Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Anticipating alien cells with alternative genetic codes: away from the alanine world!

Vladimir Kubyshkin[1,2] and Nediljko Budisa[1,2]

Can we make life with a different genetic amino acid repertoire? Can we expect organisms which would keep newly given genetic code associations permanently? To address these questions, we would like to analyze the existent genetic code amino acid repertoire as formed from derivatives of alanine. Derivation from alanine leads to the α-helix based biological world, the *Alanine World*, whereas variations in the side-chains enable tertiary folding and subsequent chemical versatility of the proteome. Proline, glycine and pyrrolysine are the rudiments in the current genetic code, indicating that the original set could be different. Furthermore, from the perspective of peptide chemistry, it shall be possible to recruit these alternative scaffolds for the construction of synthetic or alternative life. This would allow for a completely new biological world, potentially as functional and versatile as the existing one. Pursuing these options offers a strategy for a complete re-design or even *de-novo* creation of living organisms based on entirely different chemical make-up, with completely new set of solutions for both near and distant future biotechnologies.

Addresses
[1] Institute of Chemistry, Technical University of Berlin, Müller-Breslau-Str. 10, Berlin 10623, Germany
[2] Department of Chemistry, University of Manitoba, Dysart Rd .144, Winnipeg R3T 2N2, Manitoba, Canada

Corresponding authors: Kubyshkin, Vladimir (vladimir.kubyshkin@umanitoba.ca), Budisa, Nediljko (nediljko.budisa@tu-berlin.de, nediljko.budisa@umanitoba.ca)

Current Opinion in Biotechnology 2019, **60**:242–249

This review comes from a themed issue on **Chemical biotechnology**

Edited by **Sven Panke** and **Thomas Ward**

For a complete overview see the Issue and the Editorial

Available online 3rd July 2019

https://doi.org/10.1016/j.copbio.2019.05.006

> "We are about to win chemical influence on the design of the organism, and this should lead to the strangest phenomena, to changes in shape, which leave everything behind, what has been achieved by breeding and crossing."
>
> Emil Fischer, 1890 [1].

Introduction

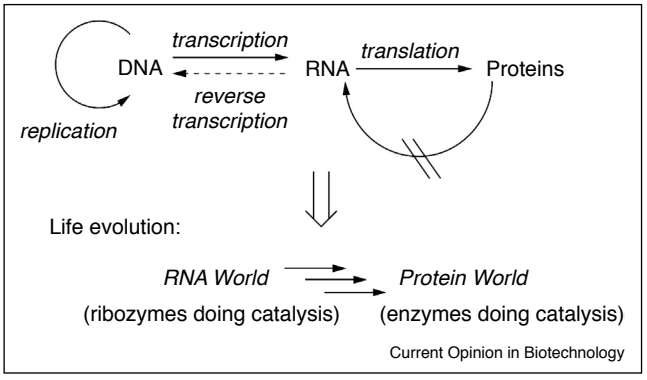
Life on Earth is a reservoir of evolutionary and adaptive innovations, which accumulated over time. In order to understand, manipulate and engineer it, we need to formulate the basic principles governing it, and rationalize relationships between the principal components of life biochemistry. We are particularly interested in the chemical identity of living systems, represented by the basic chemical composition of the biochemical components. Among these, the main focus of attention is given to the biopolymeric scaffolds, nucleic acids and polypeptides. The transition between these two biopolymeric levels or 'worlds' occurs according to the central dogma of molecular biology (Figure 1), which has been formulated as follows:

> "The Central Dogma states that once 'information' has passed into protein it cannot get out again. In more detail, the transfer of information from nucleic acid to nucleic acid or from nucleic acid to protein may be possible, but transfer from protein to protein, or from protein to nucleic acid is impossible" [2].

On the basis of the Central Dogma, the nucleic acids are mainly viewed as informational polymers, whereas the interpretation of this information occurs at the level of the proteins, where amino acid sequences fold into functional protein bodies. The hierarchy of the biopolymer scaffolds immediately suggests that the manipulation with the life systems can be performed on two levels: nucleic acids (genes) and proteins. Manipulation of the informational elements, genes, is a particular goal of synthetic biology, as a method towards new biological species [3]. The exchange and spread of the genetic information also occur naturally in the course of horizontal and vertical gene transfers, which effectively connects the living species on Earth into a large communication network [4]. Transfer of the genetic information between species is an important mechanism, which allows biological species to gain new functions and innovations from across the biosphere. Nonetheless, this mechanism is based on the common set of chemical building block elements, a universal set of chemical bases: polymeric scaffolds, nucleobases [5,6], amino acid side chains, and so on. It has been hypothesized that redesign or expansion of this set can provide an alternative way towards chemical innovations, thereby forming new species with an alternative chemical identity [7•].

Over the past few decades, a large progress has been made towards manipulations with the set of 20(+3) canonical

Figure 1



The central dogma of molecular biology states there is solely a unidirectional flow of information in living systems. The RNA World hypothesis logically follows from the central dogma.

(coded) amino acids. Nowadays, it is almost routinely possible to perform a single protein expression *in vivo* with the amino acid building blocks well beyond the canonical ones [8•]. This multitude of possibilities is even greater for *in vitro* protein translation assays, as these allow a relatively free redefinition of the genetic code set [9,10]. However, the currently available methods and approaches to emancipate or re-assign the meaning of codons in an entire living (microbial) organisms are still in the infancy stage of development. Redesign of the amino acid composition on the proteome level requires application of a pressure to organisms. The desired pressure is usually created experimentally either by constructing a stringent genetic set-up [11•,12•,13•] or through long-term cultivation in gradually changing supply medium (so-called adaptive laboratory evolution) [14••], or both. In this way, aromatic amino acids such as *para*-acetylphenyl-alanine or biphenyl-alanine or thienopyrrolyl-alanine have been appended to the canonical amino acid repertoire in genetically modified *Escherichia coli* organisms.

