA/UAG, X3 at UGA, and X2). Each is a falsifiable claim about what early or extant lineages should reveal; current candidates are discussed below and remain tentative.

Selenocysteine (Sec) occupies UGA – a stop codon in the standard code but a predicted open position in the CFC (X3). Sec and Cys are in the same protocode. The CFC raises the possibility that Sec may have preceded Cys in early coding and was later displaced as sulfur chemistry became dominant. Selenium is more reactive than sulfur and forms selenol groups (pKa ~5.2) that are deprotonated at physiological pH, making selenocysteine a superior catalytic residue in oxidative environments (Hatfield and Gladyshev, 2002).

Pyrrolysine (Pyl) occupies UAG – another stop codon, another CFC open position (X1). Pyl and Lys are in the same dominant AU protocode. Pyl is found today only in methanogenic archaea, where it is essential for methylamine metabolism – a pathway requiring NH_3. If early metabolism was heavily dependent on ammonia-based chemistry, as expected in a reducing early atmosphere, pyrrolysine may have been a critical catalytic amino acid that became dispensable when other nitrogen metabolic pathways evolved (Krzycki, 2005; Rother and Krzycki, 2010). The recently reported archaeal code in which all UAG codons encode pyrrolysine (Kivenson et al., 2025) matches the predicted X1 occupancy of UAG; it is consistent with, but does not by itself prove, the cascade interpretation.

The CFC predicts X2 as canavanine (Emmert et al., 1998), the guanidinooxy analogue of arginine. In modern legumes, canavanine functions as a major nitrogen storage compound – up to 12% of seed dry weight (Rosenthal, 1977). Arginyl-tRNA synthetase cannot distinguish canavanine from arginine, incorporating it directly into proteins. This easy interchangeability suggests that in early coding, the canavanine/arginine position was flexible, reflecting the central importance of nitrogen metabolism for emergence of the genetic code. As a falsifiable statement: the absence of canavanine or another guanidinooxy-arginine analogue in the translational apparatus of early-branching archaeal or bacterial lineages would falsify the X2 prediction; its detection would support it. This is testable now by genomic and biochemical screening.

Prediction (awaiting methods): seeding on surfaces. The CFC predicts that homogeneous monobasic triplets (AAA, UUU, GGG, CCC) self-organize into amino-acid-specific complementary pairs on catalytic mineral surfaces. This is not yet testable with existing experimental systems; it is stated as an open prediction in the Limitations (Section 4.11).

The transition from RNA to DNA as genetic material likely occurred after the main cascade. The CFC framework suggests this transition was possible only after protocode fusion produced a sufficiently complex coding system to encode the enzymes for deoxyribonucleotide synthesis (Forterre, 2005). The existence of separate AU- and GC-protocodes may reflect the chemical reality that ribonucleotides were the original coding molecules, while deoxyribonucleotides emerged as more stable storage alternatives – a transition driven by the same stability pressures that favored layered, persistent protocode mats.

4.11. Solving the zero probability problem

The space of possible coding tables – mappings of 64 codons to 20 amino acids and stop signals – is approximately 10^{83} (Novozhilov et al., 2007). The probability of arriving at the SGC through random search of this space is physically impossible, even given the age of the universe.

The CFC resolves this problem by changing its nature. The code was generated by a deterministic cascade from eight initial complementary pairs. The question shifts from "what is the probability of finding this specific code?" to "what is the probability of initiating the cascade?" The latter requires only short homogeneous RNA strands (monobasic triplets like AAA and UUU) on mineral surfaces – a vastly more tractable problem that falls within the domain of demonstrated prebiotic chemistry (Ferris, 2006; Kudella et al., 2021; Jerome et al., 2022).

This shift has a deeper implication. If the code's structure is determined by the combinatorial cascade from the beginning, then the probability question as traditionally posed is meaningless.

5. Methodological principle: first the pattern, then the explanation

Between 1609 and 1619, Kepler discovered three empirical laws of planetary motion. Newton's gravitational theory, arriving 68 years later, confirmed that Kepler's pattern pointed to a universal law – but the pattern had to come first.

In 1869, Mendeleev left empty places for undiscovered elements and predicted their properties. Gallium (1875), scandium (1879), and germanium (1886) matched his predictions. The physical explanation came only with quantum mechanics.

We invoke this parallel solely to illustrate a methodological point —

that a predictive empirical regularity can precede the discovery of its mechanism — and not as evidence for the CFC. Pattern recognition does not validate a model; it motivates the search for one.

