microbial biotechnology

Open Access

Xenomicrobiology: a roadmap for genetic code engineering

Carlos G. Acevedo-Rocha[1,2] and Nediljko Budisa[3,*]

[1]Biosyntia ApS, 2970 Hørsholm, Denmark.
[2]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970 Hørsholm, Denmark.
[3]Department of Chemistry, Technical University Berlin, Müller-Breslau-Str. 10, Berlin 10623, Germany.

Summary

Biology is an analytical and informational science that is becoming increasingly dependent on chemical synthesis. One example is the high-throughput and low-cost synthesis of DNA, which is a foundation for the research field of synthetic biology (SB). The aim of SB is to provide biotechnological solutions to health, energy and environmental issues as well as unsustainable manufacturing processes in the frame of naturally existing chemical building blocks. Xenobiology (XB) goes a step further by implementing non-natural building blocks in living cells. In this context, genetic code engineering respectively enables the re-design of genes/genomes and proteins/proteomes with non-canonical nucleic (XNAs) and amino (ncAAs) acids. Besides studying information flow and evolutionary innovation in living systems, XB allows the development of new-to-nature therapeutic proteins/peptides, new biocatalysts for potential applications in synthetic organic chemistry and biocontainment strategies for enhanced biosafety. In this perspective, we provide a brief history and evolution of the genetic code in the context of XB. We then discuss the latest efforts and challenges ahead for engineering the genetic code with focus on substitutions and additions of ncAAs as well as standard amino acid reductions. Finally, we present a roadmap for the directed evolution of artificial microbes for emancipating rare sense codons that could be used to introduce novel building blocks. The development of such xenomicroorganisms endowed with a 'genetic firewall' will

also allow to study and understand the relation between code evolution and horizontal gene transfer.

Xenobiology research directions

Biology is an analytical and informational science that is becoming increasingly dependent on chemical synthesis. The high-throughput and low-cost synthesis of DNA, for example, is the foundation of the field of synthetic biology (SB). The aim of SB is to provide urgent biotechnological-based sustainable solutions to problems in the health, energy and environmental sectors (Acevedo-Rocha, 2016). This sort of biotechnology is performed under the frame of naturally existing chemical building blocks. Compared with the rich methods of synthetic organic chemistry (Walsh *et al.*, 2005), however, the diversity of chemistries used by natural organisms is surprisingly narrow; thus, limiting a wide range of useful and chemically diverse biotransformations.

To address this issue, the field of xenobiology (XB) aims to endow biological systems with artificial chemistry absent in natural organisms (Schmidt, 2010). XB is a highly diverse area, which two main research directions (Fig. 1). One involves the design and synthesis of alternative nucleic acids (xenonucleic acids, XNAs) based upon new base pairs, sugars and modified backbones (Benner and Sismour, 2005; Pinheiro and Holliger, 2012). The second major area concerns engineering the genetic code of proteins and proteomes with non-canonical amino acids (ncAAs) (Bacher *et al.*, 2004; Budisa, 2004).

Introducing new functional members of XNA into either DNA or RNA and ncAAs into proteins breaks the universality of the genetic code of living organisms, alienating these from other natural life forms (Budisa, 2014). This alienation gives access to possible orthogonal life because both living systems cannot share information with each other, i.e. a 'genetic firewall' is built (Acevedo-Rocha and Budisa, 2011). Furthermore, these synthetic organisms are not able to grow if they are not supplied with synthetic nutrients (trophic containment) (Marliere, 2009). Combination of a genetic firewall with an 'alien ersatz' increases the degree of biocontainment (Herdewijn and Marliere, 2009).

Notably, the difference between SB and XB is that the former generates genetically modified organisms (GMOs)
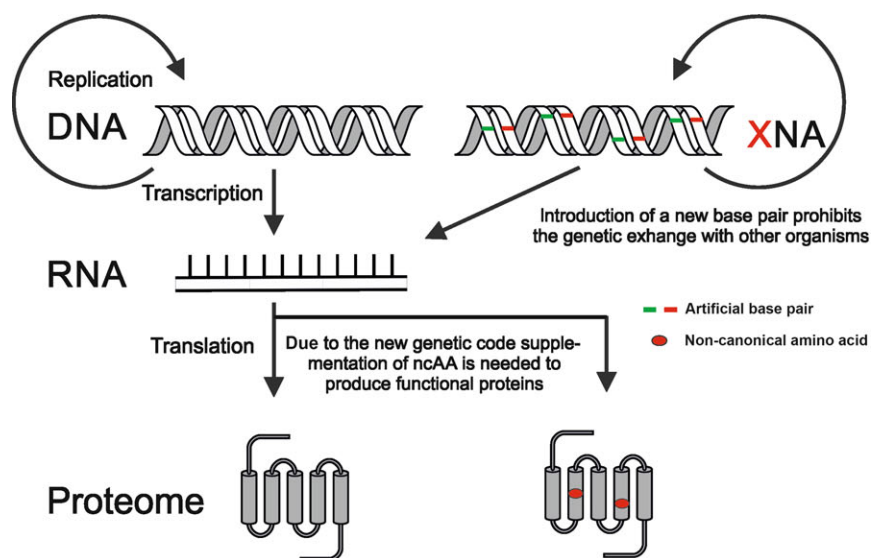
Fig. 1. Experimental approaches in xenobiology to re-design the flow of genetic information. Left: the central dogma of molecular biology is shown. DNA is replicated and transmitted to the descendants or transcribed into RNA. Translation subsequently gives rise to proteins, making up the functional proteome. Recent advances show that the introduction of artificial base pairs as XNA can also replicate and proliferate (Malyshev *et al.*, 2014). To date, however, this information cannot be transcribed or translated in vivo. Incorporation of ncAAs into the proteome was shown to be feasible if evolutionary pressure is applied (Hoesl *et al.*, 2015). Right: Methods and levels of the genetic code re-design that enables the partial or full re-design of informational flow in biology according to Budisa (2014).

based on natural chemical building blocks [e.g. canonical amino acids (cAAs)], whereas the latter aims at creating chemically modified organisms (CMOs) in which the chemical building blocks are artificial (Acevedo-Rocha, 2016). The likelihood of spreading genetic material from GMOs/CMOs into the environment is still underestimated and not well studied (Schafer *et al.*, 2011). Thus, biocontainment of GMOs/CMOs is an active research area for enhancing biosafety in XB (Schmidt and de Lorenzo, 2016), among many other SB-based research endeavours (Moe-Behrens *et al.*, 2013). The idea of 'biocontainment' is of particular interest, as it represents a tool for understanding the relationship between the evolution of the genetic code and horizontal gene transfer. This area is well-debated in the field of molecular evolution (Gogarten and Townsend, 2005).