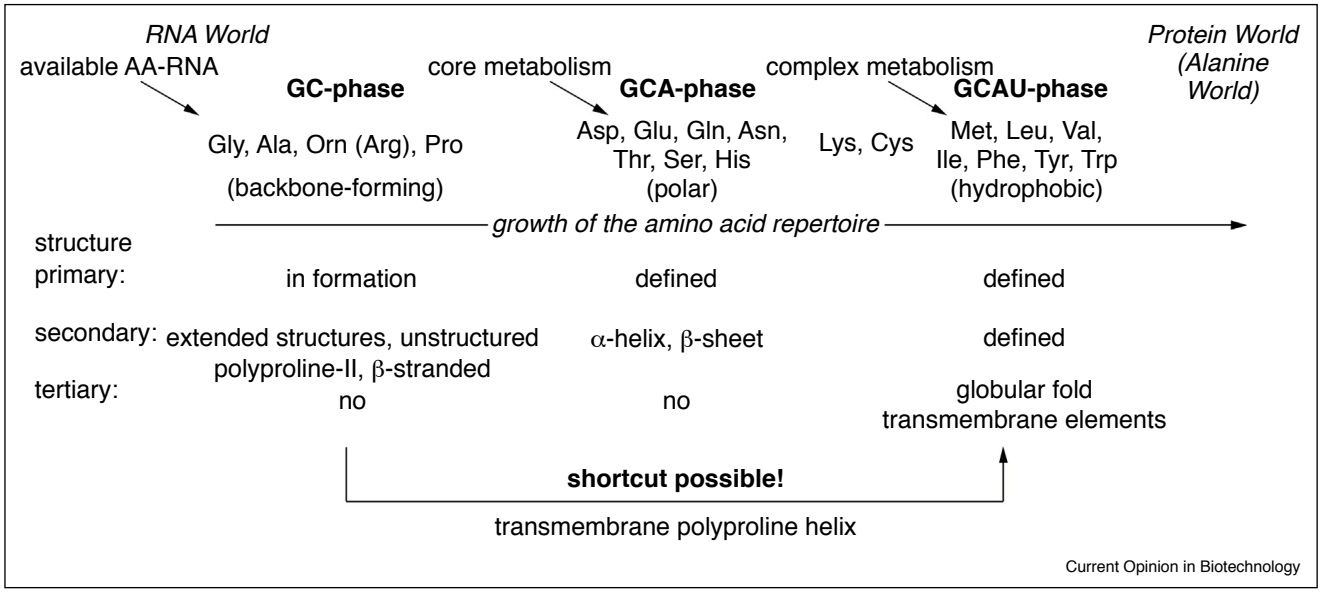
Although these achievements are impressive, we nevertheless believe that they need to be critically scrutinized, not by questioning the rigor of the experimental genetic setup, but the concept itself. The main efforts in this research field are aimed at generating genetically encoded side-chain modifications with minimal disturbance in local microenvironments for interesting academic questions or simple technological solutions [15,16]. However, in order to provide a solid basis for the design of synthetic life based on radically different chemical makeup, the experiments should be designed so that a newly introduced amino acid should provide an advantage to the synthetic cells. Ideally, they should provide/endow them with new chemical functions and processes accompanied with a complete redesign of the proteome architecture. To this end, we would like to offer a simple retrospective view on the genetic code, and formulate a strategic alternative to the existing canonical amino acid repertoire. At the core of our argument, we want to emphasize that most of the canonical amino acids are derivatives of alanine, and in this sense, we are living in the *Alanine World*. We believe that alternative potential scenarios, *Worlds*, can provide strategic solutions for the biotechnology of the distant future, enabling entirely alienated life in a test tube.

Molecular evolution and development of the genetic code

Let us make a little journey back in the history. The emergence of protein biosynthesis and the parallel development of the genetic code was the key event in evolution of life as we know it. This is due to the fact that the genetic code assigns nucleic acids with their amino acid 'meanings', thus enabling genetic development, and this is also trackable by phylogenetic analysis [17•]. The hypothesis of the *RNA World* directly follows from the central dogma [18,19•] (Figure 1). According to this hypothesis, RNA molecules were the functional molecular entities in the early phase of the development of life comprising the *RNA World*, whereas the emergence of protein biosynthesis lead to the outbreak of the *Protein World*. The ancestor molecules to those known as tRNA had associations with particular amino acids which were established beforehand, while the original amino acyl-RNA were either cofactors or metabolic components for biosynthetic purposes [20,21].

Already following from this simple picture, one can imagine that few amino acids were inevitably present in the RNA attached form, and these were available for use in translation. Glycine and aspartic acid bear carbon-skeletons, which are metabolic precursors for nucleobases, serine is a metabolic precursor of glycine and pteridine-bound C1-units (which were also available through ancient C1-metabolism), alanine is a simplest carrier of nitrogen required for nucleobase synthesis and so on. Thus, one can find simple and easy metabolic justification for the biochemical occurrence of amino acids before the protein biosynthesis has started. Perhaps, one of the weirdest initial components was proline, which is a cyclic amino acid, devoid of side-chain functional groups. However, one should not be distracted by the simplicity of this structure. Proline, in fact, may be attributed a very special role, since its amine-functional group is an efficient catalyst in reactions known as condensation reactions. Catalyzed condensations along with transaldolase/ketolase reactions are essential sources for sugars, the backbone entities of RNA biopolymers. Thus, a primitive Pro-tRNA ester could have served the role of a catalytic center; otherwise this was a part of a catalytic cascade for sugar metabolism [22]. Subsequently, the synthesis of proline would require its metabolic precursors, either glutamic acid or ornithine, the latter is also direct precursor for arginine, a component of the existing genetic code. Thereby, proline and its precursors may have been available before the advent of protein biogenesis.