The Kepler and Mendeleev cases reveal a methodological principle. Pattern-first discovery succeeds when three conditions are jointly met (Koonin, 2017): the phenomenon is universal – it applies globally, not locally (Koonin and Novozhilov, 2017); it crosses existing disciplinary boundaries, so that no single field can solve it from within; and (Koonin and Novozhilov, 2009) it concerns a fundamental law.

On these methodological criteria — universality, cross-disciplinary relevance, and fundamental importance — the genetic code is a natural candidate for a regularity-first approach. It is universal: the same code operates in virtually all life on Earth, from archaea to mammals. It crosses disciplinary boundaries: chemistry, information theory, evolutionary biology, and the philosophy of science all contribute, but none alone can explain the codon assignments. And it concerns a fundamental question: arguably the deepest unsolved problem in biology – how information first entered matter. The absence of a molecular mechanism for the CFC is historically normal and temporally bounded. Mendeleev's table waited 56 years for quantum mechanics. The CFC may wait less. A doctoral student beginning work on the genetic code today has a realistic prospect of witnessing, within their career, both the discovery of the CFC's molecular mechanism and its ultimate consequence: the solution to the origin of life.

6. Fusion rules as representation, CFC as hypothesis, and open problems

A critical distinction must be drawn between the fusion rules as representation and the CFC as hypothesis.

The fusion rules as representation. The three combinatorial rules are objective structural properties of the existing genetic code, readable directly from the standard code table without any assumption about history or mechanism. Rule 1 (second-position invariance), Rule 2 (third-position exchanges in dominant codons), and Rule 3 (first-position exchanges in recessive codons) describe the factual pattern of codon assignments. These rules are a representation – just as the periodic table describes the factual arrangement of elements. One cannot "go around" this structure: any future solution to the code's origin, regardless of molecular mechanism, must reproduce this pattern. These three simple, uniform rules are sufficient to completely describe all 64 codon–amino acid assignments. This completeness and simplicity is not self-evident: one would expect a complex mapping to require complex rules.

The CFC as hypothesis. The peculiarity of the fusion rules – their simplicity, uniformity, and completeness – generates the CFC hypothesis: competing protocodes actually fused on early Earth through the cascade. This is an empirical hypothesis – testable, in principle, and falsifiable. Evidence for or against could come from: (i) discovery or exclusion of the predicted amino acid X2 (proposed as canavanine (Emmert et al., 1998)) in primitive organisms; (ii) experimental demonstration that short homogeneous RNA triplets on mineral surfaces spontaneously organize into complementary pairs with amino acid specificity; (iii) the continued confirmation or refutation of CFC predictions about pyrrolysine, selenocysteine, and temporal order.

Open problems and limitations. First, the molecular mechanism that produced the cascade is unknown. What physical process caused dominant entities to retain their codons while recessive entities acquired new ones? These questions remain open. Second, the living membrane concept (§4.5) has not been experimentally demonstrated. No laboratory experiment has yet produced self-organizing RNA–peptide layers on mineral surfaces that exhibit protocode-like behavior – selective amino acid–codon association, competition between complementary communities, or fusion of competing codes. Third, the CFC does not explain why exactly these 20 amino acids – and not others from the much larger pool of prebiotically available non-canonical amino acids – were selected, nor does it address the homochirality problem: why exclusively L-amino acids are used in the code. Fourth, the initial conditions of the cascade — the formation of eight complementary pairs with monobasic triplets on mineral surfaces — remain hypothetical. Although Kudella et al. (2021) showed that templated ligation reduces sequence space, and Himbert et al. (2016) demonstrated self-assembly of pure nucleotide strands on montmorillonite clay, no experiment has yet generated the specific starting configuration. The selection of 20 specific L-amino acids and the initial conditions of the cascade may therefore not be independent problems but consequences of a single, as yet undiscovered, catalytic mechanism.

Fifth, the predicted amino acid X2 (canavanine) has not been found in any primitive organism's translational apparatus. Until canavanine or another guanidinooxy analogue of arginine is identified in an archaeal or early-branching bacterial lineage, this prediction remains untested.

Finally, the mechanism by which surface-associated RNA–peptide communities gave rise to membrane-bounded cells is not addressed here and remains open.