In this perspective, we briefly discuss the origin and evolution of the genetic code from the point of view of XB. We then describe the various strategies for engineering the genetic code, including the use of amino acid auxotrophic strains, suppression of stop codons, reassignment of sense codons and reduction of the standard amino acid repertoire. We also highlight the efforts on metabolic, genome and strain engineering for improving efforts for engineering the genetic code. Finally, we discuss the importance and challenges ahead for the directed evolution of xenomicroorganisms towards the understanding of evolutionary innovation and for enhancing biocontainment strategies.

Origin and features of the genetic code

The genetic code is universal in all three domains of life eukaryotes, bacteria and archaea (Woese *et al.*, 1990). It allows the transmission of genetic information stored in DNA from one generation to the next one. Transcription of protein-encoding genes into RNA enables the translation machinery to convert this information into proteins, which are the executors of the genetic information. A reliable and accurate translation of the linear RNA-sequence into a functional protein is ensured by tRNAs. These adaptor molecules read the mRNA in a three-base letter sequence code, i.e. in triplets (Crick *et al.*, 1961).

The chemistry of the universal translational apparatus is highly standardized: DNA consists of four different bases (A, T, G and C) and there are $4^3 = 64$ possible triplet combinations for encoding 20 cAAs. An important feature of the genetic code is its codon degeneracy, i.e., the genetic code is redundant because several triplets can code the same amino acid (Lagerkvist, 1978). Briefly, there are 61 sense codons coding for amino acids and 3 stop codon signals (UAA: ochre, UGA: opal and UAG: amber) that are responsible for terminating protein biosynthesis at the ribosome (Fig. 2). The assignment of one triplet to one amino acid is not random: the degeneracy of the genetic code enables an organism to choose between abundant and rare codons. The number of triplets is often directly proportional to the abundance of the

amino acid in the proteome. For example, the amino acid tryptophan, which is relatively rare (present in approx. 1.5% in the proteome of *Escherichia coli*), is only encoded by the single UGG codon, whereas arginine (around 5.5% distribution in the *E. coli* proteome) is decoded by six triplets (Lajoie *et al.*, 2013).

In *E. coli*, out of the 61 sense codons, two are unique (methionine and tryptophan) while the rest are degenerate, with amino acids being encoded by as little as two and as much as six triplets (Fig. 2). In the latter case, several of the codons are rare (e.g., arginine) and allow an organism to control the expression rate of proteins, thus providing a selective advantage when adapting to a changing environment (Subramaniam *et al.*, 2013a). For example, protein translation of abundant and rare codons seems to be equally fast when *E. coli* grows in medium rich of amino acids. In nutrient-deficient media, by contrast, translation velocity of rare codons is substantially decreased, suggesting that the intracellular amino acid concentration is important for efficient reading of rare codons and not the isoacceptor tRNA abundance (Li *et al.*, 2012; Subramaniam *et al.*, 2013b; Ling *et al.*, 2015).

An unresolved question is why the 20 cAAs were selected by evolution even though several other non-proteinogenic amino acids were available in the 'primordial soup' (Kvenvolden *et al.*, 1971) including norvaline and norleucine? Special cAAs can be added to the standard amino acid repertoire. For instance, many organisms use the 21st (selenocysteine, SeCys, or Sec) and 22nd (pyrrolysine, Pyl) amino acids (Fig. 2) for adaptive purposes (Chambers *et al.*, 1986; Srinivasan *et al.*, 2002).

The evolution of the genetic code is a historical process and we will most probably always remain agnostic about its origin and sequence of events that lead to the current codon–amino acid associations. Upon the elucidation of the genetic code in 1964 (Nirenberg and Leder, 1964), Crick postulated a 'frozen accident theory' which assumes that the association of particular amino acids to their codons was accidental until the point where any further member expansion would be lethal to cells due to functional protein destabilization (Crick, 1968). This concept has been questioned for many times during the last decades; the plasticity and evolvability of the standard genetic code is postulated to be plausible, as supported by the finding of distinct alternative genetic codes in prokaryotes (Oba *et al.*, 1991), eukaryotic nuclear genomes (Sugita and Nakase, 1999) and mitochondrial genomes (Inagaki *et al.*, 1998). Additional widely accepted concepts about the origin of the genetic code include stereochemical (Pelc, 1965), adaptive (Woese, 1965) and co-evolution (Wong, 1975, 2005) theory.

Before it was 'frozen', the early genetic code underwent evolution and expansion, which is encompassed by two main theories about codon reassignment. Figure 3 summarizes them and explains possible evolutionary mechanisms for codon reassignment to other amino

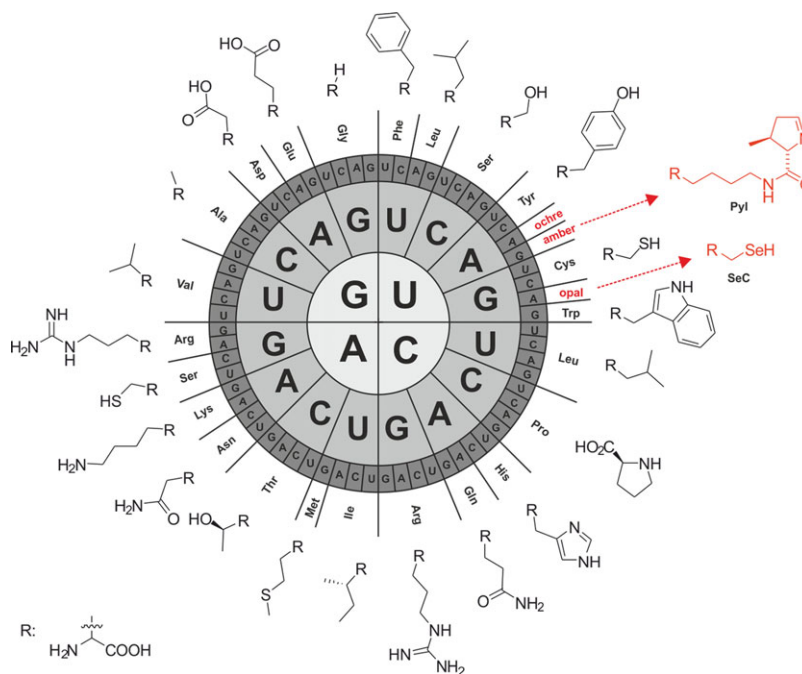**Fig. 2.** The genetic code structure in the RNA format in a radial representation. Chart from inside to outside: a triplet of mRNA (5′→3′) is assigned to one of the 20 canonical amino acids or a stop codon. The natural expansion of the genetic code of selenocysteine (SeC) at opal and pyrrolysine (Pyl) at the amber and opal stop codons is depicted (Courtesy provided by Dr Stefan Oehm) (Oehm, 2016).

acids: codon capture (Osawa *et al.*, 1992) and ambiguous intermediate (Schultz and Yarus, 1994) theory. The codon capture theory postulates that genomic composition changes due to GC pressure could impose alternative codon usage mainly due to changes at the third base position of a codon. Owing to this pressure, a potential codon can disappear, making a particular cognate tRNA obsolete. Reappearance of the codon due to random drift would lead to its recognition by another tRNA (that could also charge another amino acid) leading to codon reassignment, i.e., change the identity of the codon. Conversely, the ambiguous intermediate theory postulates an appearance of mutant tRNA in competition with the cognate tRNA. In the course of selective pressure, the latter might lose the race against the mutant tRNA, hence giving rise to another amino acid meaning [see Fig. 3 and (Budisa, 2006a)].