Figure 2



The GC-GCA-GCAU-scheme of the genetic code development as summarized by Hartman and Smith with chemical interpretation. The polypeptide structure complexity and the amino acid metabolic complexity increase from earlier to later phases.

Following the *Frozen Accident* theory of Crick [23], protein biosynthesis started with an original set of amino acids and was expanding its repertoire until the point when further recruitment would generate too many detrimental problems. Although, there are different theories, hypotheses and opinions on how the amino acid repertoire was expanding, we are particularly attracted by the concept proposed recently by Hartman and Smith [24•,25,26•]. According to the proposed scenario, the original set was restricted only to codons containing GC letters. Further expansion of the repertoire required recruitment of additional nucleobases, and subsequently the letter A was recruited, followed by the U letter.

The elegance of the GC-GCA-GCAU scheme is that it can be easily correlated to the hierarchy of protein folding [27] (Figure 2) as well as with the metabolic significances of the initial amino acids listed above. The original amino acid set coded by GC contained glycine, alanine, proline, and one cationic amino acid (now: arginine), the amino acids needed for the nucleotide biosynthesis. From the folding perspective though, polypeptides based on this set would be dominated by extended and relatively low stable structures, and subsequently would be maily disordered. However, these would already be able to adhere to polyanionic RNA molecules due to the positive net charge provided by the cationic amino acids. In the next phase, the addition of the A-letter allowed the acquisition of a number of polar amino acids. The GCA-phase enabled the formation of the α-helix, the most important secondary structure motif in common biochemistry. Formation of tertiary structure and the membrane interaction was made possible in the next GCAU-phase, after the addition of the U letter allowed recruitment of hydrophobic amino acids. The late addition of hydrophobic amino acids can be proposed to arise from the high metabolic complexity and costs of their biosynthesis. Thus, the recruitment of the RNA letters built the coding space of mRNA, while the amino acid counterparts were acquired from metabolic sources. Coevolution of these two components built up the genetic code [28].

The sequence of amino acid recruitment correlates well with the complexity of their metabolic synthesis: while the first wave of evolutionary genetic code expansion (GCA phase) has acquired the amino acids that are only a few steps away from the core metabolism, the acquisition of the amino acids in the GCAU expansion wave required development of more complex biosynthetic pathways. From this, it can be speculated that the metabolic availability of the amino acids played a key role in the development of protein biogenesis. This statement can be explained by the fact that the genetic code contains tyrosine (*para*-hydroxyphenylalanine), but does not contain γ-hydroxyproline, which is just as abundant in modern biochemistry. Simple metabolic considerations suggest that hydroxylation in the γ-position in the proline residue occurs in nature by oxidation with molecular oxygen [29], a molecular species absent in the genetic code formation phase [30]. At the same time, tyrosine biosynthesis does not require oxidation with oxygen, as this amino acid can be derived directly from prephenate, a common precursor for tyrosine and phenylalanine.

The *Alanine World* and its alternatives

From the perspective of peptide science, the set of the first amino acids, glycine, proline, alanine and one cationic species should generate a rather undefined polypeptide backbone folding, dominated by extended conformations. However, from this point further expansion of the genetic code repertoire went solely into the direction of structural derivatives of alanine. Not a single non-alanine based amino acid was recruited into the protein biosynthesis after the adoption of alanine as the preferred motif. For example, a phenyl-group containing amino acid in the genetic code repertoire is phenylalanine, and not phenylglycine, phenylproline or phenyllysine. From this point of view, the existent *Protein World* should rather be called the *Alanine World*. Very important feature of alanine is the fact that this amino acid residue exhibits the greatest α-helical propensity [31,32].

We thus would like to outline the following key attributes of the *Alanine World*: 1) the polypeptide structure is chiral as follows from chirality of the L-building blocks; 2) the amino acid building blocks have a backbone (alanine) part and the side-chain function; therefore, point mutations usually do not impact the backbone fold, but change/alter the chemical function; 3) the chemical function is close to the backbone; therefore, accumulation of the mutations can impact the secondary structure; 4) the backbone is capable of donating and accepting the hydrogen bond; therefore, the proteome is dominated by the hydrogen-bond based structures; among them α-helix is the most common. These attributes provide an empirical basis for any experiment in protein engineering. However, one should clearly realize they are essentially derived from the fact that the most canonical amino acids are structurally derived from alanine. For example, 'alanine scan' is the common approach in biochemical science that fully relies on attribute 2), and it often fails when approaching glycine or proline, since these amino acids do not share same alanine-based backbone architecture. The *Alanine World* features are so common, that they usually remain unnoticed unless these are addressed in the frame of peptide studies with radically different chemical alternatives.