7. Conclusions

The origin of the standard genetic code has resisted explanation for over sixty years. We have argued that the lack of progress is due to the underlying assumptions of the dominant framework, not insufficient effort. The belief that the code developed as a later advantage for replicators has led to structural impasses: ribozymes that degrade faster than they can be replicated non-enzymatically (§3.1), lipid membranes that are destroyed by the metal ions required for RNA chemistry and block the diffusion of substrates across their boundary (§3.2), a code that cannot be derived from physical properties (§3.3), and aaRS that are products of the code, not its architects (§3.4). The molecular chaos starting condition faces $\sim 10^{83}$ possible codes with no navigation mechanism (§3.6). Prebiotic chemistry explains which amino acids and nucleotides could have existed on early Earth, but not why UUC encodes phenylalanine and not valine (§3.7).

The CFC offers a way out. Its core consists of three combinatorial fusion rules – objective structural properties readable directly from the code table. Their remarkable simplicity and completeness call for explanation and generate the CFC hypothesis: the code arose through combinatorial fusion of competing protocodes. Unlike frameworks that reproduce only organizational patterns (the coevolution theory) or a fraction of the assignments (at most 8 of 64 in the A/U stereochemical model), the CFC reproduces the full 64-codon table, the 32/32 dominant–recessive partition, and the six-codon families (Leu, Arg, Ser) as exact, quantitative consequences of three rules, given the protocode seeding. It further explains the even/odd codon numbers and the stop-codon positions and leaves Mendeleev-like open positions (Ch. 4). The CFC is fundamentally symbiogenetic: the code arose not through gradual optimization of a single lineage but through combinatorial fusion of competing codes — placing the origin of biological information within the same principle that later produced mitochondria, chloroplasts, and the nitroplast (§4.7). The acceleration of code evolution by fusion is an independent computational result of Yarus (§4.5, §4.6), separate from the combinatorial and entropic formalization developed here. The CFC predicts that the code arose in layered RNA–peptide communities on catalytic mineral surfaces, where the code and its physical boundary are one and the same structure, so that a lipid membrane is not required at the coding stage (§4.5, §4.6). Membrane proteins are among the oldest molecular systems, supporting the protocode mat and living membrane hypothesis (§4.5, §4.6).

The methodological principle of Chapter 5 places the CFC in historical context: Mendeleev's table, Kepler's laws – in every case where the phenomenon was universal, cross-disciplinary, and fundamental, the pattern came first, the explanation followed decades later.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the authors used Claude (Anthropic) in order to improve language clarity and editing. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This work received no specific funding for the preparation of this article.

CRediT authorship contribution statement

Alexander Nesterov-Mueller: Conceptualization, Methodology. **Dimitry Schmidt:** Investigation, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Alexander Nesterov-Mueller reports was provided by Karlsruhe Institute of Technology. Alexander Nesterov-Mueller reports a relationship with Karlsruhe Institute of Technology that includes: employment. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors acknowledge the Open Access Publishing Fund of the Karlsruhe Institute of Technology.

Data availability

Data will be made available on request.