Code and metabolic engineering of auxotrophic strains

Genetic code engineering aims to use ncAAs as building blocks in proteins. It exploits the flexibility of the components involved in protein translation mainly aminoacyl-tRNA synthetases (aaRSs) and tRNAs (Fig. 4). While many aaRSs have evolved mechanisms to differentiate between naturally abundant and structurally related amino acids (especially metabolic intermediates) (Schmidt and Schimmel, 1994), most ncAA that can be

incorporated with this approach do not occur in natural environments. In the laboratory, however, a codon can be read ambiguously by mischarging tRNAs (i.e. a cognate tRNA can be charged with two different yet similar substrates). For example, methionyl-tRNA synthetase charges its cognate $tRNA^{Met}_{CAT}$ with a variety of methionine (Met) analogues, homologues and surrogates into heterologous-expressed proteins (Wiltschi *et al.*, 2009). The best know example is selenomethionine (SeMet), which can almost quantitatively replace Met in proteins and proteomes in suitable auxotrophic strains (Cowie and Cohen, 1957; Budisa *et al.*, 1995). We and others have shown many examples of potential applications of such an approach for Met analogues and beyond during the last two decades (Mohammadi *et al.*, 2001; Link and Tirrell, 2005; Acevedo-Rocha *et al.*, 2013; Bohlke and Budisa, 2014; Piotrowski *et al.*, 2015). In another example, the incorporation of norleucine (Nle) into a lipase allow the engineering of a highly active 'cold-wash' enzyme (Hoesl *et al.*, 2011).

Till date, more than 50 ncAAs (mostly Met, Trp, Tyr, Phe, Pro, Arg, Lys analogues) (Budisa, 2006b) (Fig. 4A) have been incorporated into proteins using this approach. Unfortunately, most of these experiments are limited to individual xeno-proteins overexpressed from auxotrophic strains. Although this is an important research area that can give rise to potential applications in biocatalysis, it is important to consider the costs of supplying expensive ncAAs into a big fermenter. For this reason, genetic code
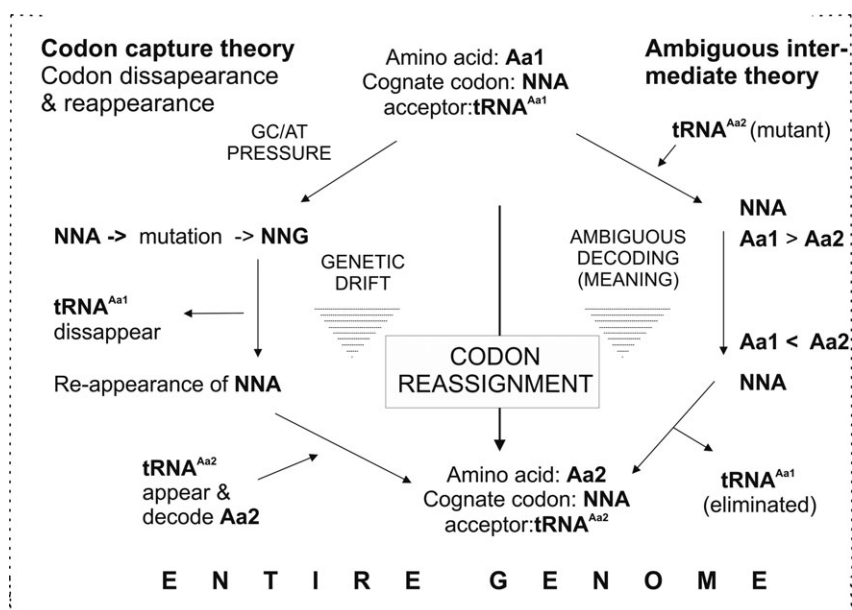
Fig. 3. The codon capture and ambiguous intermediate theory for codon reassignment in the historical development of the genetic code. While in codon capture (left) the endogenous tRNA disappears due to the removal of its designated codon, in ambiguous intermediate theory (right) the exogenous tRNA is in direct competition with the cognate one. In the end, both theories end up in a reassigned codon. Figure modified according Santos *et al.* (2004) and Budisa (2006a).

(A) Selective pressure incorporation (SPI)

(B) Stop codon suppression (SCS)

Fig. 4. Flow chart presentations of two basic in vivo approaches for incorporating non-canonical amino acids into proteins. (A) Auxotrophy-based selective pressure incorporation (SPI) method exploits the endogenous translational system to load an isostructural analogue onto a canonical tRNA l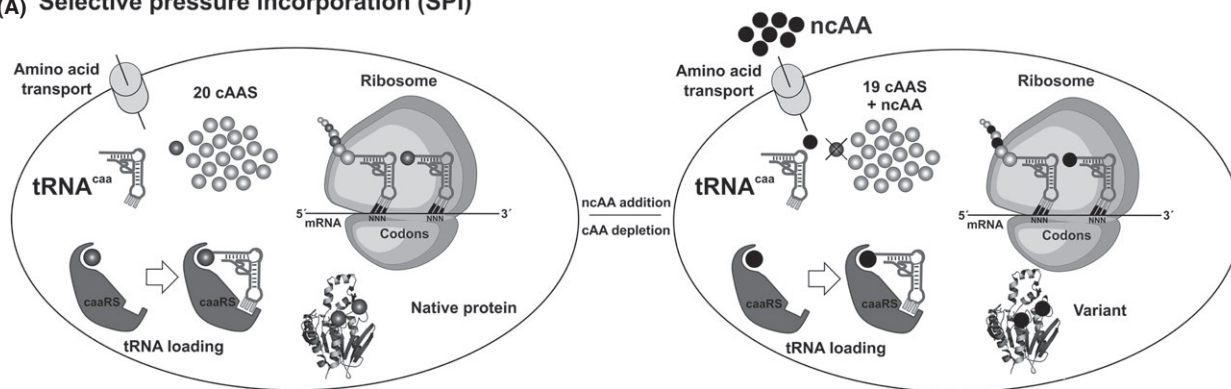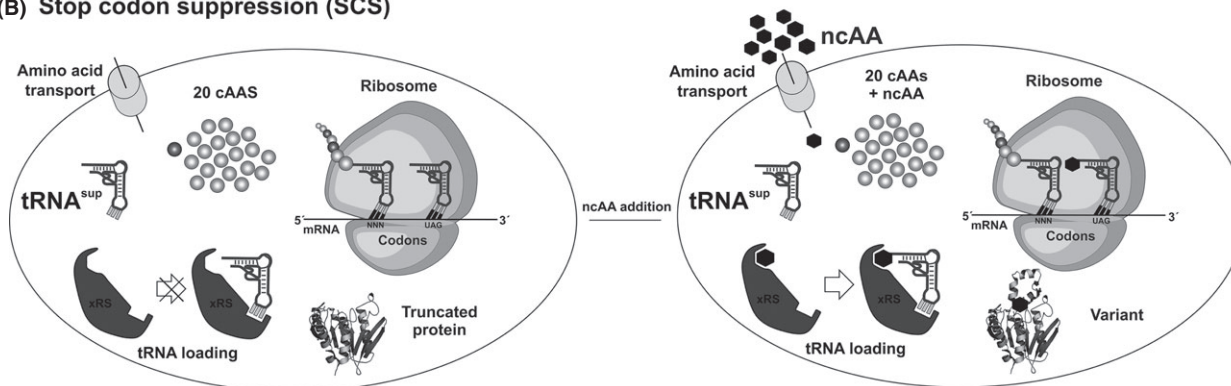eading to residue-specific incorporation, i.e. cAA→ncAA substitution. (B) Stop codon suppression (SCS) usually endows the host with a new orthogonal aaRS:tRNA pair in charge of introducing a given ncAA or more at one or more stop codons. For more details see Hoesl and Budisa (2012).