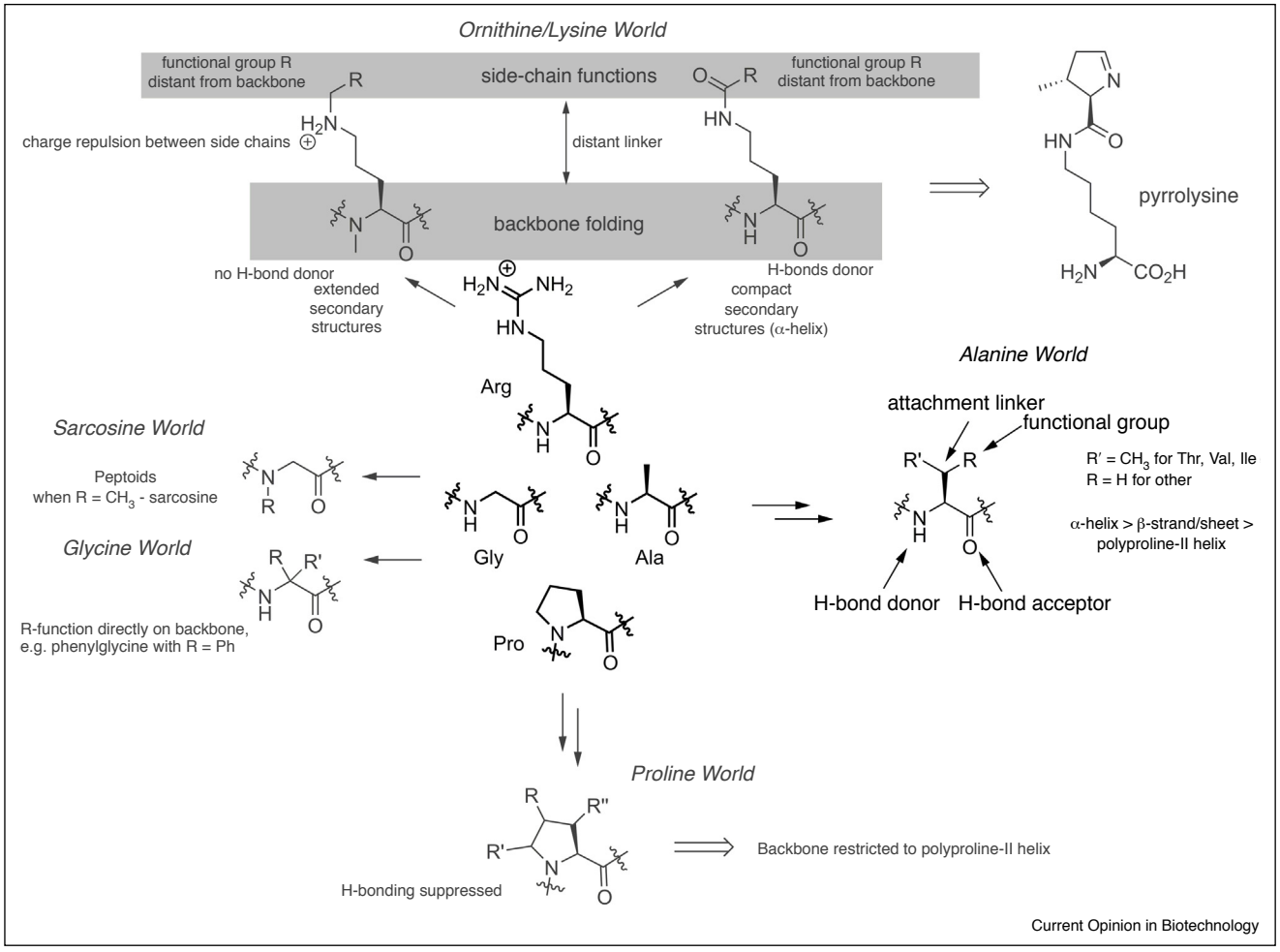
Nonetheless, from the chemical standpoint, the *Alanine World* is not the only way in which a peptide scaffold can be decorated with a rich number of functional elements. Each of the starting amino acids can propose its own alternative development of the protein chemistry as schematically illustrated on Figure 3. Especially rich is proline-based peptide chemistry. Substitutions based on the core structure of proline lead to scaffolds able to adopt extended polyproline-II helix, which is also a generic secondary structure allowed for other α-amino acids [33]. In the *Alanine World* the polyproline-II extended helix is obscured due to the competition with more stable hydrogen-bonded structures. However, this is one of the dominant structures in the so-called 'disordered' or 'denatured' state [34], as well as in the initial proteome in the GC-phase (Figure 2). Construction of polyproline-II structures based on proline analogues creates a rather stable secondary fold. Polyproline-II folded peptides have already demonstrated their ability to interact with nucleic acids [35•], secondary messengers [36], and membranes [37,38]. Conversely, assembly of polyproline-II helices into collagen triple helix has been explored by nature, and this is triggered by post-translational hydroxylation of proline residues in procollagen [39]. Alternatively, the assembly of polyproline helices into bundles has been recently described for antifreeze proteins [40,41].

Until very recently it was not clear whether the polyproline-II helix could feature in hydrophobic sequences, since the extended nature of this structure usually favors water solvation. Kubyshkin and Budisa [42,43] and others [44,45] demonstrated with the help of some proline analogues that absence of hydrophobic polyproline-II helices in living nature is not caused by fundamental limitation of the structure itself. Moreover, we recently showed that polyproline-II helix is capable of forming transmembrane elements [46••]. Thus, we created a shortcut in the Hartman-Smith scheme, by making an artificial transmembrane element bypassing development of the α-helix (Figure 2). Recently, it has also been shown that incorporation of the polyproline helices into the collagen superstructures is fully compatible with the hydrophobic environment [47•]. These solid experimental facts allow us to conclude that there are no fundamental limitations that would preclude proline and the polyproline-II helix from becoming a competent life-building constituent. Thus, the *Proline World* is a fully conceivable option, which was neglected by nature most likely due to a number of limitations arising from metabolic schemes in existing living cells.