References

Adamala, K., Szostak, J.W., 2013. Nonenzymatic template-directed RNA synthesis inside model protocells. Science 342, 1098–1100.
Allwood, A.C., Walter, M.R., Kamber, B.S., Marshall, C.P., Burch, I.W., 2006. Stromatolite reef from the early Archaean era of Australia. Nature 441 (7094), 714–718.
Łapińska, U., Glover, G., Kahveci, Z., Irwin, N.A.T., Milner, D.S., Tourte, M., Albers, S.-V., Santoro, A.E., Richards, T.A., Pagliara, S., 2023. Systematic comparison of unilamellar vesicles reveals that archaeal core lipid membranes are more permeable than bacterial membranes. PLoS Biol. 21, e3002048.
Attwater, J., Wochner, A., Holliger, P., 2013. In-ice evolution of RNA polymerase ribozyme activity. Nat. Chem. 5, 1011–1018.
Bada, J.L., 2013. New insights into prebiotic chemistry from Stanley Miller's spark discharge experiments. Chem. Soc. Rev. 42, 2186–2196.
Brooks, D.J., Fresco, J.R., Lesk, A.M., Singh, M., 2002. Evolution of amino acid frequencies in proteins over deep time. Mol. Biol. Evol. 19, 1645–1655.
Carter, C.W., Wills, P.R., 2018. Hierarchical groove discrimination by Class I and II aminoacyl-tRNA synthetases. Nucleic Acids Res. 46, 9667–9683.
Cech, T.R., 2000. Structural biology - the ribosome is a ribozyme. Science 289 (5481), 878–879.
Chen, I.A., Szostak, J.W., 2004. A kinetic study of the growth of fatty acid vesicles. Biophys. J. 87, 988–998.
Closas, J., Burcar, B.T., Herrero, M., Dotu, I., Menor-Salván, C., 2025. Polymerization and replication of primordial RNA induced by clay-water interface dynamics. Commun. Chem.
Coale, T.H., Loconte, V., Turk-Kubo, K.A., Vanslembrouck, B., Mak, W.K.E., Cheung, S., Ekman, A., Chen, J.-H., Hagino, K., Takano, Y., Nishimura, T., Adachi, M., Le Gros, M., Larabell, C., Zehr, J.P., 2024. Nitrogen-fixing organelle in a marine alga. Science 384 (6692), 217–222.
Copley, S.D., Smith, E., Morowitz, H.J., 2005. A mechanism for the association of amino acids with their codons and the origin of the genetic code. Proceed. Nat. Acad. Sciences USA 102, 4442–4447.
Di Giulio, M., 2008. An extension of the coevolution theory of the origin of the genetic code. Biol. Direct 3, 37.
Di Giulio, M., 2025. The genetic code is not universal. Biosystems 247, 105382.
Di Giulio, M., 2024a. Theories of the origin of the genetic code: strong corroboration for the coevolution theory. Biosystems 239, 105217.
Di Giulio, M., 2024b. The time of appearance of the genetic code. Biosystems 244, 105294.
Donovan, J., Copeland, P.R., 2010. The efficiency of selenocysteine incorporation. J. Mol. Biol. 400, 659–664.
Doudna, J.A., Cech, T.R., 2002. The chemical repertoire of natural ribozymes. Nature 418, 222–228.
Eigen, M., 1971. Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften 58, 465–523.
Eigen, M., Schuster, P., 1979. The Hypercycle: a Principle of Natural Self-Organization. Springer-Verlag, Berlin.
Emmert, E.A.B., Klimowicz, A.K., Thomas, M.G., Handelsman, J., 1998. Effect of canavanine from alfalfa seeds on the population biology of Bacillus cereus. Appl. Environ. Microbiol. 64, 4683–4688.
Eriani, G., Delarue, M., Poch, O., Gangloff, J., Moras, D., 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. Nature 347, 203–206.
Ferris, J.P., 2006. Montmorillonite-catalysed formation of RNA oligomers. Philos. Trans. R. Soc. B 361, 1777–1786.
Fontecilla-Camps, J.C., 2023. Reflections on the origin and early evolution of the genetic code. Chembiochem 24, e202300048.
Forterre, P., 2005. The two ages of the RNA world, and the transition to the DNA world: a story of viruses and cells. Biochimie 87 (9–10), 793–803.
Freeland, S.J., Hurst, L.D., 1998. The genetic code is one in a million. J. Mol. Evol. 47, 238–248.
Gil, R., Silva, F.J., Peretó, J., Moya, A., 2004. Determination of the core of a minimal bacterial gene set. Microbiol. Mol. Biol. Rev. 68, 518–537.
