

## Ursprung des genetischen Codes

# Die Botschaft von LUCA – der letzte universelle gemeinsame Vorfahre

ALEXANDER NESTEROV-MÜLLER, ROMAN POPOV  
 INSTITUT FÜR MIKROSTRUKTURTECHNIK, KARLSRUHER INSTITUT FÜR TECHNOLOGIE (KIT), EGGENSTEIN-LEOPOLDSHAFEN

**The standard genetic code is the only reliable information about the first living cell on Earth called the last universal common ancestor – LUCA. The central mystery of the genetic code is the assignment of the canonical amino acids to the codons – base triplets. Simple and uniform combinatorial rules indicate that the standard genetic code arose during the fusion of four primordial single base-pair codes.**

DOI: 10.1007/s12268-020-1427-x  
 © Die Autoren 2020

Es wird angenommen, dass das Leben auf der Erde vor etwa vier Milliarden Jahren aus einer hypothetischen ersten Zelle Namens LUCA entstanden ist. LUCA steht für *last universal common ancestor* – der letzte universelle gemeinsame Vorfahre. Seit Jahrzehnten versuchen die Forscher, die Natur der ersten Zelle mithilfe verschiedener Ansätze zu verstehen. Beispielsweise wird die genetische Ähnlichkeit der direkten Nachkommen von LUCA – Archaeen, Bakterien und Eukaryoten – analysiert oder der mögliche Metabolismus von LUCA unter dem Gesichtspunkt ursprünglicher geochemischer Prozesse rekonstruiert [1]. Obwohl das theoretische Konstrukt LUCA immer mehr reale Umrisse annimmt, scheint das zentrale Rätsel von LUCA um den Ursprung seines genetischen Codes unlösbar zu sein.

Der LUCA-Code ordnet einer in den Proteinen vorkommenden Aminosäure jeweils ein Triplet oder mehrere Triplets von drei aufeinanderfolgenden Nukleobasen zu. Diese Triplets heißen Codons. Sie sind bei allen Lebewesen einschließlich des Menschen (mit seltenen Ausnahmen) gleich. Der Ursprung dieser Zuordnung wurde aus verschiedenen Perspektiven betrachtet. Francis Crick, der Entdecker der DNA-Doppelhelix, betrachtete den Standardcode als *frozen accident* (eingefrorenen Zufall). Eine andere populäre Theorie erklärt die Zuordnung der Codons durch die natürliche Auslese, die auf die Optimierung der Stabilität bei der Translation von

Proteinen aus RNA-Molekülen ausgerichtet war. Diese Ansätze fanden jedoch keine überzeugenden Beweise. Das Rätsel des genetischen Codes wurde noch komplexer, als zwei nicht-kanonische, proteinogene Aminosäuren – Selenocystein (Sec) und Pyrrolysin (Pyl) – entdeckt wurden.

Trotzdem legen neuere Erkenntnisse nahe, dass das Auftreten von LUCA einfachen kombinatorischen Regeln unterlag [2]. Diese Regeln ergeben sich zwangsläufig aus der Annahme, dass der genetische Code durch Fusion von vier Einzelbasenpaar-Urcodes entstanden ist (Tab. 1). Zwei dieser Einzelbasenpaar-Urcodes sind dominant, da sie ihre ursprüngliche Codon-Zuordnung nach der Fusion behielten. Die Änderungen der Urcodes infolge der Fusion sind in Tabelle 1 rot markiert. Diese sind einfach und einheitlich: 1) In der dritten Position sind die Basen **A** und **G** sowie die Basen **U** und **C** austauschbar; 2) Die Basen in der zweiten Position bleiben unverändert; 3) In der ersten Position tritt der antisymmetrische Austausch der Basen zwischen **A** und **C** sowie **U** und **G** auf. Diese Regeln erklären die Verteilung von Aminosäuren über die Codons.

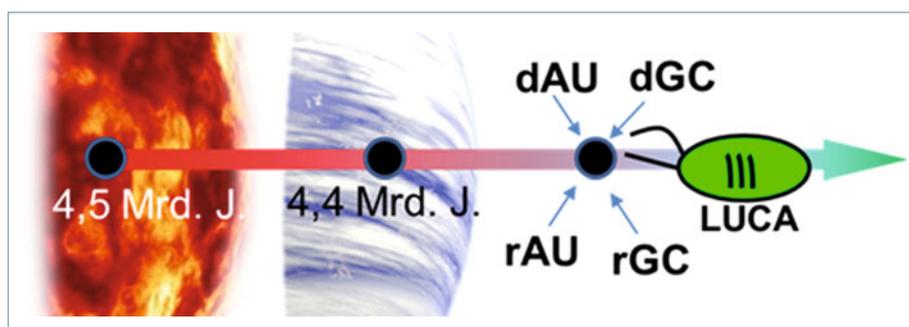
Die berühmte Geschichte von Copernicus lehrt uns, einfachen Lösungen zu vertrauen, wenn sie zur Beschreibung der Realität beitragen. Aber worauf basieren diese kombinatorischen Regeln? Sind sie die eigentliche Botschaft von LUCA an seine vielen Nachkommen? Können rezessive und dominante Codes mit identischen Basen koexistieren?

