HHS Public Access

Author manuscript

Nat Rev Microbiol. Author manuscript; available in PMC 2016 November 01.

Published in final edited form as:

Nat Rev Microbiol. 2015 November ; 13(11): 707–721. doi:10.1038/nrmicro3568.

Genetic code flexibility in microorganisms: novel mechanisms and impact on physiology

Jiqiang Ling[#1], **Patrick O'Donoghue**[#2,3], and **Dieter Söll**[4,5]

[1]Department of Microbiology and Molecular Genetics, University of Texas Health Science Center, Houston, Texas 77030, USA.

[2]Department of Biochemistry, The University of Western Ontario, London, Ontario N6A 5C1, Canada.

[3]Department of Chemistry, The University of Western Ontario, London, Ontario N6A 5C1, Canada.

[4]Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520–8114, USA.

[5]Department of Chemistry, Yale University, New Haven, Connecticut 06520–8114, USA.

[#] These authors contributed equally to this work.

Abstract

The genetic code, initially thought to be universal and immutable, is now known to contain many variations, including biased codon usage, codon reassignment, ambiguous decoding and recoding. As a result of recent advances in the areas of genome sequencing, biochemistry, bioinformatics and structural biology, our understanding of genetic code flexibility has advanced substantially in the past decade. In this Review, we highlight the prevalence, evolution and mechanistic basis of genetic code variations in microorganisms, and we discuss how this flexibility of the genetic code affects microbial physiology.

Pioneering work with *Escherichia coli* in the 1960s[1,2] 'cracked' the genetic code, which constitutes a cornerstone of molecular biology. The genetic code is established through accurate ligation of each amino acid to its correct (cognate) tRNA by a specific aminoacyl-tRNA synthetase (aaRS). The resulting aminoacyl-tRNA (aa-tRNA) products of the aaRSs read codons by codon–anticodon pairing between mRNA and tRNA on the ribosome, allowing precise translation of the genetic information from mRNA to protein (FIG. 1).

Originally, it was thought that all living organisms used a universal set of codons, with 61 of the possible 64 nucleotide triplets translating 20 amino acids (termed sense codons) and the

Correspondence to D.S. dieter.soll@yale.edu.

Competing interests statement

The authors declare no competing interests.

SUPPLEMENTARY INFORMATION

See online article: S1 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

3 remaining codons (UAA, UAG and UGA) being responsible for termination of protein synthesis (termed stop codons). However, it was later discovered that the mitochondrial genetic code of yeast deviates from the standard code, with CUN codons assigned to Thr instead of Leu, and UGA used to encode Trp[3,4]. Later sequencing, bioinformatics and biochemical studies of the protein synthesis machinery in microorganisms provided further insights into genetic code variations[5] and revealed deviations from the standard genetic code in various microorganisms, including bacteria, archaea, fungi and viruses (see Supplementary information S1 (table)). These genetic code variants retain many features of the standard code, but the known exceptions are diverse, and they have evolved through distinct biochemical mechanisms (see Supplementary information S1 (table)). In this Review, we have broadly categorized these mechanisms into four types: biased codon usage, codon reassignment, ambiguous decoding and natural genetic code expansion (FIG. 1).

Codon degeneracy allows each amino acid to be decoded by more than one codon. These synonymous codons can be recognized by different tRNA isoacceptors. Some synonymous codons are preferentially used over others at higher frequencies, leading to biased codon usage, which is found in almost all sequenced genomes. Optimal codon usage typically correlates with high protein synthesis rates, particularly for highly transcribed genes[6]. Below, we provide updated views of how codon usage affects translation and the resulting biological consequences that are only beginning to be understood[7].

Codon reassignment completely changes the meaning of a codon throughout the transcriptome. The most common example of codon reassignment occurs in microorganisms in which stop codons were reassigned to encode amino acids, but sense codons have also been reassigned. Codon reassignment events have been identified in viruses and in a wide range of microorganisms, including microbial eukaryotes and their mitochondria. Below, we focus on recently characterized codon reassignment mechanisms in yeast and bacterial systems.

Ambiguous decoding (also known as mistranslation) refers to simultaneous decoding of the same codon by two or more amino acids in one cellular compartment. Ambiguous decoding can be caused by recognition of the same tRNA by more than one aaRS or by misacylation of a tRNA with a non-cognate amino acid, which both result in different amino acids being loaded onto tRNAs recognizing the same codon. It can also be caused by ribosomal decoding errors, in which tRNAs are mismatched to their assigned codons. Because ambiguous decoding can lead to errors in protein synthesis, it is usually regarded as deleterious, but increasing evidence suggests that this is a widespread mechanism in nature and may increase microbial fitness under certain conditions, such as during stress[8]. Recent advances uncovering the causes and physiological impact of ambiguous decoding are discussed below.

Natural genetic code expansion refers to genetic codes that enable protein synthesis with more than the 20 canonical amino acids. The two known naturally evolved examples include selenocysteine (Sec) and pyrrolysine (Pyl). Whereas codon reassignment occurs at a genome-wide level, recoding refers to site-specific codon re-definition that is dependent on the mRNA sequence context[9]. One of the earliest recognized deviations of the standard

genetic code involves recoding by frameshifting[10], which changes the readout of mRNAs (for an extensive review, see REF. 9). Below, we discuss the mechanisms and physiological consequences of reassigning the stop codon UAG to Pyl and recoding the UGA codon to Sec in archaea and bacteria.

In this Review, we provide an update on recent studies of genetic code flexibility in microorganisms, with a focus on novel insights into the mechanisms of biased codon usage, codon reassignment, ambiguous decoding and recoding, and we discuss how these evolutionary events affect cellular fitness.

Biased codon usage

The most widespread mechanism of genetic code flexibility is the biased usage of synonymous codons, which is a universal feature of all microbial genomes (FIG. 2). Biased codon usage was first observed in a comparison of the degenerate codons of sequenced mRNAs, revealing that each gene in a genome exhibited similar preference for certain synonymous codons over others[11]. For example, in *E. coli*, Arg codons CGC and CGU occur over ten times more frequently than AGA and AGG. Later studies show that codon bias is present in almost all sequenced genomes[6]. Both mutation (neutral) and selection hypotheses have been proposed to explain codon usage bias (these hypotheses are reviewed in REFS 6,12). According to the mutation hypothesis, codon bias results from mutational pressure that may vary in different organisms, leading to different codon usage across various genomes. By contrast, the selection hypothesis proposes that codon bias provides a selective advantage by optimizing the level and accuracy of protein expression. A combination of both mutation and selection hypotheses has led to a more widely accepted mutation–selection–drift model[13], which posits that major (abundant) codons that are frequently used are preferred during selection, whereas mutational pressure keeps the minor (rare) codons in the genome.

Biased codon usage has been shown to affect the expression levels of endogenous and heterologous proteins, and usage of rare codons is generally considered to slow down translation[6]. However, several recent studies have reshaped our view of ‘optimal’ codon usage and protein expression. Ribosome profiling, which measures ribosomal densities at all codons in the transcriptome, of *E. coli* and *Bacillus subtilis* grown in rich media showed that rare codons recognized by tRNAs with low cellular abundance are translated at the same speed as abundant codons (FIG. 2). Instead of stalling at rare codons, ribosomes stall most often at Shine–Dalgarno-like sequences that may pair with the anti-Shine–Dalgarno sequence of the 16S ribosomal RNA[14]. Another study used a yellow fluorescent protein in *E. coli* to show that the protein synthesis rate is not affected by synonymous changes in abundant or rare codons in bacteria growing in amino acid-rich media[15]. High-throughput translation assays showed that the presence of rare codons located near the 5′ end of an open reading frame surprisingly increase protein expression about fourfold in *E. coli*[16]; this is primarily due to a decrease in the number of mRNA secondary structures, which facilitates the travelling of the ribosome along the mRNA.

Under amino acid starvation conditions, however, synonymous codons are indeed translated at distinct rates owing to the different aminoacylation levels of the corresponding tRNA isoacceptors[15] (FIG. 2). It seems that the abundance of aa-tRNA, rather than the total level of tRNA (either charged or uncharged with amino acids), correlates better with the decoding rates *in vivo*. Under nutrient-rich conditions, the supply of aa-tRNAs meets the demand of protein synthesis, and all synonymous codons appear to be translated at similar speeds. By contrast, during amino acid starvation, tRNA isoacceptors are charged at various levels, leading to distinct translation rates of synonymous codons. Similarly, other conditions that perturb the relative abundances of charged tRNA isoacceptors, such as alterations in the tRNA expression level and aminoacylation efficiency, may also affect the translation rates of synonymous codons. In addition, non-optimal codons in over-expressed genes could sequester rare tRNAs, therefore reducing the charging levels of such tRNAs to decrease total protein expression. Given that microorganisms in the natural environment are not often supplied with sufficient nutrients, maintaining optimal codons with biased usage of synonymous codons probably has an important role in maximizing the overall translational speed and growth.

Although optimized usage of synonymous codons increases the protein synthesis rate under certain conditions, non-optimal codons used at specific locations sometimes lead to increased bacterial fitness. For example, in *B. subtilis*, non-optimal Ser codons are preferentially used in the gene encoding SinR, which inhibits biofilm formation[17]. Experimental substitution of non-optimal Ser codons with optimal Ser codons increases the synthesis rate of SinR, which leads to increased protein levels of SinR and suppresses biofilm development. Under physiological conditions, depletion of cellular Ser levels during stationary phase reduces translation of the native SinR to trigger biofilm synthesis. Therefore, the apparent non-optimal Ser codons in SinR serve as a molecular sensor for the cellular levels of Ser, and this strategy enables *B. subtilis* to control the expression of biofilm-associated genes in response to environmental conditions.

In the cold-adapted cyanobacterium *Synechococcus elongatus*, the *kaiB* and *kaiC* genes are critical for the regulation of circadian rhythms, and these genes are usually highly expressed[18]. The codon adaptation index of *kaiB* and *kaiC* is below average, which means that *kaiB* and *kaiC* are enriched in non-optimal codons. Notably, experimental codon optimization of *kaiB* and *kaiC* increases the protein levels of KaiB and KaiC at low temperature, which prevents the circadian rhythm from switching off, resulting in loss of fitness. Therefore, these data demonstrate that *S. elongatus* uses non-optimal codons as a mechanism to regulate bacterial fitness (FIG. 2). Similarly, in *Neurospora crassa*, the *frq* gene, which controls the circadian clock function, also exhibits non-optimal codon usage[19]. Optimization of the codons of *frq* increases the protein levels of Frq but also leads to protein folding defects and impairs its function in the circadian feedback loop. This suggests that the non-optimal codons serve as checkpoints to slow down translation of certain regions in Frq, allowing proper folding of the protein (FIG. 2). In these examples, non-optimal codons actually provide microorganisms with selective advantages for adaptation to environmental changes.

Codon reassignment

Codon reassignment events include stop-to-sense, sense-to-stop and sense-to-sense codon changes, and occur through multiple mechanisms (see Supplementary information S1 (table)). Whereas sense-to-stop reassignments only require loss of the cognate tRNA, stop-to-sense and sense-to-sense changes require evolution of new tRNAs. Such new tRNAs may arise from modification of an existing tRNA or by tRNA duplication[20]. For example, reassignment of the UGA codon from stop to Trp is accompanied by either a mutation[21] or post-transcriptional editing (C34→U34)[22] that converts $tRNA^{Trp}_{CCA}$ to $tRNA^{Trp}_{UCA}$, which reads both the UGA and UGG codons as Trp. In bacteria and archaea, both $tRNA^{Met}_{CAU}$ and $tRNA^{Ile}_{CAU}$ contain the same CAU anticodon. Whereas $tRNA^{Met}_{CAU}$ recognizes the AUG codon, $tRNA^{Ile}_{CAU}$ is modified at position C34 by addition of lysidine[23] or agmatidine[24,25]. As a result, the modified $tRNA^{Ile}_{CAU}$ specifically recognizes AUA instead of AUG. In the mitochondria, $tRNA^{Ile}_{CAU}$ is absent and $tRNA^{Met}_{CAU}$ contains a formylcytidine (f^5C) modification at C34 (REF. 26), which allows $tRNA^{Met}_{CAU}$ to recognize A at the third position (wobble position). A recent structural study confirmed that f^5C of $tRNA^{Met}_{CAU}$ pairs with both A and G at the wobble position, enabling it to read both AUA and AUG codons as Met[27]. Consequently, AUA codons are reassigned from Ile to Met in the mitochondria.