engineering can be combined with metabolic engineering. For example, we implemented a biosynthetic route in for the biosynthesis of the ncAA azido homoalanine (Aha) by expressing in trans O-acetylhomoserine sulfhydrylase (cgOAHSS) from *Corynebacterium glutamicum* in the Met-auxotroph *E. coli* strain B834 while supplying the strain with *O*-acetyl-homoserine and inorganic sodium azide (NaN$_3$) as precursors (Ma *et al.*, 2014). This engineering effort resulted in an approach for a more economical production of Aha-labelled proteins, which can be selectively modified via 'click chemistry' with different molecules for protein immobilization or labelling (De Simone *et al.*, 2016). Likewise, the group of Wiltschi recently afforded a strain able to produce Nle-labelled proteins in sufficient economically amounts (Anderhuber *et al.*, 2016). This kind of experiments will become more important if applications of ncAAs in biocatalysis are likely to succeed. Ideally, the precursors of the ncAA should be biosynthesized starting from a cheap carbon source such as glucose. We envision more synergies between meta-bolic and genetic code engineering in the years to come.

Code engineering by stop codon suppression

In contrast to the ambiguous decoding of sense codons by ncAAs as explained above, the meaning of a stop codon can be suppressed by introducing engineered orthogonal pairs (Fig. 4B). The most common pairs are the *Methanocaldococcus jannaschii* TyrRS:tRNA$^{Tyr}_{CUA}$-pair, and PylRS:tRNA$^{Pyl}_{CUA}$-pairs from *Methanosarcina barkeri* and *Methanosarcina mazei*. This stop codon sup-pression (SCS) approach allows the site-specific incorporation of ncAAs into a target protein, which is usually overexpressed (Hoesl and Budisa, 2012). Because alteration of the protein structure is minimal, it is possible to study protein function with an unprecedented level of accuracy. For example, it is possible to study protein conformation by Förster resonance energy transfer, follow intracellular protein localization (using fluorescent ncAAs as probes) or determine protein–protein interactions by using cross-linking ncAAs (Lang *et al.*, 2015).

The SCS methodologies have matured in the last years, offering researchers with a plethora of ncAAs to

incorporate site-specifically (Wright *et al.*, 2016). However, the SCS is usually limited because the release factors (RFs), which are responsible for terminating protein biosynthesis at the ribosome, compete with the tRNAs, resulting in a mixture of labelled and truncated protein. Although the efficiency of ncAA incorporation can be improved with RF-free engineered strains (Amiram *et al.*, 2015), additional 'context effect' issues persist, which is still the most probably cause of obtaining low protein yields. Recently, Ignatova and others provided a more precise experimental description of these phenomena. The insertion of a stop codon in a particular sequence can cause large deviations from natural mRNA folding energy. Furthermore, it can affect the binding site interaction strength within the ribosome (see e.g. Gorochowski *et al.*, 2015). In the years to come, we expect to obtain better in silico models that could predict what the most suitable sites for inserting stop codons are. Hence, promoting synergies between computational chemistry and code engineering is an important research endeavour.

Code engineering by sense codon reassignment

Codon reassignment mainly includes the removal of the endogenous translation components like tRNA and aaRS, synonymous codon swaps throughout the chromosome or essential genes (Lajoie *et al.*, 2015). For example, the groups of Isaacs (Rovner *et al.*, 2015) and Church (Mandell *et al.*, 2015) engineered the genetic code of *E. coli* C321.ΔA (which lacks amber stop codons) by re-introducing amber stop codons into essential genes (devising a sort of experimental 'codon capture' approach). Because the UAG codon lacks any meaning in that distinct strain, these otherwise lethal stop codons have to be rescued with *M. jannaschii* TyrRS:tRNA$^{Tyr}_{CUA}$ for *p*-azido-L-phenylalanine (pAzF) or biphenyl-L-alanine (bipA) (Xie *et al.*, 2007) respectively. As the UAG codon has no meaning in strain C321.ΔA, both studies converted ('captured') this codon into a sense one by supplying orthogonal pairs. Although both experiments differ in their strategy and their execution, they yield the same result: a GMO that cannot grow without supplying the ncAA, therefore becoming trophically biocontained. Those synthetic auxotrophic strains show less than 6×10^{-12} escape mutants per colony forming unit in various media. Most other techniques for biocontainment purposes yield higher escape frequencies and suffer from the possibility of cross-feeding of metabolites produced from other organisms (Schmidt and de Lorenzo, 2016). Both ncAAs used (pAzF and bipA) are not available in natural environments. However, the synthetic auxotrophy only relies on the orthogonal translation system. This system might be stable under laboratory conditions, but there are many places where the system could lose its orthogonality, e.g., mutating TyrRS to accept any other cAA, introduction of a natural amber suppressors (*sup*E, *sup*D, *sup*F, *sup*Z, etc.) or simply reverting the amber sites back to their original meaning. Thus, long-term evolution experiments of strain C321.ΔA in more diverse and rich media are needed to ensure that it can be biocontained without reversing its genotype.