Further analysis demonstrates that other amino acids from the initial GC-phase set could also give rise each to its own set of other chemical solutions. This can be illustrated by the placement of the functional group in the glycine backbone bypassing the 1-carbon atom linker, for example in phenylglycine (*Glycine World*). Alternatively, a placement of a functional substituent on the nitrogen rather than Cα-atom generates structures called peptoids (*Sarcosine World*), the structures that are *per se* non-chiral (Figure 3). Although, peptoids do not usually demonstrate a defined secondary fold [48], they are prone to forming multiple *cis-trans* isomers, with tertiary fold that could potentially stabilize their conformations. Interaction of hydrophobic peptoids with membranes is also evident [49], although, no defined transmembrane peptoid has been reported, despite efforts [50].

Other alternatives could be a placement of a substituent on the distant δ-amino or ε-amino group of ornithine or

Figure 3



Some suggestions for possibilities in alternative developments of the proteome based on the initial GC-phase available amino acid set (glycine, proline, alanine and cationic amino acid).

lysine that give rise to the *Ornithine* or *Lysine World*. The idea behind the *Ornithine/Lysine World* could be uncoupling of the backbone secondary structure from the functional groups, by using a sufficiently distant linker, longer than a single methylene unit. As the result, the development of the secondary structure can be uncoupled from the chemical features of the side chain functions, and proceed in parallel. Occurrence of pyrrolysine (Figure 3), a special canonical amino acid encoded in some methanogens highlights significance of this scenario.

There are certainly more available options, which do not have rudimentary traces in the genetic code amino acid repertoire. For example, α,α-dialkyl structures (derivatives of aminoisobutyric acid) form peptide structures called peptaibols, which feature a set of secondary folds including α-helix, 3_{10}-helix and very unusual fully extended 2.0_5-helix, which is not represented in natural proteomes [51,52]. In fact, modern peptide and foldamer research allows to propose a notable number of potential backbone carriers for geometric arrangement of the biochemical functional groups; among these proline, glycine and ornithine are taken here as examples because these are preserved in the modern genetic code.

Away from the *Alanine World*!

Our analysis of the evolutionary development of the amino acid repertoire allows us to speculate that the selection of the amino acid structures was dictated by their metabolic availability in the GCA-phase, whereas in the GCAU-phase the hydrophobic motif was appended to the already established α-helix-based architecture. This is what has led to the *Alanine World* with biochemistry dominated by the α-helix and other features of the common protein architecture, as outlined above. What if the set of chemical options offered by the *Alanine World* had already been explored in the course of life's

evolution? A positive answer would mean that our attempts to add new side-chain functionalities without redesigning the core structure of the building blocks are not radical changes, rather mere variations within the *Alanine World* [53].

We should therefore be able to consider a number of strategic chemical alternatives, which should enable redesign of the genetic code repertoire without compromising chemical versatility and cellular functionality. One possible approach to accomplish this task could be to establish a stable self-sustainable system with all integrated functions based on a different type of the underlying chemical skeleton (secondary structure). Exchange of the core amino acid structure (e.g. *Proline World* or *Sarcosine World*), uncoupling of the secondary fold from the side-chain propensities (*Ornithine World*) or other options inspired from peptide and foldamer studies could open avenues for a complete redesign of protein folding, and create or evolve a completely different form of life based on these building blocks. Fortunately, modern chemical synthesis provides many sources for the amino acid analogues and related structures required for such developments. It is thus relatively easy to supply model biochemical systems with a number of man-made amino acids, thereby mimicking their metabolic availability in the cells.

We thus suggest that alternative synthetic life forms can be constructed along with this path. Thereby, we are about to change the basic chemical implementation of life but not the fundamental principles on which life is built such as defined by, for example, Gánti 'Chemoton Model' [54]. We anticipate experiments within both approaches that currently dominate synthetic biology and xenobiology, so-called top-down and bottom-up [55]. The top-down approach is usually based on adaptive laboratory evolution protocols to alienate the current genetic code, starting with building blocks that are similar to canonical ones. Over the course of generations of evolutionary adaptation and metabolic rearrangements [56], these building-block structures could be further diversified until they are completely different from those present in the original ancestral cells. Another potential direction, the bottom-up approach, should enable *de novo* design by using simple boundary systems such as artificial vesicles or compartments [57], which should evolve into a truly alternative life from scratch. This should be a life with radically different chemistries and genetic codes from the beginning — a life we do not know yet.

Finally, we believe that escaping the *Alanine World* is a very complex and rather long-term goal. However, once achieved, this will enable us to answer many fundamental questions about the origins and limits of life, at the same time providing entirely unique and innovative biotechnological solutions for medicinal chemistry and material science. In addition, dependence on noncanonical amino acid components and in general of synthetic building blocks or precursors could be considered as the built-in biological safety tool in resulting alien organisms, thus enabling complete biocontainment that is parallel life exists in complete genetic isolation from the 'old' biological world [58].

Conflict of interest statement

Nothing declared.

Acknowledgements

VK acknowledges the DFG research group 1805 for a post-doctoral position. NB acknowledges the Award for Tier 1 Canada Research Chair in Chemical Synthetic Biology by the Canadian Federal Government.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Fischer E: In *Emil Fischer - Gesammelte Werke.* Edited by Bergmann M. Springer; 1915:796-809.