A more complicated scenario involves tRNA evolution from duplication. For example, CUN codons are reassigned from Leu to Thr in *Saccharomyces cerevisiae* mitochondria owing to the loss of $tRNA^{Leu}_{UAG}$ and the presence of $tRNA^{Thr}_{UAG}$, which has an enlarged eight-nucleotide (as opposed to seven-nucleotide) anticodon loop[3]. An unmodified U at the wobble position in mitochondrial tRNAs allows pairing with all four nucleotides[28]. Interestingly, phylogenetic analyses reveal that $tRNA^{Thr}_{UAG}$ is not related to either the ancestral $tRNA^{Leu}_{UAG}$ or the isoacceptor $tRNA^{Thr}_{UGU}$ but shows high similarity to the mitochondrial $tRNA^{His}_{GUG}$ (REF. 29). Further biochemical studies verify that two nucleotide changes are sufficient to switch the $tRNA^{His}_{GUG}$ from a His-accepting to Thr-accepting molecule[29]. These results suggest that the $tRNA^{Thr}_{UAG}$ is evolutionarily derived from a precursor $tRNA^{His}_{GUG}$. In support of this model, a duplicated copy of $tRNA^{His}_{GUG}$ is found in the mitochondrial genome of *Candida albicans* and *Kluyveromyces lactis*, which are closely related to *S. cerevisiae* (FIG. 3).

In another closely related species, *Ashbya gossypii*, a $tRNA^{Ala}_{UAG}$ derived from either $tRNA^{His}_{GUG}$ or $tRNA^{Thr}_{UAG}$ reassigns mitochondrial CUA and CUU codons to Ala, and CUC and CUG codons are absent in mitochondrial genes[30]. This leaves UUG and UUA as the only Leu codons in the *A. gossypii* mitochondrial genome.

In addition to tRNA, aaRSs have evolved to enable codon reassignment, as highlighted by CUN reassignment in *S. cerevisiae* mitochondria. The *S. cerevisiae* threonyl-tRNA synthetase (ThrRS) recognizes both $tRNA^{Thr}_{UAG}$ and $tRNA^{Thr}_{UGU}$ (REF. 31), whereas *E. coli* and the mitochondrial ThrRSs from *Schizosaccharomyces pombe* and *C. albicans* do not recognize $tRNA^{Thr}_{UAG}$, suggesting that the mitochondrial ThrRS from *S. cerevisiae* has co-evolved with $tRNA^{Thr}_{UAG}$ to allow for CUN recoding (FIG. 3).

Together, these data suggest that CUN reassignment in *S. cerevisiae* and in *A. gossypii* involved several steps. First, in the mitochondrial genome of some Saccharomycetaceae species (such as *C. albicans*), the $tRNA^{His}_{GUG}$ gene was duplicated. The CUN codons and a $tRNA^{Leu}_{UAG}$ that decodes CUN then disappeared, leading to a reduced genetic code that does not involve CUN, such as in *K. lactis*. Following this reduction, in *S. cerevisiae*, one copy of $tRNA^{His}_{GUG}$ evolved to carry a UAG anticodon that reads CUN codons ($tRNA^{His}_{UAG}$). The ThrRS co-evolved with the $tRNA^{His}_{UAG}$ to recognize it as a $tRNA^{Thr}_{UAG}$. CUN codons then reappeared to complete the codon reassignment from Leu to Thr. In *A. gossypii*, secondary mutations enabled the new $tRNA^{His}_{UAG}$ to be recognized by AlaRS instead of HisRS or ThrRS, thereby generating $tRNA^{Ala}_{UAG}$ and reassigning CUN codons to Ala in this species (FIG. 3).

The driving forces and physiological consequences of codon reassignment are not completely understood. For example, reassignment of CUN codons in yeast mitochondria is considered to be neutral because the original CUN codons had been lost, new CUN codons appeared at non-conserved positions, and a new tRNA evolved, which apparently did not involve a period of ambiguous translation of mitochondrial proteins. However, a recent bioinformatic analysis of >5 trillion bp of metagenomic data revealed that stop codon reassignment is widespread in microorganisms and that differential reassignment between bacteria and phages could have physiological consequences[32]. For example, the authors found a phage from the human oral cavity that reassigned the UAG stop codon to Gln. Interestingly, the phage genes expressed at early stages contain few in-frame UAG codons and can therefore be efficiently translated by the host machinery. However, one of the phage proteins expressed during the early stages is release factor 2 (RF2), which suppresses translation of UGA codons. As the bacterial protein RF1, which suppresses translation of UAG codons, contains multiple in-frame UGA codons, expression of RF2 by the phage inhibits translation of host RF1. Inhibition of host RF1 allows the phage to translate late-stage phage genes, which are enriched in recoded UAG codons (FIG. 3). This intriguing model suggests that phages can use codon reassignment to interfere with translation of host genes without affecting translation of the phage genes, but this model needs to be experimentally validated in future studies.

Insight into the consequences of codon reassignment also comes from studies of microorganisms in which the genetic code has been re-engineered (BOX 1). Several groups have deleted RF1 in an attempt to reassign UAG to a sense codon in *E. coli* by subsequently introducing an orthogonal aaRS–tRNA[33–38] pair to incorporate a non-canonical amino acid. Notably, when the native UAG codons are present, deleting RF1 and reassigning UAG as a sense codon causes severe growth defects[34,35]. Converting all endogenous UAG codons to UAA rescues the growth phenotype[36], suggesting that the observed toxicity is due to extension of proteins beyond their intended stop. In the RF1 deletion mutant, UAG is also translated by natural amino acids (predominantly Gln, Tyr and Lys), which compete with the non-canonical amino acid insertions, resulting in impure proteins[39]. These studies indicate that radical codon reassignment may decrease fitness in some aspects, such as growth and proteome stability. Therefore, during the evolution of codon reassignment, the gained benefits may need to outweigh the potential negative effects. In one such example, phages

growing in *E. coli* cells that reassign UAG from stop to a non-canonical amino acid (3-iodotyrosine) rapidly accumulate in-frame UAG codons from mutations; such speeded codon reassignment in the phage provides benefits to the phage by increasing its titre owing to the presence of the non-canonical amino acid in the proteome[40]. An expanded genetic code can thus increase the evolvability of codons. Future studies are needed to identify the phenotypic and proteomic changes in organisms in which the genetic code has been re-engineered and to clarify the underlying molecular mechanisms of how a cell adapts to genetic code expansion.

Ambiguous decoding

In contrast to biased codon usage in which a set of codons are translated by the same amino acid, a specific codon can also be ambiguously decoded by more than one amino acid[41]. Ambiguous decoding can result from multiple aaRSs recognizing the same tRNA, which leads to different amino acids being loaded onto tRNAs recognizing the same codon; from errors in the aminoacylation reaction carried out by a specific aaRS, which loads a tRNA with a non-cognate amino acid; or from ribosomal decoding errors (FIG. 4). Ambiguous decoding leads to a statistical pool of protein products with amino acid substitutions at various positions (statistical proteome). High levels of ambiguous translation lead to accumulation of misfolded proteins that are toxic to cells.

Multiple quality control mechanisms are used to ensure fidelity during protein synthesis. For instance, some aaRSs use editing sites to hydrolyse mismatched aa-tRNAs[42]. Similarly, the ribosome selects the cognate aa-tRNA based on both preferential binding of the cognate aa-tRNA at the decoding centre and kinetic proofreading of non-cognate aa-tRNAs[43,44]. However, despite the existence of these complex mechanisms that maintain translational fidelity (reviewed in REFS 42,44), ambiguous decoding can still result from reduced fidelity during protein synthesis owing to either mutations[44] or stress[45]. For instance, mutations in the ribosome allow mismatch between the mRNA codon and tRNA anticodon, and oxidative stress decreases fidelity during aminoacylation.

Increasing evidence suggests that ambiguous decoding may be used as an adaptive mechanism by microorganisms to survive harsh environmental conditions and gain a selective advantage during evolution. Some aaRSs use editing to ensure correct pairing between amino acids and tRNAs[46], but the editing domains of ThrRS, LeuRS and PheRS are either lost or defunct in *Mycoplasma* spp. and in mitochondria from yeasts[47–49], indicating that ambiguous translation occurs in these organisms and organelles. Indeed, analyses of mass spectrometry data reveal that the intracellular bacterium *Mycoplasma mobile* encodes a statistical proteome[48], which was suggested to benefit this organism by increasing the proteomic and phenotypic diversity of the bacterium, enabling it to escape host defences[48].

Direct experimental evidence of phenotypic diversity created by a statistical proteome comes from elegant studies of CUG codon ambiguity in *C. albicans*. Altering the ratio of Ser to Leu incorporation at CUG codons introduces diverse cell and colony morphologies, as well as distinct antifungal and immune responses[50,51]. Wild-type *C. albicans* encodes a

CUG-decoding $tRNA_{CAG}$ that is recognized by both SerRS and LeuRS (FIG. 1b). The wild-type cells display colony morphologies, including smooth, ring, wrinkled and hyphae[50]. Substituting the ambiguous $tRNA_{CAG}$ with a Leu-specific tRNA abolishes CUG ambiguity as well as the smooth and ring morphologies[50] (FIG. 4). Enhanced CUG ambiguity increases cell adhesion by ambiguous translation of adhesins and reduces susceptibility to macrophage killing by decreasing surface exposure of β-glucans[51]. It has also been shown that, in *Mycobacterium smegmatis*, the level of ambiguous decoding at Asn codons increases during stationary phase and under low pH conditions[52]. Ambiguous translation increases resistance against rifampicin by producing RpoB variant proteins that are no longer recognized by this antibiotic, therefore epigenetically enhancing the fitness of *M. smegmatis*[52] (FIG. 4).

Ambiguous decoding not only increases the diversity of the proteome but may also activate stress responses. A recent study reveals that ambiguous translation caused by a mutation in the ribosome activates the bacterial general stress response and enhances tolerance to hydrogen peroxide[53]. In *E. coli*, oxidative stress induces ambiguous decoding at ACN codons, which are recognized by $tRNA^{Thr}$. ThrRS uses editing to hydrolyse misacylated Ser-$tRNA^{Thr}$, and the editing site of ThrRS contains an activated Cys that is hypersensitive to oxidation[54]. Oxidative stress thus impairs ThrRS editing and causes incorporation of both Thr and Ser into ACN codons (FIG. 4). The reactive Cys and activating residues at the ThrRS editing sites are conserved among bacteria, suggesting that oxidative-stress-induced ambiguous decoding is a highly conserved mechanism. Switching between accurate and ambiguous decoding may allow bacteria to adapt to different environmental conditions quickly, but this possibility requires additional testing in future studies.

Natural genetic code expansion

In 1966, Crick stated: “A more serious problem is whether in a normal cell a triplet can be read in more than one way.” (REF. 55). Natural genetic code expansion encompasses changes in the genetic code — which evolved in natural organisms — that enable protein synthesis with more than the 20 canonical amino acids. Sec and Pyl are the only known cases of natural genetic code expansion. The mechanisms of Sec and Pyl translation are distinct and rarely co-occur in the same organism (BOX 2).

Recoding UGA to Sec

Most UGA codons signal termination of protein synthesis. Certain organisms, including *E. coli* and humans, have a naturally expanded genetic code and recode some UGA codons to insert Sec into selenoproteins. Therefore, UGA takes on two meanings in the same cell and sometimes even in the same open reading frame[56], illustrating the “more serious problem” foreshadowed by Crick[55]. These naturally expanded genetic codes specify 21 rather than the ‘usual’ 20 amino acids.