While ambiguous decoding and SCS approaches are always in competition with the original codon meaning and function of the corresponding amino acid, sense codon reassignment aims to completely remove the pre-existent functional information of a certain codon. For example, isoleucine can be coded by codon AUA, but there is no tRNA$^{Ile}_{UAU}$ gene in *E. coli* because the third base of the tRNA$^{Met}_{UAC}$ is modified by tRNA^{Ile}-lysidine synthetase (TilS) to lysidine, yielding tRNA$^{Ile}_{LAU}$ that is read by wobbling at the third base position (Bohlke and Budisa, 2014). We generated a *tilS* knockout strain (this mutation is lethal to the cell leaving more than 5000 codons unassigned) by using a rescue plasmid expressing the *Mycoplasma mobile* IleRS:tRNA$^{Ile}_{UAU}$ pair to sustain cell viability. The rescue plasmid can be easily removed in *E. coli* ΔtilS and can be exploited to apply evolutionary pressure to reassign the Ile codon AUA to another cAA (Doring and Marliere, 1998) or even ncAA.

Another strategy was reported by Sakamoto and coworkers (Mukai *et al.*, 2015) who reassigned the rare AGG codon of arginine (Arg) to L-homoarginine (hArg) and L-N6-(1-iminoethyl)lysine (L-NIL) by using PylRS:tRNA$^{Pyl}_{CUA}$-pair. Cells can survive the deletion of Arg rare codons (*arg*W and *arg*U) only if the evolved PylRS-system is co-expressed and AGG positions in most essential genes are exchanged by synonymous codons (to avoid detrimental effects of AGG-reading as hArg or L-NIL). Although analytical data were not presented (e.g. MS/MS experiments in essential genes to verify Arg-replacement), the exchange of Arg at remaining AGG positions most probably took place given that both ncAAs used (hArg, L-NIL) are similar to the cAA Arg ('similar replaces similar') (Moroder and Budisa, 2010).

Code engineering via reduction of the standard amino acid repertoire

Widely accepted co-evolution theory about the origin of the genetic code postulates its evolution from a simplified set of amino acid and extension to recent form with the expansion of cellular metabolism (Wong, 1981). Therefore, an experimental model with the reduced genetic code could provide some insight into evolutionary trajectories of amino acids. This was accomplished by the construction of a functional enzyme using a nine-amino acid alphabet (Walter *et al.*, 2005). Namely,

Hilvert and colleagues demonstrated that *E. coli* lacking the aroQ gene for chorismate mutase is unable to grow in media where all sources for tyrosine (Tyr) and phenylalanine (Phe) are withdrawn. This synthetic auxotrophy can be exploited to screen a library of artificial chorismate mutase analogues, in which aroQ is only composed from a nine-amino acid alphabet. It was found that even this reduced set of amino acids can lead to a functional, metabolically competent and near to wild-type growing strain. However, the reduction led to a restricted diversity ending up with a destabilization of overall protein structure. These results indicate the need of more amino acids for robust proteins and explain the evolutionary force towards the increased diversity. A more general approach towards this issue was reported by Marlière and co-workers: they showed the possibility of removing Trp from the genetic code (Pezo *et al.*, 2013). Taking advantage of the conserved and essential histidine (His) residue in the active site of yeast transketolase gene *tktY*, this position was mutated to the Trp-codon UGG in *trans* in a suitably configured *E. coli* MG1655 ΔtktAB strain. Supplying this strain with the missense $tRNA_{His}^{CCA}$, the UGG codon in the essential TktY is read as a His leading to a functional gene, which is the driving force to keep the missense tRNA. Subsequently, the system was proliferated in a continuous pulse feeding regime, yielding 30 times higher mis-incorporation of His at Trp positions in proteins. This experiment suggests the possibility of engineering the genetic code applying evolutionary pressure using a reduced amino acid alphabet. This is an important research area whose potentials have not yet been extensively studied (Oehm, 2016).

Code engineering via experimental evolution

Over more than three billion years, the standard genetic standard code has not undergone any significant changes besides some local alterations (Knight *et al.*, 2001). Therefore, an utmost goal of XB is to introduce new biochemical building blocks into an organism at both the genomic and proteomic level. One option is by performing experimental evolution, which deals with the long-term cultivation of organisms under controlled conditions. Doubtless, adaptive evolution experiments with suitably designed metabolic prototypes could accelerate the propagation, manipulation and analysis of organisms in a controlled environment. Furthermore, such evolutionary adaptation is the most expedient route to generate artificial biocontained microbes. Because microbes have fast generation times and large population sizes in a reasonable and manageable dimension of cultivation space, they are ideal candidates for such experimental setups. Additionally, populations can be easily frozen for later

analysis ('fossils'). Adaptive evolution experiments have been designed to cultivate *E. coli* to adapt on glucose, glycerol and lactate minimal medium, high temperature irradiation as well as high salt or ethanol concentration (Blount *et al.*, 2012; Wiser *et al.*, 2013).

Already in the late 1950s, Cowie and Cohen (1957) demonstrated that Met auxotrophic *E. coli* grow in Met-free medium when SeMet is supplied. This proteome wide exchange of Met by producing functionally essential enzymes for cell proliferation illustrates how the amino acid composition of a proteome can be influenced by external pressure. Another report from 1963 by Rennert and Anker claims that *E. coli* cells able to use 5′,5′,5′-trifluoroleucine (TFL) as an analogue for leucine (Leu) can be isolated from a suitably designed cultivation experiment (Rennert and Anker, 1963). Wong (1983) was able to adapt a Trp deficient *Bacillus subtilis* with some few single-cell transfers to grow on 4-fluoro-L-tryptophan (4F-Trp) as Trp source. Using the mutagenic agent it was possible to obtain a derivative exhibiting preference towards the ncAA. Nevertheless, those old experiments were not accompanied by precise analytic measurements to quantify the extent of ncAA incorporation (contaminations of cAA from media preparation cannot be excluded – most likely, those sources of cAA were elemental for the observed growth). Moreover, those experiments were conducted in 'rich' medium in which standard amino acids, vitamins and nucleobases were supplemented. Therefore, only a small portion of the cellular machinery had to adapt to the incorporation of the ncAA. Bacher and Ellington (2001) adapted the *E. coli* strain C600 ΔtrpE towards 4F-Trp by serially transferring it to minimal medium while gradually increasing the 4F-Trp-to-Trp ratio. Unfortunately, the commercial source of 4F-Trp was contaminated with 0.03% Trp (the authors were not able to withdraw all cAA-source from the media). Yet they found a strain tolerating high levels of 4F-Trp to a level at which it was toxic for the parental strain. Due to the lack of next generation sequencing, only a few relevant genes were sequenced but these were unchanged (*tna*A, *trp*R, *mtr*, *trp*T, *bla*, *gap*A), while others like *trp*S, *aro*P and *tyr*R exhibited single nucleotide polymorphisms. Expression of *trp*S and *aro*P mutant proteins in *trans* in the ancestral strain led to a growth advantage in the presence of 4F-Trp. However, those mutations are not sufficient to let the C600 strain incorporate the ncAA at high levels (deep analytics were not performed). Nevertheless, this experiment improved our understanding of serial evolution experiments to change the chemical composition of the proteome.