2. Crick FHC: **On protein synthesis**. *Symposia of the Society for Experimental Biology, Number XII: The Biological Replication of Macromolecules*. Cambridge University Press; 1958:138-163.

3. Acevedo-Rocha C, Budisa N: **Xenomicrobiology: a roadmap for genetic code engineering**. *Microb Biotechnol* 2016, **9**:666-676 http://dx.doi.org/10.1111/1751-7915.12398.

4. Kubyshkin V, Acevedo-Rocha C, Budisa N: **On universal coding events in protein biogenesis**. *Biosystems* 2018, **164**:16-25 http://dx.doi.org/10.1016/j.biosystems.2017.10.004.

5. Dien VT, Morris SE, Karadeema RJ, Romesberg FE: **Expansion of the genetic code via expansion of the genetic alphabet**. *Curr Opin Chem Biol* 2018, **46**:196-202 http://dx.doi.org/10.1016/j.cbpa.2018.08.009.

6. Houlihan G, Arangundy-Franklin S, Holliger P: **Engineering and application of polymerases for synthetic genetics**. *Curr Opin Biotechnol* 2017, **48**:168-179 http://dx.doi.org/10.1016/j.copbio.2017.04.004.

7. • Marliere P: **The farther, the safer: a manifesto for securely navigating synthetic species away from the old living world**. *Syst Synth Biol* 2009, **3**:77-84 http://dx.doi.org/10.1007/s11693-009-9040-9.
This is one of the first papers that describes a fundamental concept of biocontainment. Biocontainment arises when a genetically or chemically altered organism can operate under a different genetic code, either by codon reassignment or by changing the decoding rules.

8. • Hoesl MG, Budisa N: **Recent advances in genetic code engineering in *Escherichia coli***. *Curr Opin Biotechnol* 2012, **23**:751-757 http://dx.doi.org/10.1016/j.copbio.2011.12.027.
The review provides a balanced view of currently used methods in the field of *in vivo* protein labelling with noncanonical amino acids.

9. Forster AC, Tan Z, Nalam MNL, Lin H, Qu H, Cornish VW, Blacklow SC: **Programming peptidomimetic syntheses by translating genetic codes designed *de novo***. *Proc Natl Acad Sci U S A* 2003, **100**:6353-6357 http://dx.doi.org/10.1073/pnas.1132122100.

10. Tajima K, Katoh T, Suga H: **Genetic code expansion via integration of redundant amino acid assignment by finely tuning tRNA pools**. *Curr Opin Chem Biol* 2018, **46**:212-218 http://dx.doi.org/10.1016/j.cbpa.2018.07.010.

11. • Mandell DJ, Lajoie MJ, Mee MT, Takeuchi R, Kuznetsov G, Norville JE, Gregg CJ, Stoddart BL, Church GM: **Biocontainment**

of genetically modified organisms by synthetic protein design. *Nature* 2015, **518**:55-60 http://dx.doi.org/10.1038/nature14121.
Adaptation of *E. coli* to depend on exogenously supplied 21st amino acid (*para*-biphenylalanine) introduced in essential genes.

12. Rovner AJ, Haimovich AD, Katz SR, Li Z, Grome MW,
• Gassaway BM, Amiram M, Patel JR, Gallagher RR, Rinehart J, Isaacs FJ: **Recoded organisms engineered to depend on synthetic amino acids**. *Nature* 2015, **518**:89-93 http://dx.doi.org/10.1038/nature14095.
Same as Ref. [11], but the 21st amino acid was *para*-acetylphenylanaline.

13. Xiao H, Nasertorabi F, Choi S-H, Han GW, Reed SA, Stevens RC,
• Schultz PG: **Exploring the potential impact of an expanded genetic code on protein function**. *Proc Natl Acad Sci U S A* 2015, **112**:6961-6966 http://dx.doi.org/10.1073/pnas.1507741112.
Here, the 21st amino acid was introduced in an antibiotic resistance protein, and directed evolution was applied for better performance of the resistance. Thus, the attempt is to make the 21st amino acid advantageous for survival.

14. Hoesl MG, Oehm S, Durkin P, Darmon E, Peil L, Aerni H-R,
•• Rappsilber J, Rinehart J, Leach D, Söll D, Budisa N: **Chemical evolution of a bacterial proteome**. *Angew Chem Int Ed* 2015, **54**:10030-10034 http://dx.doi.org/10.1002/anie.201502868.
The paper reports successful adaptation of *E. coli* to proteome-wide replacement of tryptophan with an analogue (thienopyrrolyl-alanine). However, the change is not permanent, and back-adaptation is possible given that indole is supplied to the cells again.

15. Budisa N: **Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire**. *Angew Chem Int Ed* 2004, **43**:6426-6463 http://dx.doi.org/10.1002/anie.200300646.

16. Wang L, Xie J, Schultz PG: **Expanding the genetic code**. *Annu Rev Biophys Biomol Struct* 2006, **35**:225-249 http://dx.doi.org/10.1146/annurev.biophys.35.101105.121507.

17. Weiss MC, Preiner M, Xavier JC, Zimorski V, Martin WF: **The last**
• **universal common ancestor between ancient earth chemistry and the onset of genetics**. *PLoS Genet* 2018, **14**:e1007518 http://dx.doi.org/10.1371/journal.pgen.1007518.
Recent genetic analysis, which proposes reconstruction of the physiology of the earliest organisms.

18. Leslie OE: **Prebiotic chemistry and the origin of the RNA World**. *Crit Rev Biochem Mol Biol* 2004, **39**:99-123 http://dx.doi.org/10.1080/10409230490460765.