Sec is chemically similar to Cys, but the lower redox potential and higher nucleophilicity of Sec explain why some selenoenzymes are far more reactive than their Cys-containing counterparts[57]. For example, mammalian Met-*R*-sulfoxide reductases containing Sec in the active site are 100-fold more active than the Cys-containing variants[58]. Sec is also found to

be more resistant to irreversible oxidation than Cys[59] and has utility as a probe for protein structure and function[60].

Sec biosynthesis occurs on its cognate tRNA, beginning with a Ser-tRNA^{Sec} precursor (reviewed in REF. 61) (BOX 2). Particular UGAs are selected for recoding to Sec by the presence of a downstream Sec insertion sequence (SECIS) (FIG. 5). A special elongation factor (SelB) that binds to both the SECIS and Sec-tRNA^{Sec} on the ribosome is also required (BOX 2). Given the complexity of UGA recoding to Sec, it is not surprising that codon recoding is rare in nature. The benefit (and significance) of recoding, as opposed to codon reassignment, is that nature found a way to expand the genetic code that does not interfere with normal protein synthesis.

Sec is not hardwired to UGA codons. Recent experiments demonstrate that the Sec machinery is able to recode nearly every codon in the genetic code table[62]. A tRNA molecule was engineered to insert Sec at the UAG codon using elongation factor Tu (EF-Tu), but even with optimal expression of the Sec synthesis components, the precursor Ser-tRNA^{Sec} was incorporated only 30% of the time[63]. Ambiguous decoding of Ser and Sec in this SECIS-independent route to selenoprotein synthesis supports the notion that the elaborate Sec recoding machinery may have evolved to enhance the fidelity of Sec insertion into proteins[64]. Further tRNA refinements created synthetic tRNAs[65,66] that, together with EF-Tu[67] mutants, led to efficient site-specific Sec incorporation[64].

Furthermore, Sec-tRNA^{Sec} is capable of wobble decoding, and a Sec-tRNA$^{Sec}_{CCA}$ mutant has been shown to efficiently decode the UGA codon in a thioredoxin reductase[68]. In *E. coli*, mutation of the tRNA^{Sec} to each of the other 63 anticodons shows that Sec can be inserted in response to 60 of the codons in the genetic code[62,66]. In fact, for 15 codons, Sec-tRNA^{Sec} can completely out-compete cognate aa-tRNA and provide selenoprotein yields that are tenfold greater than when Sec is encoded by UGA[62]. These findings indicate that microorganisms could encode Sec with a codon other than UGA and bring into question as to why UGA is the ‘chosen’ Sec codon. It is, therefore, plausible that even in nature Sec is not hardwired to UGA.

Interestingly, Sec recoding is found in all three domains of life, but not in all organisms. Sec is normally found in redox enzymes, and these selenoproteins have important roles in central energy metabolism in archaea and bacteria as well as in cellular defence against reactive oxygen species.

Approximately 20% of sequenced bacterial taxa encode Sec[69]. Some bacteria encode only a single selenoprotein, whereas others such as the δ-proteobacterium *Syntrophobacter fumaroxidans* encode 31 (REF. 69). Bacterial selenoproteins belong to 58 protein families[70]. The number of selenoprotein genes in bacterial species differs greatly, in part because replacement of Sec with Cys is common in the evolution of bacterial selenoproteins[69]. The majority of bacterial selenoproteins, including formate dehydrogenase-α subunit (FdhA), Gly reductase B and glutathione peroxidase, are homologous to thiol-based redox enzymes[69,71]. The most common bacterial selenoproteins are selenophosphate synthase (SelD) (BOX 2) and FdhA. *E. coli* formate dehydrogenase (FDH_H) is one of the most

studied selenoproteins[72]. FDH_H is a component of the formate hydrogen lyase complex and functions to reduce formate and shuttle electrons to the respiratory chain. The Sec residue in FDH_H is essential for its function, and replacement with Cys reduces the catalytic efficiency of the enzyme 200-fold[73].

Bacterial selenoproteins are involved in other central metabolic pathways, including purine degradation, and the energy-conserving acetogenesis pathway[74]. Environmental sequencing of microbial genomes revealed many new potential selenoproteins and novel selenoprotein families, such as Ser proteases and formamidase regulatory proteins, in which the role of Sec is unknown[70].

Early work demonstrated that Sec is dispensable in *E. coli* and *Salmonella enterica* subsp. *enterica* serovar Typhimurium[75], but our understanding of how natural expansion of the genetic code with Sec affects bacterial physiology is still limited. A recent report found that concentrations of sodium selenite (a necessary precursor of Sec biosynthesis) in the millimolar range were detrimental to *E. coli* growth, but concentrations of 1–10 nM stimulated cell growth[76]. Certain bacterial pathogens (such as *Staphylococcus aureus*) rely completely on the selenoprotein thioredoxin reductase to survive oxidative stress[77]. Further work will be required to understand the selective value of genetic code expansion with Sec in bacteria.

Sec-decoding archaea are confined to closely related species[78], including *Methanopyrus kandleri* and all members of the order Methanococcales. The archaeal selenophosphate synthase is a selenoprotein, but most archaeal selenoproteins are involved in methanogenesis, which is the main energy production pathway for these microorganisms[78]. Sec is found in FDH_H, formylmethano furan dehydrogenase, F_{420} reducing and non-reducing hydrogenases, heterodisulfide reductase and a HesB-like protein, which has been implicated in iron–sulfur cluster assembly[79–81]. These selenoproteins are involved in redox reactions and cofactor regeneration, facilitating the transfer of electrons in the reduction of formate to produce methane.

Genetic experiments in *Methanococcus maripaludis* indicate that archaeal selenoproteins are more active than Cys-containing homologues and that the metabolic range and efficiency of the organism decreases when their genetic code is reduced from 21 to 20 amino acids[82,83]. In *M. maripaludis* JJ, genetic inactivation of the Sec-specific elongation factor SelB leads to overexpression of Cys-containing versions of each of its selenoproteins, suggesting that the Sec version is catalytically more active[82]. *M. maripaludis* JJ cannot sustain growth on formate because there is no Cys paralogue for FDH_H, and a 15% decrease in growth rate was observed in cells consuming H_2 and CO_2.

Reassigning UAG to Pyl

Pyl was first discovered in methanogenic archaea and identified as the twenty-second genetically encoded amino acid[84,85] (BOX 2). A recent search of the US National Center for Biotechnology Information and the US Joint Genome Institute sequence databases shows that there are 46 microbial species (21 archaea and 25 bacteria) that contain all of the genes that are necessary to synthesize Pyl-containing proteins. Recently, a new clade of Pyl-

decoding methanogens was discovered in Methanomassiliicoccales species related to the euryarchaeal order Thermoplasmatales[86]. The sequences hint at potential new roles for Pyl in CRISPR-associated protein Cas1 and digeranylgeranylglyceryl phosphate synthase[87], but it is still debatable whether all UAG codons in Pyl-decoding organisms lead to Pyl insertion.

The biochemical mechanism required to genetically encode Pyl is distinct and orthogonal to the system that evolved to insert Sec into proteins. Free Pyl is biosynthesized in the cell from two molecules of Lys by three enzymes that are the products of the *pylB, pylC* and *pylD* genes[88]. Structural and biochemical work on Pyl biosynthesis was recently reviewed[89]. Pyl is then ligated onto $tRNA^{Pyl}$ by PylRS[90,91]. The CUA anticodon of $tRNA^{Pyl}$ reads the UAG codon and reassigns the meaning of UAG from stop to Pyl (FIG. 5).

Natural incorporation of Pyl was first experimentally verified in a monomethylamine methyltransferase (MtmB)[85], when crystallographic structures of Pyl-containing MtmB (purified from *Methanosarcina acetivorans*) provided an atomic-level view of Pyl in the active site[84]. A hypothetical structure-based reaction mechanism suggested that Pyl may function to properly orient the methylamine substrate for transfer of the methyl group to a cognate corrinoid protein (MtmC). Related bacterial methyltransferases that lack Pyl are now known to use Gly betaine as a substrate[92], a quaternary amine and a precursor of the tertiary amine trimethylamine (TMA). The authors argue that, because Gly betaine is a quaternary amine, it is already an ‘activated’ methyl donor, possibly providing a clue as to the role of Pyl in the methylamine methyltransferases. In the proposed reaction scheme[84], Pyl may form an adduct with TMA, converting TMA into a quaternary amine intermediate that is activated for methyl transfer.

Pyl is, however, not restricted to the active site of methyltransferases and was also identified and characterized in the tRNA-editing enzyme $tRNA^{His}$ guanylyltransferase (Thg1) from *M. acetivorans*[93], in which Pyl is only required to read through an in-frame UAG codon in the Thg1 mRNA to generate a full-length protein; thus, Pyl serves as a ‘normal’ amino acid without a catalytic role. Furthermore, a recent proteomic investigation confirmed UAG translation or Pyl insertion in eight additional proteins in *M. acetivorans*[94]. These proteins included methylcorrinoid:coenzyme M (CoM) methyltransferase (MtaA), His kinase, tRNA endonuclease, hypothetical proteins and methylornithine synthase (PylB), which is the enzyme that catalyses the initial step of Pyl biosynthesis. PylB is folded[95] and active without Pyl[88], but the 21 amino acid extension following Pyl readthrough may modulate PylB activity or participate in sensing or regulating Pyl synthesis.

Translation of UAG codons in *M. acetivorans* may involve competition between release factor and $PyltRNA^{Pyl}$ readthrough, and it is possible that some UAG codons are translated more efficiently as Pyl than others. In a *M. acetivorans* $tRNA^{Pyl}$ deletion mutant, three instances of UAG codons serving as stop codons were detected by mass spectrometry, with none detected in the wild-type strain[94]. The fact that UAG codes for stop in the $tRNA^{Pyl}$ deletion mutant indicates that the *M. acetivorans* release factor has activity towards the UAG codons. Indeed, an *in vitro* study using chimeric *Methanosarcina barkeri*–human release factor in an *in vitro* translation system found that the *M. barkeri* release factor had activity towards all three stop codons (UAG, UAA and UGA), but significantly reduced

activity for UAG[96]. Although no natural *M. acetivorans* transcripts have been shown to use UAG as a stop codon in wild-type cells, previous work using an *E. coli* β-glucuronidase gene reporter with a UAG codon showed translation with Pyl (20%) and stopping at UAG (80%) in wild-type *M. acetivorans*[97]. Competition with release factor is not unique to Pyl. Trp decoding at UGG is well known to compete with release factor[62,98], and release factor is also active at SECIS-recoded UGA codons that specify Sec[62,99]. Despite the fact that competition exists between translational readthrough and termination, Trp and Sec codons are not considered to be stop codons. The above studies indicate that UAG codons are not recoded to Pyl at specific loci, as is the case for UGA codons recoded to Sec (see above). Future studies will provide a more detailed picture of Pyl decoding, but the preponderance of the evidence indicates that in native *M. acetivorans* transcripts UAG is reassigned to Pyl.

Pyl recoding has physiological impacts on both archaea and bacteria. For example, in genetic studies, a *M. acetivorans* strain lacking $tRNA^{Pyl}$ was unable to grow on TMA[94,100]. This observation reaffirmed the notion that Pyl is evolutionarily connected to methylamine metabolism[101]. The fact that all Pyl-decoding organisms encode at least one methylamine methyltransferase with an in-frame UAG codon (placing Pyl in the active site of these enzymes)[84] suggests that the selective value of Pyl may be related to its role in methylamine metabolism. However, all organisms that putatively encode Pyl also have UAG codons in mRNAs that direct production of other proteins. Protein families encoded in the *M. acetivorans* genome and enriched with Pyl residues include transposases, recombinases, methylamine and methylcorrinoid:CoM methyltransferases, methyltransferases (with unknown substrates), radical *S*-adenosylmethionine (SAM) enzymes and His kinases involved in two-component signalling[94].