Die moderne Biotechnologie bejaht diese Frage. Kürzlich wurde ein vollständig orthogonales System zur Proteinsynthese in Bakterienzellen von einem Forscherteam um Alexander Mankin demonstriert [3]. In diesem Fall wurde die orthogonale Proteinsynthese durch das modifizierte Ribosom durchgeführt, wobei beide seiner Untereinheiten durch dasselbe RNA-Molekül miteinander verbunden waren. Die Entwicklung solcher Systeme verspricht interessante Möglichkeiten für die rekombinante Produktion von Proteinen, ohne den Hauptapparat der Zelle zu stören. Es ist also wahrscheinlich, dass die Vorläufer von LUCA bis zu ihrer physischen Fusion zunächst als orthogonale AU- und GC-Einzelbasencodes koexistierten (Abb. 1).

Die kombinatorischen Regeln von LUCA öffnen den Schleier über den Ursprung von Stopcodons, bei denen die Translation der Peptidkette durch das Ribosom unterbrochen wird. Es stellt sich heraus, dass die Stopcodons das Ergebnis des Ausschlusses nicht-kanonischer Aminosäuren aus dem genetischen Code sind [2]. Beispielsweise wird

Tab. 1: Kombinatorische Fusion des genetischen Standardcodes

Dominante Codes					
Aminosäure	AU-Urcode	Standardcode	Aminosäure	GC-Urcode	Standardcode
Lys	AAA	AAA, AAG	Gly	GGG	GGG, GGA
Asn	AAU	AAU, AAC	Gly	GGC	GGC, GGU
Ile	AUU	AUU, AUC	Ala	GCC	GCC, GCU
Phe	UUU	UUU, UUC	Ala	GCG	GCG, GCA
Leu	UUA	UUA, UUG	Pro	CCC	CCC, CCU
Tyr	UAU	UAU, UAC	Pro	CCG	CCG, CCA
Ile (Met)	AUA	AUA, AUG	Arg	CGC	CGC, CGU
Stop (Pyl)	UAA	UAA, UAG	Arg	CGG	CGG, CGA
Rezessive Codes					
Aminosäure	AU-Urcode	Standardcode	Aminosäure	GC-Urcode	Standardcode
Gln	AAA	CAA, CAG	Trp, Stop (Sec)	GGG	UGG, UGA
His	AAU	CAU, CAC	Cys	GGC	UGC, UGU
Leu	AUU	CUU, CUC	Ser	GCC	UCC, UCU
Val	UUU	GUU, GUC	Ser	GCG	UCG, UCA
Val	UUA	GUA, GUG	Thr	CCC	ACC, ACU
Asp	UAU	GAU, GAC	Thr	CCG	ACG, ACA
Leu	AUA	CUA, CUG	Ser (X1)	CGC	AGC, AGU
Glu	UAA	GAA, GAG	Arg (X2)	CGG	AGG, AGA



▲ **Abb. 1:** Der Ursprung des letzten universellen gemeinsamen Vorfahren LUCA als Fusion der dominanten und rezessiven Einzelbasenpaar-Codes. Entstehung der Erde vor 4,5 Mrd. Jahren, Entstehung der Ozeane vor 4,4 Mrd. Jahren, Entstehung von LUCA (grün) vor ca. 4 Mrd. Jahren. Die Buchstaben d und r stehen für „dominant“ und „rezessiv“.

- [2] Nesterov-Mueller A, Popov R (2020) On the origin of the standard genetic code as a fusion of prebiotic single-base-pair codes. Preprint, arXiv:2004.09230
- [3] Aleksashin NA, Szal T, d'Aquino AE et al. (2020) A fully orthogonal system for protein synthesis in bacterial cells. *Nat Commun* 11:1858
- [4] Byun BJ, Kang YK (2011) Conformational Preferences and pK(a) value of selenocysteine residue. *Biopolymers* 95:345–353
- [5] Jukes TH, Osawa S (1990) The genetic code in mitochondria and chloroplasts. *Experientia* 46:1117–1126
- [6] Gardini S, Cheli S, Baroni S et al. (2016) On nature's strategy for assigning genetic code multiplicity. *Plos One* 11:e0148174
- [7] Emmert EAB, Milner JL, Lee JC et al. (1998) Effect of canavanine from alfalfa seeds on the population biology of *Bacillus cereus*. *Appl Environ Microb* 64:4683–4688
- [8] Kamo T, Sakurai S, Yamanashi T et al. (2015) Cyanamide is biosynthesized from L-canavanine in plants. *Sci Rep* 5:10527

**Funding** Open Access funding provided by Projekt DEAL.

**Open Access** Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden. Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen. Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

#### Korrespondenzadresse:

apl. Prof. Dr. A. Nesterov-Müller  
Institut für Mikrostrukturtechnik  
Karlsruher Institut für Technologie  
Hermann-von-Helmholtz-Platz 1  
D-76344 Eggenstein-Leopoldshafen  
Alexander.Nesterov-Mueller@kit.edu

#### AUTOREN



##### Alexander Nesterov-Müller

Physikstudium an der Lomonosov Universität Moskau, Russland. 2003–2006 Promotion an der Universität Heidelberg. 2007–2010 Postdoc am Deutschen Krebsforschungszentrum (DKFZ), Heidelberg. Seit 2011 Forschungsgruppenleiter und apl. Professor an der Fakultät für Maschinenbau am Karlsruher Institut für Technologie.