19. Neveu M, Kim H-J, Benner SA: **The "strong" RNA World**
• **hypothesis: fifty years old**. *Astrobiology* 2013, **13** http://dx.doi.org/10.1089/ast.2012.0868.
Recent evolutionary considerations about the RNA world, which also considers possible chemical alternatives of original life development.

20. Szathmáry E: **Coding coenzyme handles: a hypothesis for the origin of the genetic code**. *Proc Natl Acad Sci U S A* 1993, **90**:9916-9920 http://dx.doi.org/10.1073/pnas.90.21.9916.

21. Kun Á, Radványi Á: **The evolution of the genetic code: impasses and challenges**. *Biosystems* 2018, **164**:217-225 http://dx.doi.org/10.1016/j.biosystems.2017.10.006.

22. Hein JE, Blackmond DG: **On the origin of single chirality of amino acids and sugars in biogenesis**. *Acc Chem Res* 2012, **45**:2045-2054 http://dx.doi.org/10.1021/ar200316n.

23. Söll D, RajBhandary UL: **The genetic code – thawing the 'frozen accident'**. *J Biosci* 2006, **31**:459-463 http://dx.doi.org/10.1007/BF02705185.

24. Hartman H, Smith T: **The evolution of the ribosome and the**
• **genetic code**. *Life* 2014, **4**:227-249 http://dx.doi.org/10.3390/life4020227.
Proposal for the GC-GCA-GCAU scheme of the genetic code development during the RNA-to-Protein World transition at the early stage of life evolution.

25. Smith T, Hartman H: **The evolution of class II aminoacyl-tRNA synthetases and the first code**. *FEBS Lett* 2015, **589**:3499-3507 http://dx.doi.org/10.1016/j.febslet.2015.10.006.

26. Grosjean H, Westhof E: **An integrated, structure- and energy-**
• **based view of the genetic code**. *Nucl Acid Res* 2016, **44**:8020-8040 http://dx.doi.org/10.1093/nar/gkw608.
Analysis of codon-anticodon interaction development and associated post-transcriptional modifications. The conclusions here fully support the GC-GCA-GCAU model.

27. Kubyshkin V, Budisa N: **Synthetic alienation of microbial organisms by using genetic code engineering: why and how?** *Biotechnol J* 2017, **12**:1600097 http://dx.doi.org/10.1002/biot.201600097.

28. Wong JT-F, Ng S-K, Mat W-K, Hu T, Xue H: **Coevolution theory of the genetic code at age forty: pathway to translation and synthetic life**. *Life* 2016, **6**:12 http://dx.doi.org/10.3390/life6010012.

29. Gorres KL, Raines RT: **Prolyl 4-hydroxylase**. *Crit Rev Biochem Mol Biol* 2010, **45**:106-124 http://dx.doi.org/10.3109/10409231003627991.

30. Lyons TW, Reinhard CT, Planavsky NJ: **The rise of oxygen in earth's early ocean and atmosphere**. *Nature* 2014, **506**:307-315 http://dx.doi.org/10.1038/nature13068.

31. Blaber M, Zhang X-J, Matthews BW: **Structural basis of amino acid α helix propensity**. *Science* 1993, **260**:1637-1640 http://dx.doi.org/10.1126/science.8503008.

32. Pace CN, Scholtz JM: **A helix propensity scale based on experimental studies of peptides and proteins**. *Biophys J* 1998, **75**:422-427 http://dx.doi.org/10.1016/S0006-3495(98)77529-0.

33. Adzhubei AA, Sternberg MJE, Makarov AA: **Polyproline-II helix in proteins: structure and function**. *J Mol Biol* 2013, **425**:2100-2132 http://dx.doi.org/10.1016/j.jmb.2013.03.018.

34. Shi Z, Chen K, Liu Z, Kallenbach NR: **Conformation of the backbone in unfolded proteins**. *Chem Rev* 2006, **106**:1877-1897 http://dx.doi.org/10.1021/cr040433a.

35. Nanda M, Ganesh K: **4(*R*/*S*)-Guanidinylprolyl collagen peptides:**
• **on-resin synthesis, complexation with plasmid DNA, and the role of peptides in enhancement of transfection**. *J Org Chem* 2012, **77**:4131-4135 http://dx.doi.org/10.1021/jo300070p.
The paper addresses for the first time interactions between cationic polypeptides built on polyproline-II scaffold and nucleic acids. The interaction between these two components was also essential in the original phase of primitive polypeptide biosynthesis according to the GC-GCA-GCAU scheme.

36. Foletti C, Kramer RA, Mauser H, Jenal U, Bleicher KH, Wennemers H: **Functionalized proline-rich peptides bind the bacterial second messenger c-di-GMP**. *Angew Chem Int Ed* 2018, **57**:7729-7733 http://dx.doi.org/10.1002/anie.201801845.

37. Pujals S, Giralt E: **Proline-rich, amphipathic cell-penetrating peptides**. *Adv Drug Deliv Rev* 2008, **60**:473-484 http://dx.doi.org/10.1016/j.addr.2007.09.012.

38. Nepal M, Mohamed MF, Blade R, Eldesouky HE, Anderson TN, Seleem MN, Chmielewski J: **A library approach to cationic amphiphilic polyproline helices that target intracellular pathogenic bacteria**. *ACS Infect Dis* 2018, **4**:1300-1305 http://dx.doi.org/10.1021/acsinfecdis.8b00124.

39. Shoulders M, Raines RT: **Collagen structure and stability**. *Annu Rev Biochem* 2009, **78**:929-958 http://dx.doi.org/10.1146/annurev.biochem.77.032207.120833.