In *M. acetivorans*, there are 267 in-frame UAG codons in annotated genes, and Pyl is intricately linked with the proteome, cellular metabolism and fitness. Growth of a $tRNA^{Pyl}$ deletion strain in minimal medium containing methanol as the sole carbon source resulted in significant increases in generation and lag time compared with the wild-type strain[94]. This phenotypic defect corresponded to ~350 differentially abundant (mostly non-Pyl-containing) proteins. In the deletion mutant, the abundance of several enzymes required for methanogenesis from methanol were reduced 20-fold, indicating lower metabolic efficiency with methanol. Methyltransferases involved in methanogenesis from dimethylsulfides[102] increased 100-fold, suggesting an altered or compensatory metabolism[94]. Although anabolic or other defects in the deletion mutant cannot be ruled out, the data show Pyl has a far-reaching impact on the composition of the proteome and suggest that *M. acetivorans* is metabolically less efficient when the organism has a reduced genetic code.

The Pyl-decoding trait exists in 25 anaerobic bacteria from the Firmicutes phylum (in classes Clostridia and Negativicutes) and the δ-Proteobacteria class. Early work in the field showed that *E. coli* expressed Pyl-containing proteins with a recombinant Pyl-decoding system, indicating that bacteria are capable of decoding UAG codons as Pyl[90,103]. However, there has been little work to show whether Pyl is actively decoded in bacteria in which the Pyl trait naturally occurs.

An investigation into several of these bacteria uncovered a novel genetic code variation referred to as dynamic genetic code expansion[104]. Several bacterial strains thought to genetically encode Pyl and at least one TMA methyltransferase (MttB) with one in-frame UAG codon were examined for the ability to produce Pyl-tRNA^{Pyl} and Pyl-containing MttB. The Pyl-decoding bacterium *Acetohalobium arabaticum* genetically encodes 20 amino acids when grown on pyruvate; however, when grown on TMA, the cells dynamically expand their genetic code to 21 amino acids with Pyl[104]. Only in the presence of TMA did *A. arabaticum* synthesize Pyl-tRNA^{Pyl}, produce Pyl-containing MttB and consume TMA. In the absence of TMA, *A. arabaticum* transcriptionally silences the gene encoding PylRS and downregulates *pylB*, *pylC* and *pylD* transcription. The mechanism by which *A. arabaticum* senses TMA and initiates the expression of the Pyl-decoding system remains unknown. Interestingly, despite encoding a functional PylRS[105,106] and tRNA^{Pyl} pair[107], *Desulfitobacterium hafniense* did not show detectable production of Pyl-tRNA^{Pyl} or expression of MttB. *Desulfitobacterium dehalogenans* expressed transcripts for tRNA^{Pyl} and PylRS, but no detectable transcription of the Pyl biosynthesis genes (*pylB*, *pylC* and *pylD*) was observed. This suggests that Pyl is either not decoded at all or simply not decoded under the conditions tested. Collectively, these data show two different ways that codon usage can be adapted to genetic code expansion with Pyl. Pyl-decoding archaea severely limit their use of UAG codons (~5% of all annotated genes 'stop' at a UAG codon)[87,104], whereas many other organisms, including all Pyl-decoding bacteria[104], use UAG codons abundantly (~25% of all annotated genes 'stop' at a UAG codon). Furthermore, the recently sequenced Thermoplasmatales-related methanogens display an extremely reduced use of UAG codons, as low as 1.6%[87]. The Pyl machinery is thought to be the evolutionary driving force that maintains low UAG codon usage in these species[87]. The archaeal strategy seems to be to limit UAG codons to locations where Pyl is required or at least tolerated. By contrast, the bacteria examined so far seem to be able to silence Pyl when Pyl-containing proteins are not needed, which may allow these species to use higher levels of UAG codons.

Conclusion and outlook

Recent developments in genome sequencing, biochemistry, bioinformatics and structural biology have facilitated the discovery of new genetic codes and provided crucial tools to analyse and gain insight into the physiological impact of genetic code flexibility. In this Review, we categorized the mechanisms of genetic code variation into four types — biased codon usage, codon reassignment, ambiguous decoding and natural genetic code expansion (FIG. 1b) — but this is certainly an incomplete list of the genetic code diversity in the living world. Increasing evidence suggests that genetic code flexibility is a selected trait during evolution to benefit microorganisms under certain conditions. However, little is known about the molecular mechanisms that lead to the improved fitness as a result of genetic code variation.

Our improved understanding of natural genetic code variations has also fostered a new exciting research area in expanding the genetic code with non-canonical amino acids and engineering synthetic organisms with new genetic codes (BOX 1). Single-cell model microorganisms with advanced genetic systems, such as *E. coli* and *S. cerevisiae*, serve as major vehicles for such engineering efforts. However, even in these well-studied organisms,

we are only beginning to understand the physiological changes brought about by engineered genetic code variations.

For each genetic code variant described above, far more is known about how the genetic code changed (in terms of evolutionary or biochemical mechanisms) than about why the genetic code changed or the potential selection pressures that drive and maintain these changes. Therefore, these are areas where there is still great potential for future experimentation. Molecular genetic approaches have already revealed much about genetic code variants in terms of effects on growth, proteome status and cellular metabolism. Furthermore, engineered synthetic organisms with very different genetic codes may lay the basis for systematic experiments to explore the selective value and physiological impact of genetic code evolution.

Our knowledge of decoding mechanisms is far from complete. These gaps in knowledge have presented major obstacles to understanding the evolution of the genetic code and engineering synthetic cells with expanded codes. We envision that with advances in single-cell genome sequencing and systems biology, further genetic code variations will be uncovered. This will provide new insights into genetic code flexibility and novel tools for engineering organisms with altered genetic codes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Work in the authors' laboratories was supported by grants from the US National Institute of General Medical Sciences (GM022854 to D.S.; and GM115431 to J.L.), from the Natural Sciences and Engineering Research Council of Canada (RGPIN 04282–2014 to P.O.), from the Canadian Institutes of Health Research Tier 2 Canada Research Chair (950-229917 to P.O.), and from The University of Texas Health Science Center at Houston start-up fund (to J.L.). The authors are grateful to I. Heinemann for discussions and a critical reading of the manuscript.

Glossary

Aminoacyl-tRNA	(aa-tRNA). A tRNA molecule with an amino acid attached to the 3′ end. It is used as a substrate by the ribosome to synthesize proteins.
Codon–anticodon pairing	During translation, the bases of the mRNA codon and the tRNA anticodon need to match each other. Watson–Crick pairing (A–U and G–C) in the first and second positions of the codon is required for efficient decoding, whereas the third position allows more flexible pairing, for example, between G and U or using modified bases.
Synonymous codons	Different triplet nucleotide sequences that decode the same amino acid.
tRNA isoacceptors	Different tRNA species recognized by the same aminoacyl-tRNA synthetase and ligated with the same amino acid.

Misacylation	Incorrect pairing of an amino acid and tRNA by an aminoacyl-tRNA synthetase. Errors resulting from misacylation, if left uncorrected, reduce the overall translational fidelity.
Frameshifting	Change in the reading frame during translation due to mutations in the DNA, errors during transcription or translation or specific mRNA structures, leading to new protein sequences.
Shine–Dalgarno-like sequences	mRNA sequences that share high similarity with the Shine–Dalgarno sequence, which pairs with the anti-Shine–Dalgarno sequence of the ribosomal RNA.
Codon adaptation index	A method for analysing usage bias of synonymous codons using a set of highly expressed genes from a species as a reference to assign a score to each gene.
Wobble position	The third position of a codon, which is more flexibly recognized by the tRNA compared with other positions.
Phages	(Also called bacteriophages). Viruses that infect and propagate within bacteria. Phages contain their own genome but hijack the translational machinery of the bacterial host for protein synthesis.
RpoB	The β-subunit of the bacterial RNA polymerase and target of the antibiotic rifampicin.
Nucleophilicity	The property to donate an electron in chemical reactions.
Elongation factor Tu	(EF-Tu). A bacterial elongation factor that delivers aminoacyl-tRNAs to the ribosome during peptide synthesis. The counterpart of EF-Tu in archaea and eukaryotes is EF-1A

References

1. Nirenberg M, et al. RNA codewords and protein synthesis, VII. On the general nature of the RNA code. Proc. Natl Acad. Sci. USA. 1965; 53:1161–1168. [PubMed: 5330357]
2. Söll D, et al. Studies on polynucleotides, XLIX. Stimulation of the binding of aminoacyl-sRNA's to ribosomes by ribotrinucleotides and a survey of codon assignments for 20 amino acids. Proc. Natl Acad. Sci. USA. 1965; 54:1378–1385. [PubMed: 5325653]
3. Li M, Tzagoloff A. Assembly of the mitochondrial membrane system: sequences of yeast mitochondrial valine and an unusual threonine tRNA gene. Cell. 1979; 18:47–53. [PubMed: 389433]
4. Macino G, Coruzzi G, Nobrega FG, Li M, Tzagoloff A. Use of the UGA terminator as a tryptophan codon in yeast mitochondria. Proc. Natl Acad. Sci. USA. 1979; 76:3784–3785. [PubMed: 226981] First discovery of codon reassignment in microorganisms.
5. Ambrogelly A, Palioura S, Söll D. Natural expansion of the genetic code. Nat. Chem. Biol. 2007; 3:29–35. [PubMed: 17173027]
6. Plotkin JB, Kudla G. Synonymous but not the same: the causes and consequences of codon bias. Nat. Rev. Genet. 2011; 12:32–42. [PubMed: 21102527]
7. Quax TE, Claassens NJ, Söll D, van der Oost J. Codon bias as a means to fine-tune gene expression. Mol. Cell. 2015; 59:149–161. [PubMed: 26186290]

8. Pan T. Adaptive translation as a mechanism of stress response and adaptation. Annu. Rev. Genet. 2013; 47:121–137. [PubMed: 23988117] Excellent review on benefits of ambiguous decoding under stress conditions.
9. Atkins JF, Baranov PV. The distinction between recoding and codon reassignment. Genetics. 2010; 185:1535–1536. [PubMed: 20713743]
10. Atkins JF, Gesteland RF, Reid BR, Anderson CW. Normal tRNAs promote ribosomal frameshifting. Cell. 1979; 18:1119–1131. [PubMed: 391405] One of the first studies to demonstrate frameshifting and genetic code variability.
11. Grantham R, Gautier C, Gouy M, Mercier R, Pave A. Codon catalog usage and the genome hypothesis. Nucleic Acids Res. 1980; 84:r49–r62. [PubMed: 6986610]
12. Shabalina SA, Spiridonov NA, Kashina A. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic Acids Res. 2013; 41:2073–2094. [PubMed: 23293005]
13. Bulmer M. The selection-mutation-drift theory of synonymous codon usage. Genetics. 1991; 129:897–907. [PubMed: 1752426]
14. Li GW, Oh E, Weissman JS. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. Nature. 2012; 484:538–541. [PubMed: 22456704]
15. Subramaniam AR, Pan T, Cluzel P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. Proc. Natl Acad. Sci. USA. 2013; 110:2419–2424. [PubMed: 23277573]
16. Goodman DB, Church GM, Kosuri S. Causes and effects of N-terminal codon bias in bacterial genes. Science. 2013; 342:475–479. [PubMed: 24072823]
17. Subramaniam AR, et al. A serine sensor for multicellularity in a bacterium. eLife. 2013; 2:e01501. [PubMed: 24347549]
18. Xu Y, et al. Non-optimal codon usage is a mechanism to achieve circadian clock conditionality. Nature. 2013; 495:116–120. [PubMed: 23417065] Suggested that non-optimal codons are used as a regulatory mechanism.
19. Zhou M, et al. Non-optimal codon usage affects expression, structure and function of clock protein FRQ. Nature. 2013; 495:111–115. [PubMed: 23417067] Suggested that non-optimal codons are used as a regulatory mechanism.
20. Lang, BF.; Lavrov, D.; Beck, N.; Steinberg, SV. Organelle Genetics. Bullerwell, CE., editor. Springer; 2012. p. 431-474.
21. Sengupta S, Yang X, Higgs PG. The mechanisms of codon reassignments in mitochondrial genetic codes. J. Mol. Evol. 2007; 64:662–688. [PubMed: 17541678]
22. Alfonzo JD, Blanc V, Estevez AM, Rubio MA, Simpson L. C to U editing of the anticodon of imported mitochondrial $tRNA^{Trp}$ allows decoding of the UGA stop codon in *Leishmania tarentolae*. EMBO J. 1999; 18:7056–7062. [PubMed: 10601027] Demonstrated that RNA editing is responsible for UGA reassignment in *Leishmania tarentolae* mitochondria.
23. Muramatsu T, et al. A novel lysine-substituted nucleoside in the first position of the anticodon of minor isoleucine tRNA from *Escherichia coli*. J. Biol. Chem. 1988; 263:9261–9267. [PubMed: 3132458]
24. Mandal D, et al. Agmatidine, a modified cytidine in the anticodon of archaeal $tRNAI^{Ile}$, base pairs with adenosine but not with guanosine. Proc. Natl Acad. Sci. USA. 2010; 107:2872–2877. [PubMed: 20133752]
25. Ikeuchi Y, et al. Agmatine-conjugated cytidine in a tRNA anticodon is essential for AUA decoding in archaea. Nat. Chem. Biol. 2010; 6:277–282. [PubMed: 20139989]
26. Tomita K, et al. Codon reading patterns in *Drosophila melanogaster* mitochondria based on their tRNA sequences: a unique wobble rule in animal mitochondria. Nucleic Acids Res. 1999; 27:4291–4297. [PubMed: 10518623]
27. Cantara WA, Murphy FV, Demirci H, Agris PF. Expanded use of sense codons is regulated by modified cytidines in tRNA. Proc. Natl Acad. Sci. USA. 2013; 110:10964–10969. [PubMed: 23781103]