##### Roman Popov

Masterstudium Physikalische Chemie an der Nationalen Forschungsuniversität Samara, Russland, weiteres Studium an der Karlsruhe School of Optics and Photonics, Karlsruher Institut für Technologie (KIT). 2015–2018 Promotion in Ingenieurwissenschaften am KIT. Seit 2018 Postdoc am KIT.

Pyrrolysin (Pyl) mit dem gleichen Codon (UAG) wie das Stopcodon codiert. Gemäß **Tabelle 1** war Pyrrolysin im dominanten AU-Urcode mit seinem ursprünglichen Codon (UAA) vorhanden. Nach Ausschluss von Pyrrolysin aus dem Urcode und Fusion wurden beide Codons (UAA und UAG) zu Stopcodons. Der Grund für das Verschwinden von Pyrrolysin könnte ein Mangel an seinem Vorläufer Lysin (Lys) sein, der aufgrund seiner intensiven Verwendung für die Proteinsynthese für weitere Reaktionen im selben Code nicht mehr verblieb.

Eine weitere aus dem rezessiven GC-Urcode verschwundene Aminosäure ist Selenocystein (Sec) (**Tab. 1**). Selenocystein ließ Platz für die zwei Stopcodons UGA und UGG. Grund für das Verschwinden von Selenocystein könnte seine hohe Reaktivität sein, die es ihm nicht erlaubt, in freier Form in Zellen zu verbleiben [4]. Nach der Fusion wurde das UGG-Codon von Tryptophan (Trp) besetzt. Da Tryptophan nach der Code-Fusion als neue Aminosäure in den genetischen Standardcode aufgenommen wurde, ist sie die seltenste Aminosäure in Wildtyp-*Escherichia coli*. Es ist bemerkenswert, dass das Trp-Codon UGG weiterhin die Stopcodon-Funktion in den Mitochondrien aller bisher untersuchten Organismen ausübt [5].

Die rezessiven Codons von Arginin (AGG und AGA) wirken als Stopcodons in den Mitochondrien von Säugetieren und *Drosophila* [5]. Tatsächlich gehörten diese Codons zu der ursprünglichen nicht-kanonischen Aminosäure X2 mit ähnlichen Eigenschaften wie Arginin. Diese Substitution erklärt die Tatsache, dass Arginin die maximale Anzahl von Codons hat, gleichzeitig aber eine durchschnittliche Häufigkeit in Proteinen aufweist [6]. Vermutlich verdrängte Arginin sein Guanidinoxy-Analogon – die schwächere

Base Canavanin. Canavanin kommt in einigen Hülsenfrüchten vor und ist für eine Reihe von Organismen toxisch [7]. Es scheint mit dem Translationssystem lebender Organismen sehr gut vertraut zu sein und integriert sich anstelle von Arginin leicht in Proteine. Seine Rolle bei Mikroorganismen, insbesondere bei der Stickstofffixierung, ist noch lange nicht verstanden [8].

Die Entstehung von LUCA ist für das Design neuer molekularer Gerüste und makromolekularer Enzyme aus Peptiden und RNA mithilfe der Aminosäuren aus den einzelnen Urcodes interessant. Zum Beispiel hatte jeder der vier Urcodes von LUCA seine eigene positiv geladene Aminosäure – Lysin, Histidin, Arginin und Canavanin (X2). Möglicherweise war dies notwendig, um eine gute Adhäsion an das negativ geladene RNA-Rückgrat sicherzustellen. In diesem Fall waren die zwei einzigen negativ geladenen Aminosäuren (Asparaginsäure und Glutaminsäure) nur im rezessiven AU-Urcode präsent.

Die Menschheit hört bekanntlich nicht auf, von der Besiedlung anderer Planeten zu träumen. Das führt unvermeidlich zur Frage nach den synthetischen Mikroorganismen, welche für die Besiedlung der neuen Umgebung am besten geeignet sind. Die Botschaft von LUCA kann hier helfen. Seine Anpassung an die Umweltbedingungen auf der Erde begann bereits auf der Ebene von vier koexistierenden Urcodes.

#### Danksagung

Die Autoren danken Friederike Götz für die Korrekturen. Für die finanzielle Unterstützung danken wir der DFG (AOBJ 655892). ■

#### Literatur

- [1] Weiss MC, Preiner M, Xavier JC et al. (2018) The last universal common ancestor between ancient Earth chemistry and the onset of genetics. *Plos Genet* 14:e1007518