Goldman, A.D., Fournier, G.P., Kaçar, B., 2026. Universal paralogs provide a window into evolution before LUCA. Cell Genom., 101140
Grewal, D.S., Dasgupta, R., Sun, C., Tsuno, K., Costin, G., 2019. Delivery of carbon, nitrogen, and sulfur to the silicate Earth by a giant impact. Sci. Adv. 5, eaau3669.
Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., Altman, S., 1983. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. Cell 35, 849–857.
Haig, D., Hurst, L.D., 1991. A quantitative measure of error minimization in the genetic code. J. Mol. Evol. 33, 412–417.
Hartman, H., Smith, T.F., 2014. The evolution of the ribosome and the genetic code. Life 4, 227–249.
Hatfield, D.L., Gladyshev, V.N., 2002. How selenium has altered our understanding of the genetic code. Mol. Cell Biol. 22 (11), 3565–3576.
Hegde, R.S., Keenan, R.J., 2024. A unifying model for membrane protein biogenesis. Nat. Struct. Mol. Biol. 31 (7), 1009–1017.
Higgs, P.G., 2009. A four-column theory for the origin of the genetic code. Biol. Direct 4, 16.
Higgs, P.G., Pudritz, R.E., 2009. A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. Astrobiology 9, 483–490.
Himbert, S., Chapman, M., Deamer, D.W., Rheinstädter, M.C., 2016. Organization of nucleotides in different environments and the formation of pre-polymers. Sci Rep-Uk 6, 31285.
Hoesl, M.G., Budisa, N., 2012. Recent advances in genetic code engineering in Escherichia coli. Curr. Opin. Biotechnol. 23 (5), 751–757.
Horning, D.P., Joyce, G.F., 2016. Amplification of RNA by an RNA polymerase ribozyme. Proceed. Nat. Acad. Sciences USA 113, 9786–9791.
Jenne, F., Berezkin, I., Tempel, F., Schmidt, D., Popov, R., Nesterov-Mueller, A., 2023. Screening for primordial RNA-peptide interactions using high-density peptide arrays. Life 13, 796.
Jerome, C.A., Kim, H.J., Mojzsis, S.J., Benner, S.A., Biondi, E., 2022. Catalytic synthesis of polyribonucleic acid on prebiotic rock glasses. Astrobiology 22, 629–636.
Kivenson, V., Paul, B.G., Valentine, D.L., Bhatt, A.S., 2025. An archaeal genetic code with all TAG codons as pyrrolysine. Science 390 eadu2404.
Koonin, E.V., 2017. Frozen accident pushing 50: stereochemistry, expansion, and chance in the evolution of the genetic code. Life 7, 22.
Koonin, E.V., Novozhilov, A.S., 2009. Origin and evolution of the genetic code: the universal enigma. IUBMB Life 61, 99–111.
Koonin, E.V., Novozhilov, A.S., 2017. Origin and evolution of the universal genetic code. Annu. Rev. Genet. 51, 45–62.
Krzycki, J.A., 2005. The direct genetic encoding of pyrrolysine. Curr. Opin. Microbiol. 8 (6), 706–712.
Kubyshkin, V., Budisa, N., 2019. The alanine world model for the development of the amino acid repertoire in protein biosynthesis. Int. J. Mol. Sci. 20, 5507.
Kudella, P.W., Tkachenko, A.V., Salditt, A., Maslov, S., Braun, D., 2021. Structured sequences emerge from random pool when replicated by templated ligation. Proceed. Nat. Acad. Sciences USA 118, e2018830118.
Lei, L., Burton, Z.F., 2020. Evolution of life on Earth: trna, Aminoacyl-tRNA synthetases and the genetic code. Life 10, 21.
Li, Y., Breaker, R.R., 1999. Kinetics of RNA degradation by specific base catalysis of transesterification involving the 2'-hydroxyl group. J. Am. Chem. Soc. 121, 5364–5372.
Luu, H.T.L., Karbalaei-Heidari, H.R., Budisa, N., 2025. Essential logic and facts behind the expansion of the genetic code: a critical assessment. ChemCatChem 17 (20).
Mansy, S.S., Schrum, J.P., Krishnamurthy, M., Tobé, S., Treco, D.A., Szostak, J.W., 2008. Template-directed synthesis of a genetic polymer in a model protocell. Nature 454, 122–125.
Margulis, L., 1970. Origin of Eukaryotic Cells. Yale University Press, New Haven, CT.
Margulis, L., 2010. Symbiogenesis. A new principle of evolution rediscovery of boris mikhaylovich kozo-polyansky (1890-1957). Paleontol. J. 44 (12), 1525–1539.