28. Marniemi J, Parkki MG. Radiochemical assay of glutathione *S*-epoxide transferase and its enhancement by phenobarbital in rat liver *in vivo*. Biochem. Pharmacol. 1975; 24:1569–1572. [PubMed: 9]
29. Su D, et al. An unusual $tRNA^{Thr}$ derived from $tRNA^{His}$ reassigns in yeast mitochondria the CUN codons to threonine. Nucleic Acids Res. 2011; 39:4866–4874. [PubMed: 21321019]
30. Ling J, Daoud R, Lajoie MJ, Söll D, Lang BF. Natural reassignment of CUU and CUA sense codons to alanine in *Ashbya* mitochondria. Nucleic Acids Res. 2014; 42:499–508. [PubMed: 24049072]
31. Ling J, et al. Yeast mitochondrial threonyl-tRNA synthetase recognizes tRNA isoacceptors by distinct mechanisms and promotes CUN codon reassignment. Proc. Natl Acad. Sci. USA. 2012; 109:3281–3286. [PubMed: 22343532]
32. Ivanova NN, et al. Stop codon reassignments in the wild. Science. 2014; 344:909–913. [PubMed: 24855270] Uncovered widespread stop codon reassignment events in microorganisms.
33. Mukai T, et al. Codon reassignment in the *Escherichia coli* genetic code. Nucleic Acids Res. 2010; 38:8188–8195. [PubMed: 20702426]
34. Johnson DB, et al. Release factor one is nonessential in *Escherichia coli*. ACS Chem. Biol. 2012; 7:1337–1344. [PubMed: 22662873]
35. Heinemann IU, et al. Enhanced phosphoserine insertion during *Escherichia coli* protein synthesis via partial UAG codon reassignment and release factor 1 deletion. FEBS Lett. 2012; 586:3716–3722. [PubMed: 22982858]
36. Lajoie MJ, et al. Genomically recoded organisms expand biological functions. Science. 2013; 342:357–360. [PubMed: 24136966] Created the first synthetic microorganism with complete removal of UAG stop codons.
37. Mukai T, et al. Highly reproductive *Escherichia coli* cells with no specific assignment to the UAG codon. Sci. Rep. 2015; 5:9699. [PubMed: 25982672]
38. Mukai T, et al. Reassignment of a rare sense codon to a non-canonical amino acid in *Escherichia coli*. Nucleic Acids Res. 2015; 43 http://dx.doi.org/10.1093/nar/gkv787.
39. Aerni HR, Shifman MA, Rogulina S, O'Donoghue P, Rinehart J. Revealing the amino acid composition of proteins within an expanded genetic code. Nucleic Acids Res. 2015; 43:e8. [PubMed: 25378305]
40. Hammerling MJ, et al. Bacteriophages use an expanded genetic code on evolutionary paths to higher fitness. Nat. Chem. Biol. 2014; 10:178–180. [PubMed: 24487692]
41. Schultz DW, Yarus M. Transfer RNA mutation and the malleability of the genetic code. J. Mol. Biol. 1994; 235:1377–1380. [PubMed: 8107079]
42. Ling J, Reynolds N, Ibba M. Aminoacyl-tRNA synthesis and translational quality control. Annu. Rev. Microbiol. 2009; 63:61–78. [PubMed: 19379069]
43. Rodnina MV, Wintermeyer W. Ribosome fidelity: tRNA discrimination, proofreading and induced fit. Trends Biochem. Sci. 2001; 26:124–130. [PubMed: 11166571]
44. Zaher HS, Green R. Fidelity at the molecular level: lessons from protein synthesis. Cell. 2009; 136:746–762. [PubMed: 19239893]
45. Netzer N, et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. Nature. 2009; 462:522–526. [PubMed: 19940929]
46. Mascarenhas, AP.; An, S.; Rosen, AE.; Martinis, SA.; Musier-Forsyth, K. Protein Engineering. RajBhandary, UL.; Köhrer, C., editors. Springer; 2008. p. 153-200.
47. Roy H, Ling J, Alfonzo J, Ibba M. Loss of editing activity during the evolution of mitochondrial phenylalanyl-tRNA synthetase. J. Biol. Chem. 2005; 280:38186–38192. [PubMed: 16162501]
48. Li L, et al. Naturally occurring aminoacyl-tRNA synthetases editing-domain mutations that cause mistranslation in *Mycoplasma* parasites. Proc. Natl Acad. Sci. USA. 2011; 108:9378–9383. [PubMed: 21606343] Suggested that *Mycoplasma* spp. may use ambiguous decoding to defend against the host immune response.
49. Yadavalli SS, Ibba M. Selection of tRNA charging quality control mechanisms that increase mistranslation of the genetic code. Nucleic Acids Res. 2013; 41:1104–1112. [PubMed: 23222133]

50. Bezerra AR, et al. Reversion of a fungal genetic code alteration links proteome instability with genomic and phenotypic diversification. Proc. Natl Acad. Sci. USA. 2013; 110:11079–11084. [PubMed: 23776239]
51. Miranda I, et al. *Candida albicans* CUG mistranslation is a mechanism to create cell surface variation. mBio. 2013; 4:e00285–13. [PubMed: 23800396]
52. Javid B, et al. Mycobacterial mistranslation is necessary and sufficient for rifampicin phenotypic resistance. Proc. Natl Acad. Sci. USA. 2014; 111:1132–1137. [PubMed: 24395793] Demonstrated that ambiguous decoding increases resistance to an antibiotic in mycobacteria.
53. Fan Y, et al. Protein mistranslation protects bacteria against oxidative stress. Nucleic Acids Res. 2015; 43:1740–1748. [PubMed: 25578967]
54. Wu J, Fan Y, Ling J. Mechanism of oxidant-induced mistranslation by threonyl-tRNA synthetase. Nucleic Acids Res. 2014; 42:6523–6531. [PubMed: 24744241]
55. Crick FH. The Croonian lecture, 1966. The genetic code. Proc. R. Soc. Lond. B. 1967; 167:331–347. [PubMed: 4382798]
56. Turanov AA, et al. Genetic code supports targeted insertion of two amino acids by one codon. Science. 2009; 323:259–261. [PubMed: 19131629]
57. Arner ES. Selenoproteins — what unique properties can arise with selenocysteine in place of cysteine? Exp. Cell Res. 2010; 316:1296–1303. [PubMed: 20206159]
58. Kim HY, Gladyshev VN. Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-*R*-sulfoxide reductases. PLoS. Biol. 2005; 3:e375. [PubMed: 16262444]
59. Snider GW, Ruggles E, Khan N, Hondal RJ. Selenocysteine confers resistance to inactivation by oxidation in thioredoxin reductase: comparison of selenium and sulfur enzymes. Biochemistry. 2013; 52:5472–5481. [PubMed: 23865454]
60. Metanis N, Hilvert D. Natural and synthetic selenoproteins. Curr. Opin. Chem. Biol. 2014; 22:27–34. [PubMed: 25261915]
61. Yoshizawa S, Böck A. The many levels of control on bacterial selenoprotein synthesis. Biochim. Biophys. Acta. 2009; 1790:1404–1414. [PubMed: 19328835] **An excellent review of bacterial selenoproteins.**
62. Bröcker MJ, Ho JM, Church GM, Söll D, O'Donoghue P. Recoding the genetic code with selenocysteine. Angew. Chem. Int. Ed. Engl. 2014; 53:319–323. [PubMed: 24511637]
63. Aldag C, et al. Rewiring translation for elongation factor Tu-dependent selenocysteine incorporation. Angew. Chem. Int. Ed. Engl. 2013; 52:1441–1445. [PubMed: 23193031]
64. Su D, et al. How an obscure archaeal gene inspired the discovery of selenocysteine biosynthesis in humans. IUBMB Life. 2009; 61:35–39. [PubMed: 18798524]
65. Thyer R, Robotham SA, Brodbelt JS, Ellington AD. Evolving $tRNA^{Sec}$ for efficient canonical incorporation of selenocysteine. J. Am. Chem. Soc. 2015; 137:46–49. [PubMed: 25521771]
66. Miller C, et al. A synthetic tRNA for EF-Tu mediated selenocysteine incorporation *in vivo* and *in vitro*. FEBS Lett. 2015; 589:2194–2199. [PubMed: 26160755]
67. Haruna K, Alkazemi MH, Liu Y, Söll D, Englert M. Engineering the elongation factor Tu for efficient selenoprotein synthesis. Nucleic Acids Res. 2014; 42:9976–9983. [PubMed: 25064855]
68. Xu J, Croitoru V, Rutishauser D, Cheng Q, Arner ES. Wobble decoding by the *Escherichia coli* selenocysteine insertion machinery. Nucleic Acids Res. 2013; 41:9800–9811. [PubMed: 23982514]
69. Zhang Y, Romero H, Salinas G, Gladyshev VN. Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. Genome Biol. 2006; 7:R94. [PubMed: 17054778] Although many new sequences are now available, this is still the definitive resource that documents the evolutionary replacement of Cys with Sec residues in bacterial proteins.
70. Zhang Y, Gladyshev VN. Trends in selenium utilization in marine microbial world revealed through the analysis of the global ocean sampling (GOS) project. PLoS Genet. 2008; 4:e1000095. [PubMed: 18551170]
71. Stadtman TC. Selenocysteine. Annu. Rev. Biochem. 1996; 65:83–100. [PubMed: 8811175]