40. Pentelute BL, Gates ZP, Tereshko V, Dashnau JL, Vanderkooi JM, Kossiakoff AA, Kent SBH: **X-ray structure of snow flea antifreeze protein determined by racemic crystallization of synthetic protein enantiomers**. *J Am Chem Soc* 2008, **130**:9695-9701 http://dx.doi.org/10.1021/ja8013538.

41. Treviño MÁ, Pantoja-Uceda D, Menéndez M, Gomez MV, Mompeán M, Laurents DV: **The singular NMR fingerprint of a polyproline II helical bundle**. *J Am Chem Soc* 2018, **140**:16988-17000 http://dx.doi.org/10.1021/jacs.8b05261.

42. Kubyshkin V, Budisa N: **Construction of a polyproline structure with hydrophobic exterior using octahydroindole-2-carboxylic acid**. *Org Biomol Chem* 2017, **15**:619-627 http://dx.doi.org/10.1039/c6ob02306a.

43. Kubyshkin V, Budisa N: **Exploring hydrophobicity limits of polyproline helix with oligomeric octahydroindole-2-carboxylic acid**. *J Pept Sci* 2018, **24**:e3076 http://dx.doi.org/10.1002/psc.3076.

44. Martin C, Lebrun A, Martinez J, Cavelier F: **Synthesis of homopolypeptides with PPII structure**. *J Polym Sci A: Polym Chem* 2013, **51**:3103-3109 http://dx.doi.org/10.1002/pola.26705.

45. Martin C, Legrand B, Lebrun A, Berthomieu D, Martinez J, Cavelier F: **Silaproline helical mimetics selectively form an all-*trans* PPII helix**. *Chem Eur J* 2014, **20**:14240-14244 http://dx.doi.org/10.1002/chem.201404820.

46. •• Kubyshkin V, Grage SL, Bürck J, Ulrich AS, Budisa N: **Transmembrane polyproline helix**. *J Phys Chem Lett* 2018, **9**:2170-2174 http://dx.doi.org/10.1021/acs.jpclett.8b00829.
Nature is very conservative in constructing membrane-spanning protein motifs. Here we designed a new transmembrane helix, which is yet built on a completely natural structure. This fact demonstrates that there is rich potential chemical space non-explored by natural evolution.

47. • Kubyshkin V, Budisa N: **Promotion of the collagen triple helix in a hydrophobic environment**. *Org Biomol Chem* 2019, **17**:2502-2507 http://dx.doi.org/10.1039/c9ob00070d.
In this work, we make the next step towards the alternative backbone life, and show that assembly of the collagen triple helix is suitable for hydrophobic media. The work suggests that collagen can actually be converted into a membrane protein, thus it may open an entirely new avenue for architecting synthetic membrane proteins.

48. Gangloff N, Ulbricht J, Lorson T, Schlaag H, Luxenhofer R: **Peptoids and polypeptoids at the frontier of supra- and macromolecular engineering**. *Chem Rev* 2016, **116**:1753-1802 http://dx.doi.org/10.1021/acs.chemrev.5b00201.

49. Schettini R, Costabile C, Della Sala G, Buirey J, Tosolini M, Tecilla P, Vaccaro MC, Bruno I, De Riccardis F, Izzo I: **Tuning the biomimetic performances of 4-hydroxyproline-containing cyclic peptoids**. *Org Biomol Chem* 2018, **16**:6708-6717 http://dx.doi.org/10.1039/c8ob01522h.

50. Tang Y-C, Deber CM: **Aqueous solubility and membrane interactions of hydrophobic peptides with peptoid tags**. *Pept Sci* 2002, **76**:110-118 http://dx.doi.org/10.1002/bip.10566.

51. Das S, Salah KBH, Djibo M, Inguimbert N: **Peptaibols as a model for the insertions of chemical modifications**. *Arc Biochem Biophys* 2018, **658**:16-30 http://dx.doi.org/10.1016/j.abb.2018.09.016.

52. Crisma M, Formaggio F, Alemán C, Torras J, Ramakrishnan C, Kalmankar N, Balaram P, Toniolo C: **The fully-extended conformation in peptides and proteins**. *Peptide Sci* 2019, **110**:e23100 http://dx.doi.org/10.1002/bip.23100.

53. Diwo C, Budisa N: **Alternative biochemistries for alien life: basic concepts and requirements for the design of a robust biocontainment system in genetic isolation**. *Genes (Basel)* 2018, **10**:E17 http://dx.doi.org/10.3390/genes10010017.

54. Gánti T: *Chemoton Theory: Theory of Living Systems*. Kluwer Academic/Plenum Publishers; 2003. ISBN 9780306477850.

55. Luisi PL: *The Emergence of Life: From Chemical Origins to Synthetic Biology*. Cambridge University Press; 2006. ISBN 9781107092396.

56. Völler J-S, Budisa N: **Coupling genetic code expansion and metabolic engineering for synthetic cells**. *Curr Opin Biotechnol* 2017, **48**:1-7 http://dx.doi.org/10.1016/j.copbio.2017.02.002.

57. Reinkemeier CD, Girona GE, Lemke EA: **Designer membraneless organelles enable codon reassignment of selected mRNAs in eukaryotes**. *Science* 2019, **363**:eaaw2644 http://dx.doi.org/10.1126/science.aaw2644.

58. Budisa N: **Xenobiology, new-to-nature synthetic cells and genetic firewall**. *Curr Org Chem* 2014, **18**:936-943 http://dx.doi.org/10.2174/1385272818081406161543O1.