Based on the experiences gathered on earlier long-term *E. coli* cultivation experiments by the groups of Lenski, Ellington, Bacher and Wong, we also set out to adapt this bacterium towards an ncAA. In particular, our

group completely substituted Trp by the analogue Thienopyrrole-alanine ([3,2]Tpa) in a Trp auxotrophic strain (Hoesl *et al.*, 2015). To avoid contamination of Trp traces in commercial ncAA preparations, the strains were engineered to produce the surrogate by using the enzyme tryptophan synthase (TrpBA), which condensates the indole analogue [3,2]Thienopyrrole with serine yielding [3,2]Tpa (Budisa *et al.*, 2001). Rigorous and high-precision analytics confirmed the almost completely quantitative replacement of Trp by the surrogate at 20 899 sense codons in *E. coli* W3110 genome. Hence, the engineered strain can be defined as 'trophically reassigned' (i.e., the meaning of a codon is redefined throughout the whole proteome). However, supplementing these cells with Trp reverses them to 'natural' ones given that they still favour the incorporation of the canonical building block. To achieve a Trp-independent reassignment (i.e. 'real' codon reassignment) at all UGG codons across *E. coli*'s genome – an experimental strategy for biocontainment still needs to be developed and executed (Acevedo-Rocha and Schulze-Makuch, 2015).

Roadmap to equip xenomicroorganisms with a 'genetic firewall'

In the previous sections, we highlighted various approaches for genetic code engineering. The most popular one is perhaps the genome-wide replacements of stop or rare sense codons with synonymous alternatives in order to design genomes with radically altered genetic codes, which can be used for biocontainment purposes. However, owing to the complexity of experimental genome re-design, it is difficult to avoid changes that impair cellular fitness, which creates fragile strains. Among other features, rare codons fine tune translation to facilitate the biogenesis of the encoded protein and their synonymous replacements can lethally impact mRNA secondary structure or ribosome binding sites which is expected to result in dramatically decreased fitness during the genome assembly process (Plotkin and Kudla, 2011). Furthermore, widely used orthogonal pairs are not as active and accurate as natural aaRSs (Nehring *et al.*, 2012). An alternative strategy for experimental genetic code evolution relies on the global substitution of cAAs with ncAAs (or an addition of ncAAs) assisted with simple metabolic optimization following an 'ambiguous intermediate' engineering concept. In other words, we propose a novel strategy that relies on liberation of rare sense codons of the genetic code (i.e. 'codon emancipation') from their natural decoding functions (Bohlke and Budisa, 2014). This approach consists of long-term cultivation of bacterial strains coupled with the design of orthogonal pairs for sense codon decoding. In

particular, directed evolution of bacteria should be designed to enforce ambiguous decoding of target codons using genetic selection. In this system, viable mutants with improved fitness towards missense suppression can be selected from large bacterial populations that can be automatically cultivated in suitably designed turbidostat devices. Once 'emancipation' is performed, full codon reassignment can be achieved with suitably designed orthogonal pairs. Codon emancipation will likely induce compensatory adaptive mutations that will yield robust descendants tolerant to disruptive amino acid substitutions in response to codons targeted for reassignment. We envision this strategy as a promising experimental road to achieve sense codon reassignment – the ultimate prerequisite to achieve stable 'biocontainment' as an emergent feature of xenomicroorganisms equipped with a 'genetic firewall'.

Conclusions

In summary, genetic code engineering with ncAA by using amino acid auxotrophic strains, SCS and sense codon reassignment has provided invaluable tools to study accurately protein function as well as many possible applications in biocatalysis. Nevertheless, to fully realize the power of synthetic organic chemistry in biological systems, we envision synergies with metabolic, genome and strain engineering in the next years to come. In particular, we believe that the experimental evolution of strains with ncAAs will allow the development of 'genetic firewall' that can be used for enhanced biocontainment and for studying horizontal gene transfer. Additionally, these efforts could allow the production of new-to-nature therapeutic proteins and diversification of difficult-to-synthesize antimicrobial compounds for fighting against 'super' pathogens (McGann *et al.*, 2016).

Yet the most fascinating aspect of XB is perhaps to understand the genotype–phenotype changes that lead to artificial evolutionary innovation. To what extent is innovation possible? What emergent properties are going to appear? Will these help us to re-examine the origin of the genetic code and life itself? During evolution, the choice of the basic building blocks of life was dictated by (i) the need for specific biological functions; (ii) the abundance of elements and precursors in past habitats on earth and (iii) the nature of existing solvent(s) and available energy sources in the prebiotic environment (Budisa, 2014). Thus far, there are no detailed studies on proteomics and metabolomics of engineered xenomicrobes, let alone systems biology models that could integrate the knowledge from such efforts. In 2020, we expect to have such kind of analysis in order to have a clearer picture of the possibilities that artificial

evolution can offer. We believe that expanding the repertoire of the genetic code beyond the canonical 20 (+2) amino acids will not only change the chemical makeup of life, but also allow to re-examine our understanding and current concepts about the origin of the genetic code and life itself.

Acknowledgements

We thank Dr Stefan Oehm and Ms. Federica Agostini for help in the preparation of figures and scientific discussions during manuscript preparation.

Conflict of interest

None declared.

References

Acevedo-Rocha, C. (2016) The synthetic nature of biology. In *Ambivalences of Creating Life Societal and Philosophical Dimensions of Synthetic Biology*. Kristin Hagen, Margret Engelhard, Georg Toepfer (eds.). Switzerland: Springer, pp. 9–53.

Acevedo-Rocha, C.G., and Budisa, N. (2011) On the road towards chemically modified organisms endowed with a genetic firewall. *Angew Chem Int Ed Engl* **50:** 6960–6962.

Acevedo-Rocha, C.G., and Schulze-Makuch, D. (2015) How many biochemistries are available to build a cell? *ChemBioChem* **16:** 2137–2139.

Acevedo-Rocha, C.G., Hoesl, M.G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., *et al.* (2013) Non-canonical amino acids as a useful synthetic biological tool for lipase-catalysed reactions in hostile environments. *Catal Sci Technol* **3:** 1198–1201.

Amiram, M., Haimovich, A.D., Fan, C.G., Wang, Y.S., Aerni, H.R., Ntai, I., *et al.* (2015) Evolution of translation machinery in recoded bacteria enables multi-site incorporation of nonstandard amino acids. *Nat Biotechnol* **33:** 1272–1275.

Anderhuber, N., Fladischer, P., Gruber-Khadjawi, M., Striedner, G., Mairhofer, J., and Wiltschi, B. (2016) High-level biosynthesis of norleucine in *E. coli* for the economic labeling of proteins. *J Bacteriol* (in press). doi: 10.1016/j.jbiotec.2016.04.033

Bacher, J.M., and Ellington, A.D. (2001) Selection and characterization of *Escherichia coli* variants capable of growth on an otherwise toxic tryptophan analogue. *J Bacteriol* **183,** 5414–5425.