72. Jormakka M, Byrne B, Iwata S. Formate dehydrogenase — a versatile enzyme in changing environments. Curr. Opin. Struct. Biol. 2003; 13:418–423. [PubMed: 12948771]
73. Axley MJ, Böck A, Stadtman TC. Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. Proc. Natl Acad. Sci. USA. 1991; 88:8450–8454. [PubMed: 1924303]
74. Stock T, Rother M. Selenoproteins in Archaea and Gram-positive bacteria. Biochim. Biophys. Acta. 2009; 1790:1520–1532. [PubMed: 19344749]
75. Stadtman TC, Davis JN, Zehelein E, Böck A. Biochemical and genetic analysis of *Salmonella typhimurium* and *Escherichia coli* mutants defective in specific incorporation of selenium into formate dehydrogenase and tRNAs. Biofactors. 1989; 2:35–44. [PubMed: 2679652]
76. Tetteh AY, et al. Transcriptional response of selenopolypeptide genes and selenocysteine biosynthesis machinery genes in *Escherichia coli* during selenite reduction. Int. J. Microbiol. 2014; 2014:394835. [PubMed: 24839442]
77. Lu J, Holmgren A. The thioredoxin antioxidant system. Free Radic. Biol. Med. 2014; 66:75–87. [PubMed: 23899494]
78. Rother M, Krzycki JA. Selenocysteine, pyrrolysine, and the unique energy metabolism of methanogenic archaea. Archaea. 2010; 2010:453642. [PubMed: 20847933]
79. Stock T, Selzer M, Rother M. *In vivo* requirement of selenophosphate for selenoprotein synthesis in archaea. Mol. Microbiol. 2010; 75:149–160. [PubMed: 19919669]
80. Kryukov GV, Gladyshev VN. The prokaryotic selenoproteome. EMBO Rep. 2004; 5:538–543. [PubMed: 15105824]
81. Cobucci-Ponzano B, Rossi M, Moracci M. Translational recoding in archaea. Extremophiles. 2012; 16:793–803. [PubMed: 23015064]
82. Rother M, Mathes I, Lottspeich F, Böck A. Inactivation of the *selB* gene in *Methanococcus maripaludis*: effect on synthesis of selenoproteins and their sulfur-containing homologs. J. Bacteriol. 2003; 185:107–114. [PubMed: 12486046] A seminal study on the phenotypic impact of removing Sec from the genetic code of the model archaeaon *M. maripaludis*.
83. Hohn MJ, Palioura S, Su D, Yuan J, Söll D. Genetic analysis of selenocysteine biosynthesis in the archaeon *Methanococcus maripaludis*. Mol. Microbiol. 2011; 81:249–258. [PubMed: 21564332]
84. Hao B, et al. A new UAG-encoded residue in the structure of a methanogen methyltransferase. Science. 2002; 296:1462–1466. [PubMed: 12029132] Discovery of the twenty-second genetically encoded amino acid, Pyl.
85. Srinivasan G, James CM, Krzycki JA. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science. 2002; 296:1459–1462. [PubMed: 12029131]
86. Borrel G, et al. Genome sequence of '*Candidatus Methanomethylophilus alvus*' Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. J. Bacteriol. 2012; 194:6944–6945. [PubMed: 23209209]
87. Borrel G, et al. Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine. BMC Genomics. 2014; 15:679. [PubMed: 25124552]
88. Gaston MA, Zhang L, Green-Church KB, Krzycki JA. The complete biosynthesis of the genetically encoded amino acid pyrrolysine from lysine. Nature. 2011; 471:647–650. [PubMed: 21455182] Elucidated activities of the biosynthetic route to Pyl.
89. Krzycki JA. The path of lysine to pyrrolysine. Curr. Opin. Chem. Biol. 2013; 17:619–625. [PubMed: 23856058]
90. Blight SK, et al. Direct charging of $tRNA_{CUA}$ with pyrrolysine *in vitro* and *in vivo*. Nature. 2004; 431:333–335. [PubMed: 15329732] **Elucidated the mechanism of Pyl decoding.**
91. Polycarpo C, et al. An aminoacyl-tRNA synthetase that specifically activates pyrrolysine. Proc. Natl Acad. Sci. USA. 2004; 101:12450–12454. [PubMed: 15314242] **Elucidated the mechanism of Pyl decoding.**
92. Ticak T, Kountz DJ, Girosky KE, Krzycki JA, Ferguson DJ Jr. A nonpyrrolysine member of the widely distributed trimethylamine methyltransferase family is a glycine betaine methyltransferase. Proc. Natl Acad. Sci. USA. 2014; 111:e4668–e4676. [PubMed: 25313086]

93. Heinemann IU, et al. The appearance of pyrrolysine in $tRNA^{His}$ guanylyltransferase by neutral evolution. Proc. Natl Acad. Sci. USA. 2009; 106:21103–21108. [PubMed: 19965368]
94. O'Donoghue P, et al. Reducing the genetic code induces massive rearrangement of the proteome. Proc. Natl Acad. Sci. USA. 2014; 111:17206–17211. [PubMed: 25404328] Provided proteome-level view of the phenotypic impact of removing Pyl from the genetic code of *M. acetivorans*.
95. Quitterer F, List A, Eisenreich W, Bacher A, Groll M. Crystal structure of methylornithine synthase (PylB): insights into the pyrrolysine biosynthesis. Angew. Chem. Int. Ed. Engl. 2012; 51:1339–1342. [PubMed: 22095926]
96. Alkalaeva E, et al. Translation termination in pyrrolysine-utilizing archaea. FEBS Lett. 2009; 583:3455–3460. [PubMed: 19796638]
97. Longstaff DG, Blight SK, Zhang L, Green-Church KB, Krzycki JA. *In vivo* contextual requirements for UAG translation as pyrrolysine. Mol. Microbiol. 2007; 63:229–241. [PubMed: 17140411]
98. Freistroffer DV, Kwiatkowski M, Buckingham RH, Ehrenberg M. The accuracy of codon recognition by polypeptide release factors. Proc. Natl Acad. Sci. USA. 2000; 97:2046–2051. [PubMed: 10681447]
99. Mansell JB, Guevremont D, Poole ES, Tate WP. A dynamic competition between release factor 2 and the $tRNA^{Sec}$ decoding UGA at the recoding site of *Escherichia coli* formate dehydrogenase H. EMBO J. 2001; 20:7284–7293. [PubMed: 11743004]
100. Mahapatra A, et al. Characterization of a *Methanosarcina acetivorans* mutant unable to translate UAG as pyrrolysine. Mol. Microbiol. 2006; 59:56–66. [PubMed: 16359318]
101. Krzycki JA. Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases. Curr. Opin. Chem. Biol. 2004; 8:484–491. [PubMed: 15450490]
102. Oelgeschlager E, Rother M. *In vivo* role of three fused corrinoid/methyl transfer proteins in *Methanosarcina acetivorans*. Mol. Microbiol. 2009; 72:1260–1272. [PubMed: 19432805]
103. Polycarpo CR, et al. Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. FEBS Lett. 2006; 580:6695–6700. [PubMed: 17126325]
104. Prat L, et al. Carbon source-dependent expansion of the genetic code in bacteria. Proc. Natl Acad. Sci. USA. 2012; 109:21070–21075. [PubMed: 23185002] Demonstrated natural Pyl decoding in bacteria and revealed first example of dynamic genetic code expansion.
105. Jiang R, Krzycki JA. PylSn and the homologous N-terminal domain of pyrrolysyl-tRNA synthetase bind the tRNA that is essential for the genetic encoding of pyrrolysine. J. Biol. Chem. 2012; 287:32738–32746. [PubMed: 22851181]
106. Katayama H, Nozawa K, Nureki O, Nakahara Y, Hojo H. Pyrrolysine analogs as substrates for bacterial pyrrolysyl-tRNA synthetase *in vitro* and *in vivo*. Biosci. Biotechnol. Biochem. 2012; 76:205–208. [PubMed: 22232266]
107. Nozawa K, et al. Pyrrolysyl-tRNA synthetase-$tRNA^{Pyl}$ structure reveals the molecular basis of orthogonality. Nature. 2009; 457:1163–1167. [PubMed: 19118381]
108. O'Donoghue P, Ling J, Wang YS, Söll D. Upgrading protein synthesis for synthetic biology. Nat. Chem. Biol. 2013; 9:594–598. [PubMed: 24045798]
109. Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annu. Rev. Biochem. 2010; 79:413–444. [PubMed: 20307192]
110. Chin JW. Reprogramming the genetic code. EMBO J. 2011; 30:2312–2324. [PubMed: 21602790]
111. Johnson JA, Lu YY, Van Deventer JA, Tirrell DA. Residue-specific incorporation of non-canonical amino acids into proteins: recent developments and applications. Curr. Opin. Chem. Biol. 2010; 14:774–780. [PubMed: 21071259]
112. Chin JW. Expanding and reprogramming the genetic code of cells and animals. Annu. Rev. Biochem. 2014; 83:379–408. [PubMed: 24555827] An excellent review on engineering protein synthesis for genetic code expansion in diverse expression systems.
113. Rovner AJ, et al. Recoded organisms engineered to depend on synthetic amino acids. Nature. 2015; 518:89–93. [PubMed: 25607356]
114. Mandell DJ, et al. Biocontainment of genetically modified organisms by synthetic protein design. Nature. 2015; 518:55–60. [PubMed: 25607366]

115. Campbell JH, et al. UGA is an additional glycine codon in uncultured SR1 bacteria from the human microbiota. Proc. Natl Acad. Sci. USA. 2013; 110:5540–5545. [PubMed: 23509275]
116. Wang HH, et al. Programming cells by multiplex genome engineering and accelerated evolution. Nature. 2009; 460:894–898. [PubMed: 19633652]
117. Isaacs FJ, et al. Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement. Science. 2011; 333:348–353. [PubMed: 21764749] First example of a genome engineered with 62 codons by mutation of all TAGs to TAA.
118. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]
119. Wiedenheft B, Sternberg SH, Doudna JA. RNA-guided genetic silencing systems in bacteria and archaea. Nature. 2012; 482:331–338. [PubMed: 22337052]
120. Gibson DG, et al. Creation of a bacterial cell controlled by a chemically synthesized genome. Science. 2010; 329:52–56. [PubMed: 20488990] First demonstration of genome transplantation with a synthetic genome.
121. Annaluru N, et al. Total synthesis of a functional designer eukaryotic chromosome. Science. 2014; 344:55–58. [PubMed: 24674868]
122. Budiman ME, et al. Eukaryotic initiation factor 4a3 is a selenium-regulated RNA-binding protein that selectively inhibits selenocysteine incorporation. Mol. Cell. 2009; 35:479–489. [PubMed: 19716792]
123. Hohn MJ, Park HS, O'Donoghue P, Schnitzbauer M, Söll D. Emergence of the universal genetic code imprinted in an RNA record. Proc. Natl Acad. Sci. USA. 2006; 103:18095–18100. [PubMed: 17110438]
124. Bilokapic S, et al. Structure of the unusual seryl-tRNA synthetase reveals a distinct zinc-dependent mode of substrate recognition. EMBO J. 2006; 25:2498–2509. [PubMed: 16675947]
125. Itoh Y, et al. Decameric SelA•tRNA^{Sec} ring structure reveals mechanism of bacterial selenocysteine formation. Science. 2013; 340:75–78. [PubMed: 23559248]
126. Sherrer RL, O'Donoghue P, Söll D. Characterization and evolutionary history of an archaeal kinase involved in selenocysteinyl-tRNA formation. Nucleic Acids Res. 2008; 36:1247–1259. [PubMed: 18174226]
127. Carlson BA, et al. Identification and characterization of phosphoseryl-tRNA$^{[Ser]Sec}$ kinase. Proc. Natl Acad. Sci. USA. 2004; 101:12848–12853. [PubMed: 15317934]
128. Palioura S, Sherrer RL, Steitz TA, Söll D, Simonovic M. The human SepSecS–tRNA^{Sec} complex reveals the mechanism of selenocysteine formation. Science. 2009; 325:321–325. [PubMed: 19608919]
129. Copeland PR, Fletcher JE, Carlson BA, Hatfield DL, Driscoll DM. A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. EMBO J. 2000; 19:306–314. [PubMed: 10637234]
130. Bifano AL, Atassi T, Ferrara T, Driscoll DM. Identification of nucleotides and amino acids that mediate the interaction between ribosomal protein L30 and the SECIS element. BMC Mol. Biol. 2013; 14:12. [PubMed: 23777426]
131. Allmang C, Wurth L, Krol A. The selenium to selenoprotein pathway in eukaryotes: more molecular partners than anticipated. Biochim. Biophys. Acta. 2009; 1790:1415–1423. [PubMed: 19285539]
132. Yoshizawa S, et al. Structural basis for mRNA recognition by elongation factor SelB. Nat. Struct. Mol. Biol. 2005; 12:198–203. [PubMed: 15665870]
133. Yuan J, et al. Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems. FEBS Lett. 2010; 584:342–349. [PubMed: 19903474]
134. Eggertsson G, Söll D. Transfer ribonucleic acid-mediated suppression of termination codons in *Escherichia coli*. Microbiol. Rev. 1988; 52:354–374. [PubMed: 3054467]
135. Ambrogelly A, et al. Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. Proc. Natl Acad. Sci. USA. 2007; 104:3141–3146. [PubMed: 17360621]
136. Kavran JM, et al. Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. Proc. Natl Acad. Sci. USA. 2007; 104:11268–11273. [PubMed: 17592110]