Martin, W., Russell, M.J., 2003. On the origins of cells: a hypothesis for the evolutionary transitions. Philos. Trans. R. Soc. B 358, 59–85.
Massey, S.E., 2016. The neutral emergence of error minimized genetic codes superior to the standard genetic code. J. Theor. Biol. 408, 237–242.
Miller, S.L., 1953. A production of amino acids under possible primitive Earth conditions. Science 117, 528–529.
Müller, F., Escobar, L., Xu, F., Węgrzyn, E., Nainytė, M., Amatov, T., Chan, C.-Y., Pichler, A., Carell, T., 2022. A prebiotically plausible scenario of an RNA-peptide world. Nature 605, 279–284.
Monnard, P.A., Deamer, D.W., 2002. Membrane self-assembly processes: steps toward the first cellular life. Anat. Rec. 268, 196–207.
Moody, E.R.R., Álvarez-Carretero, S., 2024. The nature of the last universal common ancestor and its impact on the early Earth system. Nat. Ecol. Evol. 8, 1654–1666.
Muchowska, K.B., Varma, S.J., Moran, J., 2019. Synthesis and breakdown of universal metabolic precursors promoted by iron. Nature 569, 104–107.
Nesterov-Mueller, A., 2025. The combinatorial fusion Cascade as a neural network. AI 6, 23.
Nesterov-Mueller, A., Popov, R., Seligmann, H., 2021. Combinatorial fusion rules to describe codon assignment in the standard genetic code. Life 11, 4.
Nesterov-Mueller, A., Popov, R., 2021. The combinatorial fusion Cascade to generate the standard genetic code. Life 11, 975.
Novozhilov, A.S., Wolf, Y.I., Koonin, E.V., 2007. Evolution of the genetic code: partial optimization of a random code for robustness. Biol. Direct 2, 24.
Papineau, D., She, Z., Dodd, M.S., Iacoviello, F., Slack, J.F., Hauri, E., Shearing, P., Little, C.T.S., 2022. Metabolically diverse primordial microbial communities in Earth's oldest seafloor-hydrothermal jasper. Sci. Adv. 8 eabm2296.
Patel, B.H., Percivalle, C., Ritson, D.J., Duffy, C.D., Sutherland, J.D., 2015. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. Nat. Chem. 7, 301–307.
Pizzarello, S., Shock, E., 2010. The organic composition of carbonaceous meteorites. Cold Spring Harbor Perspect. Biol. 2, a002105.
Preiner, M., Asche, S., 2019. Catalysts, autocatalysis and the origin of metabolism. Interface Focus 9, 20190072.
Radakovic, A., DasGupta, S., Wright, T.H., Aitken, H.R.M., Szostak, J.W., 2024. A potential role for RNA aminoacylation prior to its role in peptide synthesis. Proceed. Nat. Acad. Sciences USA 121, e2410206121.
Rosenthal, G.A., 1977. The biological effects and mode of action of L-canavanine, a structural analogue of L-arginine. QRB (Q. Rev. Biol.) 52, 155–178.
Rother, M., Krzycki, J.A., 2010. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. Archaea, 2010.
Schreiber, U., Locker-Grütjen, O., Mayer, C., 2012. Hypothesis: origin of life in deep-reaching tectonic faults. Orig. Life Evol. Biosph. 42, 47–54.
Srinivasan, G., James, C.M., Krzycki, J.A., 2002. Pyrrolysine encoded by UAG in archaea. Science 296, 1459–1462.
Stairs, S., Nobile, A., Lane, N., 2018. Life as a guide to prebiotic nucleotide synthesis. Nat. Commun. 9, 5176.
Szathmáry, E., Santos, M., 2018. The evolution of the genetic code: impasses and challenges. Biosystems 164, 217–225.
Trifonov, E.N., 2000. Consensus temporal order of amino acids and evolution of the triplet code. Gene 261, 139–151.
Vetsigian, K., Woese, C., Goldenfeld, N., 2006. Collective evolution and the genetic code. Proceed. Nat. Acad. Sciences USA 103, 10696–10701.
Wang, Z., Becker, H., 2013. Ratios of S, Se and Te in the silicate Earth require a volatile-rich late veneer. Nature 499, 328–331.
Wehbi, S., Wheeler, L.C., Engqvist, M.K.M., Campbell, I.M., Harms, M.J., 2024. Order of amino acid recruitment into the genetic code resolved by LUCA's protein domains. Proceed. Nat. Acad. Sciences USA 121.
Weiss, M.C., Preiner, M., Xavier, J.C., Zimorski, V., Martin, W.F., 2018. The last universal common ancestor between ancient Earth chemistry and the onset of genetics. PLoS Genet. 14, e1007518.
Woese, C.R., 2002. On the evolution of cells. P. Natl. Acad. Sci. USA 99 (13), 8742–8747.
Woese, C.R., Olsen, G.J., Ibba, M., Söll, D., 2000. Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. Microbiol. Mol. Biol. Rev. 64, 202–236.
Wolf, Y.I., Koonin, E.V., 2007. On the origin of the translation system and the genetic code in the RNA world. Biol. Direct 2, 14.
Wong, J.T., 1975. A co-evolution theory of the genetic code. Proceed. Nat. Acad. Sciences USA 72, 1909–1912.
Yarus, M., 2022. A crescendo of competent coding (c3) contains the standard genetic code. RNA 28, 1337–1347.
Yarus, M., 2023. The genetic code assembles via division and fusion, basic cellular events. Life 13, 2069.
Yarus, M., 2024. Ordering events in a developing genetic code. RNA Biol. 21, 256–263.
Yarus, M., Widmann, J.J., Knight, R., 2009. RNA-amino acid binding: a stereochemical era for the genetic code. J. Mol. Evol. 69, 406–429.