Bacher, J.M., Hughes, R.A., Wong, J.T.F., and Ellington, A.D. (2004) Evolving new genetic codes. *Trends Ecol Evol* **19:** 69–75.

Benner, S.A., and Sismour, A.M. (2005) Synthetic biology. *Nat Rev Genet* **6:** 533–543.

Blount, Z.D., Barrick, J.E., Davidson, C.J., and Lenski, R.E. (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* **489:** 513–518.

Bohlke, N., and Budisa, N. (2014) Sense codon emancipation for proteome-wide incorporation of noncanonical amino acids: rare isoleucine codon AUA as a target for genetic code expansion. *FEMS Microbiol Lett* **351:** 133–144.

Budisa, N. (2004) Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew Chem Int Ed Engl* **43:** 6426–6463.

Budisa, N. (2006a) Amino acids and codons – code organisation and protein structure. In *Engineering the Genetic Code*. Nediljko Budisa (ed.). Weinheim, Germany: WILEY-VHC, pp. 66–85.

Budisa, N. (2006b) Reprogramming the cellular translation machinery. In *Engineering the Genetic Code*. Nediljko Budisa (ed.). Weinheim, Germany: WILEY-VHC, pp. 90–165.

Budisa, N. (2014) Xenobiology, new-to-nature synthetic cells and genetic firewall. *Curr Org Chem* **18:** 936–943.

Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J., and Huber, R. (1995) High-level biosynthetic substitution of methionine in proteins by Its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in *Escherichia coli*. *Eur J Biochem* **230:** 788–796.

Budisa, N., Alefelder, S., Bae, J.H., Golbik, R., Minks, C., Huber, R., and Moroder, L. (2001) Proteins with beta-(thienopyrrolyl)alanines as alternative chromophores and pharmaceutically active amino acids. *Protein Sci* **10:** 1281–1292.

Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W., and Harrison, P.R. (1986) The structure of the mouse glutathione peroxidase gene: the selenocysteine in the active site is encoded by the 'termination' codon, TGA. *EMBO J* **5:** 1221–1227.

Cowie, D.B., and Cohen, G.N. (1957) Biosynthesis by *s* of active altered proteins containing selenium instead of sulfur. *Biochim Biophys Acta* **26:** 252–261.

Crick, F.H. (1968) The origin of the genetic code. *J Mol Biol* **38:** 367–379.

Crick, F.H., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961) General nature of the genetic code for proteins. *Nature* **192:** 1227–1232.

De Simone, A., Hoesl, M.G., and Budisa, N. (2016) Engineering lipases with an expanded genetic code. In *Applied Biocatalysis – From Fundamental Science to Industrial Applications*. Liese, A.H., Kettling, U. and Antranikian, G. (eds). Weinheim, Germany: WILEY-VHC, pp. 3–12.

Doring, V., and Marliere, P. (1998) Reassigning cysteine in the genetic code of *Escherichia coli*. *Genetics* **150:** 543–551.

Gogarten, J.P., and Townsend, J.P. (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* **3:** 679–687.

Gorochowski, T.E., Ignatova, Z., Bovenberg, R.A.L., and Roubos, J.A. (2015) Trade-offs between tRNA abundance and mRNA secondary structure support smoothing of translation elongation rate. *Nucleic Acids Res* **43:** 3022–3032.

Herdewijn, P., and Marliere, P. (2009) Toward safe genetically modified organisms through the chemical diversification of nucleic acids. *Chem Biodivers* **6:** 791–808.

Hoesl, M.G., and Budisa, N. (2012) Recent advances in genetic code engineering in *Escherichia coli*. *Curr Opin Biotechnol* **23:** 751–757.

Hoesl, M.G., Acevedo-Rocha, C.G., Nehring, S., Royter, M., Wolschner, C., Wiltschi, B., *et al.* (2011) Lipase congeners designed by genetic code engineering. *ChemCatChem* **3:** 213–221.

Hoesl, M.G., Oehm, S., Durkin, P., Darmon, E., Peil, L., Aerni, H.-R., *et al.* (2015) Chemical s. *Angew Chem Int Ed Engl* **54:** 10030–10034.

Inagaki, Y., Ehara, M., Watanabe, K.I., Hayashi-Ishimaru, Y., and Ohama, T. (1998) Directionally evolving genetic code: the UGA codon from stop to tryptophan in mitochondria. *J Mol Evol* **47:** 378–384.

Knight, R.D., Freeland, S.J., and Landweber, L.F. (2001) Rewiring the keyboard: evolvability of the genetic code. *Nat Rev Genet* **2:** 49–58.

Kvenvolden, K.A., Lawless, J.G., and Ponnamperuma, C. (1971) Nonprotein Amino Acids in the Murchison Meteorite. *Proc Natl Acad Sci USA* **68:** 486–490.

Lagerkvist, U. (1978) "Two out of three": an alternative method for codon reading. *Proc Natl Acad Sci USA* **75:** 1759–1762.

Lajoie, M.J., Kosuri, S., Mosberg, J.A., Gregg, C.J., Zhang, D., and Church, G.M. (2013) Probing the limits of genetic recoding in essential genes. *Science* **342:** 361–363.

Lajoie, M.J., Söll, D., and Church, G.M. (2015) Overcoming challenges in engineering the genetic code. *J Mol Biol* **428:** 1004–1021.

Lang, K., Davis, L., and Chin, J.W. (2015) Genetic encoding of unnatural amino acids for labeling proteins. *Methods Mol Biol* **1266:** 217–228.

Li, G.W., Oh, E., and Weissman, J.S. (2012) The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484:** 538–541.

Ling, J., O'Donoghue, P., and Söll, D. (2015) Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology. *Nat Rev Microbiol* **13:** 707–721.

Link, A.J., and Tirrell, D.A. (2005) Reassignment of sense codons in vivo. *Methods* **36:** 291–298.

Ma, Y., Biava, H., Contestabile, R., Budisa, N., and Di Salvo, M.L. (2014) Coupling bioorthogonal chemistries with artificial metabolism: Intracellular biosynthesis of azidohomoalanine and its incorporation into recombinant proteins. *Molecules* **19:** 1004–1022.

Malyshev, D.A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J.M., *et al.* (2014) A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509:** 385–388.

Mandell, D.J., Lajoie, M.J., Mee, M.T., Takeuchi, R., Kuznetsov, G., Norville, J.E., *et al.* (2015) Biocontainment of genetically modified organisms by synthetic protein design. *Nature* **518:** 55–60.

Marliere, P. (2009) The farther, the safer: a manifesto for securely navigating synthetic species away from the old living world. *Syst Synth Biol* **3:** 77–84.