137. Wan W, et al. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. Angew. Chem. Int. Ed. Engl. 2010; 49:3211–3214. [PubMed: 20340150]

138. Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N^{ϵ}-acetyllysine in recombinant proteins. Nat. Chem. Biol. 2008; 4:232–234. [PubMed: 18278036]

139. Umehara T, et al. *N*-acetyl lysyl-tRNA synthetases evolved by a CcdB-based selection possess *N*-acetyl lysine specificity *in vitro* and *in vivo*. FEBS Lett. 2012; 586:729–733. [PubMed: 22289181]

140. Yanagisawa T, Umehara T, Sakamoto K, Yokoyama S. Expanded genetic code technologies for incorporating modified lysine at multiple sites. Chembiochem. 2014; 15:2181–2187. [PubMed: 25179816]

141. Guo LT, et al. Polyspecific pyrrolysyl-tRNA synthetases from directed evolution. Proc. Natl Acad. Sci. USA. 2014; 111:16724–16729. [PubMed: 25385624]

142. Lobanov AV, et al. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. Genome Biol. 2007; 8:r198. [PubMed: 17880704]

143. Grobe T, Reuter M, Gursinsky T, Sohling B, Andreesen JR. Peroxidase activity of selenoprotein GrdB of glycine reductase and stabilisation of its integrity by components of proprotein GrdE from *Eubacterium acidaminophilum*. Arch. Microbiol. 2007; 187:29–43. [PubMed: 17009022]

144. Hurley JM, Dunlap JC. Cell biology: a fable of too much too fast. Nature. 2013; 495:57–58. [PubMed: 23417066]

Box 1

Emerging biotechnology for engineering the genetic code

The flexibility of the genetic code as illustrated by natural codon reassignments has prompted researchers to expand the genetic code with non-canonical amino acids[108]. Both site-specific[109,110] and residue-specific[111] approaches have been developed to co-translationally insert non-canonical amino acids into proteins of interest. The site-specific approach uses an orthogonal pair of aminoacyl-tRNA synthetase (aaRS)–tRNA to reassign a stop or quadruplet codon to a non-canonical amino acid. By contrast, the residue-specific approach allows ambiguous decoding of a sense codon by both the native amino acid and a non-canonical amino acid. Site-specific sense codon recoding is also being explored as a means to further expand the genetic code[62]. Genetic code expansion with non-canonical amino acids has been applied to site-specifically label proteins with fluorophores, study post-translational modifications and identify newly synthesized proteins[109]. Given the flexibility of the active site and orthogonality in bacteria and eukaryotes, tyrosyl-tRNA synthetase (TyrRS) and PylRS have been most widely used for site-specific insertion of non-canonical amino acids (reviewed in REFS 108,112). Recently, non-canonical amino acids have been used as safeguards to contain recoded bacteria[113,114]. In such synthetic microorganisms, stop codons are introduced into essential genes that require translation with a non-canonical amino acid. Therefore, the viability of the recoded bacteria depends on the presence of the non-canonical amino acid, preventing them from proliferating in natural ecosystems. As an additional safety layer, stop-codon-interrupted genes should be less susceptible to horizontal gene transfer (HGT), thus helping to prevent transmission of genes from synthetic organisms to natural genomes. A similar mechanism may also occur in nature. The bacterium SR1 encodes Gly with five codons, including the UGA ‘stop’ codon. This genetic code variation was hypothesized to bias HGT, because 85% of SR1’s genes contain UGA sense codons, which would be interpreted as stop codons by recipient microorganisms[115].

Recent advances in genome editing and DNA synthesis provide powerful tools to engineer the genetic code and understand the evolution of codon reassignments. The goal of these efforts is to remove codons from the genome so that they can be reassigned to new and non-canonical amino acids. Multiplex automated genome engineering (MAGE)[116] and conjugative assembly genome engineering (CAGE)[117] approaches were developed to allow simultaneous and genome-wide editing of the *Escherichia coli* genome. MAGE uses synthetic single-stranded DNA fragments to simultaneously target multiple chromosomal sites and introduce precise genomic changes by recombination-based genetic engineering, whereas CAGE uses a conjugation strategy to move large regions of the chromosome between different strains. A combination of MAGE and CAGE has allowed complete reassignment of UAG to UAA in an *E. coli* strain[36]. Another potent genome-editing tool is based on the CRISPR system, which permits convenient genome editing in bacterial and eukaryotic cells[118,119]. Complete *de novo* synthesis of a *Mycoplasma* genome[120] and a yeast chromosome[121] have also been achieved. Together, these technological advances set the stage for synthetic organisms

(with potentially radically different genetic codes) that will provide novel opportunities for biotechnology and new insights into the evolution of the genetic code.

Box 2

Natural genetic code expansion

Selenocysteine (Sec) biosynthesis evolved twice independently[122] (FIG. 5). This is not in itself a unique feature of Sec, as two independently evolved aminoacylation systems exist for Lys, Gln, Asn, Gly, Cys[123] and possibly Ser[124]. The fact that Sec is exclusively biosynthesized on its tRNA, however, is unusual. Sec is also the only amino acid for which a specific aminoacyl-tRNA synthetase (aaRS) does not exist. Initially, the normal seryl-tRNA synthetase (SerRS) ligates Ser onto $tRNA^{Sec}$, and the Ser-$tRNA^{Sec}$ product is converted to Sec-$tRNA^{Sec}$ (FIG. 5). This reaction is catalysed by a single pyridoxal-phosphate (PLP)-dependent enzyme in bacteria (SelA), which forms a stunning decameric ring structure that simultaneously binds to ten $tRNA^{Sec}$ molecules[125]. In archaea[126] and eukaryotes[127], a unique kinase (phosphoseryl(Sep)-$tRNA^{Sec}$ kinase (PSTK)) phosphorylates the tRNA-bound Ser to form phosphoseryl (Sep)-$tRNA^{Sec}$, and a distant relative of SelA, a dimeric enzyme known as Sep-$tRNA^{Sec}$:Sec-$tRNA^{Sec}$ synthetase (SepSecS), produces the Sec-$tRNA^{Sec}$ substrate required to insert Sec into proteins. SepSecS is dimeric, like most members of the PLP enzyme family, and binds to two $tRNA^{Sec}$ molecules[128]. Both SelA and SepSecS require the selenium donor selenophosphate, which is the product of selenophosphate synthase (SelD), to complete the Sec synthesis reaction.

Adding to the complexity of selenoprotein biosynthesis, a specialized elongation factor (SelB in *Escherichia coli*, EF-Sec in eukaryotes) and a RNA recoding signal (Sec insertion sequence (SECIS)) are also required to convert the meaning of UGA from stop to Sec. In archaea and bacteria, the SECIS is found in the reading frame of the selenoprotein transcript, within a few nucleotides of the recoded UGA codon. In eukaryotes, the SECIS (of which there are two structurally divergent types) is found in the 3′ untranslated region (UTR), and additional proteins are required for Sec insertion (including SECIS-binding protein 2 (REF. 129)) or for regulation of the recoding event (including ribosomal protein L30 (REF. 130) and initiation factor 4a3 (REF. 122)). Despite numerous biochemical and structural studies on Sec synthesis and recoding in bacteria[61] and eukaryotes[131], the structural basis of UGA recoding on the ribosome remains unresolved. A structural model of the bacterial system[132] indicates that the SECIS–SelB–Sec-$tRNA^{Sec}$ complex may fit between the ‘body’ and ‘head’ regions of the 30S ribosomal subunit, which could alter the conformation of the decoding centre on the translating ribosome.

Following the discovery of Sec, pyrrolysine (Pyl) is the second known case of natural genetic code expansion. Genetic code expansion with Sec and Pyl arose by distinct evolutionary and biochemical mechanisms[133]. Unlike Pyl, which appears in proteins as either a catalytic[84] or non-catalytic[93] residue, Sec nearly always has essential roles in the active sites of redox enzymes[28]. Another feature that distinguishes Pyl from Sec is that incorporation of Sec requires a SECIS mRNA structure, whereas Pyl can effectively provide informational suppression[134] of UAG codons without a co-evolved mRNA secondary structure[97]. The flexibility of $tRNA^{Pyl}$ in UAG translation[135], together with the natural orthogonality of the PylRS–$tRNA^{Pyl}$ pair in bacterial and eukaryotic hosts,

and the large amino acid-binding pocket of PylRS[107,136], prompted multiple groups to engineer PylRS as a facile system for genetic incorporation of non-canonical amino acids in bacterial and mammalian cells[109,112,137–141]. The Sec system was recently shown to be a powerful mechanism for recoding sense codons[62]. We anticipate that both systems will continue to drive further genetic code engineering that is inspired by nature (BOX 1).

Fascinatingly, Sec and Pyl may coexist in some organisms. Two genome sequences, that of *Acetohalobium arabaticum* and one from metagenomic sequencing of a bacterial symbiont of the worm *Olavius algarvensis*[142] show evidence of genetic codes with 22 amino acids. Indeed, two Gly reductase selenoprotein B (*grdB*) genes in *A. arabaticum* contain both a SECIS-recoded UGA codon for Sec incorporation as well as one or two UAG codons, which may be translated as Pyl in the presence of trimethylamine. *A. arabaticum* was shown to decode UAG as Pyl[104] and because Sec in GrdB is required for Gly reductase activity[143], this organism may produce the only known natural proteins that contain 22 genetically encoded amino acids.

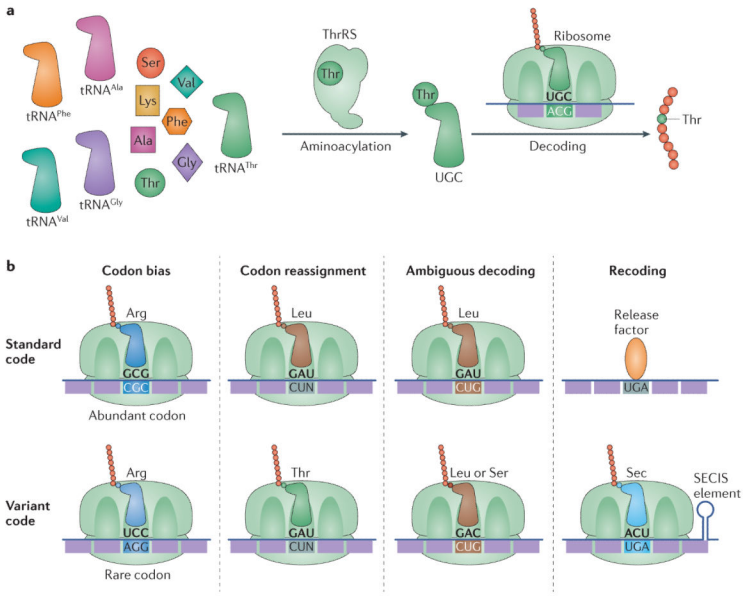

Figure 1. Mechanisms of genetic code flexibility

a | Each amino acid is attached to the corresponding tRNA by a specialized aminoacyl-tRNA synthetase (aaRS), in a reaction called aminoacylation. For example, threonyl-tRNA synthetase (ThrRS) selects Thr out of the amino acid pool and ligates it on to the 3′ end of $tRNA^{Thr}$. The resulting aminoacyl-tRNAs are then delivered to the ribosome by initiation or elongation factors to decode the matching codon. **b** | There are multiple mechanisms of genetic code flexibility. Codon bias refers to selective usage of synonymous codons to encode the same amino acid. The frequency of codons in a given organism typically matches the cellular abundance of the corresponding tRNA. Codon reassignment requires evolution of a new tRNA to decode sense codons with a new amino acid, or a new tRNA that can decode stop codons with an amino acid. Ambiguous decoding refers to simultaneous decoding of the same codon by two or more amino acids in one cellular compartment; this could be caused by recognition of the same tRNA by more than one aaRS, by misacylation of a tRNA or by ribosomal decoding errors. Recoding traditionally refers to partial codon reassignment that is context dependent. For instance, in certain bacteria and eukaryotes, a subset of UGA stop codons with a nearby selenocysteine (Sec) insertion sequence (SECIS) element, in the presence of SelB, are recoded to Sec, whereas other UGA stop codons retain their ability to signal translational termination.