McGann, P., Bunin, J.L., Snesrud, E., Singh, S., Maybank, R., Ong, A.C., *et al.* (2016) Real time application of whole genome sequencing for outbreak investigation – What is an achievable turnaround time? *Diagn Microbiol Infect Dis* **85:** 277–282.

Moe-Behrens, G.G., Davis, R., and Haynes, K.A. (2013) Preparing synthetic biology for the world. *Front Microbiol* **4:** 10.

Mohammadi, F., Prentice, G.A., and Merrill, A.R. (2001) Protein-protein interaction using tryptophan analogues: novel spectroscopic probes for toxin-elongation factor-2 interactions. *Biochemistry* **40:** 10273–10283.

Moroder, L., and Budisa, N. (2010) Synthetic biology of protein folding. *ChemPhysChem* **11:** 1181–1187.

Mukai, T., Hoshi, H., Ohtake, K., Takahashi, M., Yamaguchi, A., Hayashi, A., *et al.* (2015) Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. *Sci Rep* **5:** 9699.

Nehring, S., Budisa, N., and Wiltschi, B. (2012) Performance analysis of orthogonal pairs designed for an expanded eukaryotic genetic code. *PLoS ONE* **7:** 13.

Nirenberg, M., and Leder, P. (1964) RNA codewords and protein synthesis. The effect of trinucleotides upon the binding of sRNA to ribosomes. *Science* **145:** 1399–1407.

Oba, T., Andachi, Y., Muto, A., and Osawa, S. (1991) CGG: an unassigned or nonsense codon in *Mycoplasma capricolum*. *Proc Natl Acad Sci USA* **88:** 921–925.

Oehm, S. (2016) *Adaptation of E. coli* towards tryptophan analogue usage. PhD Thesis. Berlin, Germany: Department of Chemistry of TU Berlin.

Osawa, S., Jukes, T.H., Watanabe, K., and Muto, A. (1992) Recent evidence for evolution of the genetic code. *Microbiol Rev* **56:** 229–264.

Pelc, S.R. (1965) Correlation between coding-triplets and amino-acids. *Nature* **207:** 597–599.

Pezo, V., Louis, D., Guérineau, V., Le Caer, J.-P., Gaillon, L., Mutzel, R., and Marlière, P. (2013) A metabolic prototype for eliminating tryptophan from the genetic code. *Sci Rep* **3:** 1359.

Pinheiro, V.B., and Holliger, P. (2012) The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Curr Opin Chem Biol* **16:** 245–252.

Piotrowski, C., Ihling, C.H., and Sinz, A. (2015) Extending the cross-linking/mass spectrometry strategy: facile incorporation of photo-activatable amino acids into the model protein calmodulin in *Escherichia coli* cells. *Methods* **89:** 121–127.

Plotkin, J.B., and Kudla, G. (2011) Synonymous but not the same: the causes and consequences of codon bias. *Nat Rev Genet* **12:** 32–42.

Rennert, O., and Anker, H. (1963) On the incorporation of 5', 5', 5'-trifluoroleucine into proteins of *E. coli*. *Biochemistry* **2:** 471–476.

Rovner, A.J., Haimovich, A.D., Katz, S.R., Li, Z., Grome, M.W., Gassaway, B.M., *et al.* (2015) Recoded organisms engineered to depend on synthetic amino acids. *Nature* **518:** 89–93.

Santos, M., Moura, G., Massey, S.E., and Tuite, M.F. (2004) Driving change: the evolution of alternative genetic codes. *Trends Genet* **20:** 95–102.

Schafer, M.G., Ross, A.A., Londo, J.P., Burdick, C.A., Lee, E.H., Travers, S.E., *et al.* (2011) The establishment of genetically engineered canola populations in the U.S. *PLoS ONE* **6,** 8–11.

Schmidt, M. (2010) Xenobiology: a new form of life as the ultimate biosafety tool. *BioEssays* **32:** 322–331.

Schmidt, M., and de Lorenzo, V. (2016) Synthetic bugs on the loose: containment options for deeply engineered (micro)organisms. *Curr Opin Biotechnol* **38:** 90–96.

Schmidt, E., and Schimmel, P. (1994) Mutational isolation of a sieve for editing in a transfer RNA synthetase. *Science* **264:** 265–267.

Schultz, D.W., and Yarus, M. (1994) Transfer RNA mutation and the malleability of the genetic code. *J Mol Biol* **235:** 1377–1380.

Srinivasan, G., James, C.M., and Krzycki, J.A. (2002) Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. *Science* **296:** 1459–1462.

Subramaniam, A.R., Deloughery, A., Bradshaw, N., Chen, Y., O'Shea, E., Losick, R., and Chai, Y. (2013a) A serine sensor for multicellularity in a bacterium. *eLife* **2,** e01501.

Subramaniam, A.R., Pan, T., and Cluzel, P. (2013b) Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc Natl Acad Sci USA* **110:** 2419–2424.

Sugita, T., and Nakase, T. (1999) Non-universal usage of the leucine CUG codon and the molecular phylogeny of the genus Candida. *Syst Appl Microbiol* **22:** 79–86.

Walsh, C.T., Garneau-Tsodikova, S., and Gatto, G.J. (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew Chem Int Ed Engl* **44:** 7342–7372.

Walter, K.U., Vamvaca, K., and Hilvert, D. (2005) An active enzyme constructed from a 9-amino acid alphabet. *J Biol Chem* **280:** 37742–37746.

Wiltschi, B., Merkel, L., and Budisa, N. (2009) Fine tuning the N-terminal residue excision with methionine analogues. *ChemBioChem* **10:** 217–220.

Wiser, M.J., Ribeck, N., and Lenski, R.E. (2013) Long-term dynamics of adaptation in asexual populations. *Science* **342:** 1364–1367.

Woese, C.R. (1965) On the evolution of the genetic code. *Proc Natl Acad Sci USA* **54:** 1546–1552.

Woese, C.R., Kandler, O., and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* **87:** 4576–4579.

Wong, J. (1975) A co-evolution theory of the genetic code. *Proc Natl Acad Sci USA* **72:** 1909–1912.

Wong, J. (1981) Coevolution of genetic code and amino acid biosynthesis. *TIBS* **6:** 33–36.

Wong, J. (1983) Membership mutation of the genetic code: loss of fitness by tryptophan. *Proc Natl Acad Sci USA* **80:** 6303–6306.

Wong, J. (2005) Coevolution theory of the genetic code at age thirty. *BioEssays* **27:** 416–425.

Wright, T.H., Vallee, M.R.J., and Davis, B.G. (2016) From chemical mutagenesis to post-expression mutagenesis: a 50 year odyssey. *Angew Chem Int Ed Engl* **55:** 5896–5903.

Xie, J., Liu, W., and Schultz, P.G. (2007) A genetically encoded bidentate, metal-binding amino acid. *Angew Chem Int Ed Engl* **46:** 9239–9242.