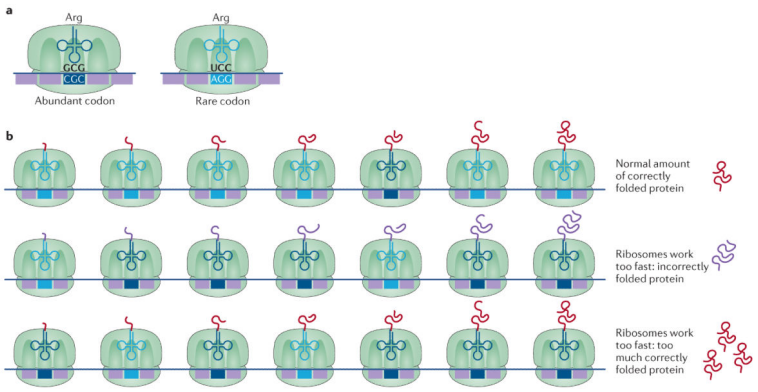

Figure 2. Biased codon usage

a | Biased codon usage occurs when synonymous codons are decoded by different tRNA isoacceptors. As some synonymous codons are used at higher frequencies than others, this leads to biased codon usage. **b** | Although the optimized usage of synonymous codons increases the protein synthesis rate under most conditions, non-optimal codon usage can improve bacterial fitness under certain conditions. For example, codon optimization of the *frq* gene in *Neurospora crassa* results in a loss of fitness. *frq* controls the circadian clock function, and optimizing the codons of *frq* increases the expression of Frq but leads to defects in folding of this protein. These folding defects impair Frq function in the regulation of the circadian feedback loop (middle panel). Similarly, the *kaiB* and *kaiC* genes in the cold-adapted cyanobacterium *Synechococcus elongatus* are also enriched for non-optimal codons. The *kaiB* and *kaiC* genes are critical for regulation of circadian rhythms, and codon optimization increases the protein levels of KaiB and KaiC and prevents the circadian rhythm from switching off, resulting in loss of fitness by the cyanobacteria (bottom panel). Part **b** of the image adapted from REF. 144, Nature Publishing Group.

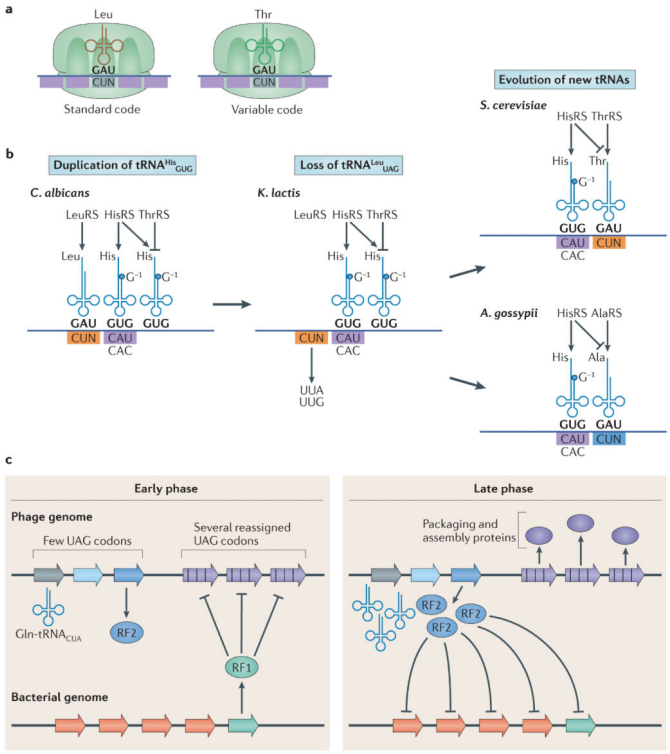

Figure 3. Codon reassignment

a | CUN codons are read as Leu in the standard genetic code through the use of tRNA$^{Leu}_{UAG}$. However, in some microorganisms, CUN codons have been reassigned to Thr. **b** | CUN reassignment in mitochondria involved several steps. In the mitochondrial genome of some Saccharomycetaceae species (for example, *Candida albicans*), the tRNA$^{His}_{GUG}$ gene was duplicated. The CUN codons and a tRNA$^{Leu}_{UAG}$ that decodes CUN then disappeared, leading to a reduced genetic code, as in *Kluyveromyces lactis*. In *Saccharomyces cerevisiae*, one copy of tRNA$^{His}_{GUG}$ evolved to carry an anticodon UAG that reads CUN codons (tRNA$^{His}_{UAG}$). The threonyl-tRNA synthetase (ThrRS) co-evolved with the tRNA$^{His}_{UAG}$ to recognize it as a tRNA$^{Thr}_{UAG}$. CUN codons then reappeared to complete the codon reassignment from Leu to Thr. In *Ashbya gossypii*, secondary mutations enabled the new tRNA$^{His}_{UAG}$ to be recognized by AlaRS instead of HisRS or ThrRS, therefore generating tRNA$^{Ala}_{UAG}$ and reassigning CUN codons to Ala in this species. **c** | Codon reassignment can affect bacteria and phage physiology. In a phage from the human oral cavity, reassigning UAG codons to Gln may allow the phage to interfere with translation of host genes without affecting translation of the phage genes. In the early stages of infection, the phage early genes contain few in-frame UAG codons and are therefore efficiently translated by the host machinery. Among these genes, the phage expresses release factor 2 (RF2), which suppresses translation of UGA codons, including recoded ones. As the bacterial protein RF1, which suppresses translation of UAG codons, contains multiple in-frame UGA codons, expression of RF2 by the phage inhibits translation of host RF1. This inhibition allows the phage to translate late-stage phage genes, which are enriched in recoded UAG codons, while modifying the translation of host genes. Part **b** is adapted from Su, D., Lieberman, A., Lang, B. F., Simonovic, M., Söll, D. & Ling, J., An unusual tRNA^{Thr}

derived from $tRNA^{His}$ reassigns in yeast mitochondria the CUN codons to threonine, *Nucleic Acids Res.*, 2011, **39,** 11, 4866–4874, by permission of Oxford University Press. Part **c** is adapted from Ivanova, N. N. *et al.* Stop codon reassignments in the wild. ***Science*** **344**, 909–913 (2014). Reprinted with permission from AAAS.

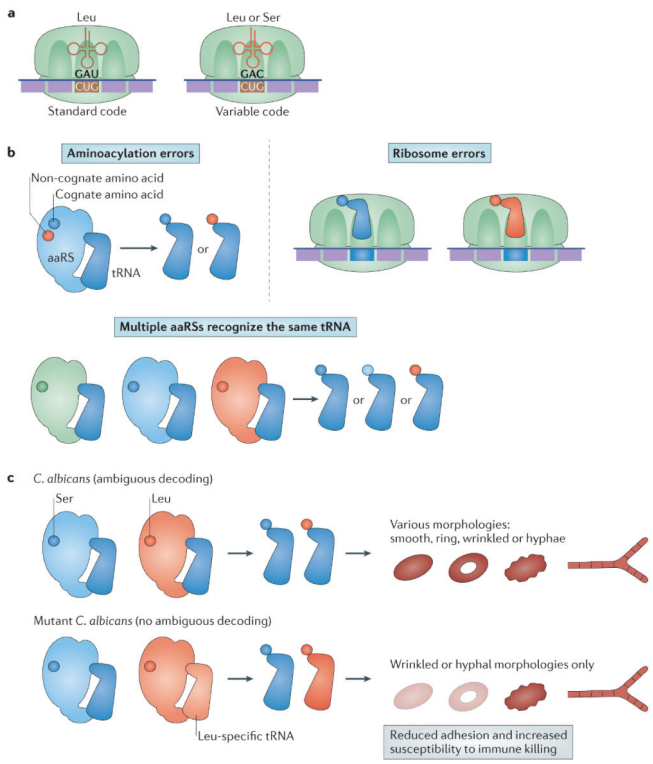

Figure 4. Ambiguous decoding

a | Ambiguous decoding refers to the process by which the same codon gives rise to incorporation of different amino acids in a nascent polypeptide chain. **b** | Ambiguous decoding can result from errors in the aminoacylation reaction carried out by a specific aminoacyl-tRNA synthetase (aaRS), which loads a tRNA with a non-cognate amino acid; by multiple aaRSs recognizing the same tRNA, which leads to different amino acids being loaded onto tRNAs recognizing the same codon; or by ribosomal decoding errors. **c** | In *Candida albicans*, altering the ratio of Ser to Leu incorporation at CUG codons introduces diverse cell and colony morphologies. Wild-type *C. albicans* encodes a CUG-decoding $tRNA_{CAG}$ that is recognized by both seryl- and leucyl-tRNA synthetases. This ability of ambiguous decoding enables wild-type cells to display different colony morphologies, including smooth, ring, wrinkled and hyphae. Eliminating the ambiguity by the substitution of the $tRNA_{CAG}$ with a Leu-specific tRNA results in loss of the smooth and ring morphologies. These changes also reduce cell adhesion and increase fungal susceptibility to macrophage killing by immune cells.

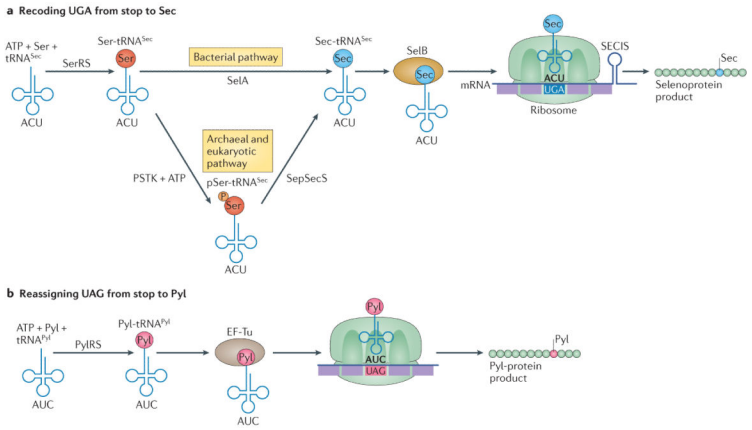

Figure 5. Expanding the genetic code with Sec and Pyl

a | Protein synthesis with selenocysteine (Sec). Sec is biosynthesized on its tRNA. This occurs in multiple steps, beginning with Ser-tRNA^{Sec} formation catalysed by the normal seryl-tRNA synthetase (SerRS). In bacteria, Ser-tRNA^{Sec} is converted to Sec-tRNA^{Sec} by the Sec synthase (SelA). In archaea and eukaryotes, Ser-tRNA^{Sec} is first phosphorylated by phosphoseryl(Sep)-tRNA^{Sec} kinase (PSTK) to generate pSer-tRNA^{Sec} and then an enzyme related to SelA known as Sep-tRNA^{Sec}:Sec-tRNA^{Sec} synthase (SepSecS) converts pSer-tRNA^{Sec} species to Sec-tRNA^{Sec}. A specialized elongation factor (SelB) simultaneously binds to Sec-tRNA^{Sec} as well as the Sec insertion sequence (SECIS) element to direct recoding on the ribosome of specific UGA codons in selenoprotein mRNAs. **b** | Protein synthesis with pyrrolysine (Pyl). Pyl is biosynthesized as a free amino acid in the cell. PylRS ligates Pyl to tRNA^{Pyl}, which contains an anticodon (5′-CUA-3′) that reads UAG codons. Like canonical aminoacyl-tRNAs, Pyl-tRNA^{Pyl} is bound by elongation factor Tu (EF-Tu), enabling UAG translation with Pyl on the ribosome. The Pyl system does not require a specialized elongation factor.